

Facultad de Ingeniería – Universidad de la República

# Data Warehouse de estadísticas vitales y de salud del Ministerio de Salud Pública

---

Carmela Beiro – Sofía Rosano

Tutor: Adriana Marotta

Informe de Proyecto de Grado presentado al Tribunal Evaluador como requisito de graduación de la carrera Ingeniería en Computación

Montevideo, Uruguay

Julio 2016





## Resumen

El Ministerio de Salud Pública (MSP) cuenta con varios sistemas de información para la gestión de datos de salud a nivel nacional. Entre estos sistemas se destacan dos que dan soporte a la generación de estadísticas vitales e indicadores de salud; el sistema Certificado de Nacido Vivo Electrónico, donde se registran todos los nacimientos vivos en el territorio nacional, y el sistema de registro de defunciones, Certificado de Defunción.

La motivación para la realización de este proyecto surge a raíz de la necesidad del MSP de contar con un sistema que resuelva la recolección, integración y depuración de los datos de dichos sistemas de gestión, y permita al personal del área estadística y social acceder a los mismos en forma adecuada para los análisis que desean realizar.

Como resultado principal del proyecto se obtiene un prototipo de sistema de Data Warehouse, poblado con los datos de los nacimientos y defunciones manejados por el MSP luego de realizar un análisis y mejora de la calidad de los mismos, que permite la elaboración de indicadores de salud y estadísticas vitales, tales como cantidad de nacimientos y tasa de mortalidad infantil. A su vez, se obtiene como resultado una aplicación web que provee las funcionalidades necesarias para obtener las estadísticas e indicadores preestablecidos, así como confeccionar consultas y reportes a medida. Además, ofrece distintas opciones para la carga y actualización del sistema, con el fin de facilitar su mantenimiento. Tanto el sistema de Data Warehouse como la aplicación web fueron desarrollados utilizando herramientas de código abierto, fundamentalmente Pentaho Community y Java EE.

**Palabras claves:** Data Warehouse, estadísticas vitales, nacidos vivos, mortalidad, indicadores de salud.



# Índice

Resumen .....	3
Capítulo 1 - Introducción .....	9
1.1 Contexto.....	9
1.2 Problema planteado.....	9
1.3 Objetivos .....	10
1.3.1 Objetivos específicos.....	10
1.3.2 Resultados esperados .....	10
1.4 Organización del documento .....	10
Capítulo 2 - Conceptos básicos .....	13
2.1 Introducción a los sistemas de Data Warehouse.....	13
2.2 Arquitectura .....	14
2.3 Desarrollo de sistemas de Data Warehouse .....	16
2.3.1 Especificación de requerimientos.....	16
2.3.2 Diseño conceptual.....	17
2.3.3 Diseño lógico.....	25
2.3.4 Diseño físico e implementación .....	27
2.4 Calidad de datos.....	28
2.4.1 Dimensiones.....	28
2.4.2 Modelo de calidad .....	30
2.4.3 Limpieza de datos .....	30
Capítulo 3 – Análisis .....	31
3.1 Descripción de la realidad.....	31
3.2 Análisis de requerimientos .....	33
3.2.1 Requerimientos funcionales .....	33
3.2.2 Requerimientos no funcionales .....	36
3.3 Descripción de las fuentes .....	37
3.3.1 Problemas encontrados .....	41
Capítulo 4 – Diseño .....	43
4.1 Diseño conceptual.....	43

4.1.1 Dimensiones y jerarquías .....	43
4.1.2 Medidas.....	51
4.1.3 Relaciones dimensionales .....	53
4.1.4 Estudio de aditividad de medidas .....	54
4.1.5 Verificación del diseño conceptual .....	60
4.2 Diseño lógico .....	62
4.2.1 Relación dimensional Nacimientos .....	62
4.2.2 Relación dimensional Mortalidad .....	64
4.2.3 Relación dimensional Mortalidad Infantil.....	65
4.3 Diseño físico .....	66
Capítulo 5 – Implementación.....	67
5.1 Herramientas utilizadas .....	67
5.1.1 Pentaho BI Suite Community Edition.....	67
5.2 Carga y actualización.....	69
5.2.1 Calidad de datos.....	69
5.2.2 Extracción, transformación y carga.....	70
5.2.3 Verificación de los procesos de ETL .....	88
5.3 Implementación de cubos.....	89
5.3.1 Implementación de cubos en Mondrian.....	89
5.3.2 Cubos implementados .....	93
5.3.3 Verificación de la implementación de cubos .....	100
Capítulo 6 – Interfaz de usuario.....	103
6.1 Inicio.....	104
6.2 Consultas.....	105
6.3 Carga .....	107
6.4 Verificación de la interfaz de usuario .....	108
Capítulo 7 - Conclusiones.....	111
7.1 Trabajo a futuro .....	112
Bibliografía .....	115
Anexo I - Calidad de datos .....	117

Anexo II – Carga y actualización.....	139
Anexo III – Consultas para verificación de ETL.....	151
Anexo IV – Implementación de cubos .....	157
Anexo V – Manual de usuario .....	165



# Capítulo 1 - Introducción

Actualmente la gran mayoría de las organizaciones estatales generan grandes volúmenes de datos que luego son utilizados como insumo al momento de tomar decisiones estratégicas y evaluar el desempeño de las mismas. Dada la importancia de dichas tareas, resulta fundamental que los datos utilizados sean confiables y se encuentren libres de errores. A su vez, para tomar ciertas decisiones resulta de gran utilidad realizar análisis enmarcados dentro del contexto global de la organización, lo cual suele requerir de la integración de datos de distintos sistemas, tanto internos como externos a la misma.

De lo antedicho se desprende la necesidad de contar con un sistema que permita realizar análisis de datos de manera eficiente, integrando datos de distintas fuentes y asegurando la calidad de los mismos. Este proyecto está orientado a brindar un sistema con dichas características al Ministerio de Salud Pública.

## 1.1 Contexto

El presente proyecto fue propuesto en el contexto de un trabajo entre el grupo Concepción de Sistemas de Información (CSI) del Instituto de Computación (InCo) y la Unidad de Información Nacional en Salud (UINS) del Ministerio de Salud Pública (MSP) a través de la Fundación Ricaldoni.

El principal cometido de la UINS es confeccionar las estadísticas vitales del país, a partir de los Certificados de Nacido Vivo y Defunción, documentos expedidos para todo recién nacido con signos vitales y fallecido respectivamente. Adicionalmente, se encarga de elaborar indicadores relativos a la gestación y el parto, a partir de los datos del Sistema de Información Perinatal, así como otros indicadores y reportes solicitados por autoridades y organizaciones, entre las que se encuentra la Organización Mundial de la Salud.

## 1.2 Problema planteado

En la actualidad la UINS elabora manualmente las estadísticas e indicadores requeridos. Esto supone un gran esfuerzo por parte del personal, ya que además de realizar los cálculos pertinentes y elaborar reportes a medida, debe resolver varios problemas de calidad e integrar datos de distintas fuentes de forma manual. Si bien existen trabajos previos que resuelven parcialmente algunos de los problemas que se plantean, no se tiene hasta el momento, un sistema que les provea una solución integral al problema.

## 1.3 Objetivos

El objetivo general del presente proyecto es desarrollar un prototipo de sistema de Data Warehouse que permita la elaboración de estadísticas vitales e indicadores de salud, principalmente los referidos a nacimientos y defunciones.

### 1.3.1 Objetivos específicos

Los objetivos específicos planteados son:

- Relevamiento de los requerimientos de análisis y de las fuentes de datos del MSP.
- Diseño conceptual y lógico del Data Warehouse.
- Implementación del prototipo de sistema de Data Warehouse.
- Diseño e implementación de interfaz a medida que permita realizar consultas y reportes personalizados, así como facilitar la carga y actualización del Data Warehouse.

### 1.3.2 Resultados esperados

Una vez finalizado el proyecto, se espera que el resultado final esté compuesto por:

- Diseño e implementación del prototipo.
- Documentación detallada del trabajo realizado.
- Implementación de la interfaz de usuario a medida.
- Guía de instalación y manual de usuario de la interfaz desarrollada.

## 1.4 Organización del documento

A continuación se describe la estructura del presente informe.

En el capítulo 2 se brindan conceptos fundamentales sobre sistemas de Data Warehouse, necesarios para el entendimiento del proyecto. Se incluye una introducción que expone a grandes rasgos la motivación detrás de los mismos y sus principales características, una breve descripción de su arquitectura y un recorrido por las fases que componen al proceso de desarrollo. Adicionalmente, se presentan conceptos básicos de calidad de datos.

En el capítulo 3 se describe la realidad en la que se enmarca el proyecto, se especifican los requerimientos, tanto funcionales como no funcionales, y se presentan las fuentes de datos provistas.

En el capítulo 4 se detalla el diseño tanto conceptual, como lógico y físico del Data Warehouse definido para la realidad planteada. Adicionalmente, se presenta la verificación del diseño conceptual.

En el capítulo 5 se tratan los detalles de implementación. Se incluye una descripción de las herramientas utilizadas, del desarrollo de los componentes del Data Warehouse y su posterior verificación.

En el capítulo 6 se exponen a grandes rasgos las funcionalidades provistas por la interfaz de usuario desarrollada a medida, así como la verificación de la misma.

En el capítulo 7 se presentan las conclusiones y lineamientos para trabajos futuros.

Al final del documento se encuentra el Anexo I, que detalla los resultados del análisis de calidad realizado sobre las fuentes, los Anexos II, III y IV, que contienen archivos referentes a la implementación y verificación de los componentes del Data Warehouse, y el Anexo V, que presenta el manual de usuario de la interfaz desarrollada.



## Capítulo 2 - Conceptos básicos

En el presente capítulo se introducen los conceptos fundamentales necesarios para el entendimiento del proyecto.

### 2.1 Introducción a los sistemas de Data Warehouse

El análisis de datos en las organizaciones se ha vuelto cada vez más importante a través de los años, ya que estas deben mejorar sus procesos de toma de decisiones para poder mantener su ventaja competitiva.

Las bases de datos tradicionales, conocidas como bases de datos operacionales, no satisfacen los requerimientos para el análisis de datos. Están diseñadas con el fin de soportar las operaciones diarias de una organización de forma eficiente, asegurando el rápido acceso a los datos a múltiples usuarios. Típicamente almacenan datos detallados, no contienen datos históricos y, como en general se encuentran normalizadas, tienen un bajo rendimiento a la hora de ejecutar consultas muy complejas que deben unir varias tablas relacionales o agregar grandes volúmenes de datos. Además, cuando se desea analizar el comportamiento de la organización como un todo, comúnmente se deben integrar datos de varios sistemas operacionales y esto puede ser una tarea difícil de llevar a cabo, ya que pueden existir diferencias en las definiciones y el contenido de los datos [1].

Los sistemas de Data Warehouse fueron propuestos para responder de forma más adecuada a la creciente necesidad de contar con un sistema que permita realizar análisis de datos de manera eficiente y de soporte los procesos de toma de decisiones. Según Inmon [2] un Data Warehouse se puede definir como *una colección de datos orientados a temas, integrados, no volátiles e históricos, organizados para soportar un proceso de toma de decisiones*. A continuación se detallan dichos aspectos distintivos:

- **Orientado a temas:** un Data Warehouse se organiza en torno a temas particulares de análisis, que varían en función del tipo de actividades realizadas por la organización. No así las bases de datos operacionales, que tienen como foco principal las funciones específicas que deben llevarse a cabo por la aplicación.
- **Integrado:** comúnmente para analizar el comportamiento de una organización es necesario obtener datos de varios sistemas operacionales. Esto implica que un Data Warehouse deba tener la capacidad de almacenar de forma homogénea datos que provengan de distintas fuentes y presenten ciertas diferencias. Entre estas diferencias se pueden destacar el formato, la codificación, la existencia de sinónimos (campos con nombres diferentes pero los mismos datos), homónimos (campos con el mismo nombre pero con distinto significado) y la multiplicidad de ocurrencias de los datos.
- **No volátil:** un Data Warehouse debe almacenar datos durante largos períodos de tiempo, asegurando que los mismos persistan sin ser modificados.

- **Histórico:** para datos que en la realidad varían a lo largo del tiempo, se deben registrar los sucesivos cambios, almacenando el período de validez de los mismos.

## 2.2 Arquitectura

En la Figura 2.1 se puede observar la arquitectura típica de un Data Warehouse según Malinowski [1], que consta de los siguientes componentes:

- **Fuentes de datos:** contienen los datos que se van a utilizar para poblar el Data Warehouse. Las mismas pueden ser tanto internas como externas a la organización.
- **Capa back-end:** está compuesta por herramientas de extracción, transformación y carga (ETL) de datos, utilizadas para alimentar el Data Warehouse. A continuación se detallan los tres pasos de este proceso:
  - **Extracción:** recopilación de datos de diversas fuentes. Las mismas pueden ser bases de datos relacionales u otras fuentes con distintos formatos.
  - **Transformación:** conversión de los datos del formato de la fuente al formato del Data Warehouse. Incluye tareas de depuración, filtrado, integración y agregación de los mismos.
  - **Carga:** población del Data Warehouse con los datos transformados. Además, comprende tareas de actualización periódica del mismo.

Adicionalmente, puede incluir una base de datos temporal sobre la cual se ejecutan todos los procesos de integración y transformación datos, antes de realizar la carga del Data Warehouse. Esto es lo que se conoce como *data staging*.

- **Capa Data Warehouse:** está compuesta por un Data Warehouse corporativo y/o varios Data Marts, y un repositorio de metadata que almacena información sobre el Data Warehouse y su contenido. Un Data Warehouse corporativo es un Data Warehouse centralizado que abarca todas las áreas funcionales y departamentos de una organización. Por otra parte, un Data Mart es un Data Warehouse especializado, orientado hacia un área funcional particular o un grupo de usuarios en una organización.
- **Capa OLAP (On-Line Analytical Processing):** servidor OLAP que soporta tanto datos multidimensionales como sus operaciones. Se pueden distinguir distintos tipos de servidores OLAP, según la forma en que almacenan los datos:
  - **ROLAP (Relational OLAP):** almacena los datos en un motor relacional.
  - **MOLAP (Multidimensional OLAP):** trabaja sobre almacenamiento especializado.
  - **HOLAP (Hybrid OLAP):** combina ambas estrategias.

- **Capa front-end:** está compuesta por herramientas clientes, como por ejemplo herramientas OLAP, de estadísticas y generación de reportes, que permiten visualizar y analizar interactivamente el contenido del Data Warehouse.

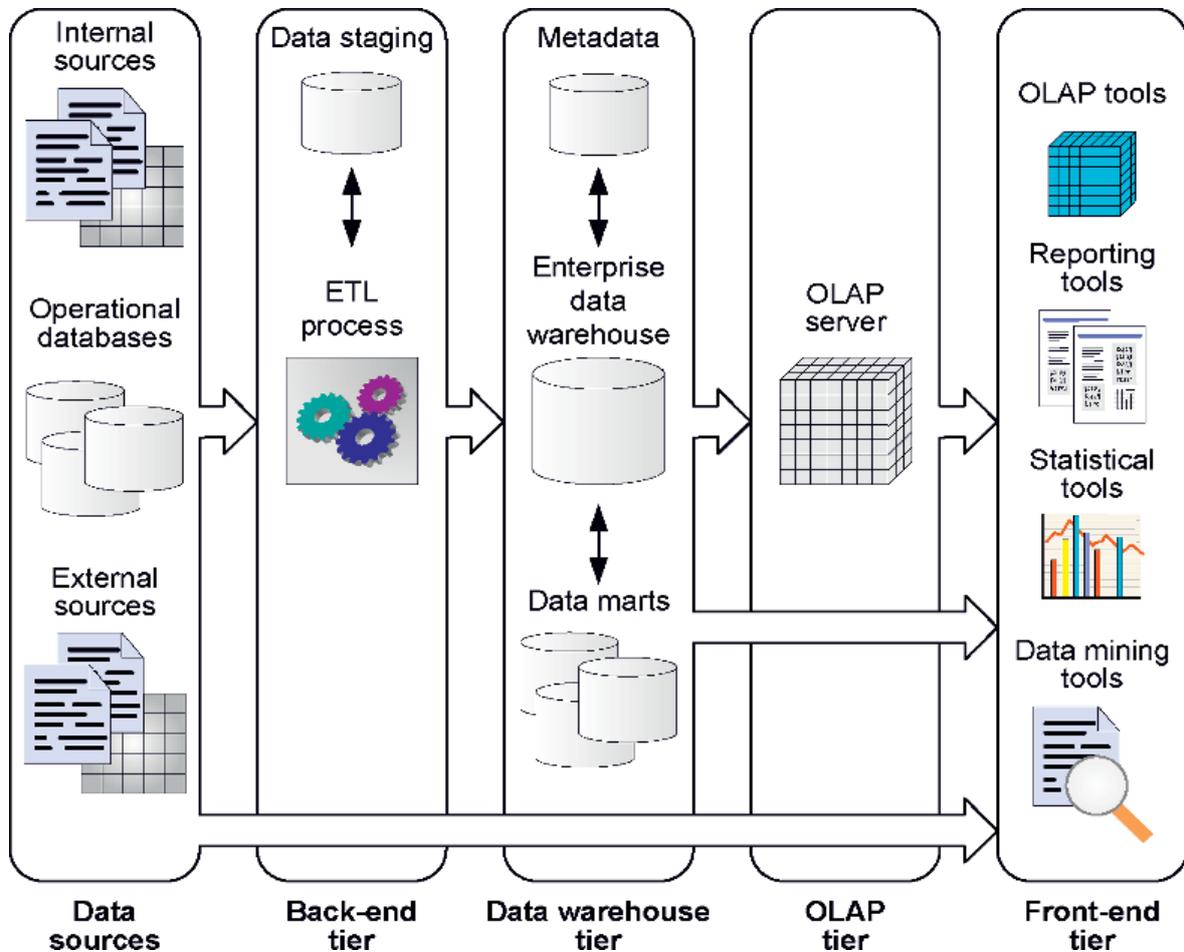


Figura 2.1 - Arquitectura típica de un Data Warehouse [1]

## 2.3 Desarrollo de sistemas de Data Warehouse

Uno de los enfoques más generales del proceso de desarrollo de un sistema de Data Warehouse, ilustrado en la Figura 2.2, consta de las siguientes cuatro fases:

- Especificación de requerimientos
- Diseño conceptual
- Diseño lógico
- Diseño físico e implementación

Durante la fase de diseño se establece un esquema conceptual para el Data Warehouse, que luego es transformado en un modelo de datos lógico, que es la base para la implementación del mismo.

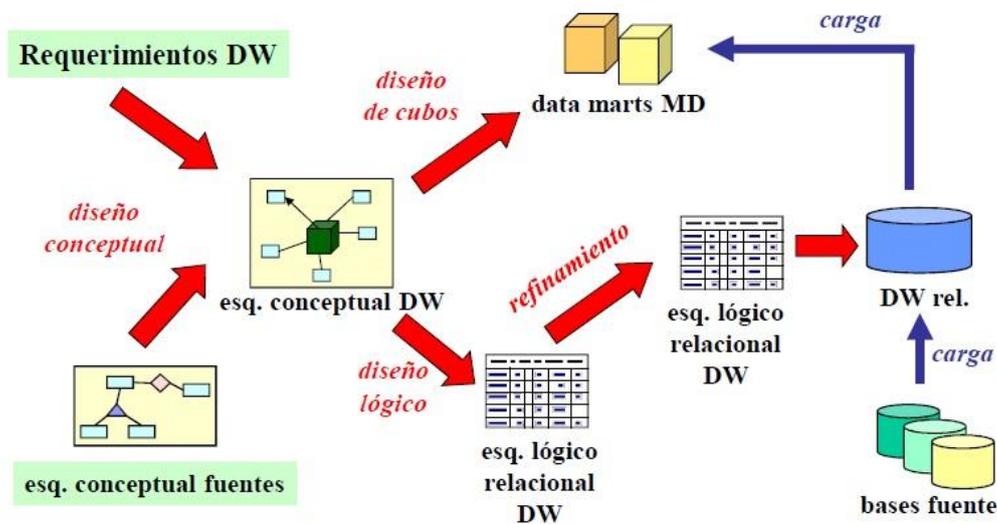


Figura 2.2 - Enfoque general de desarrollo de un Data Warehouse [3]

### 2.3.1 Especificación de requerimientos

Esta etapa del desarrollo tiene como objetivo especificar las metas, limitaciones y prioridades del sistema de Data Warehouse. Es de gran importancia establecer requerimientos precisos durante la misma, ya que estos son la base de las etapas siguientes del desarrollo.

Existen distintos enfoques para la especificación de requerimientos. Puede ser un proceso orientado al usuario, donde estos determinan criterios y objetos de análisis relevantes; orientado al negocio, donde las estructuras del Data Warehouse se derivan del análisis de los procesos o requerimientos de negocio; orientado a las fuentes de datos; o un enfoque combinado.

## 2.3.2 Diseño conceptual

El propósito de esta fase es representar los requerimientos especificados en la etapa anterior de manera clara y concisa, para que pueda ser entendida por los usuarios del sistema y por los diseñadores en las siguientes etapas de diseño.

No existe actualmente un consenso acerca de cómo implementar el diseño conceptual a partir de los requerimientos. Sin embargo se pueden distinguir dos enfoques, uno basado en los requerimientos y otro en las fuentes. En el primero, los requerimientos son el universo de información y las fuentes se relacionan luego. El modelo CMDM (*Conceptual MultiDimensional Model*), propuesto como tesis de Maestría por el grupo CSI del InCo [4] utiliza este enfoque. En el segundo, que suele ser de utilidad cuando los requerimientos están poco claros, el Data Warehouse se obtiene realizando transformaciones sobre las fuentes. Este enfoque es utilizado en el modelo MultiDim propuesto por Golfarelli [5].

Tanto CMDM como MultiDim se basan en el modelo multidimensional, que será explicado en la siguiente sección. Luego, se detallará el modelo CMDM que fue utilizado para realizar el diseño conceptual de este proyecto.

### *Modelo multidimensional*

Las consultas utilizadas para la toma de decisiones suelen ser complejas, poco intuitivas de escribir en los lenguajes tradicionales y suelen tener un tiempo de respuesta demasiado largo si son ejecutadas en bases de datos operacionales. Por esta razón, los sistemas de Data Warehouse utilizan un modelo multidimensional que permite un mejor rendimiento y entendimiento de los datos para su análisis.

En una base de datos multidimensional los datos se representan como matrices n-dimensionales denominadas  **cubos** . Estos se definen a partir de sus  **dimensiones**  y  **medidas** . Las dimensiones son los criterios de análisis y se encuentran en los ejes del cubo; mientras que las medidas son los valores o indicadores que se desean analizar, y son las celdas resultantes de la intersección de las dimensiones en el cubo. En la Figura 2.3 se puede observar un ejemplo de un cubo tridimensional que analiza datos sobre ventas de una cadena de supermercados. Sus dimensiones son Vendedores, Productos y Tiempo, mientras que la medida a analizar es la cantidad de ventas.

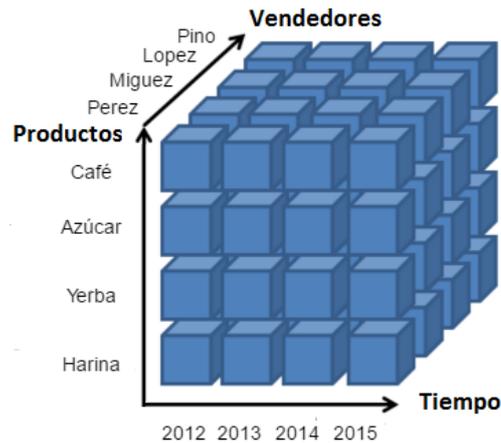


Figura 2.3 - Ejemplo cubo Ventas

Dentro de una dimensión existen **jerarquías** que permiten analizar las medidas según distintas granularidades. Estas relacionan los **niveles** más bajos, que son los más detallados, con los más altos, más generales. En la Figura 2.4 se muestra una posible jerarquía para la dimensión Vendedores. Todas las ventas están relacionadas con un vendedor, que trabaja en una ciudad, la cual pertenece a una región. Aunque no está señalado en la figura, en toda jerarquía los miembros del nivel más alto son agrupados bajo el miembro *All*, que representa la jerarquía completa. En el ejemplo de la Figura 2.3, si se analiza al cubo según el nivel *All* de la dimensión Vendedores se obtiene el total de ventas para todos los vendedores de todas las regiones.

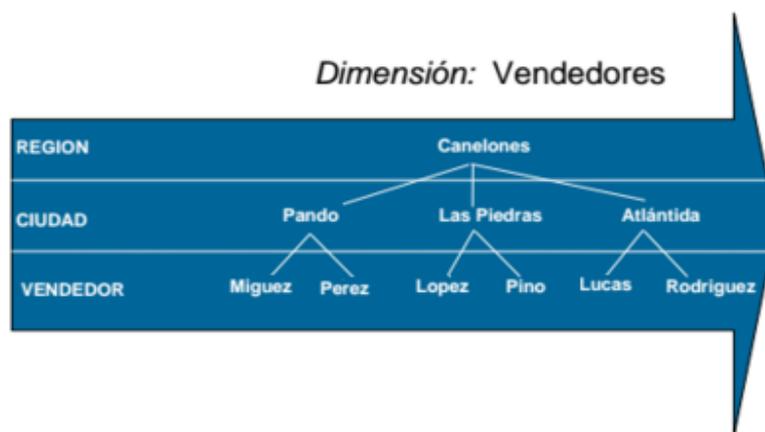


Figura 2.4 - Ejemplo de jerarquía en dimensión Vendedores [3]

## Operaciones OLAP

Existen operaciones OLAP que permiten visualizar los datos de manera interactiva desde distintas perspectivas y según distintos niveles utilizando las dimensiones y sus jerarquías. A continuación se detalla el conjunto de las más comúnmente utilizadas [3].

### *Slice*

Golfarelli [6] define *slice* como la operación que reduce la cantidad de dimensiones por las cuales se desea analizar una medida, asignando a una dimensión un valor fijo. En la Figura 2.5 se puede observar el resultado de realizar la operación *slice* sobre el cubo de Figura 2.3, asignando un valor fijo a la dimensión Vendedores.

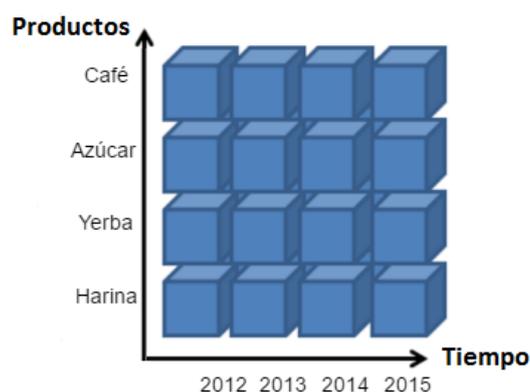


Figura 2.5 - Ejemplo operación slice

### *Dice*

Esta operación realiza una selección de secciones del cubo dependiendo del valor de ciertas dimensiones. En la Figura 2.6 se muestra el resultado de realizar la operación *dice* sobre el cubo de Figura 2.3, eligiendo los valores Café y Azúcar en la dimensión Productos.

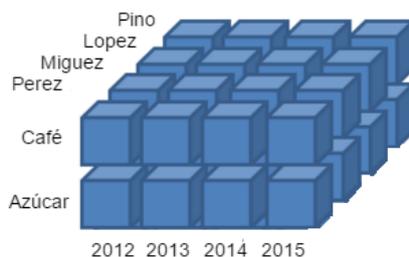


Figura 2.6 - Ejemplo operación dice

### Pivot

La operación *pivot* rota los ejes del cubo para permitir distintas presentaciones del mismo. En la Figura 2.7 se puede observar el resultado de realizar dicha operación sobre el cubo de Figura 2.3.

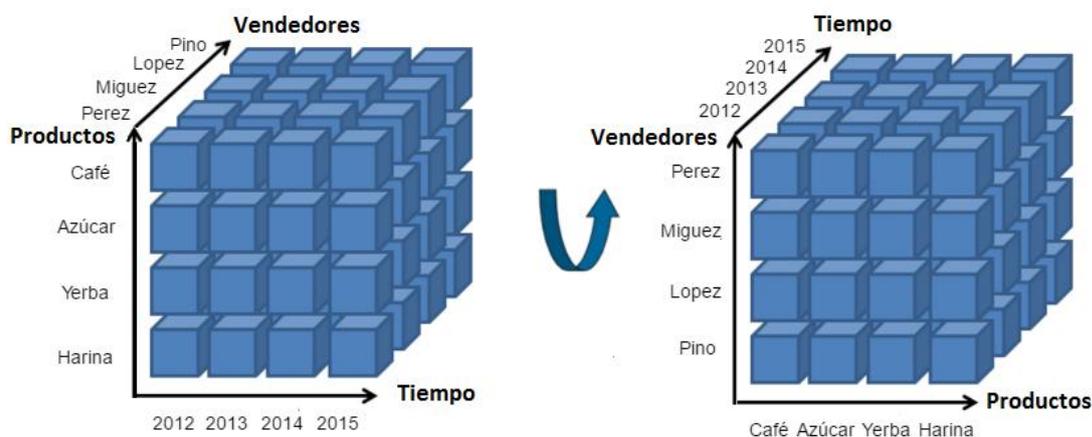


Figura 2.7 - Ejemplo operación pivot

### Drill-Up y Drill-Down

Estas operaciones implican movimientos en una jerarquía de una dimensión, son agrupamientos y desagrupamientos respectivamente. *Drill-up* se mueve desde un nivel más detallado a uno más general de una jerarquía, mientras que *drill-down* realiza lo contrario. La Figura 2.8 ilustra el resultado de realizar estas operaciones sobre la dimensión Vendedores del cubo de Figura 2.3.

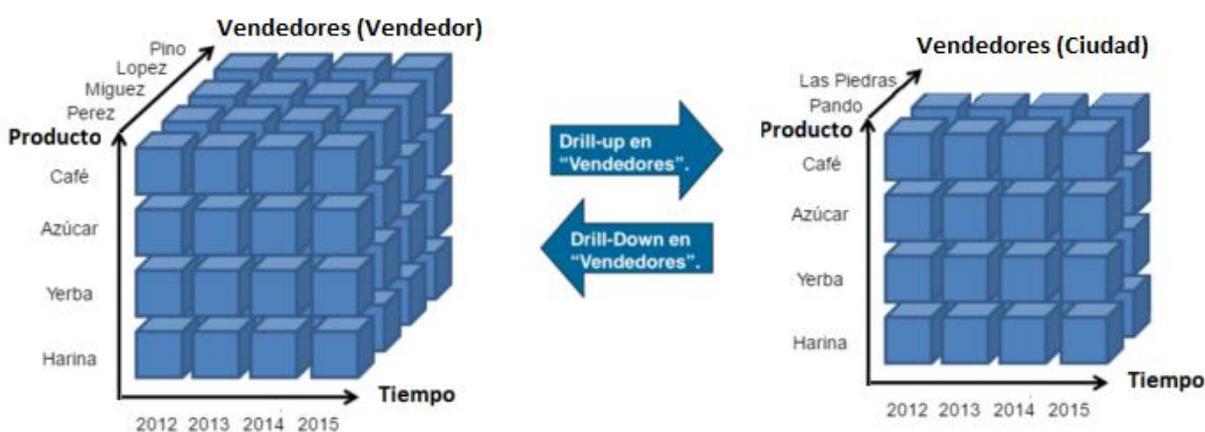


Figura 2.8 - Ejemplo operaciones drill-down y drill-up

### Roll-Up

Cuando se realiza un *drill-up* la operación *roll-up* calcula las medidas en función de los agrupamientos. Para utilizarla se debe especificar cómo se calcularán los nuevos valores de las medidas. Dicha especificación generalmente se detalla mediante una tabla, llamada cuadro *roll-up*, en donde se establece para cada medida y cada operación *roll-up* posible, la función de agregación a utilizar. Esta función puede obtenerse a partir de operaciones aritméticas como la suma o promedio, u otros procedimientos más complejos. En la Figura 2.9 se muestra el resultado de aplicar la operación *roll-up* utilizando la suma al realizar un *drill-up* sobre la dimensión Vendedores en el cubo de la Figura 2.3.

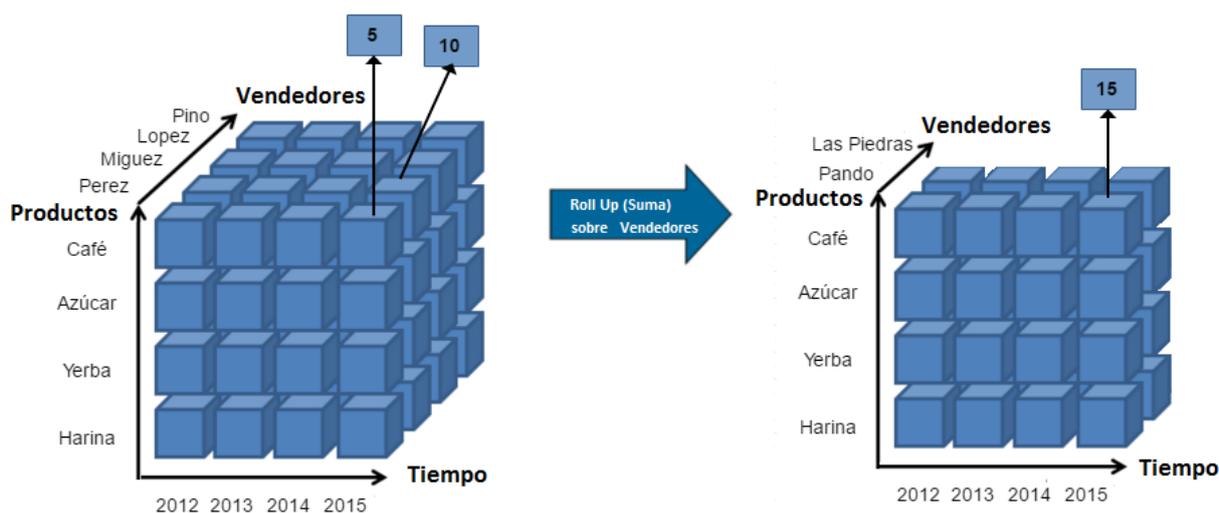


Figura 2.9 - Ejemplo operación roll-up

### Drill-Across

Esta operación relaciona dos cubos y permite ejecutar consultas que involucren a ambos. Cabe destacar que ambos cubos deben tener alguna dimensión en común. Como ejemplo, se podría realizar un *drill-across* entre el cubo de la Figura 2.3 y un cubo que permita analizar datos sobre costos en el tiempo para compararlos.

## *Agregación de medidas*

Cuando se realizan operaciones OLAP es de vital importancia manipular adecuadamente las medidas para asegurar la correctitud de los resultados obtenidos. La agregación de medidas se da al cambiar el nivel de abstracción o granularidad a la cual se están analizando los datos, es decir, al moverse de un nivel a otro dentro de una jerarquía. Para asegurar que la agregación de las medidas sea correcta, es necesario que se satisfagan las siguientes condiciones:

- **Instancias disjuntas:** la agrupación de instancias de un mismo nivel con respecto a sus padres en el siguiente nivel debe resultar en conjuntos disjuntos. Es decir, una instancia de un nivel no puede pertenecer a más de una instancia del nivel padre. En el ejemplo presentado anteriormente, en la dimensión Vendedores de la Figura 2.4, un vendedor no puede pertenecer a más de una ciudad. Cumplir con esta condición evita el problema del doble conteo.
- **Completitud:** todas las instancias deben estar incluidas en una jerarquía y cada una de ellas debe estar relacionada a una instancia del siguiente nivel. En el ejemplo de la Figura 2.4, la dimensión Vendedores contiene a todos los vendedores y cada vendedor tiene asignada una ciudad.
- **Uso correcto de funciones de agregación:** es necesario tener en cuenta el tipo de medida al momento de elegir la función de agregación. Existen distintos tipos de medidas de acuerdo a su grado de aditividad. Por una parte, se encuentran las medidas aditivas, que son aquellas que conservan la semántica al aplicar la operación *roll-up*, con función de agregación suma, por cualquier dimensión. En el ejemplo presentado anteriormente, cuyo cubo se puede observar en la Figura 2.3, la medida cantidad de ventas es una medida aditiva. Por otra parte, se encuentran las medidas semi-aditivas, que son aquellas que son aditivas en algunas dimensiones y en otras no. Y finalmente se encuentran las no aditivas que no conservan la semántica al aplicar la operación *roll-up*, con función de agregación suma, en ninguna dimensión [3].

## Modelo CMDM

El objetivo principal del modelo CMDM [4] es especificar de manera directa, expresiva y precisa determinada realidad en términos multidimensionales. Para ello, presenta las siguientes estructuras básicas:

- Niveles:** representan un conjunto de datos del mismo tipo. Existe una analogía entre un nivel y una entidad en el Modelo Entidad Relación. Para representar el esquema de un nivel se utiliza un rectángulo que contiene el nombre y la estructura del mismo. En la Figura 2.10 se muestra un ejemplo de cómo quedaría representado el nivel Vendedor visto en el ejemplo anterior.



Figura 2.10 - Ejemplo Nivel CMDM

- Dimensiones:** una dimensión está compuesta por un conjunto de niveles organizados en jerarquías. En cada una de ellas se tiene una relación 1:N entre objetos de nivel superior y nivel inferior. El esquema de una dimensión está representado por un rectángulo dentro del cual aparece su nombre y un grafo dirigido donde los nodos son los niveles que participan de esa dimensión. Cabe señalar que las medidas también se representan como dimensiones. En la Figura 2.11 queda ejemplificado cómo se debe representar la dimensión Vendedores. Por otra parte, en la Figura 2.12 se muestra la representación de una posible estructuración de las dimensiones restantes y de la medida cantidad de ventas del cubo de la Figura 2.3.

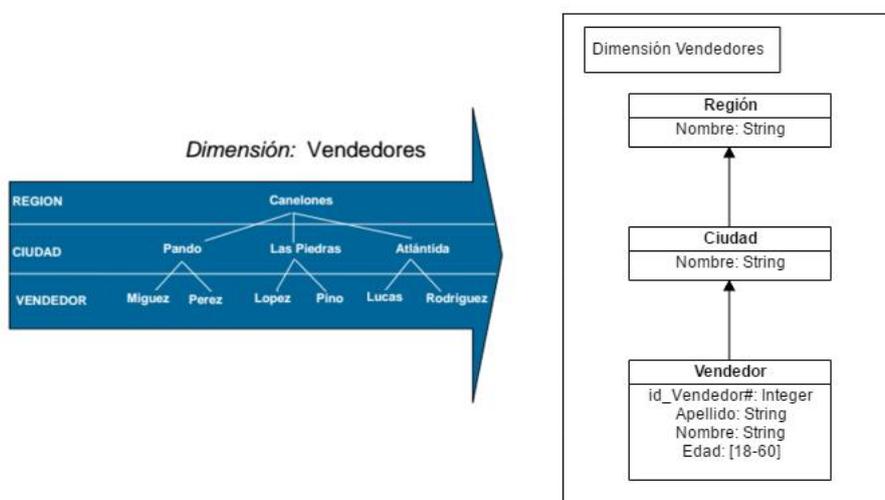


Figura 2.11 - Ejemplo Dimensión Vendedores CMDM

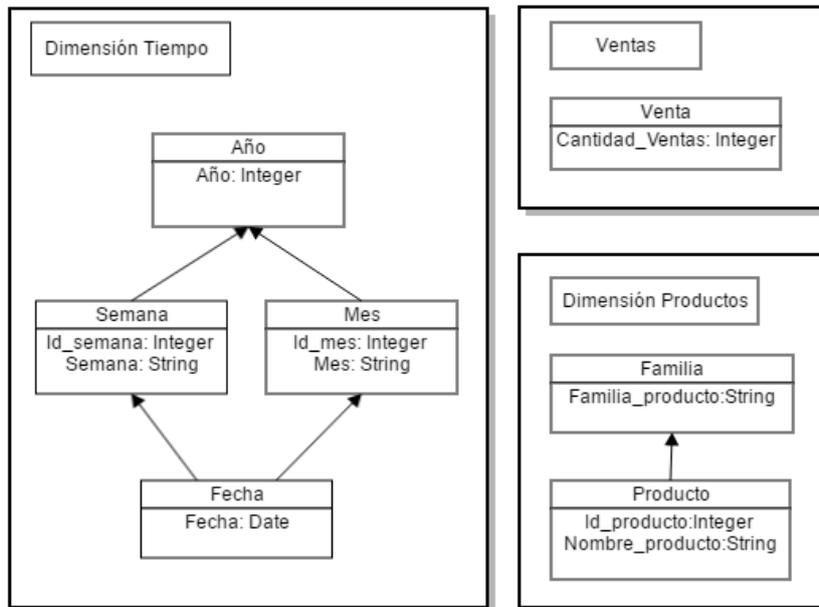


Figura 2.12 - Ejemplo Dimensiones y Medida CDMM

- **Relaciones dimensionales:** representan cruzamientos entre dimensiones. Se tiene un elemento en el conjunto relación si y sólo si hay un cruzamiento. Esto obliga a que las dimensiones participantes realmente sean cruzables. En otras palabras, *representa un conjunto de cubos, tomado del conjunto de todos los cubos que se pueden construir a partir de los niveles de un conjunto dado de dimensiones* [4]. El cubo de la Figura 2.3 se puede ver como una instancia de una relación dimensional, como queda ejemplificado en la Figura 2.13. El esquema de una relación dimensional está dado por un grafo en forma de estrella. El nodo central es de forma oval y tiene el nombre de la relación dimensional, y los nodos “satélite” son rectangulares y tienen el nombre de cada una de las dimensiones que participan de la relación.

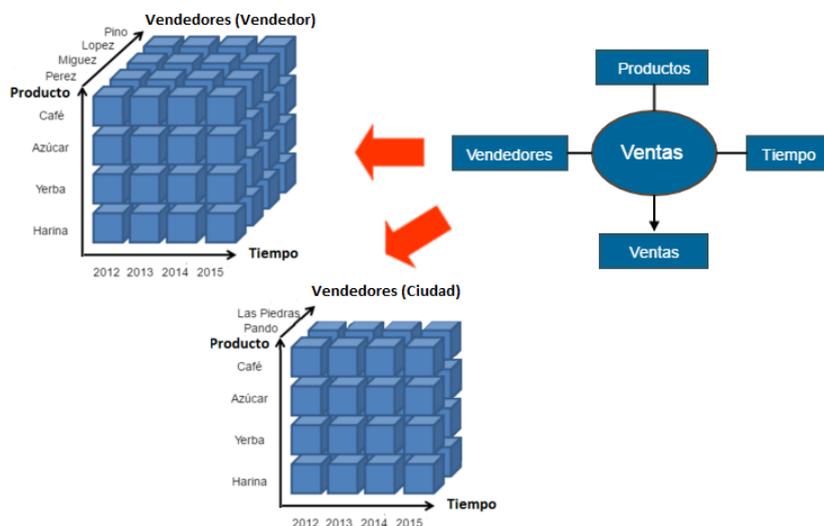


Figura 2.13 - Ejemplo Relación Dimensional Ventas CDMM

### 2.3.3 Diseño lógico

Esta fase del proceso de desarrollo toma como insumo el resultado de la fase anterior, el diseño conceptual, y en base a este determina el modo de almacenamiento, las estructuras y el esquema lógico del Data Warehouse. Uno de los principales problemas a resolver es la transformación del esquema conceptual, que es abstracto, al esquema lógico, que lo implementa. Obtener las estructuras adecuadas para la realización de análisis eficientes y resolución de consultas complejas, también es uno de los objetivos primordiales en esta fase.

Como ya se mencionó en la sección *Arquitectura*, existen tres modos de almacenamiento para un Data Warehouse. Las características de los mismos según [3] son:

- **ROLAP:** el Data Warehouse se almacena en una base de datos relacional. De esta forma se trabaja con un modelo estandarizado, tanto en su estructura como en su lenguaje de manipulación, SQL. Este modo de almacenamiento brinda flexibilidad y potencial para realizar consultas *ad-hoc*, especialmente consultas por clave y por condición. Sin embargo, requiere de una transformación compleja de datos y operaciones multidimensionales hacia estructuras y operaciones relacionales.
- **MOLAP:** el Data Warehouse se almacena en una base de datos multidimensional. Los datos y las operaciones se mapean directamente y se almacenan los datos en estructuras especializadas. Esto permite realizar consultas multidimensionales con alta performance. No obstante, el modelo es menos estandarizado, ya que no hay estándares de estructura pero sí de lenguaje de consulta, MDX. Por otra parte, requiere de más espacio en disco que la tecnología ROLAP.
- **HOLAP:** este modo de almacenamiento combina las dos tecnologías, ROLAP y MOLAP, y de esta forma logra un balance entre la performance y el espacio en disco. Por ejemplo, suele utilizarse ROLAP para los datos más detallados y MOLAP para los datos agregados.

La herramienta OLAP utilizada para la implementación de este proyecto, Pentaho Community [7], requiere que el Data Warehouse esté almacenado en una base de datos relacional. Por esta razón el tipo de almacenamiento utilizado fue ROLAP.

El modelo relacional, utilizado en el diseño de bases de datos relacionales, resulta poco adecuado para realizar consultas multidimensionales, ya que implica *joins* entre varias tablas y sumalizaciones que son costosas y poco optimizables. Por lo tanto, para los sistemas de Data Warehouse ROLAP se utiliza el modelo multidimensional, que está orientado a consultas OLAP y representa los conceptos multidimensionales sobre estructuras relacionales. Este intenta minimizar *joins* y totalizaciones, utilizando redundancias y desnormalizaciones. Dicho modelo presenta dos estructuras básicas; las **tablas de hecho**, que son las tablas principales del modelo y almacenan las medidas numéricas del negocio, resultantes de la intersección de las dimensiones; y las **tablas de dimensión**, donde se guardan las descripciones de las dimensiones.

Existen varios tipos de esquemas para el diseño lógico de un Data Warehouse relacional que especifican la forma de relacionar una tabla de hecho con las tablas de dimensión. Los más utilizados son el esquema estrella y el esquema copo de nieve.

El esquema estrella presenta sus tablas de hecho unidas a todas las tablas de dimensión correspondientes por relaciones 1:N. Estas últimas deben estar desnormalizadas e incluir sus jerarquías embebidas, ocasionando de esta forma redundancia en los datos almacenados. Dada una tabla de hecho, su clave primaria está compuesta por la concatenación de las claves primarias de todas las tablas de dimensión relacionadas a esta. En la Figura 2.14 se puede observar el resultado de aplicar el esquema estrella para el diseño lógico de la relación dimensional de la Figura 2.13.

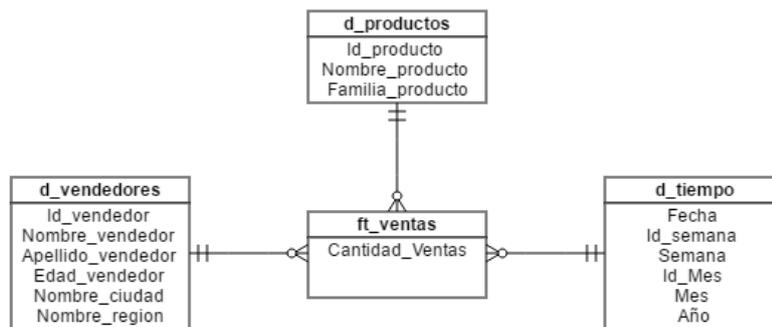


Figura 2.14 - Ejemplo esquema estrella

El esquema copo de nieve surge de normalizar las jerarquías de cada dimensión en un esquema estrella. Es decir, se descompone cada dimensión en sus respectivos niveles y se tiene una tabla por cada uno de ellos relacionados a través de claves foráneas. Además, cada tabla de hecho está unida con la tabla correspondiente al nivel más bajo de cada dimensión mediante una relación 1:N. En la Figura 2.15 se puede observar el resultado de aplicar el esquema copo de nieve para el diseño lógico de la relación dimensional de la Figura 2.13.

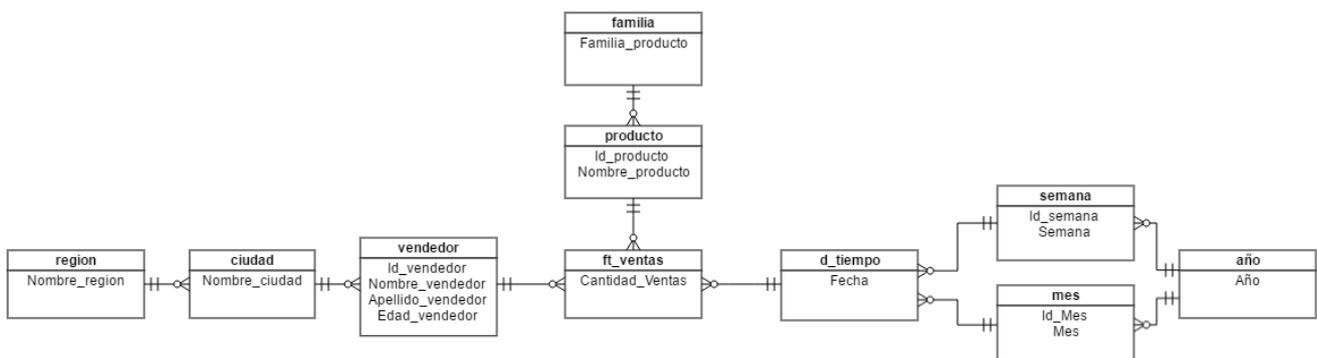


Figura 2.15 - Ejemplo esquema copo de nieve

La decisión de qué esquema escoger para realizar el diseño lógico depende de qué características del Data Warehouse se deseen priorizar. Por una parte, el esquema estrella provee un menor tiempo de respuesta al realizar consultas debido a la redundancia contenida en sus tablas de dimensión, mientras que por otra, el esquema copo de nieve optimiza el espacio utilizado en disco y permite mantener las tablas más fácilmente.

### 2.3.4 Diseño físico e implementación

La etapa de diseño físico está estrechamente relacionada con el diseño lógico ya que en ambas se toman decisiones sobre el almacenamiento de los datos y las estructuras del Data Warehouse. Sin embargo, a diferencia del diseño lógico, el diseño físico se centra en mejorar la performance de las consultas y facilitar el mantenimiento del Data Warehouse. Es por esto que durante el mismo las decisiones a tomar pueden incluir el uso del particionamiento de tablas, tablas agregadas, vistas materializadas, claves subrogadas y/o índices.

El diseño físico depende del DBMS a utilizar. En base a esto, se define cómo se almacenan internamente las tablas, por ejemplo particionada o en una única unidad; qué índices se deben utilizar para optimizar las consultas más importantes del sistema y realizar *joins* de manera eficiente; y qué representación interna al DBMS se debe usar para los datos. Por ejemplo, con respecto a este último punto, es importante elegir correctamente los tipos de datos, para lo cual se deben seleccionar los más pequeños del DBMS que permitan representar los datos.

De esta forma, se culmina el proceso de diseño. Una vez finalizado dicho proceso se procede a la fase de carga y actualización, cuyo objetivo es mantener el Data Warehouse cargado con los datos correspondientes. Esta fase incluye procesos de extracción, transformación y carga (ETL), descritos previamente en la sección *Arquitectura*. Durante estos procesos se extraen los datos de distintas fuentes, se integran, transforman, limpian, filtran y cargan en el Data Warehouse.

## 2.4 Calidad de datos

Los datos constituyen la materia prima para la toma de decisiones operativas y estratégicas por lo que es necesario que estén libres de errores y sean confiables. Por esta razón, la calidad de datos es de vital importancia, pudiendo tener consecuencias serias y de gran trascendencia para la eficiencia y efectividad de las organizaciones que los utilizan.

En general, se reduce el concepto de calidad de datos a “exactitud de datos”. Sin embargo, abarca más que esto. Otros atributos importantes como completitud, consistencia y frescura son necesarios para poder caracterizarlo de forma completa, siendo por lo tanto un concepto multifacético [8].

A continuación serán descritos conceptos y actividades que son necesarios para el análisis y la posible mejora de la calidad de los datos en un Data Warehouse.

### 2.4.1 Dimensiones

Como se mencionó anteriormente, la calidad se caracteriza vía múltiples dimensiones o atributos que ayudan a calificar los datos. Una dimensión captura una faceta a alto nivel de la calidad.

A continuación se presentan varios conceptos de calidad relacionados con las dimensiones:

- **Factor:** representa un aspecto particular de una dimensión de calidad. Una dimensión puede verse como un agrupamiento de factores de calidad que tienen el mismo propósito.
- **Métrica:** instrumento que define la forma de medir un factor de calidad. Para esta se debe definir la semántica (cómo se mide), las unidades de medición y la granularidad de la medida. Un mismo factor de calidad puede medirse con diferentes métricas.
- **Método de medición:** proceso que implementa una métrica. La implementación del método es dependiente de la aplicación en concreto y de la estructura de la base de datos. Una misma métrica puede ser medida por diferentes métodos.

Las dimensiones que fueron de interés para el desarrollo de este proyecto son Exactitud, Completitud y Consistencia. A continuación se presenta una breve descripción de las mismas junto con sus factores y posibles métricas.

## *Exactitud*

La exactitud se relaciona con la correctitud, tanto intrínseca como de representación, y con la precisión con la que están representados los datos en un sistema de información.

El factor que resulta relevante estudiar en este proyecto, junto con sus posibles medidas y mediciones, es la correctitud sintáctica. Dicho factor mide si los valores del sistema de información corresponden a valores válidos del dominio, sin importar si son valores reales o no.

Las posibles métricas para la correctitud sintáctica son un booleano, que indica si un dato es sintácticamente correcto o no, y la desviación de la correctitud sintáctica, que provee la distancia a los datos correctos más parecidos. Estas se pueden medir de dos formas distintas; por extensión, que implica comparar con un diccionario que representa el dominio, o por comprensión, que conlleva chequear si se satisfacen las reglas de sintaxis.

Para medir la exactitud de un conjunto de datos, por ejemplo tablas, en función de la exactitud de cada dato se utilizan ratios y promedios, tanto para booleanos como para desviaciones.

## *Completitud*

La completitud puede ser definida genéricamente como el grado de amplitud, profundidad y alcance de los datos necesario para la tarea en cuestión. Recubre aspectos extensionales e intensionales del sistema de información. Extensionales ya que analiza la cantidad de entidades de la realidad representados, e intensionales ya que estudia la cantidad de datos sobre cada entidad.

El factor de calidad dentro de la dimensión completitud que resulta de interés para el proyecto es la densidad, que determina cuánta información se tiene y cuánta falta sobre las entidades. Una forma común de medirlo es utilizando un ratio que determine el porcentaje de valores no nulos del sistema.

## *Consistencia*

La consistencia captura la satisfacción de reglas semánticas definidas sobre los datos. Estas pueden ser tanto reglas de integridad para una base de datos como reglas de los usuarios.

Los factores de consistencia de interés son: la integridad de dominio, que evalúa la satisfacción de reglas sobre el contenido de un atributo; la integridad intra-relación, que estudia la satisfacción de reglas entre atributos de una misma tabla; y la integridad inter-relación, que mide la satisfacción de reglas entre atributos de varias tablas. Una métrica existente para medir estos factores es un booleano de consistencia, que determina si el dato satisface o no las reglas, y que puede tener granularidad celda o conjunto de celdas.

## 2.4.2 Modelo de calidad

Previo a evaluar y mejorar la calidad de un sistema de información es necesario definir un modelo de calidad adecuado a las necesidades y prioridades de los consumidores de los datos del sistema. Se deben determinar y especificar dimensiones y factores a medir, métricas y datos o grupos de datos donde estas se aplican, y base de metadatos de calidad donde se registrarán los resultados de la medición.

Existen distintos métodos de trabajo para determinar las dimensiones y factores de calidad. Uno de estos métodos se basa en definirlos a partir de los requerimientos de los usuarios. Esto se puede realizar de dos formas distintas. La primera consiste en definir las dimensiones, factores y métricas a partir de los datos prioritarios; mientras que la segunda consta de definir primero las dimensiones y los factores, luego los datos y por último las métricas.

Otro método que se puede aplicar es Data Profiling, que consiste en obtener las dimensiones y factores de calidad a partir del perfil de los datos, ya que estos proveen una noción de por dónde están las fallas. Este método no requiere de la intervención de los usuarios finales.

## 2.4.3 Limpieza de datos

La limpieza de datos consiste en identificar y eliminar inconsistencias, discrepancias y errores en los datos, para garantizar tanto la consistencia como la correctitud de los mismos, mejorando de esta forma la calidad de un sistema de información. Durante el proceso de implementación de un Data Warehouse esta actividad se realiza como parte del proceso de extracción, transformación y carga, en particular durante la etapa de transformación. Previo a realizar esta actividad, se debe realizar una medición de la calidad de los datos a partir del modelo creado para localizar los errores y así poder corregirlos.

## Capítulo 3 – Análisis

A continuación se presenta el contexto en el cual se enmarca el proyecto, así como el análisis del mismo.

### 3.1 Descripción de la realidad

La Unidad de Información Nacional en Salud (UINS) del Ministerio de Salud Pública (MSP) es la encargada de elaborar las estadísticas vitales del país. Debe proveer indicadores establecidos por la Organización Mundial de la Salud (OMS) tales como la tasa de mortalidad infantil, la cantidad de nacimientos por año y la razón de mortalidad materna, así como estadísticas a demanda requeridas por autoridades y organizaciones. A su vez, conforme a las políticas de datos abiertos que han sido impulsadas en el último tiempo, algunos indicadores con sus criterios de análisis preestablecidos son publicados en la página web del Ministerio, permitiendo de esta forma que cualquier usuario pueda consultarlos cuando así lo desee. En la actualidad, en todos los casos mencionados, las estadísticas elaboradas se presentan en cuadros con formato ODS.

La información requerida para realizar estas tareas es obtenida a partir de cuatro sistemas de información que pertenecen al MSP: el Certificado de Nacido Vivo Electrónico (CNVe), que da soporte al proceso de emisión de los certificados que se expiden a todo recién nacido con signos vitales; el Certificado de Defunción Electrónico (CDe), que brinda el soporte necesario para la expedición de los certificados ante todo fallecimiento; el Certificado de Defunción en Papel (CDp), que permite ingresar a un sistema paralelo los certificados de defunción que son emitidos en papel; y el Sistema de Información Perinatal (SIP), que registra información durante el embarazo, parto y puerperio de la madre y también brinda información sobre el recién nacido.

Resulta de interés destacar que existen dos sistemas que gestionan información acerca de las defunciones debido a que por el momento el sistema CDe no se encuentra operativo en todos los centros de salud y estudios forenses. Por esta razón, para aproximadamente 1.000 de los 33.000 fallecimientos anuales, el certificado se emite directamente desde el sistema CDe mientras que el resto aún se emite en papel y es luego ingresado al sistema correspondiente por personal del MSP.

Debido a la falta de validación en el ingreso de los datos, la poca legibilidad y errores de digitación en algunos certificados de defunción en papel, y las dificultades que se les presentan a los usuarios para ingresar códigos en el formato estándar, entre otros, las bases de datos correspondientes a los sistemas CNVe, CDe y CDp presentan problemas de calidad referentes a la correctitud, completitud y consistencia. Por esta razón, las mismas no son utilizadas directamente en la generación de las estadísticas, sino que se utiliza una base de datos mejorada, cuya estructura se detallará en la sección Descripción de las fuentes, que resulta de aplicar un procedimiento de limpieza y unificación de las bases de datos fuente. Esta fue creada como parte de un proceso de mejora de calidad

puesto en práctica en el marco de un convenio entre el MSP y la Fundación Julio Ricaldoni en el año 2012 [9].

El equipo encargado de generar los cuadros estadísticos mencionados anteriormente está compuesto por personal del área de computación y personal del área estadística y social. La primera instancia de este proceso es llevada a cabo por el área de computación y comprende la extracción de todos los datos relevantes a partir de la base de datos mejorada y la base de datos del SIP. Dicha extracción se realiza mediante scripts SQL y el resultado final de la misma es volcado en planillas. Luego, el personal del área estadística y social lleva a cabo un análisis de los datos obtenidos, realizando cruzamientos entre los datos de la base mejorada y el SIP para corregir inconsistencias, completar datos faltantes y eliminar datos erróneos. Una vez finalizada la etapa de correcciones y mejoras de los datos, se procede a elaborar manualmente los cuadros con la información y los criterios de análisis requeridos.

Tal como se puede apreciar, el proceso de elaboración de los cuadros estadísticos requiere de varias etapas complejas que se realizan manualmente pero eventualmente podrían ser automatizadas. De ahí se desprende la necesidad de contar con un sistema que realice de forma automática la confección de los cuadros. Este sistema debería realizar tareas tales como la extracción e integración de datos provenientes de distintas fuentes, la mejora de la calidad de los mismos, el almacenamiento de los datos en una base de datos adecuada para este tipo de manejo, además de permitir el mantenimiento de dicha base y la confección de los cuadros. Adicionalmente, sería deseable que dicho sistema provea interfaces tanto para personal especializado del MSP, como para cualquier usuario que necesite acceder a los datos abiertos que proporciona el MSP sin necesidad de contar con conocimientos técnicos.

## 3.2 Análisis de requerimientos

A continuación se presentan los requerimientos, tanto funcionales como no funcionales, que fueron establecidos en conjunto con la UINS de acuerdo a la realidad planteada en la sección anterior.

### 3.2.1 Requerimientos funcionales

Diseñar y construir un Data Warehouse que permita obtener el subconjunto de los Indicadores Básicos en Salud (IBS) establecidos por la Organización Panamericana de la Salud (OPS), división regional de la OMS, presentado en el Cuadro 3.1.

ID	Indicador	Construcción del indicador
I1	<p><b>Tasa de mortalidad infantil, por 1.000 nacidos vivos (menores de 1 año)</b></p> <p><i>Cociente entre el número de nacidos vivos que mueren antes de cumplir 1 año de edad en un determinado año y el número de nacidos vivos en el mismo año, expresado por 1.000 nacidos vivos.</i></p>	(Nº de defunciones de 0 a 364 días en el año acaecido / Nº total de nacidos vivos en el año acaecido) x 1.000
I2	<p><b>Número de defunciones infantiles</b></p> <p><i>Número de muertes de niños menores de 1 año de edad en un determinado año.</i></p>	Número total de defunciones de 0 a 364 días en el año acaecido
I3	<p><b>Tasa de mortalidad infantil, por 1.000 nacidos vivos neonatal (de 0 a 27 días)</b></p> <p><i>Cociente entre el número de nacidos vivos que mueren antes de cumplir 28 días de edad en un determinado año y el número de nacidos vivos en el mismo año, expresado por 1.000 nacidos vivos.</i></p>	(Nº de defunciones de 0 a 27 días en el año acaecido / Nº de nacidos vivos en el año acaecido) x 1.000
I4	<p><b>Tasa de mortalidad infantil, por 1.000 nacidos vivos postneonatal (de 28 días a &lt; 1 año)</b></p> <p><i>Cociente entre el número de nacidos vivos que mueren antes de cumplir 1 año de edad pero han vivido 28 días o más en un determinado año y el número de nacidos vivos en el mismo año, expresado por 1.000 nacidos vivos.</i></p>	(Nº de defunciones 28 a 364 días en el año acaecido / Nº de nacidos vivos en el año acaecido) x 1.000
I5	<p><b>Mortalidad de menores de 5 años reportada</b></p> <p><i>Cociente entre número de muertes de niños menores de 5 años de edad y el número de nacidos vivos en el</i></p>	(Nº de defunciones en menores de 5 años en el año acaecido / Nº de nacidos vivos en el año acaecido) x 1.000

	<i>mismo año, expresado por 1.000 nacidos vivos.</i>	
16	<p><b>Razón de mortalidad materna por 100.000 nacidos vivos</b></p> <p><i>Cociente entre el número de muertes maternas en un determinado año y el número de nacidos vivos en el mismo año, expresado por 100.000 nacidos vivos.</i></p>	(Nº de defunciones maternas en el año acaecido / Nº de nacidos vivos en el año acaecido) x 100.000
17	<p><b>Número de defunciones maternas</b></p> <p><i>Número de muertes maternas en un determinado año. Muerte materna se define como la defunción de una mujer mientras está embarazada o dentro de los 42 días siguientes a la terminación de su embarazo.</i></p>	Número de muertes maternas en el año acaecido
18	<p><b>Número de nacidos vivos</b></p> <p><i>Número de nacidos vivos en un determinado año. Nacido vivo es la expulsión o extracción completa de un producto de la concepción del cuerpo de la madre, independientemente de la duración del embarazo, que después de dicha separación respire o de cualquier otra señal de vida.</i></p>	Número de nacidos vivos en el año acaecido
19	<p><b>Proporción de bajo peso al nacer</b></p> <p><i>Cociente entre el número de recién nacidos vivos con peso inferior a 2.500 gramos, medido al momento de nacimiento o dentro de las primeras horas de vida, para un año dado y el número de nacidos vivos en el mismo año, expresado por 100 nacidos vivos.</i></p>	(Nº de nacidos vivos con peso < 2500g en el año acaecido / Nº de nacidos vivos en el año acaecido) x 100

Cuadro 3.1 – Indicadores [10]

Adicionalmente, debe ser posible confeccionar los cuadros estadísticos publicados en la sección *Estadísticas* de la página web del MSP [11], que contienen los indicadores mencionados analizados según diferentes criterios. En base a los criterios establecidos en dichos cuadros y los indicadores presentados anteriormente, se definieron los requerimientos de análisis descritos en el Cuadro 3.2

ID	Requerimiento	Indicadores
RF1	Cantidad nacidos vivos por tiempo, grupo etario gestacional, sexo, peso, edad de la madre, nivel educativo de la madre, estado conyugal de la madre, cantidad de consultas prenatales, tipo de parto, lugar de parto, asistencia del parto, zona geográfica de ocurrencia y residencia materna (localidad, departamento, Montevideo/Interior, país, tipo zona - rural/urbano) y lugar de nacimiento (establecimiento, institución).	18
RF2	<p>Proporción por peso al nacer, definida como:</p> $(Cantidad\ de\ nacidos\ vivos\ con\ determinado\ peso / Cantidad\ de\ nacidos) \times 100$ <p>Debe ser posible analizar dicha proporción por tiempo, grupo etario gestacional, sexo, peso, edad de la madre, nivel educativo de la madre, estado conyugal de la madre, cantidad de consultas prenatales, tipo de parto, lugar de parto, asistencia del parto, zona geográfica de ocurrencia y residencia materna (localidad, departamento, Montevideo/Interior, país, tipo zona - rural/urbano) y lugar de nacimiento (establecimiento, institución).</p>	19
RF3	Cantidad de defunciones por tiempo, zona geográfica de ocurrencia y residencia (localidad, departamento, Montevideo/Interior, país, tipo zona - rural/urbano), causas (categoría, grupo, capítulo), grupo etario y sexo. Además, interesa analizar la cantidad de defunciones maternas (durante el embarazo, parto o puerperio).	12, 17
RF4	<p>Tasa de mortalidad infantil, definida como:</p> $(Cantidad\ de\ defunciones\ de\ menores\ de\ 5\ años / Cantidad\ de\ nacidos\ vivos) \times 1000$ <p>Debe ser posible analizar dicha tasa por tiempo, zona geográfica de ocurrencia y residencia materna (localidad, departamento, Montevideo/Interior, país, tipo zona - rural/urbano), grupo etario, sexo y</p>	11, 13, 14, 15

	causa de defunción.	
RF5	<p>Razón de mortalidad materna, definida como:</p> $(Cantidad\ de\ defunciones\ maternas / Cantidad\ de\ nacidos\ vivos) \times 100000$ <p>Debe ser posible analizar dicha razón por tiempo, zona geográfica de ocurrencia y residencia (localidad, departamento, Montevideo/Interior, país, tipo zona - rural/urbano), grupo etario y causa de defunción.</p>	16

*Cuadro 3.2 - Requerimientos funcionales*

### 3.2.2 Requerimientos no funcionales

En el Cuadro 3.3 se especifican los requerimientos no funcionales establecidos.

ID	Requerimientos no funcionales
RNF1	Implementar la solución utilizando la suite de Pentaho Community Edition sobre MySQL.
RNF2	Implementar un programa para la realización de la carga y actualización del Data Warehouse que resulte amigable y no requiera de amplio conocimiento técnico.
RNF3	Desarrollar una interfaz gráfica sencilla, clara e intuitiva para llevar a cabo consultas y estudios estadísticos internos del MSP.
RNF4	Proveer documentación detallada del sistema y su funcionamiento para su posterior mantenimiento por parte del MSP.

*Cuadro 3.3 - Requerimientos no funcionales*

Adicionalmente, si bien no fue un requerimiento solicitado por el MSP, se considera necesario realizar un análisis de calidad junto con un proceso de limpieza y mejora de los datos, ya que se entiende que le dará un gran valor agregado al producto final.

### 3.3 Descripción de las fuentes

Luego de firmar un acuerdo de confidencialidad sobre los datos, fueron provistas por el MSP la base mejorada, que como se mencionó anteriormente contiene datos de los sistemas CNVe, CDe y CDp, y la base de datos del SIP. No obstante no fue posible utilizar esta última, que era necesaria tanto para la confección de algunos indicadores, como para la realización de chequeos de calidad de los datos correspondientes al CNVe en la base mejorada. Esto se debió a que las fuentes proporcionadas contenían un valor fijo en el atributo cédula del recién nacido en todas las tuplas, lo cual hacía imposible trabajar con ambas bases en conjunto ya que las mismas debían relacionarse mediante dicho atributo, y por otra parte, los contenidos de muchas de sus variables eran en su mayoría nulos o vacíos. Por otro lado, la nomenclatura utilizada era poco descriptiva y no se contaba con documentación de la misma. Por estas razones algunos indicadores de interés para el MSP no pudieron ser incluidos en los requerimientos funcionales.

Por consiguiente, se trabajó únicamente con la base mejorada, la cual permitió elaborar todos los indicadores presentados en la sección anterior. Los datos provistos en dicha base corresponden a los nacimientos ocurridos entre enero de 2013 y setiembre de 2015, y a las defunciones acaecidas entre enero de 2012 y setiembre de 2015. Con respecto al esquema de la misma, es posible agrupar las tablas en tres conjuntos de acuerdo a su contenido; por un lado las tablas que contienen datos de los Certificados de Nacidos Vivos (CNV), por otro las que tienen datos de los Certificados de Defunción (CD) y finalmente las codigueras. Si bien estas últimas son utilizadas para hacer referencia a datos tanto de CNV como de CD, se agrupan en un conjunto distinto con el fin de simplificar y dar más claridad a los diagramas presentados a continuación.

La Figura 3.1 presenta el esquema lógico de las tablas que contienen datos relacionados con los CNV. Tal como se puede observar, existe una tabla principal (*certificados\_nacidos\_vivos*) que almacena los datos necesarios para identificar a un certificado, y relacionadas con esta se encuentran tablas que contienen los datos del recién nacido, la madre, el embarazo, el parto, el padre y el certificador que realizó el CNV.

Por otra parte, en la Figura 3.2 se puede observar el esquema lógico de las tablas referentes a los CD. De forma análoga a los CNV, existe una tabla principal (*certificados\_defuncion*) que almacena los datos básicos de identificación de un CD, y vinculadas a esta se encuentran tablas que tienen los datos del fallecido y la defunción. Además, existen otras tablas relacionadas con la tabla principal que se utilizan en los casos en que el fallecido es menor para almacenar datos, por ejemplo, de la madre y el padre.

Finalmente, en la Figura 3.3 se muestran las codigueras referenciadas por las tablas tanto de CNV como de CD. Estas almacenan datos referentes a la ubicación geográfica (domicilios, calles, localidades, departamentos, entre otros), centros médicos y sus dependencias (establecimientos e instituciones), información referente a los partos (tipo de lugar en que ocurrió el parto, tipo de asistencia del parto), estados civiles y nivel de estudios, entre otros.

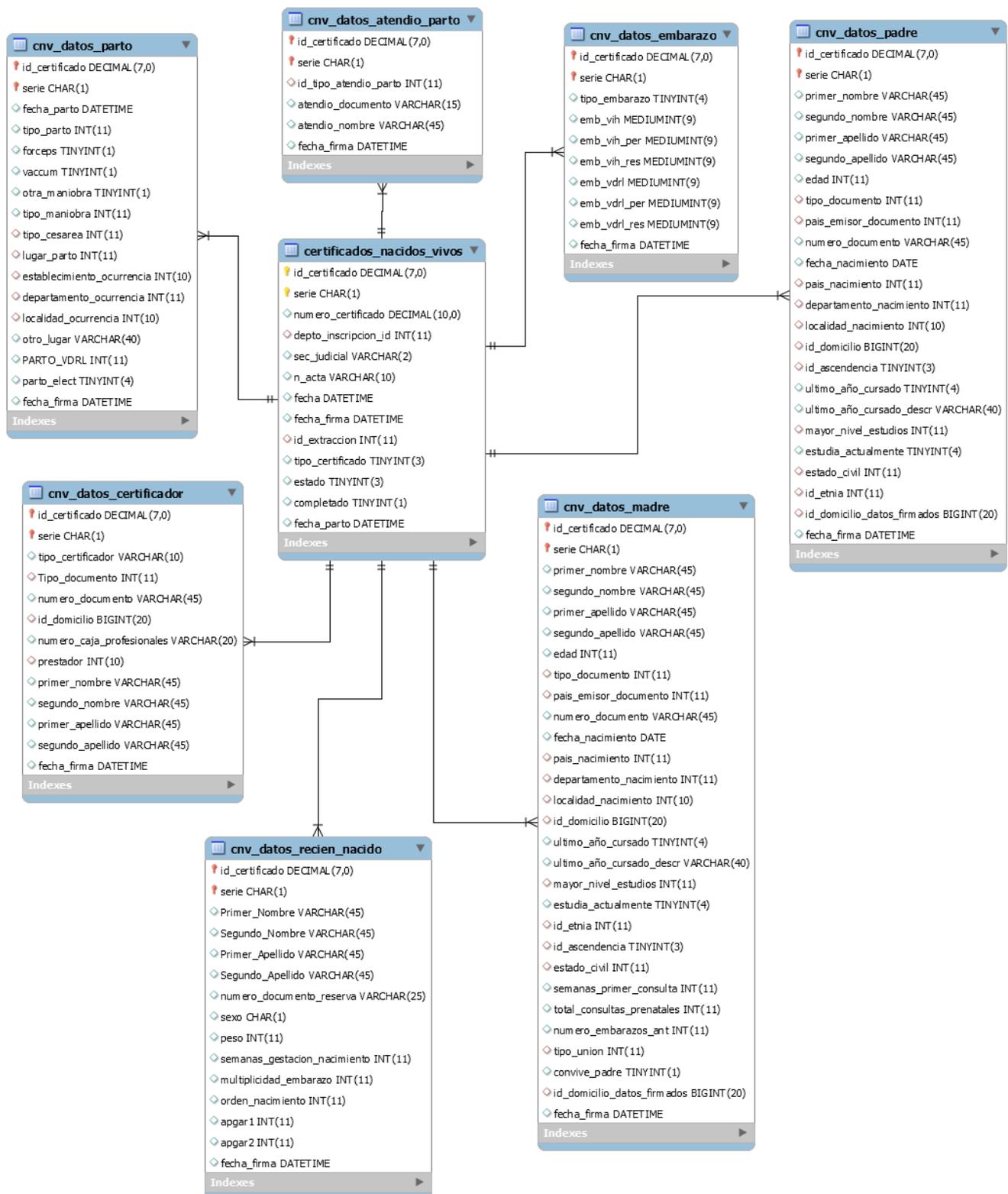


Figura 3.1 - Base mejorada - Esquema lógico CNV

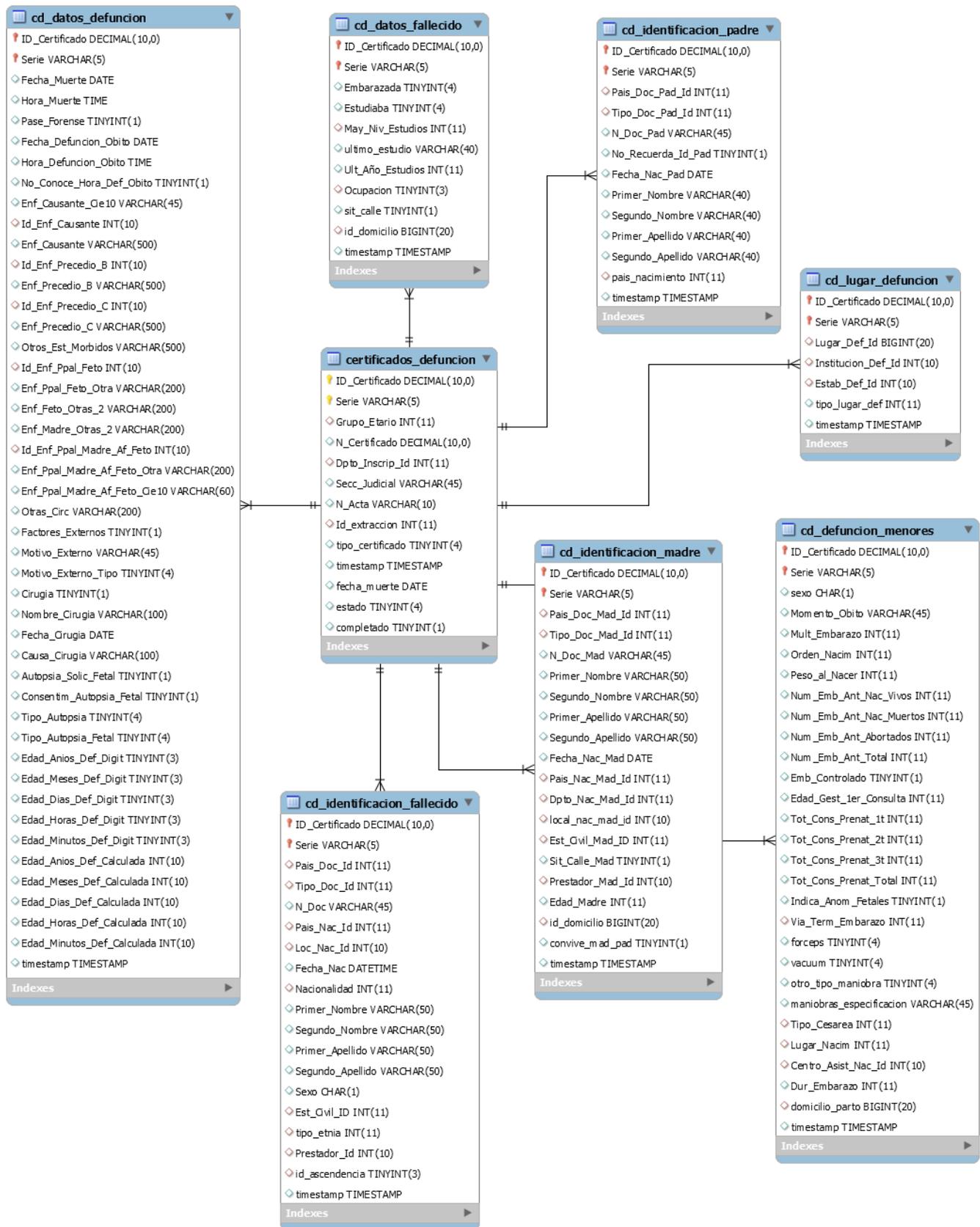


Figura 3.2 - Base mejorada - Esquema lógico CD

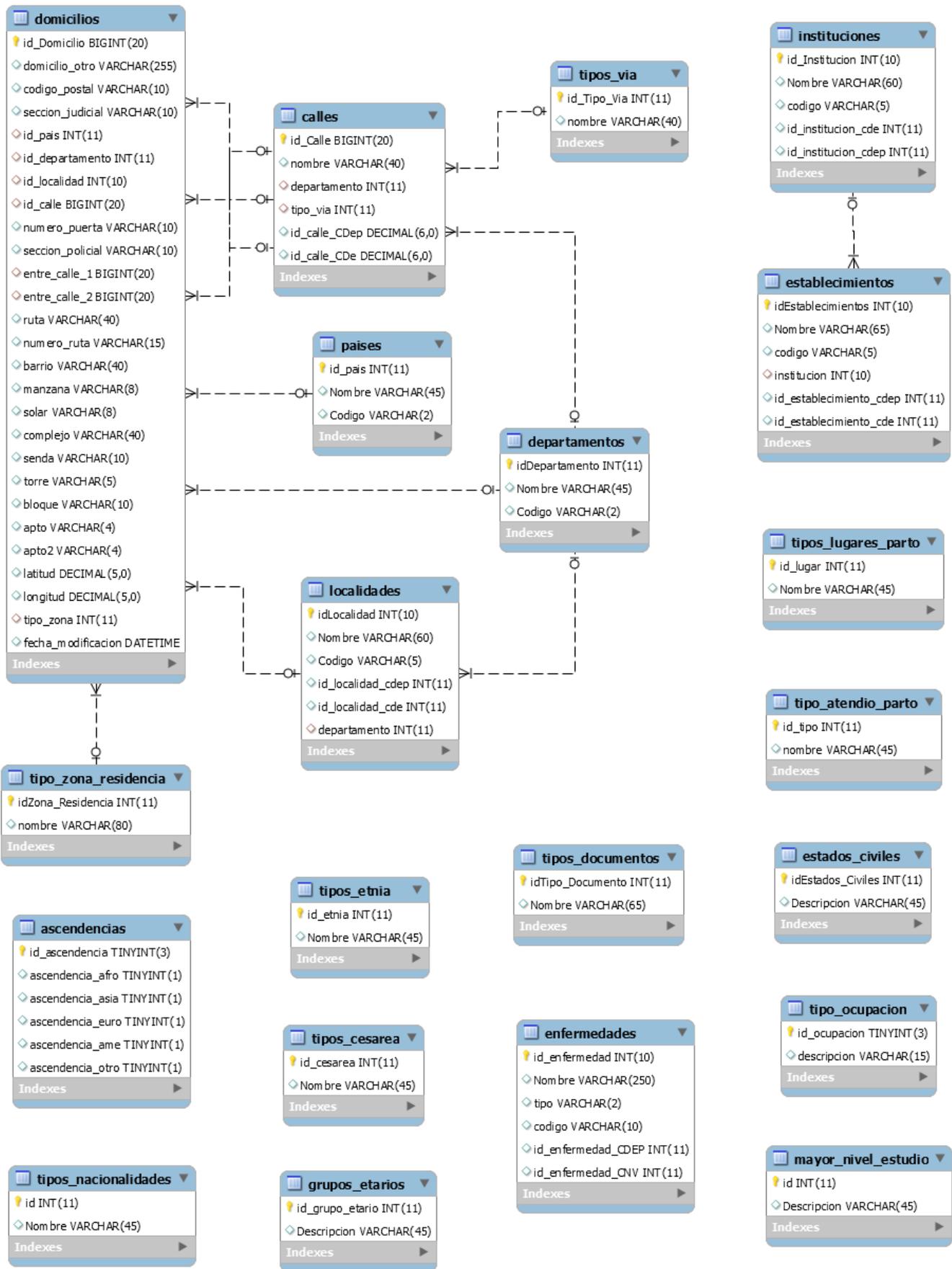


Figura 3.1 - Base mejorada - Codigueras utilizadas

### 3.3.1 Problemas encontrados

Además de la principal dificultad ya mencionada de no poder utilizar la base de datos del SIP, surgieron otras problemáticas al momento de trabajar con las fuentes provistas. Una de ellas fue la existencia de atributos codificados dentro de la base mejorada que no tenían ninguna codiguera asociada. La solución puesta en práctica, en la mayoría de estos casos, por el equipo del área informática de la UINS es asignar las descripciones de dichos códigos al momento de realizar la extracción de los datos para volcarlos en las planillas, mediante scripts SQL. Fue necesario tener esto en cuenta durante la implementación del proceso de carga y actualización del Data Warehouse, ya que dichas descripciones debieron ser obtenidas de forma manual examinando el script. Esto ocurre, por ejemplo, con el atributo *tipo\_parto* de la tabla *cnv\_datos\_parto*.

Sin embargo, en el caso del atributo *Enf\_Causante\_CIE10* de la tabla *cd\_datos\_defuncion*, para el cual no existe la codiguera correspondiente, no se realiza ningún mapeo en el script y no fue posible obtener la documentación pertinente para realizarlo. Este atributo hace referencia a la causa de muerte codificada según la *Clasificación Internacional de Enfermedades*, décima versión (CIE-10) [12] de la OMS. Al ser una representación estándar, fue posible obtener de la página web del Instituto Médico Forense del Gobierno Argentino [13] un archivo Excel que contiene las descripciones y categorizaciones de las causas. El mismo fue utilizado como base durante el proceso de carga y actualización y fue corroborado manualmente con la información provista en la página oficial de la clasificación [12].



## Capítulo 4 – Diseño

En el presente capítulo se detalla el diseño realizado en base al análisis de la realidad presentada.

### 4.1 Diseño conceptual

A continuación se presenta el diseño conceptual basado en los requerimientos previamente expuestos. Se decidió utilizar este enfoque ya que los requerimientos eran el universo de información, al no contarse en un principio con las bases de datos fuente, además de ser claros y concisos.

Adicionalmente, se incluye la representación gráfica de los distintos componentes del diseño conceptual utilizando el modelo CMDM descrito en el Capítulo 2.

#### 4.1.1 Dimensiones y jerarquías

Para modelar la realidad del problema fue necesario construir las siguientes dimensiones:

- **Tiempo:** esta dimensión, ilustrada en la Figura 4.1, está compuesta por una única jerarquía que se diseñó de forma tal que fuera posible agregar los indicadores por mes y por año.

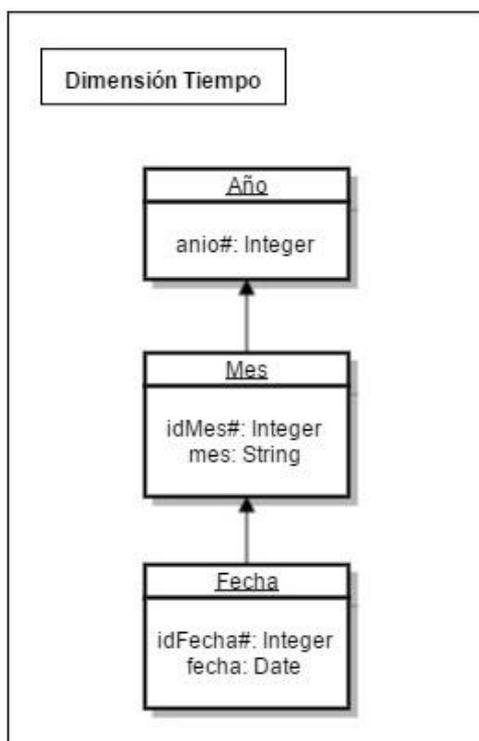


Figura 4.1 - Dimensión Tiempo

- **Edades:** representa la edad en años. Se utiliza para analizar datos tanto de nacimientos como de fallecimientos. En el primer caso, representa la edad de la madre del nacido vivo, mientras que en el segundo, representa la edad del fallecido. Por este motivo es que contiene dos jerarquías, que se pueden observar en la Figura 4.2. Una de ellas, GrupoEtario, es la más genérica y diseñada principalmente con el propósito de agregar edades de fallecidos, ya que agrupa las edades de 0 a 110 en rangos de 5 años. Por otra parte, GrupoEtarioMaterno fue diseñada más específicamente para la agregación de datos por edad materna ya que agrupa las edades que van de 14 a 44 en rangos de 5 años mientras que las demás edades simplemente las clasifica en menores de 14 o mayores de 44 años.

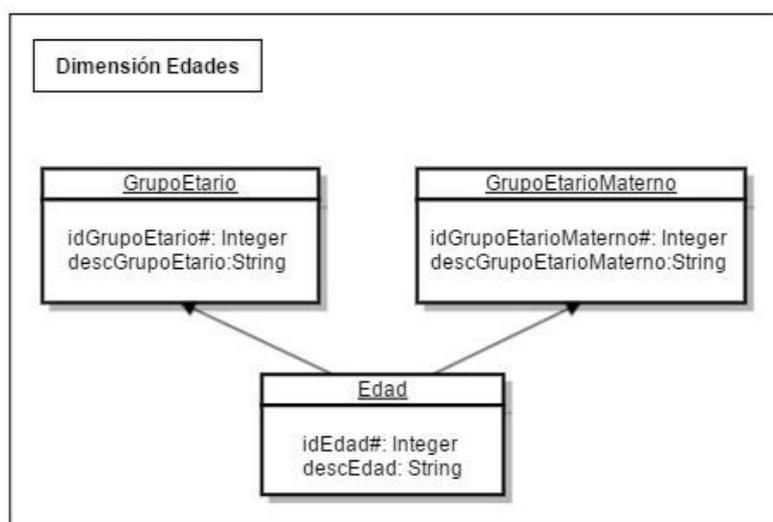


Figura 4.2 - Dimensión Edades

- **Sexo:** representa el sexo tanto de los nacidos vivos como de los fallecidos. Puede tomar los siguientes valores: Femenino, Masculino, Indefinido, Desconocido o Erróneo. Su representación gráfica se muestra en la Figura 4.3.

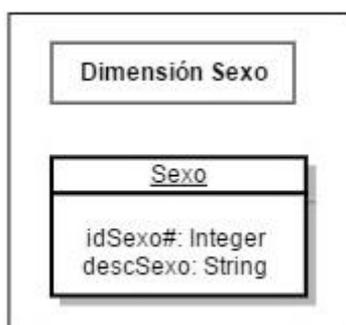


Figura 4.3 - Dimensión Sexo

- **Domicilios y Localidades:** estas dimensiones modelan las zonas geográficas de ocurrencia tanto de los nacimientos como los fallecimientos, y de residencia de las madres y los fallecidos. Originalmente se diseñó una sola dimensión llamada ubicaciones geográficas pero luego, al estudiar las fuentes, surgió la necesidad de dividirla en dos ya que para los nacimientos no se contaba con el domicilio de ocurrencia, sino que solamente con la localidad de ocurrencia. Como se puede observar en la Figura 4.4, la dimensión Domicilios contiene dos jerarquías que permiten agregar los datos por tipo de zona (rural, urbana, suburbana, etc) y por país. Por otra parte, la dimensión Localidades permite agrupar por departamento y por Montevideo / Interior.

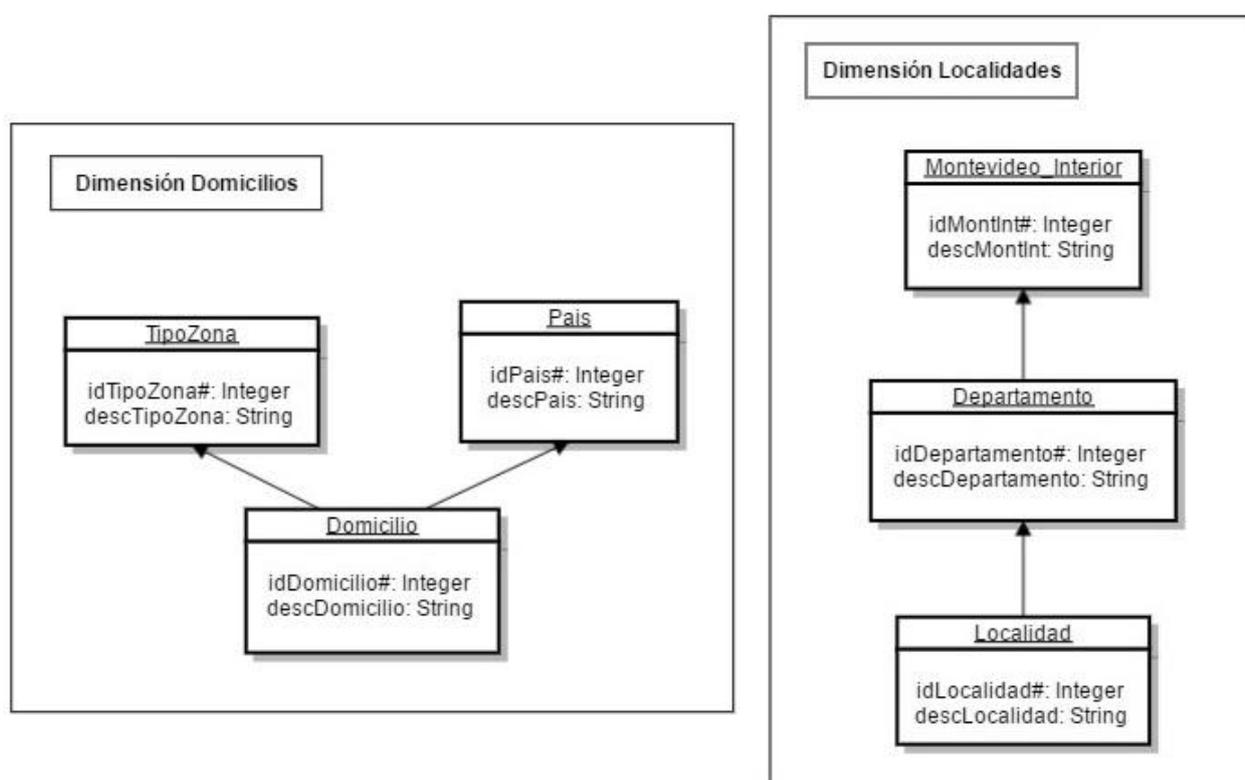


Figura 4.4 - Dimensiones Domicilios y Localidades

- **CausasCIE10:** representa las enfermedades causantes de los fallecimientos según la *Clasificación Internacional de Enfermedades* (CIE-10). Dentro de dicha clasificación existen diferentes niveles de agrupación de las enfermedades (Categoría, Grupo y Capítulo) que fueron agregados a la jerarquía, tal como se muestra en la Figura 4.5.

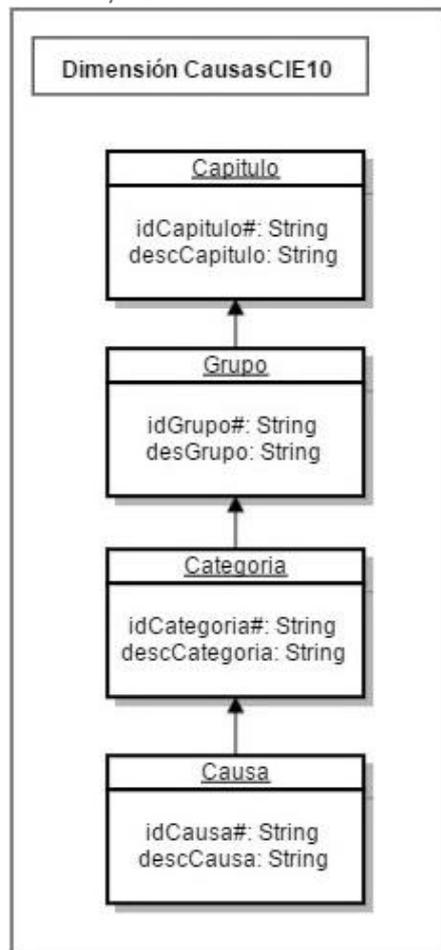


Figura 4.5 - Dimensión CausasCIE10

- **Establecimientos:** representa los establecimientos donde ocurren los nacimientos. Contiene una jerarquía que permite agrupar los establecimientos por institución. La misma se puede observar en la Figura 4.6.

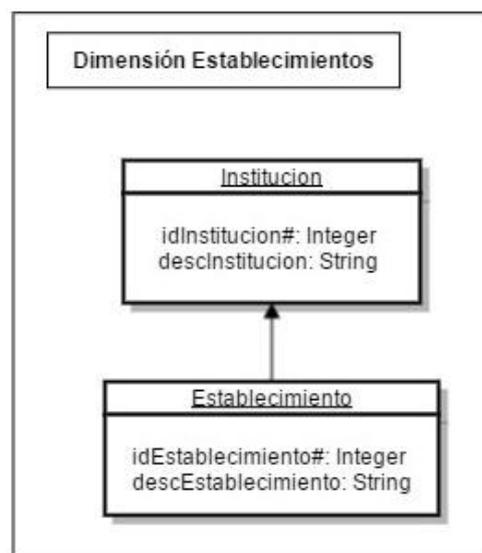


Figura 4.6 - Dimensión Establecimientos

- **Nacidos Vivos:** esta dimensión, ilustrada en la Figura 4.7, representa a los recién nacidos vivos. Contiene dos jerarquías que permiten analizar los nacimientos según la edad gestacional en semanas y el peso al nacer en gramos del nacido vivo. En cuanto a la jerarquía GrupoEtario, la misma contiene dos niveles: EdadGestacional, que permite analizar los datos por semanas de gestación del nacido vivo; y GrupoEtarioGestacional que permite agrupar los nacimientos por edad gestacional de la siguiente manera:
  - Entre 20 y 31
  - Entre 32 y 33
  - Entre 34 y 35
  - 36
  - 37
  - 38
  - 39
  - 40
  - 41
  - Más de 41

Por otra parte, la jerarquía Peso está compuesta por tres niveles que permiten realizar un estudio de acuerdo a los siguientes criterios: peso, rango de peso y categoría de peso al nacer. En cuanto a los rangos de peso, pueden tomar los siguientes valores:

- Menos de 1000
- De 1000 a 1499
- De 1500 a 1999
- De 2000 a 2499
- De 2500 a 2999
- De 3000 a 3499
- De 3500 a 3999
- Más de 4000

Asimismo, las categorías de peso son las siguientes:

- Extremado bajo peso: menos de 1000gr
- Bajo peso: entre 1000 y 2499gr
- Normal: 2500gr o más

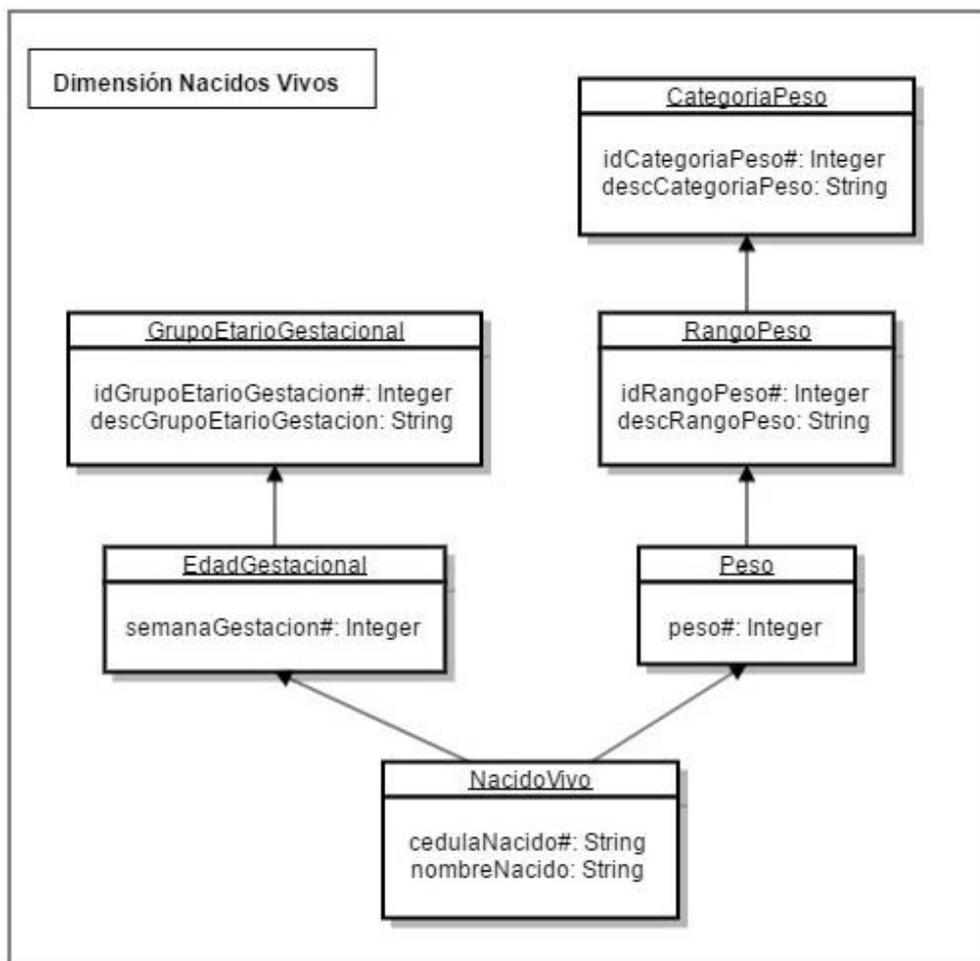


Figura 4.7 - Dimensión Nacidos Vivos

- **Madres:** esta dimensión modela a las madres de los nacimientos. Cada nacimiento está relacionado con una única instancia de madre; es decir, es la madre de ese nacimiento y no la madre como entidad independiente. Si una misma mujer fuera madre de más de un nacido vivo, existirían tantas instancias de madre como nacimientos. A su vez, esta dimensión incluye los datos del parto de la madre, ya que tanto parto como madre son instancias relacionadas al nacimiento y no entidades por sí solas.

Como se muestra en la Figura 4.8, esta dimensión contiene varias jerarquías que permiten analizar la información de los nacimientos de acuerdo a los siguientes criterios: tipo de parto, asistencia a la madre durante el parto, lugar del parto, cantidad de consultas prenatales a las que asistió la madre, nivel educativo y estado conyugal de la madre.

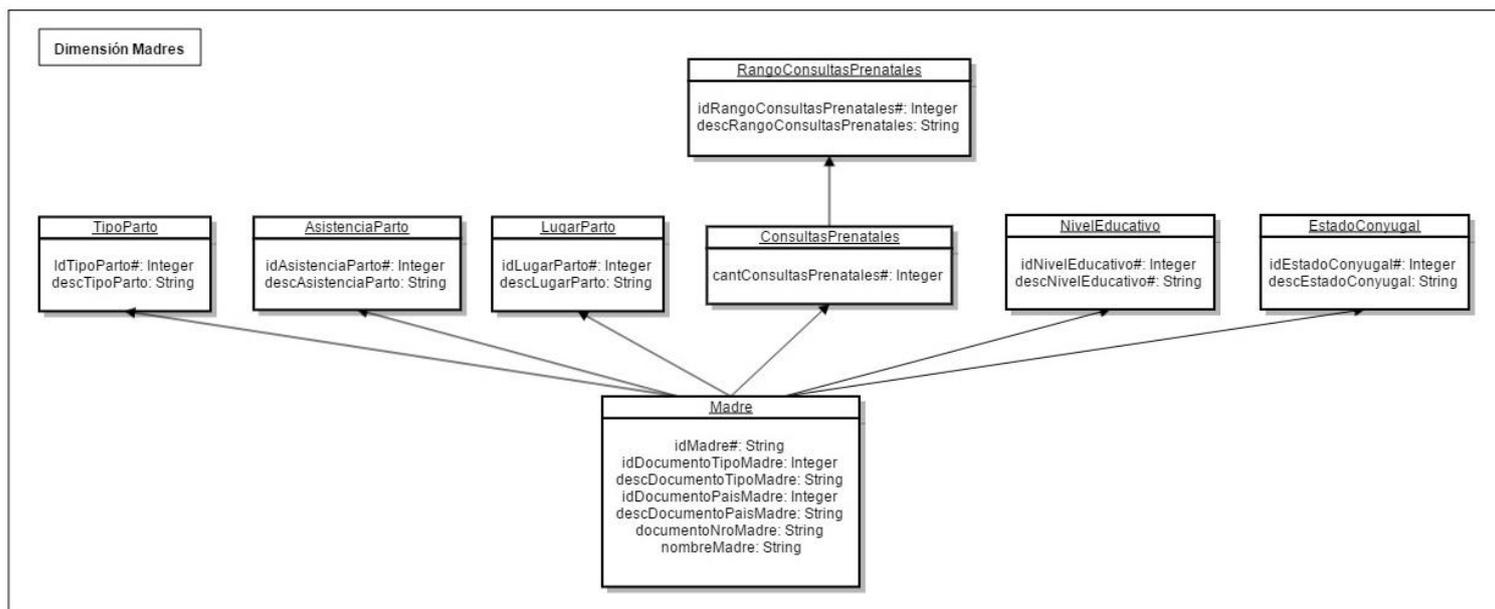


Figura 4.8 - Dimensión Madres

- **Fallecidos**: esta dimensión, que representa a las personas fallecidas, contiene una jerarquía que permite realizar análisis relacionados con la mortalidad materna. Dicha jerarquía, que se puede observar en la Figura 4.9, contiene un nivel llamado Embarazada que puede tomar los siguientes valores: Sí (indica que la persona estaba embarazada o en parto o en puerperio), No, No Corresponde, No Aplica o Se Ignora.

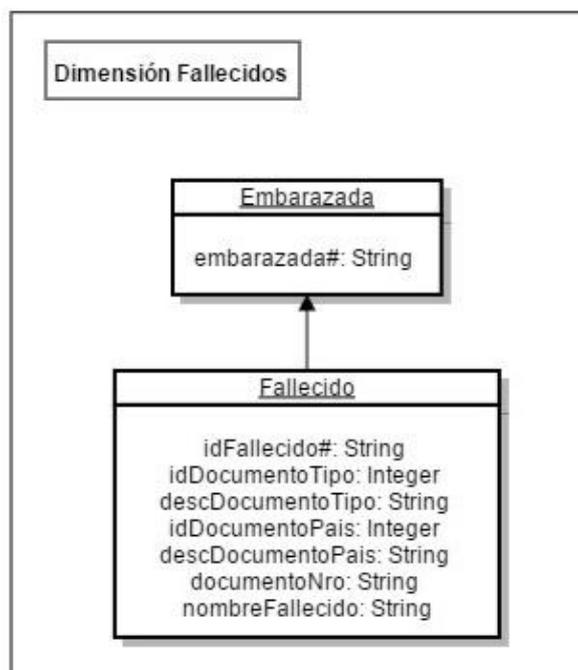


Figura 4.9 - Dimensión Fallecidos

- **Fallecidos Menores:** esta dimensión, cuya representación se encuentra en la Figura 4.10, modela los fallecidos menores de 5 años y permite realizar un análisis por edad del fallecido. Está compuesta por una sola jerarquía, que contiene los niveles EdadMenor, que representa la edad en días en el rango que va de 0 a 1.826 días (equivalente a 5 años), y GrupoEtarioMenor. Este último permite realizar agrupaciones por edad del fallecido menor de la siguiente manera:

- Neonatal precoz: de 0 a 7 días
- Neonatal tardía: de 8 a 27 días
- Postneonatal: de 28 a 364 días
- 1 año o más

Se decidió definir esta dimensión, además de la anteriormente presentada Fallecidos, ya que es de interés realizar análisis enfocados en los menores de 5 años (que representan una porción pequeña del universo de fallecidos), por lo que resulta conveniente modelarlos por separado.

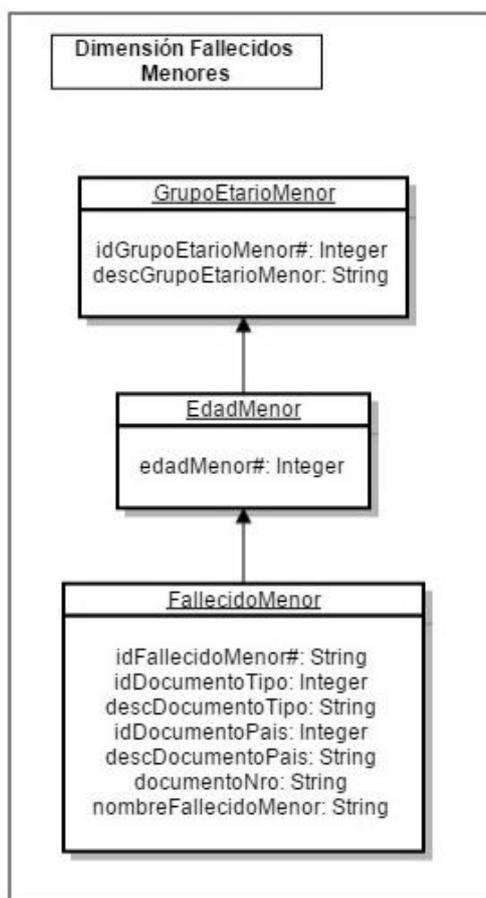


Figura 4.10 - Dimensión Fallecidos Menores

## 4.1.2 Medidas

En la siguiente sección se presentan las medidas definidas para el análisis y su relación con los indicadores presentados en los requerimientos.

### *Nacimientos*

Como se puede observar en la Figura 4.11, en relación a los Nacimientos, es posible analizar tanto la cantidad de nacimientos como la proporción por peso. Esta última cumple con la definición planteada en el requerimiento RF2, y por lo tanto, resulta en una generalización del indicador *19 Proporción de bajo peso al nacer*. Esto se debe a que no solo permite calcular la proporción de bajo peso sino que permite calcularla para cualquier nivel de la jerarquía Peso de la dimensión Nacidos Vivos, así como para cualquier nivel de la dimensión Tiempo, sin restringirse al nivel Año.

Por otra parte, la medida cantidad de nacimientos puede considerarse una generalización del indicador *18 Número de nacidos vivos*, puesto que también puede ser analizada por cualquier nivel de la dimensión Tiempo.

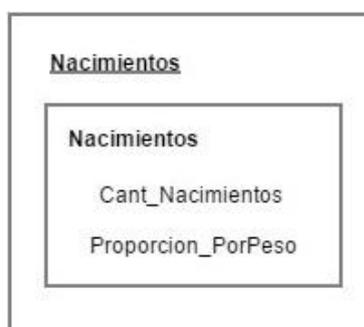


Figura 4.11 - Medidas - Nacimientos

### *Mortalidad*

En relación a la Mortalidad, tal como se muestra observar en la Figura 4.12, es posible analizar tanto la cantidad de fallecimientos como la razón de mortalidad materna. La medida cantidad de fallecimientos puede considerarse una generalización del indicador *17 Número de defunciones maternas*, ya que permite analizar todas las defunciones por cualquier nivel de la dimensión Tiempo, sin limitarse a las defunciones maternas en determinado año.

En cuanto a la razón de mortalidad materna, esta cumple con la especificación del requerimiento RF5, resultando en una generalización del indicador *16 Razón de mortalidad materna por 100.000 nacidos vivos*, debido a que puede ser analizada por cualquier nivel de la dimensión Tiempo. Cabe mencionar que para calcularla es necesario contar tanto con la medida Cant\_Fallecimientos como con la medida Cant\_Nacimientos presentada anteriormente.



Figura 4.12 - Medidas - Mortalidad

### ***Mortalidad Infantil***

Con respecto a la Mortalidad Infantil, tal como se puede observar en la Figura 4.13, es posible analizar la cantidad de fallecimientos de menores de 5 años, así como la tasa de mortalidad infantil. La primera puede considerarse una generalización del indicador *I2 Número de defunciones infantiles*, ya que permite analizar todas las defunciones de menores de 5 años por cualquier nivel de la dimensión Tiempo, sin restringirse a las defunciones de menores de 1 año en determinado año.

Por otro lado, la medida tasa de mortalidad infantil, definida de acuerdo al requerimiento RF4, puede ser calculada para cualquier nivel de la jerarquía GrupoEtarioMenor de la dimensión Fallecidos Menores y para cualquier nivel de la dimensión Tiempo. Por este motivo, la misma resulta en una generalización de los indicadores referentes a la *Tasa de mortalidad infantil, por 1.000 nacidos vivos* (indicadores I1, I3 a I5). Es importante señalar que para calcularla es necesario contar con las medidas **Cant\_Fallecimientos\_Menores** y **Cant\_Nacimientos** conjuntamente.



Figura 4.13 - Medidas - Mortalidad Infantil

### 4.1.3 Relaciones dimensionales

En las Figuras 4.14, 4.15 y 4.16 se presentan las relaciones dimensionales a partir de las cuales se confeccionaron los cubos para el análisis.

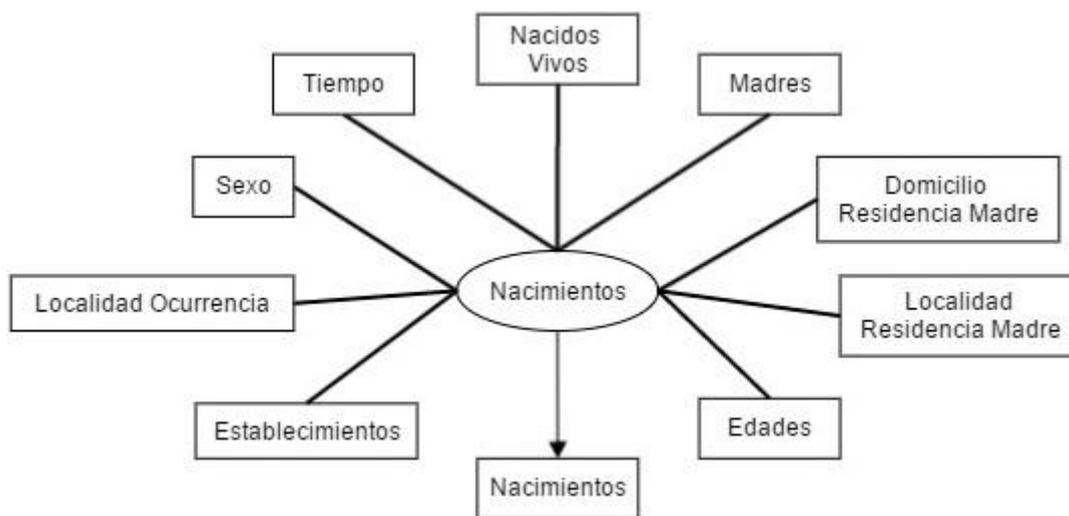


Figura 4.14 - Relación Dimensional Nacimientos

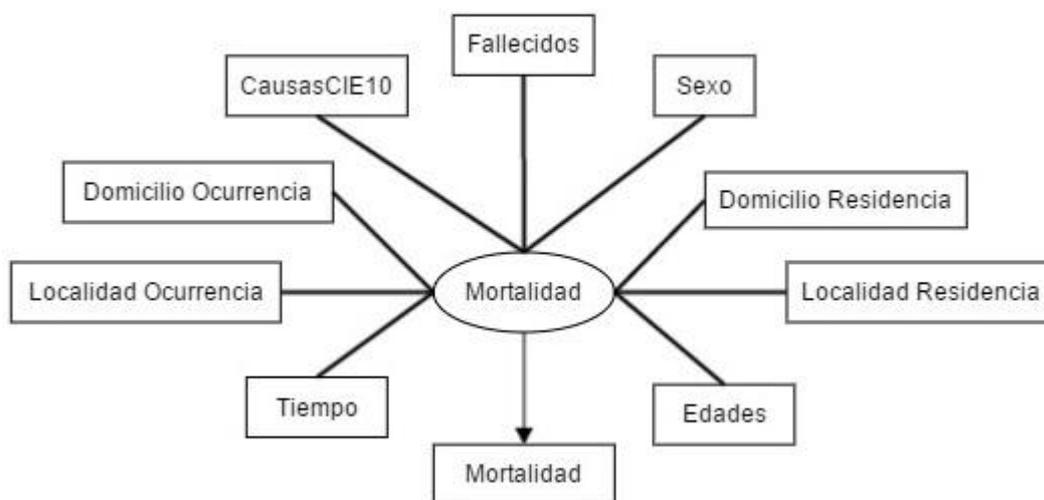


Figura 4.15 - Relación Dimensional Mortalidad

Cabe destacar que para obtener la medida *Razon\_MortalidadMaterna* es necesario realizar un *drill-across* entre las relaciones dimensionales *Mortalidad* y *Nacimientos*, ya que para calcularla es necesario contar tanto con la medida *Cant\_Fallecimientos* de la relación *Mortalidad*, como con la medida *Cant\_Nacimientos* de la relación *Nacimientos*.

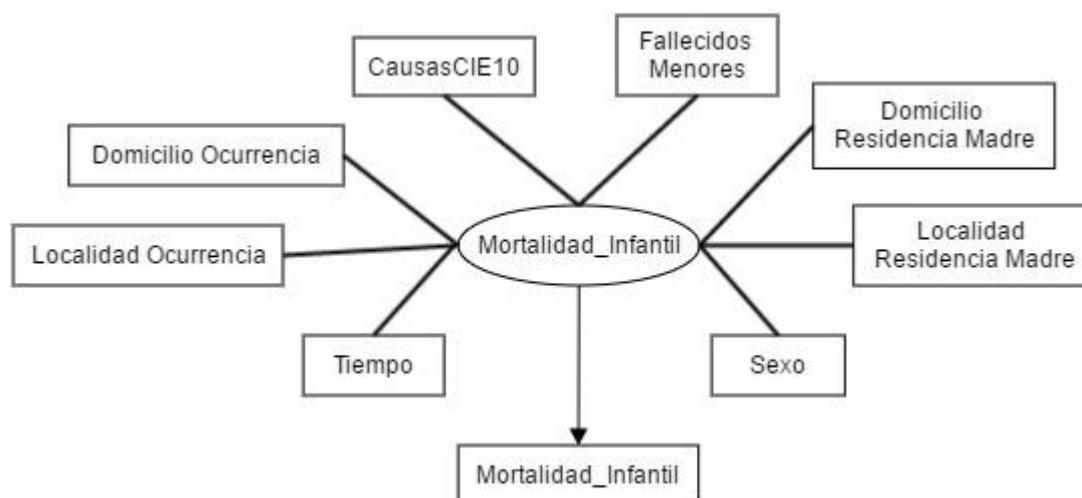


Figura 4.16 - Relación Dimensional Mortalidad Infantil

De forma análoga al caso de la medida *Razon\_MortalidadMaterna*, para calcular la medida *Tasa\_MortalidadInfantil* es necesario realizar un *drill-across* entre las relaciones dimensionales *Mortalidad\_Infantil* y *Nacimientos*, ya que para calcularla es necesario contar tanto con la medida *Cant\_Fallecimientos\_Menores* de la relación *Mortalidad\_Infantil*, como con la medida *Cant\_Nacimientos* de la relación *Nacimientos*.

#### 4.1.4 Estudio de aditividad de medidas

A continuación se presenta para cada relación dimensional el cuadro *roll-up* de las medidas que se analizan en cada una de ellas. Este cuadro muestra para cada operación *roll-up* qué función de agregación se debe utilizar. Se estructura de forma tal que en la segunda columna se encuentran todas las operaciones *drill-up* posibles dentro de las dimensiones vinculadas a la relación, señaladas en la primera columna. Luego, existe una columna por cada medida que se desea estudiar y en ella se indica, para cada operación de la segunda columna, la función de agregación que se debe utilizar. El símbolo **N/A** (no aplica) indica que no es posible realizar la operación *roll-up*, mientras que el símbolo **+** representa la suma y el símbolo **t** la siguiente operación de agregación que aplica a las medidas tasa, razón y proporción:

*El numerador y el denominador se agregan utilizando la operación suma. Luego se realiza el cociente entre ellos y la multiplicación por el amplificador correspondiente. Dicho amplificador para la realidad planteada indica por cuantos nacidos vivos es el cálculo. Por ejemplo, en el caso del indicador I1, el amplificador es 1.000.*

Por otra parte, el símbolo **t2** representa una operación de agregación que también aplica a las medidas tasa, razón y proporción, cuyo cálculo es el siguiente:

*El numerador se agrega utilizando la operación suma, mientras que el denominador permanece incambiado. Luego se realiza el cociente entre ellos y la multiplicación por el amplificador correspondiente.*

Este último cálculo se utiliza cuando la definición de la medida requiere analizar únicamente el numerador por el criterio de análisis seleccionado, tal como ocurre con *Proporcion\_PorPeso*. Por otra parte, dicho cálculo también resulta de utilidad en los casos en que las medidas se obtienen realizando operaciones *drill-across* entre dos relaciones dimensionales, y la dimensión por la cual se quiere realizar el análisis no es compartida entre ambas relaciones. Por ejemplo, cuando se desea analizar la medida *Tasa\_MortalidadInfantil*, cuyo cálculo requiere de una operación *drill-across* entre las relaciones dimensionales *Mortalidad\_Infantil* y *Nacimientos*, según la dimensión *CausasCIE10* que se encuentra únicamente en la relación *Mortalidad\_Infantil*.

La definición de cada uno de los cuadros *roll-up* para las relaciones dimensionales *Nacimientos*, *Mortalidad* y *Mortalidad Infantil* se pueden encontrar en los Cuadros 4.1, 4.2 y 4.3 respectivamente.

Como se puede observar en el Cuadro 4.3, correspondiente a la relación dimensional *Mortalidad\_Infantil*, la primera operación que se desea estudiar es el *drill-up* desde el nivel *Fecha* al nivel *Mes* de la dimensión *Tiempo*. Al realizar esta operación, la función de agregación a utilizar para la medida *Cant\_Fallecimientos\_Menores* es la suma, y para la medida *Tasa\_MortalidadInfantil* es el cálculo **t**. Suponiendo que hubo 2000 nacimientos en el mes de abril de 2015, y fallecieron 10 menores de 5 años durante este mes, entonces el valor de la medida *Tasa\_MortalidadInfantil* para el mes de abril de ese año, resultante de aplicar el cálculo **t**, es  $5 = (10/2000*1000)$ .

Por otra parte, en el mismo cuadro, cuando se desea estudiar el *drill-up* desde el nivel *EdadMenor* al nivel *GrupoEtarioMenor* de la dimensión *Fallecidos Menores* la operación de agregación a utilizar para la medida *Tasa\_MortalidadInfantil* es el cálculo **t2**. Suponiendo que hubo 2000 nacimientos en el mes de abril de 2015 y fallecieron 5 niños de 28 a 364 días de edad en dicho mes, entonces el valor de la medida *Tasa\_MortalidadInfantil* para el grupo etario postneonatal en abril de 2015, resultante de aplicar el cálculo **t2**, es  $2,5 = (5/2000*1000)$ .

En cuanto a la aditividad de las medidas, se puede observar que *Cant\_Nacimientos*, *Cant\_Fallecimientos* y *Cant\_Fallecimientos\_Menores* son aditivas ya que la función de agregación en todos los casos es la suma. Por otra parte, las demás medidas son no aditivas ya que en ningún caso es posible aplicar la función de agregación suma y conservar la semántica de las mismas. Además, cabe destacar que al momento de realizar este estudio se verificó que se cumplan todas las condiciones presentadas en el Capítulo 2, que aseguran la correcta agregación de las medidas.

## Nacimientos

Dimensión	Operaciones <i>drill-up</i>	Cant_ Nacimientos	Proporcion_ PorPeso
<b>Tiempo</b>	Fecha -> Mes	+	t
	Mes -> Año	+	t
	Año -> All	+	t
<b>Madres</b>	Madre -> TipoParto	+	t
	Madre -> AsistenciaParto	+	t
	Madre -> LugarParto	+	t
	Madre -> ConsultasPrenatales	+	t
	ConsultasPrenatales -> RangoConsultasPrenatales	+	t
	Madre -> NivelEducativo	+	t
	Madre -> EstadoConyugal	+	t
	TipoParto / AsistenciaParto / LugarParto / RangoConsultasPrenatales / NivelEducativo / EstadoConyugal -> All	+	t
<b>Localidad Ocurrencia</b>	Localidad -> Departamento	+	t
	Departamento -> Montevideo/Interior	+	t
	Montevideo/Interior -> All	+	t
<b>Establecimientos</b>	Establecimiento -> Institucion	+	t
	Institucion -> All	+	t
<b>Nacidos Vivos</b>	NacidoVivo -> EdadGestacional	+	t
	EdadGestacional -> GrupoEtarioGestacional	+	t
	GrupoEtarioGestacional -> All	+	t
	NacidoVivo -> Peso	+	t2
	Peso -> RangoPeso	+	t2
	RangoPeso -> CategoriaPeso	+	t2
	CategoriaPeso -> All	+	t2
<b>Domicilio Residencia Madre</b>	Domicilio -> TipoZona	+	t
	Domicilio -> Pais	+	t
	Pais / TipoZona -> All	+	t

<b>Localidad Residencia Madre</b>	Localidad -> Departamento	+	t
	Departamento -> Montevideo/Interior	+	t
	Montevideo/Interior -> All	+	t
<b>Edades</b>	Edad -> Grupo Etario	+	t
	Edad -> Grupo EtarioMaterno	+	t
	GrupoEtario / GrupoEtarioMaterno -> All	+	t
<b>Sexo</b>	Sexo -> All	+	t

Cuadro 4.1 - Cuadro roll-up Nacimientos

## Mortalidad

Dimensión	Operaciones <i>drill-up</i>	Cant_ Fallecimientos	Razon_Mortalidad Materna
<b>Tiempo</b>	Fecha -> Mes	+	t
	Mes -> Año	+	t
	Año -> All	+	t
<b>CausasCIE10</b>	Causa -> Categoría	+	t2
	Categoría -> Grupo	+	t2
	Grupo -> Capítulo	+	t2
	Capítulo -> All	+	t2
<b>Localidad Ocurrencia</b>	Localidad -> Departamento	+	t
	Departamento -> Montevideo/Interior	+	t
	Montevideo/Interior -> All	+	t
<b>Domicilio Ocurrencia</b>	Domicilio -> TipoZona	+	t2
	Domicilio -> País	+	t2
	País / TipoZona-> All	+	t2
<b>Fallecidos</b>	Fallecido -> Embarazada	+	N/A
	Embarazada -> All	+	N/A
<b>Domicilio Residencia</b>	Domicilio -> TipoZona	+	t
	Domicilio -> País	+	t
	País / TipoZona -> All	+	t
<b>Localidad</b>	Localidad -> Departamento	+	t

<b>Residencia</b>	Departamento -> Montevideo/Interior	+	t
	Montevideo/Interior -> All	+	t
<b>Edades</b>	Edad -> GrupoEtario	+	t
	Edad -> GrupoEtarioMaterno	+	t
	GrupoEtario / GrupoEtarioMaterno -> All	+	t
<b>Sexo</b>	Sexo -> All	+	t

Cuadro 4.2 - Cuadro roll-up Mortalidad

### Mortalidad Infantil

Dimensión	Operaciones <i>drill-up</i>	Cant_ Fallecimientos_ Menores	Tasa_ Mortalidad Infantil
<b>Tiempo</b>	Fecha -> Mes	+	t
	Mes -> Año	+	t
	Año -> All	+	t
<b>CausasCIE10</b>	Causa -> Categoria	+	t2
	Categoria -> Grupo	+	t2
	Grupo -> Capítulo	+	t2
	Capitulo -> All	+	t2
<b>Localidad Ocurrencia</b>	Localidad -> Departamento	+	t
	Departamento -> Montevideo/Interior	+	t
	Montevideo/Interior -> All	+	t
<b>Domicilio Ocurrencia</b>	Domicilio -> TipoZona	+	t2
	Domicilio -> Pais	+	t2
	Pais/TipoZona -> All	+	t2
<b>Fallecidos Menores</b>	Fallecido_Menor -> EdadMenor	+	t2
	EdadMenor -> GrupoEtarioMenor	+	t2
	GrupoEtarioMenor -> All	+	t2
<b>Domicilio</b>	Domicilio -> TipoZona	+	t

<b>Residencia Madre</b>	Domicilio -> País	+	t
	País / TipoZona -> All	+	t
<b>Localidad Residencia Madre</b>	Localidad -> Departamento	+	t
	Departamento -> Montevideo/Interior	+	t
	Montevideo/Interior -> All	+	t
<b>Sexo</b>	Sexo -> All	+	t

*Cuadro 4.3 - Cuadro roll-up Mortalidad Infantil*

## 4.1.5 Verificación del diseño conceptual

En la presente sección se analiza si es posible confeccionar los indicadores definidos durante el Análisis de requerimientos aplicando operaciones OLAP sobre el modelo multidimensional diseñado. Para ello, se listan dichos indicadores y luego, para cada uno de ellos, la secuencia de operaciones necesarias para obtenerlos.

- *I1 Tasa de mortalidad infantil, por 1.000 nacidos vivos (menores de 1 año) e I2 Número de defunciones infantiles*
  1. **Roll-ups** sucesivos en la relación dimensional Mortalidad\_Infantil sobre la dimensión Fallecidos Menores hasta alcanzar el nivel GrupoEtarioMenor, aplicando a las medidas Cant\_Fallecimientos\_Menores y Tasa\_MortalidadInfantil la operación de agregación correspondiente.
  2. **Dice** seleccionando los valores *Neonatal precoz*, *Neonatal tardía* y *Postneonatal* del nivel GrupoEtarioMenor de la dimensión Fallecidos Menores.
  3. **Slice** de todas las dimensiones, asignando el valor *All*, excepto la dimensión Tiempo.
  4. **Roll-ups** sucesivos sobre la dimensión Tiempo hasta alcanzar el nivel Año, aplicando a las medidas Cant\_Fallecimientos\_Menores y Tasa\_MortalidadInfantil la operación de agregación correspondiente.
  
- *I3 Tasa de mortalidad infantil, por 1.000 nacidos vivos neonatal (de 0 a 27 días)*
  1. **Roll-ups** sucesivos en la relación dimensional Mortalidad\_Infantil sobre la dimensión Fallecidos Menores hasta alcanzar el nivel GrupoEtarioMenor, aplicando a la medida Tasa\_MortalidadInfantil la operación de agregación correspondiente.
  2. **Dice** seleccionando los valores *Neonatal precoz* y *Neonatal tardía* del nivel GrupoEtarioMenor de la dimensión Fallecidos Menores.
  3. **Slice** de todas las dimensiones, asignando el valor *All*, excepto la dimensión Tiempo.
  4. **Roll-ups** sucesivos sobre la dimensión Tiempo hasta alcanzar el nivel Año, aplicando a la medida Tasa\_MortalidadInfantil la operación de agregación correspondiente.
  
- *I4 Tasa de mortalidad infantil, por 1.000 nacidos vivos postneonatal (de 28 días a < 1 año)*
  - Análogo al indicador I3, con la diferencia de que en el paso 2 se debe seleccionar el valor *Postneonatal*, en lugar de *Neonatal precoz* y *Neonatal tardía*.
  
- *I5 Mortalidad de menores de 5 años reportada*
  1. En la relación dimensional Mortalidad\_Infantil, **slice** de todas las dimensiones, asignando el valor *All*, excepto la dimensión Tiempo.
  2. **Roll-ups** sucesivos sobre la dimensión Tiempo hasta alcanzar el nivel Año, aplicando a la medida Tasa\_MortalidadInfantil la operación de agregación correspondiente.

- *16 Razón de mortalidad materna por 100.000 nacidos vivos*
  1. En la relación dimensional Mortalidad, **slice** de todas las dimensiones, asignando el valor *All*, excepto la dimensión Tiempo.
  2. **Roll-ups** sucesivos sobre la dimensión Tiempo hasta alcanzar el nivel Año, aplicando a la medida *Razon\_MortalidadMaterna* la operación de agregación correspondiente.
  
- *17 Número de defunciones maternas*
  1. En la relación dimensional Mortalidad, **roll-up** sobre la dimensión Fallecidos hacia el nivel Embarazada, aplicando a la medida *Cant\_Fallecimientos* la operación de agregación correspondiente.
  2. **Dice** seleccionando el valor *Sí* del nivel Embarazada de la dimensión Fallecidos.
  3. **Slice** de todas las dimensiones, asignando el valor *All*, excepto la dimensión Tiempo.
  4. **Roll-ups** sucesivos sobre la dimensión Tiempo hasta alcanzar el nivel Año, aplicando a la medida *Cant\_Fallecimientos* la operación de agregación correspondiente.
  
- *18 Número de nacidos vivos*
  1. En la relación dimensional Nacimiento, **slice** de todas las dimensiones, asignando el valor *All*, excepto la dimensión Tiempo.
  2. **Roll-ups** sucesivos sobre la dimensión Tiempo hasta alcanzar el nivel Año, aplicando a la medida *Cant\_Nacimientos* la operación de agregación correspondiente.
  
- *19 Proporción de bajo peso al nacer*
  1. En la relación dimensional Nacimientos, **roll-ups** sucesivos sobre la dimensión Nacidos Vivos hasta alcanzar el nivel *CategoriaPeso*, aplicando a la medida *Proporcion\_PorPeso* la operación de agregación correspondiente.
  2. **Dice** seleccionando los valores *Extremado bajo peso* y *Bajo peso* del nivel *CategoriaPeso* de la dimensión Nacidos Vivos.
  3. **Slice** de todas las dimensiones, asignando el valor *All*, excepto la dimensión Tiempo.
  4. **Roll-ups** sucesivos sobre la dimensión Tiempo hasta alcanzar el nivel Año, aplicando a la medida *Proporcion\_PorPeso* la operación de agregación correspondiente.

De esta forma, queda validado el diseño conceptual presentado, ya que a partir de él se pueden obtener todos los indicadores solicitados.

## 4.2 Diseño lógico

El objetivo de esta sección es presentar la metodología aplicada para obtener el esquema lógico a partir del esquema conceptual y mostrar los resultados obtenidos.

Tal como se mencionó en el Capítulo 2 el tipo de almacenamiento elegido fue ROLAP, ya que así lo requería la herramienta Pentaho, utilizada para la implementación del proyecto. El esquema utilizado para definir la estructura de la base de datos relacional del Data Warehouse fue el esquema estrella, ya que se priorizó el tiempo de respuesta sobre el espacio utilizado en disco.

Como primer paso para traducir el esquema conceptual al esquema lógico es necesario identificar las tablas de dimensión y la tabla de hecho en una relación dimensional. En este tipo de esquema las tablas de dimensión se encuentran desnormalizadas y contienen sus jerarquías embebidas. Posteriormente se debe unir la tabla de hecho con cada dimensión por una relación 1:N y asignar como clave de la tabla de hecho la unión de las claves de las dimensiones relacionadas. Para añadir claridad al esquema, se utilizó el prefijo *d\_* para los nombres de las tablas de dimensión y *ft\_* para las tablas de hecho.

Durante la etapa de diseño conceptual se definieron tres relaciones dimensionales: Nacimientos, Mortalidad y Mortalidad Infantil. A continuación se detalla el resultado de traducir cada relación dimensional a su correspondiente esquema lógico utilizando el procedimiento descrito.

### 4.2.1 Relación dimensional Nacimientos

Como se puede apreciar en la Figura 4.17, el esquema lógico que se obtiene a partir de la relación dimensional Nacimientos se estructura en torno a la tabla de hecho *ft\_nacimientos*. En cuanto a las medidas, cabe señalar que no existen atributos que se correspondan con estas. En el caso de *Cant\_Nacimientos* esto se debe a que es posible calcularla simplemente contando la cantidad de registros en la tabla de hecho. Si por el contrario se hubiera optado por almacenar un atributo correspondiente a dicha medida, el mismo tomaría siempre el valor 1, por lo que tendría el mismo efecto que contar la cantidad de registros, requiriendo espacio de almacenamiento adicional. Por otra parte, con respecto a la medida *Proporcion\_PorPeso* esta se calcula en base a la medida *Cant\_Nacimientos* utilizando funcionalidades de la herramienta Pentaho, tal como se detallará en el Capítulo 5, por lo que no es necesario que esté almacenada en la tabla.

En cuanto a las tablas de dimensión es fácil notar que contienen los atributos de todas sus jerarquías embebidos en ellas. Como ejemplo se puede observar la tabla de la dimensión tiempo, *d\_tiempo*, que contiene los atributos *idFecha* y *fecha* pertenecientes al nivel Fecha, *idMes* y *mes* pertenecientes al nivel Mes y *anio* perteneciente al nivel Anio. Por otra parte, es fácil notar que la tabla *ft\_nacimientos* tiene como clave la unión de las claves de cada dimensión con la cual está relacionada.

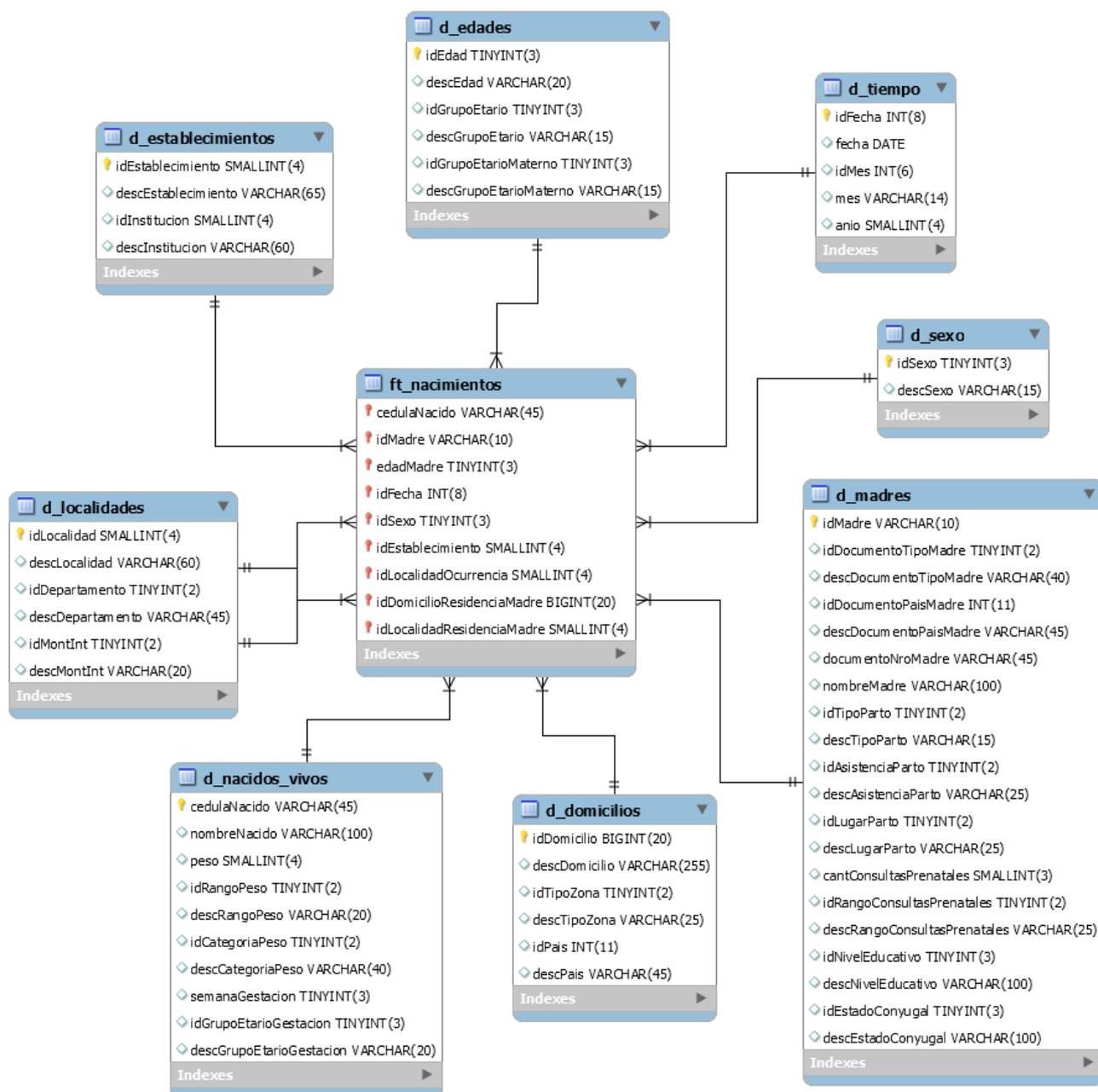


Figura 4.17 - Nacimientos

## 4.2.2 Relación dimensional Mortalidad

En la Figura 4.18 se muestra el esquema lógico obtenido a partir de la relación dimensional Mortalidad. Este cuenta con una tabla de hecho, *ft\_mortalidad*, cuyos atributos son las claves foráneas a las tablas dimensionales relacionadas. De forma análoga a la medida *Cant\_Nacimientos* de la relación dimensional Nacimientos, no es necesario contar con un atributo correspondiente a la medida *Cant\_Fallecimientos* ya que la misma también se calcula contando la cantidad de registros de la tabla de hecho. Con respecto a la medida *Razon\_MortalidadMaterna* tampoco es necesario almacenarla en un atributo ya que esta se calcula en base a las medidas *Cant\_Nacimientos* y *Cant\_Fallecimientos* utilizando funcionalidades de la herramienta Pentaho.

Se debe tener en cuenta que las tablas de dimensión *d\_tiempo*, *d\_domicilios*, *d\_localidades*, *d\_edades* y *dsexo* presentes en los tres esquemas son las mismas ya que las dimensiones están compartidas entre las tres relaciones dimensionales.

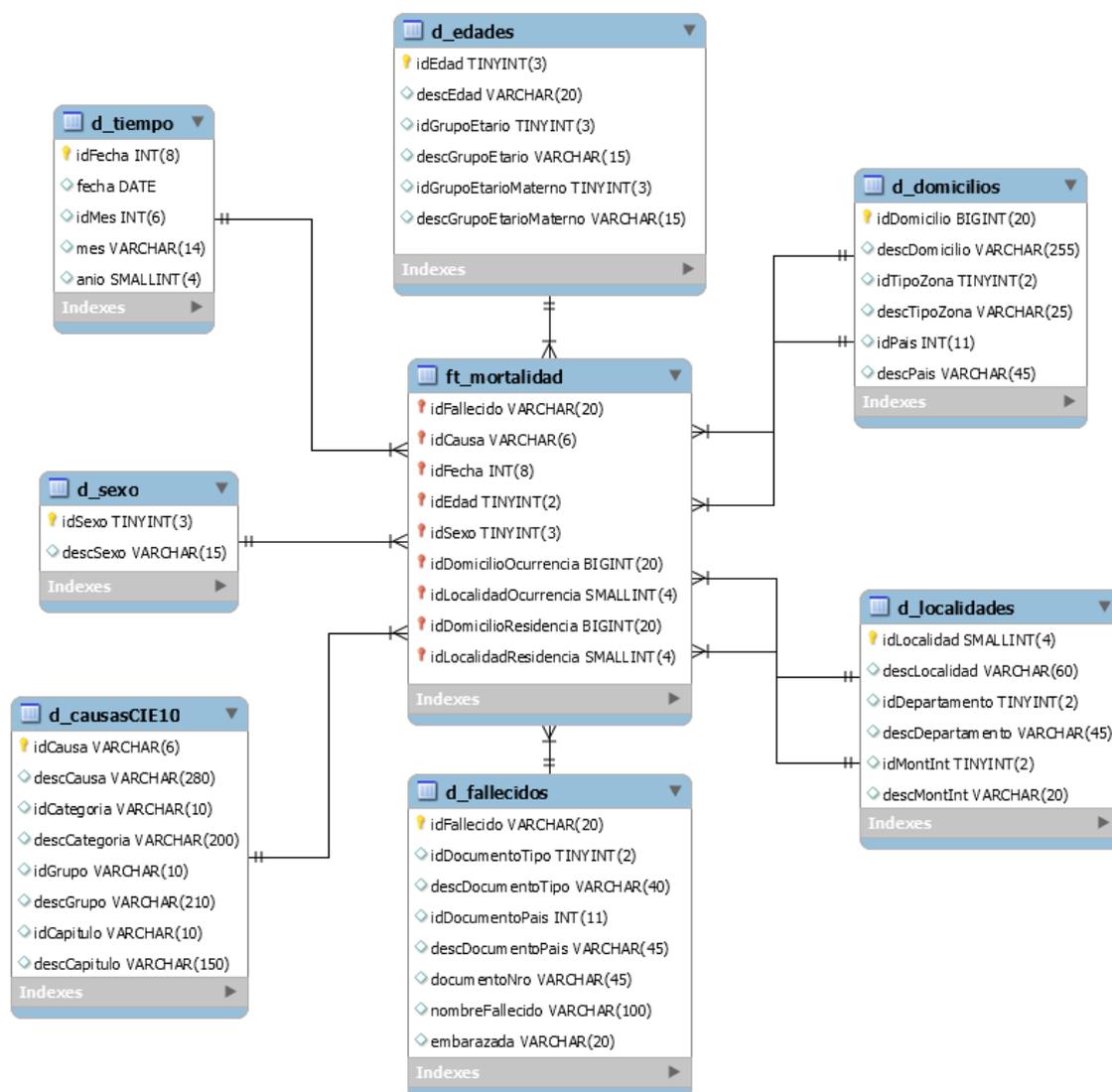


Figura 4.18 - Mortalidad

### 4.2.3 Relación dimensional Mortalidad Infantil

El esquema lógico resultante de traducir la relación dimensional Mortalidad Infantil se muestra en la Figura 4.19. Este está compuesto por la tabla de hecho *ft\_mortalidad\_infantil* y por las tablas de dimensión correspondientes a cada dimensión de la relación. De forma análoga al caso de Mortalidad, no es necesario contar con atributos correspondientes a las medidas.

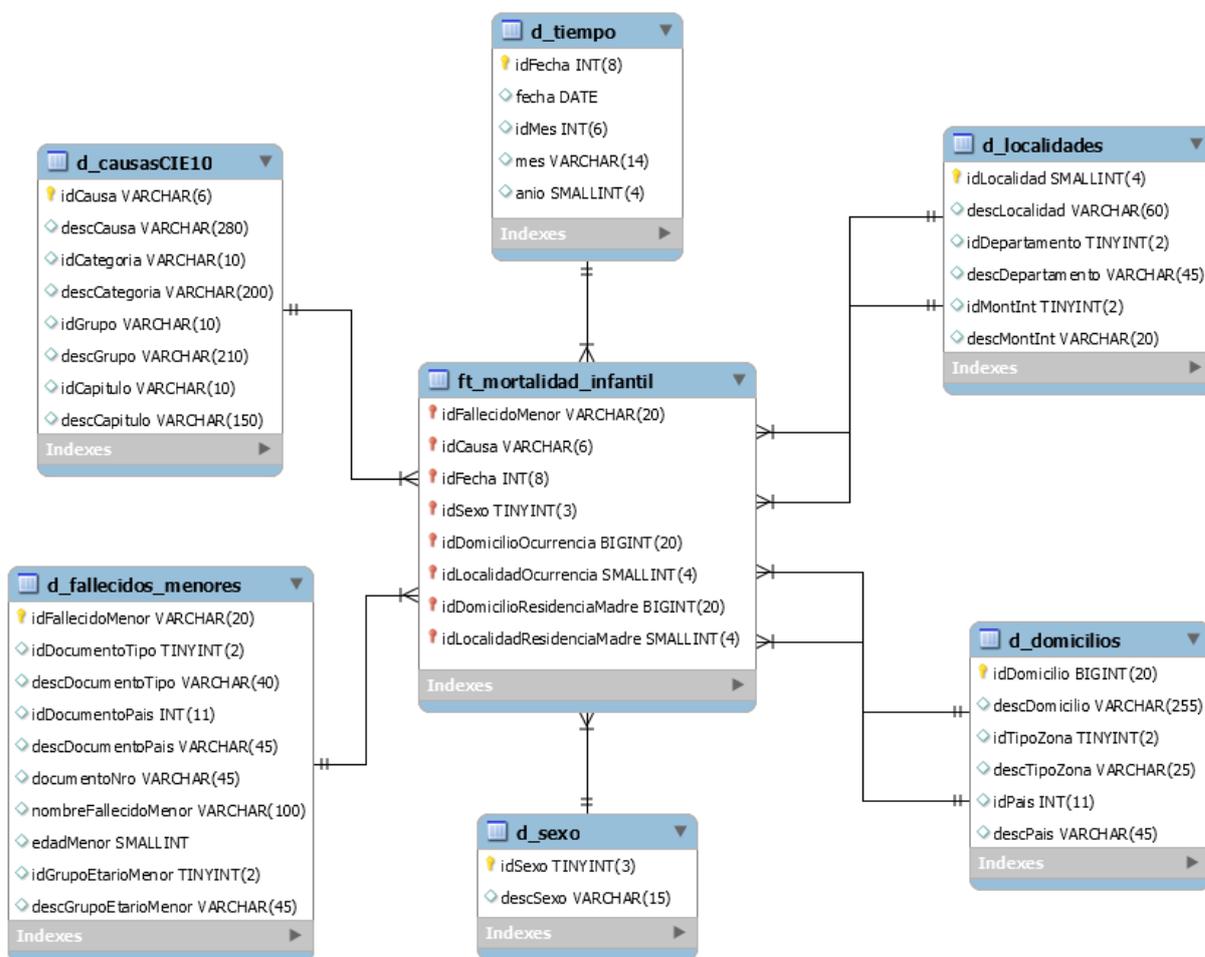


Figura 4.19 - Mortalidad Infantil

## 4.3 Diseño físico

A continuación se presentan las decisiones tomadas referentes al diseño físico del Data Warehouse de forma tal de mejorar la performance del sistema. Tal como se especificó en los requerimientos no funcionales, el manejador de base de datos empleado fue MySQL.

Como se explica en la documentación de MySQL [14], el tamaño ocupado en disco por una base de datos influye de manera directa en la eficiencia del sistema. Cuanto más pequeñas sean las tablas, menor va a ser el tiempo de respuesta de las consultas al manipularse de esta forma menos memoria. Por esta razón, teniendo como objetivo minimizar el espacio ocupado en disco, para cada atributo se seleccionaron los tipos de datos más pequeños del DBMS que permitieran representarlos. MySQL provee varios tipos de datos con los que se puede trabajar, facilitando de esta forma el diseño eficiente de la base de datos. Algunos de estos se muestran en la tabla de la Figura 4.20. Este aspecto del diseño físico se ve reflejado, por ejemplo, en el atributo *idEdad* de la tabla *d\_edades*, que puede contener valores entre 0 y 151. Como se puede observar en la tabla, el tipo de datos más pequeño que permite representarlo es *TINYINT*, por lo tanto este fue el elegido al momento de definir la tabla.

[MySQL 5.6 Reference Manual]

Type	Storage (Bytes)	Minimum Value (Signed/Unsigned)	Maximum Value (Signed/Unsigned)
TINYINT	1	-128	127
		0	255
SMALLINT	2	-32768	32767
		0	65535
MEDIUMINT	3	-8388608	8388607
		0	16777215
INT	4	-2147483648	2147483647
		0	4294967295

Figura 4.20 - Ejemplo tipos de datos MySQL [3]

Adicionalmente se declararon los atributos como *NOT NULL*, ya que esto reduce el espacio de almacenamiento en un bit por columna. Esto fue considerado durante el proceso de carga de forma tal de no permitir que se almacenen valores nulos para ningún atributo. Conjuntamente, en los casos en que fue posible, se prefirió utilizar tipos de datos numéricos por sobre tipos de datos de texto ya que los primeros se pueden almacenar utilizando menor cantidad de bytes.

Por otra parte, la clave de la tabla *d\_tiempo*, *idFecha*, es una clave subrogada. Este tipo claves se diferencian de las claves naturales, como lo son las claves del resto de las tablas, en que son generadas artificialmente al cargar el Data Warehouse, de tipo numérico sin signo y no tienen ningún significado de negocio en la base de datos fuente. En este caso, el valor almacenado para la fecha 01/11/2015 será 20151101, en lugar de utilizarse el tipo de datos *DATE*. De esta forma se preserva la relación de orden, mejorando así la performance al momento de realizar *joins* por este atributo entre la tabla *d\_tiempo* y las tablas de hecho del sistema.

# Capítulo 5 – Implementación

En el presente capítulo se describe la implementación de la solución propuesta.

## 5.1 Herramientas utilizadas

A continuación se presenta una breve descripción de las herramientas utilizadas para el desarrollo del proyecto.

### 5.1.1 Pentaho BI Suite Community Edition

Pentaho BI Suite es una colección de herramientas que trabajan en conjunto para brindar soluciones en el área de *Business Intelligence*. Existen dos ediciones disponibles para el usuario, Enterprise y Community. La edición Enterprise tiene un costo mensual y provee más funcionalidades que la edición Community, mientras que esta última se puede utilizar de forma gratuita, es de código abierto y se encuentra licenciada bajo la Licencia Pública General GNU versión 3.

La edición Community versión 5.3 fue la elegida para la realización de este proyecto, no sólo por la posibilidad de utilizarla de forma gratuita sino también por el interés de la UINS de emplear una herramienta de código abierto. Las herramientas provistas por la suite Pentaho Community [7] que fueron utilizadas para la implementación del Data Warehouse se presentan a continuación.

#### *Pentaho Business Analytics Server*

El servidor Business Analytics de Pentaho [15] procesa el contenido de reportes, análisis y *dashboards*. Es gestionado a través de una herramienta web, la Consola de Usuario, que permite crear y editar dichas visualizaciones de contenido. Esta herramienta posibilita la realización de análisis multidimensionales ya que el servidor tiene integrado el motor OLAP **Mondrian** [16], que permite ejecutar consultas escritas en el lenguaje de consulta MDX (*MultiDimensional EXpressions*), leyendo datos de una base relacional, y presentar los resultados en un formato multidimensional. Mondrian cuenta con una capa dimensional, que es la encargada de analizar, validar y ejecutar las consultas MDX, y una capa de almacenamiento que se corresponde con la base de datos relacional.

#### *Pentaho Data Integration*

La herramienta Pentaho Data Integration [17], también denominada Kettle, es el componente de la suite Pentaho encargado de los procesos de Extracción, Transformación y Carga. Las soluciones creadas con la misma están construidas a partir de dos tipos de objetos, las transformaciones y los

trabajos. Está compuesta por un motor de integración de datos que es capaz de interpretar y ejecutar este tipo de objetos. Además, provee el siguiente conjunto de herramientas y utilidades:

- **Spoon:** IDE gráfico para crear transformaciones y trabajos.
- **Kitchen:** herramienta de línea de comandos para ejecutar trabajos.
- **Pan:** herramienta de línea de comandos para ejecutar transformaciones.
- **Carte:** servidor liviano para ejecutar trabajos y transformaciones desde un host remoto.

### *Pentaho Schema Workbench*

Pentaho Schema Workbench [18] es una interfaz gráfica de diseño que permite crear y testear los esquemas que representan la base multidimensional de Mondrian. Dichos esquemas se definen mediante archivos XML que contienen los metadatos que describen el modelo dimensional, y cómo este se relaciona con el modelo relacional.

### *Saiku Analytics Community Edition*

Saiku Analytics Community Edition [19] es una *suite* de código abierto de análisis OLAP que permite visualizar datos y crear gráficas. Se utilizó instalándolo como un *plugin* para el servidor Business Analytics de Pentaho, accediendo a sus funcionalidades mediante la Consola de Usuario.

## 5.2 Carga y actualización

En la presente sección se describe, en alto nivel, el proceso de carga y actualización del Data Warehouse, así como el estudio de calidad realizado sobre las fuentes de datos.

### 5.2.1 Calidad de datos

Dada la importancia de la calidad de los datos en un sistema de Data Warehouse, se decidió realizar una etapa de limpieza durante el proceso de extracción, transformación y carga. Para ello fue necesario definir un modelo y realizar un análisis de calidad sobre la base mejorada provista por el MSP a partir del mismo, con el fin de eliminar y corregir los errores encontrados, así como prevenir posibles errores que pudieran surgir en futuras actualizaciones.

El principal insumo utilizado para el estudio de calidad fue la documentación generada durante la implementación del plan de mejora de la calidad de las bases de datos de los sistemas CNVe, CDe y CDp [9], puesto en práctica en el año 2012 por la Fundación Ricaldoni.

Las dimensiones de calidad incluidas en el modelo definido fueron exactitud, completitud y consistencia. Dentro de la dimensión exactitud, el factor analizado fue correctitud sintáctica, mientras que dentro de completitud se analizó la densidad y dentro de consistencia los factores integridad de dominio e integridad inter-relación.

En el Anexo I se puede encontrar la especificación del modelo elegido para realizar la medición de la calidad de datos en la base fuente, los resultados luego de analizarla, y las soluciones implementadas para corregir y eliminar los errores detectados.

En términos generales, a partir del estudio realizado se puede concluir que aproximadamente el 99% de los datos se encuentran libres de errores sintácticos, no contienen valores nulos, se encuentran dentro de los dominios establecidos (en los casos que corresponde) y cumplen con las relaciones establecidas. Sin embargo, se pueden observar algunas excepciones, como es el caso del atributo *embarazada* de la tabla *cd\_datos\_fallecido* para el cual aproximadamente el 66% de los datos contienen errores sintácticos. Otras dos excepciones son el caso de los atributos *id\_calle* y *tipo\_zona* de la tabla *domicilios*, cuyo análisis indica que el 28,42% y el 56,19% de los datos respectivamente contienen valores nulos.

En la siguiente sección, donde se presenta el proceso de carga y actualización del Data Warehouse, se mencionan las tareas de limpieza implementadas. Entre ellas se encuentra la creación de tuplas *dummy* con valores *Erróneo* o *No Indicado*, el primero, para el manejo de valores fuera del dominio establecido, y el segundo, para reemplazar valores nulos o vacíos y así evitar jerarquías incompletas; y la eliminación de tuplas cuya fecha, ya sea de parto o de muerte, sea nula ya que se considera que este atributo es de vital importancia y sin él la tupla carece de sentido. Otras de las tareas implementadas incluyen el estudio de atributos secundarios en los casos en que el atributo principal

se encuentra vacío o contiene información errónea; y estudios particulares, como por ejemplo, en el caso del atributo *embarazada* cuando este toma el valor nulo se consulta el atributo *sexo* y la edad del fallecido para decidir qué valor cargar en el Data Warehouse.

## 5.2.2 Extracción, transformación y carga

El proceso de carga de cada una de las dimensiones y tablas de hecho del Data Warehouse se realiza mediante la ejecución de un trabajo en Pentaho, cuya representación en alto nivel se presenta en la Figura 5.1. El flujo detallado del trabajo se puede encontrar en el Anexo II.

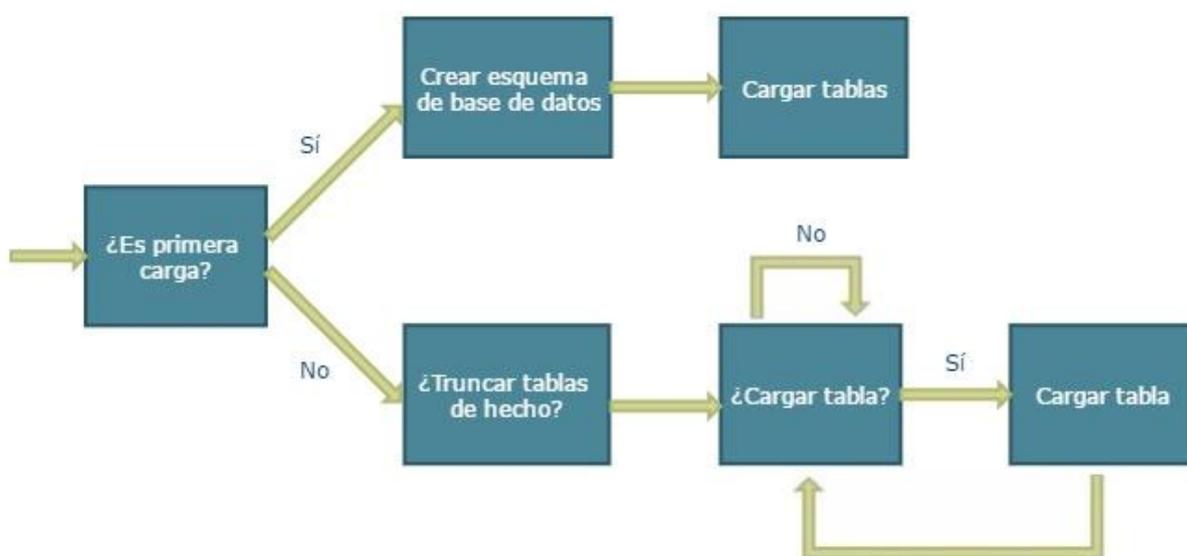


Figura 5.1 - Carga del Data Warehouse

En el diagrama se pueden distinguir dos posibles escenarios, la primera carga y la actualización. El proceso recibe un parámetro de entrada *primerCarga*, de tipo booleano, que determina el contexto. En caso de que se desee realizar la primera carga, se procede a crear el esquema de la base de datos y luego se cargan las tablas de dimensión y de hecho del Data Warehouse. Para ello se hace referencia a las transformaciones en Pentaho que se encargan de realizarlo, cuyos diagramas se detallan más adelante. En caso contrario, se evalúa el parámetro *truncarTablas*. Este último, también de tipo booleano, determina si se deben vaciar las tablas de hecho y las tablas de dimensión *d\_nacidos\_vivos*, *d\_madres*, *d\_fallecidos* y *d\_fallecidos\_menores*.

Por otra parte, el proceso también recibe un parámetro por cada tabla de la base de datos, excepto para las siguientes tablas: *d\_localidades*, que se carga en conjunto con la tabla *d\_domicilios*; *d\_madres* y *d\_nacidos\_vivos* que se cargan junto a la tabla *ft\_nacimientos*; y *d\_fallecidos* y *d\_fallecidos\_menores* que se cargan junto a las tablas *ft\_mortalidad* y *ft\_mortalidad\_infantil*

respectivamente. Dichos parámetros son de tipo booleano y determinan si se desea volver a cargar la o las tablas correspondientes al momento de realizar una actualización. Si se desea actualizar alguna dimensión se deben vaciar las tablas de hecho, ya que sino es posible que queden datos inconsistentes. Luego, para cada tabla o grupo de tablas se evalúa el parámetro correspondiente, y en caso de que su valor sea *verdadero* se realiza la carga. En caso de que se desee actualizar la tabla correspondiente a una dimensión, la misma se vacía y luego se carga; en caso contrario, es decir, si se desea actualizar una tabla de hecho, se procede directo a la carga, ya que en las transformaciones correspondientes si los datos ya existen se sobrescriben y de otro modo se insertan. De esta forma se permite que la actualización de las tablas de hecho se realice de manera incremental y así evitar tener que borrar las tablas y volver a cargar todos los datos nuevamente cada vez que se desee agregar datos nuevos.

Adicionalmente, se reciben como parámetros configuraciones asociadas a la conexión de la base de datos, el nombre del esquema fuente, una fecha de inicio y una fecha de fin para la carga y/o actualización de las tablas *d\_madres*, *d\_nacidos\_vivos*, *d\_fallecidos*, *d\_fallecidos\_menores*, *ft\_nacimientos*, *ft\_mortalidad* y *ft\_mortalidad\_infantil*, que son pasados como parámetros a las transformaciones correspondientes.

A continuación se describen en alto nivel los procesos de carga de cada una de las dimensiones y tablas de hecho del Data Warehouse a los que se hace referencia en el proceso descrito anteriormente. Los flujos detallados de dichos procesos también se pueden encontrar en el Anexo II.

## Dimensión Tiempo

En la Figura 5.2 se ilustra, en términos generales, el proceso de carga de la dimensión Tiempo.



Figura 5.2 - Carga d\_tiempo

En primera instancia, se generan 10.950 filas (cantidad de días que hay en 30 años) con fechas consecutivas a partir de una fecha base. La fecha base configurada en la ETL, elegida en base a los datos recibidos para la construcción del Data Warehouse, es 01/01/2012.

Una vez generadas las fechas, se calculan los demás campos que forman parte de la jerarquía. Para identificar la fecha, se utiliza una clave subrogada de 8 dígitos con el siguiente formato AAAAMMDD, donde AAAA representa el año, MM el mes y DD el día. En cuanto al nivel *Mes* se calculan los siguientes dos campos: el identificador, que es un número de 6 dígitos con el siguiente formato

AAAAMM (análogo al identificador de la fecha) y la descripción, que consiste del nombre del mes. Finalmente, se calcula el año, completando así todos los niveles de la jerarquía.

En última instancia, se guardan los datos generados en la tabla *d\_tiempo*.

### *Dimensión Edades*

Los pasos para realizar la carga de la dimensión Edades se pueden resumir tal como se muestra en la Figura 5.3.



Figura 5.3 - Carga *d\_edades*

En primer lugar, se generan filas con edades comprendidas en el rango que va de 0 a 150 años y una fila adicional, con valor 151. Esta última se genera con el fin de ser utilizada al momento de cargar las tablas de hecho para los casos en que el cálculo de la edad en años, ya sea de una madre o un fallecido, resulta en un valor fuera del rango establecido. El identificador utilizado para el nivel *Edad* es el valor numérico generado.

En segundo lugar se calcula la descripción de la edad. Si la edad está comprendida entre 0 y 150, la descripción coincide con el valor numérico, mientras que si la edad es 151, se le asigna la descripción *Erróneo*. Luego, se calculan los niveles *GrupoEtario* y *GrupoEtarioMaterno* tal como se muestra en el Cuadro 5.1. En dicho cuadro se pueden observar las descripciones de ambos grupos etarios, ya que los identificadores utilizados son autoincrementales. Además, ambos niveles tienen un grupo *Erróneo* para las edades que están fuera de rango. El nivel *GrupoEtarioMaterno* tiene un elemento *No Indicado*, utilizado para los casos en que no se tiene el dato de la edad.

GrupoEtario	GrupoEtarioMaterno
0-4	No Indicado
5-9	14 o menos
10-14	15-19
15-19	20-24
20-24	25-29
25-29	30-34
30-34	35-40
35-40	40-44
40-44	45 o más

45-49	Erróneo
50-54	
55-59	
60-64	
65-69	
70-74	
75-79	
80-84	
85-89	
90-94	
95-99	
100-104	
105-109	
110 o más	
Erróneo	

Cuadro 5.1 - Dimensión Edades - GruposEtarios

Finalmente, se guardan los datos generados en la tabla *d\_edades*.

### Dimensión Sexo

El diagrama de la Figura 5.4 ilustra, en términos generales, el proceso de carga de la dimensión Sexo.



Figura 5.4 - Carga dsexo

En el primer paso del proceso, se generan 5 filas con identificadores que van del 1 al 5. Luego, se asignan las descripciones a los registros tal como se muestra en el Cuadro 5.2.

Identificador	Descripción
1	Femenino
2	Masculino
3	Indefinido
4	Desconocido
5	Erróneo

Cuadro 5.2 - Dimensión Sexo

En la última etapa se almacenan los datos generados en la tabla *d\_sexo*.

## *Dimensión Establecimientos*

La Figura 5.5 representa el proceso de carga de la dimensión Establecimientos.



*Figura 5.5 - Carga d\_establecimientos*

En primera instancia se extraen los datos del esquema fuente, más específicamente de las tablas *Establecimientos* e *Instituciones*. A continuación se asocian las Instituciones a los Establecimientos correspondientes y, con el fin de evitar que quede incompleta la jerarquía, a los establecimientos que no tienen institución asociada se les asigna la institución *No Indicado*.

Para identificar tanto los establecimientos como las instituciones se utilizan los identificadores extraídos de las fuentes. Seguidamente se agrega un establecimiento adicional, que lleva la descripción *No Indicado* y que se utiliza en los casos en que se desconoce el establecimiento de ocurrencia.

En última instancia, se persiste la información generada en la tabla *d\_establecimientos*.

## Dimensión Domicilios

En la Figura 5.6 se presentan las etapas generales del proceso de carga de la dimensión Domicilios.

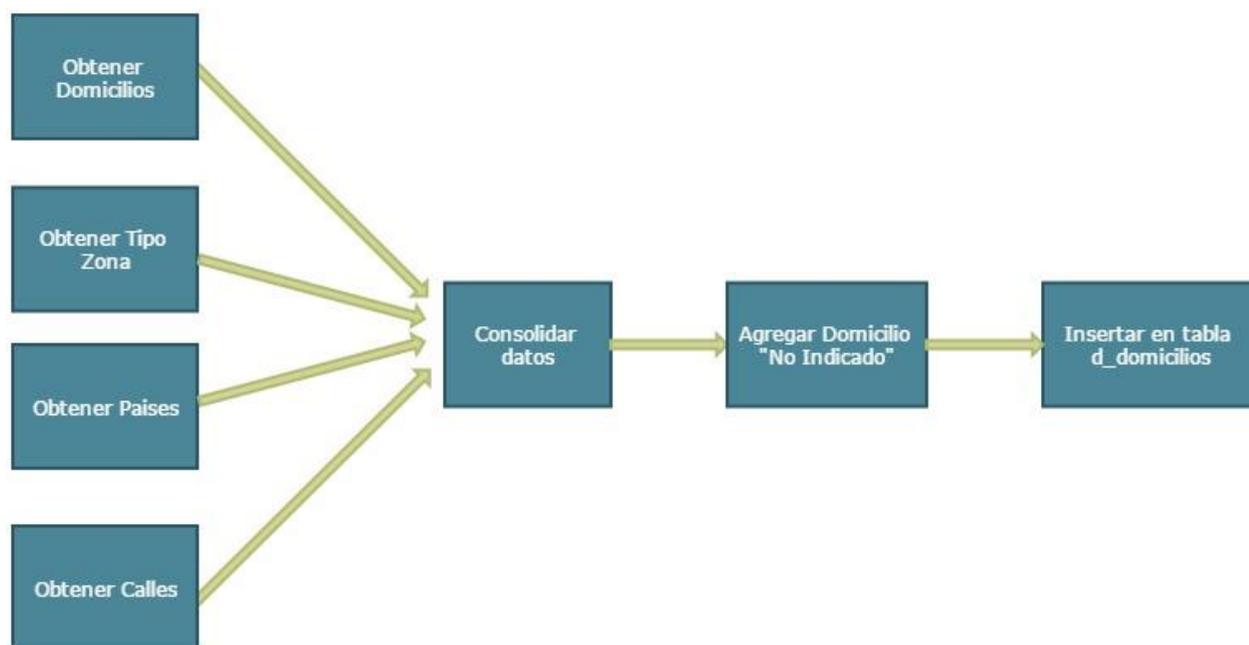


Figura 5.6 - Carga d\_domicilios

Durante las primeras etapas del proceso se obtienen los datos necesarios para cargar la dimensión Domicilios a partir de las tablas *domicilios*, *tipo\_zona\_residencia*, *países* y *calles* de la base de datos fuente. Luego se consolidan los datos, asignando el valor *No Indicado* para los casos en que no se tiene el tipo de zona y/o el país y/o la calle, de forma tal que no se formen jerarquías incompletas.

Para identificar los domicilios se utiliza un número entero de 20 dígitos que coincide con el identificador del domicilio en las fuentes. Además, se guarda una descripción del domicilio, formada por el nombre de la calle y el número de puerta.

A continuación se genera un domicilio de descripción *No Indicado* para ser utilizado en los casos en que el domicilio se desconoce. Para finalizar, se insertan los datos en la tabla *d\_domicilios*.

## Dimensión Localidades

Los pasos que componen el proceso de carga de la dimensión Localidades se pueden encontrar en la Figura 5.7.

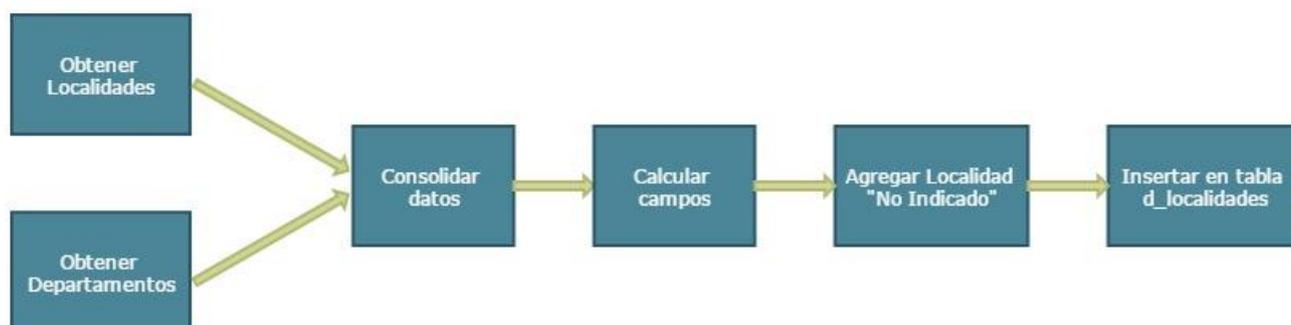


Figura 5.7 - Carga d\_localidades

Durante el primer paso del proceso se recuperan los datos de las tablas fuentes *localidades* y *departamentos*. Dichos datos son asociados, completando el nivel *Departamento* con el valor *No Indicado* en los casos en que el mismo es nulo. Como identificador del nivel *Localidad* se utiliza un número de 4 dígitos que toma el mismo valor que el identificador de la localidad en las fuentes.

Para continuar, se calculan los valores del nivel *Montevideo\_Interior* en función del departamento asociado a la localidad de acuerdo al Cuadro 5.3.

Departamento	Nivel <i>Montevideo_Interior</i>
Montevideo	Montevideo
Desconocido	Desconocido
No Indicado	No Indicado
Cualquier valor que no se encuentre en los listados anteriormente	Interior

Cuadro 5.3 - Dimensión Localidades - *Montevideo\_Interior*

En el siguiente paso se agrega una nueva localidad con descripción *No Indicado* y finalmente se persisten los datos en la tabla *d\_localidades*.

## Dimensión CausasCIE10

La Figura 5.8 ilustra las etapas que conforman el proceso de carga de la dimensión CausasCIE10.



Figura 5.8 - Carga d\_causascie10

El proceso de carga comienza con la extracción de los datos de las causas a partir de una planilla Excel utilizada como fuente. En dicha planilla se encuentran los datos, compuestos por un identificador alfanumérico y una descripción, tanto de las causas CIE10, como de sus categorías, grupos y capítulos asociados.

Una vez obtenidos dichos datos, se agregan las causas *dummy* con descripción *No Indicado* y *Erróneo*, y se procede a insertar los datos en la tabla *d\_causascie10*.

## Dimensión Nacidos Vivos

El diagrama de la Figura 5.9 representa el flujo de carga y/o actualización de la dimensión Nacidos Vivos.

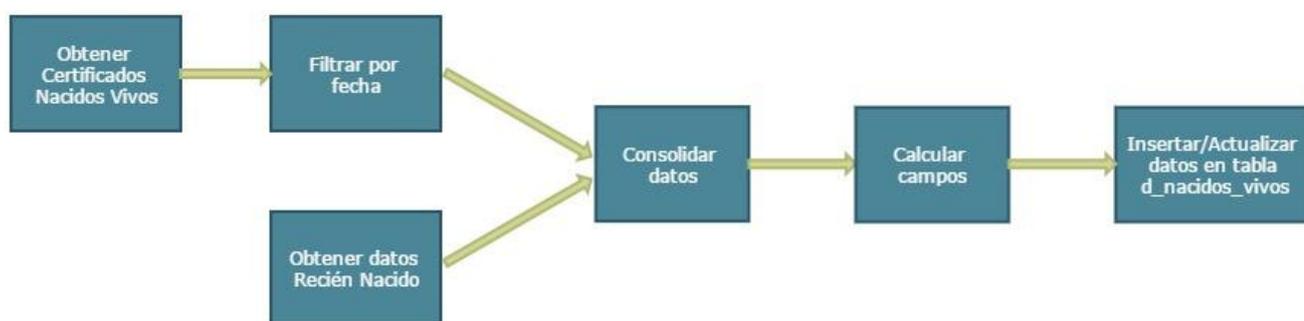


Figura 5.9 - Carga d\_nacidos\_vivos

En el primer paso se recuperan los certificados de nacimiento de la base de datos fuente, puntualmente de la tabla *certificados\_nacidos\_vivos*. Esto se realiza con el fin de filtrar los datos a cargar y/o actualizar, de acuerdo al período seleccionado para la carga y la fecha de ocurrencia del parto.

Una vez obtenidos dichos datos, se obtiene el período de fechas ingresado por parámetro y, en caso de que el mismo no sea vacío, se procede a filtrar los nacimientos ocurridos en dicho período. Paralelamente se recuperan los datos de los recién nacidos de la tabla *cnv\_datos\_recien\_nacido*.

Luego, se asocian los datos de los recién nacidos y los certificados de nacimiento filtrados, de forma tal que el resultado esté compuesto únicamente por los datos de los recién nacidos cuyo nacimiento haya ocurrido dentro del período de carga ingresado.

Como identificador del nivel *NacidoVivo* se utiliza un arreglo de caracteres que contiene el número de cédula reservado para el nacido vivo, que se extrae del atributo *numero\_documento\_reserva*. Por otra parte, se guarda también el nombre del recién nacido, y en los casos en los que este campo se encuentra vacío en las fuentes, se le asigna el valor *No Indicado*.

A continuación se calculan los demás miembros de la dimensión. El *GrupoEtarioGestacional* se define en base a la *EdadGestacional* (que coincide con el atributo *semanas\_gestacion\_nacimiento*), tal como se muestra en el Cuadro 5.4.

Intervalo de EdadGestacional	GrupoEtarioGestacional
0	No Indicado
20-31	Entre 20 y 31
32-33	Entre 32 y 33
34-35	Entre 34 y 35
36	36
37	37
38	38
39	39
40	40
41	41
42	Más de 41
Cualquier valor que no se encuentre en los listados previamente	Erróneo

Cuadro 5.4 - Dimensión Nacidos Vivos - GrupoEtario

En cuanto al *Peso*, el mismo coincide con el atributo *peso* y en base a este, se calcula el nivel *RangoPeso*. Una vez obtenido el *RangoPeso*, se define el nivel *CategoriaPeso*. La definición de estos dos niveles se puede observar en el Cuadro 5.5. Cabe señalar que para los casos en que el peso registrado es menor a 500gr o mayor a 8000gr se considera que el dato ingresado es erróneo.

RangoPeso	CategoriaPeso
No Indicado	No Indicado
Menos de 1000	Extremado bajo peso (menos de 1000gr)
De 1000 a 1499	Bajo peso (entre 1000 y 2499gr)
De 1500 a 1999	

De 2000 a 2499	Normal (2500gr o mas)
De 2500 a 2999	
De 3000 a 3499	
De 3500 a 3999	
De 4000 a 8000	
Erróneo	Erróneo

Cuadro 5.5 - Dimensión Nacidos Vivos - Jerarquía Peso

Para culminar, según corresponda, se insertan o actualizan los datos en la tabla *d\_nacidos\_vivos*.

### Dimensión Madres

El proceso de carga y actualización de la dimensión Madres se puede esquematizar tal como se muestra en la Figura 5.10.

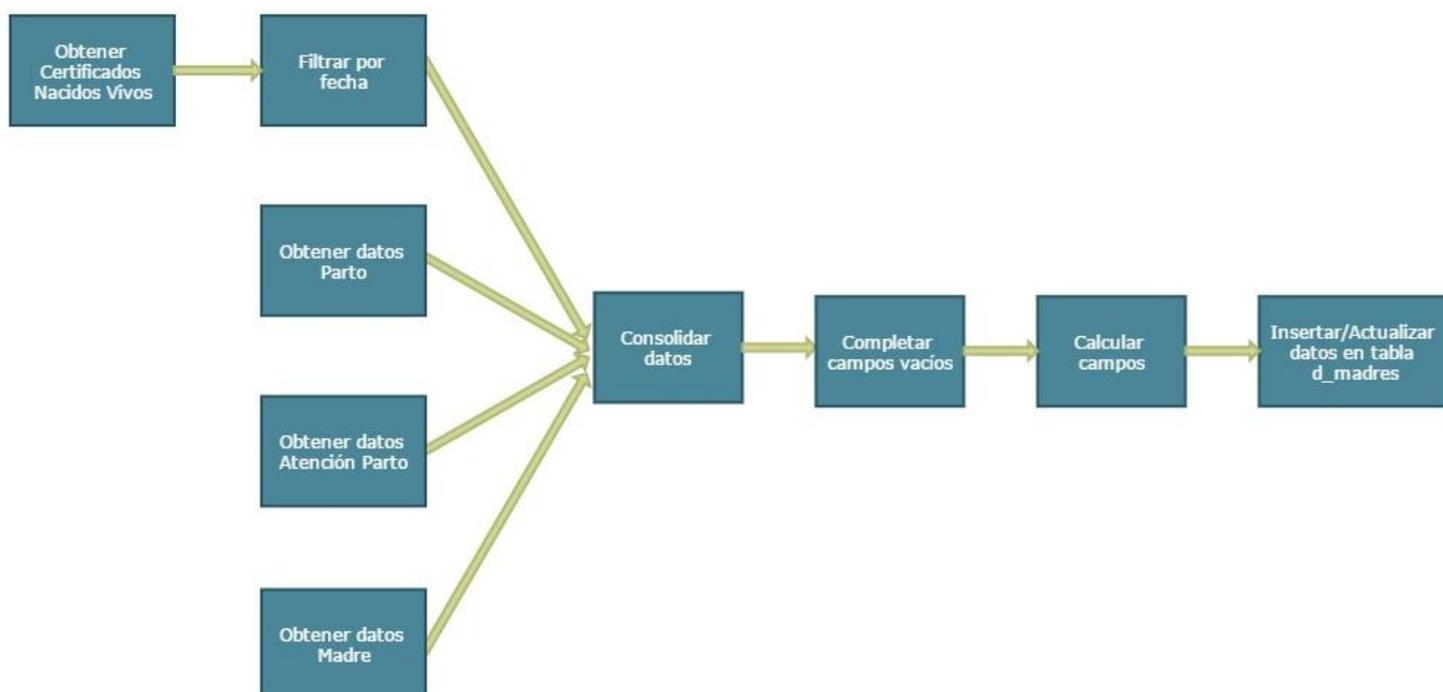


Figura 5.10 - Carga *d\_madres*

De forma análoga al proceso de carga de la dimensión Nacidos Vivos, en el primer paso de este proceso se obtienen los datos de los certificados de nacimiento de la tabla *certificados\_nacidos\_vivos* de las fuentes y se filtran de acuerdo al rango de fechas seleccionado para la carga/actualización. Concurrentemente, se extraen también los datos de los partos, de atención de los partos y de las madres desde las tablas fuente *cnv\_datos\_parto*, *cnv\_atendio\_parto* y *cnv\_datos\_madre* respectivamente.

Una vez obtenidos todos los datos, se asocian de modo que permanezcan sólo los datos de los nacimientos ocurridos en el período seleccionado; siempre y cuando las fechas ingresadas no sean vacías. En caso de que lo sean, simplemente se asocian todos los datos obtenidos. Luego, los campos vacíos se reemplazan por tuplas *dummy* con descripción *No Indicado*, con el fin de evitar trabajar con jerarquías incompletas.

A continuación, se calculan los restantes campos de la dimensión. Para identificar el nivel *Madre*, se utiliza un arreglo de 10 caracteres compuesto por la concatenación del identificador y la serie del certificado. A su vez, se guarda también el nombre de la madre, siendo este la concatenación de los atributos *primer\_nombre* y *segundo\_nombre*, y los datos del documento de la misma (tipo, país y número).

En cuanto al nivel *ConsultasPrenatales*, su identificador coincide con el atributo *total\_consultas\_prenatales*. En base a este nivel, se calcula el nivel *RangosConsultasPrenatales* de acuerdo al Cuadro 5.6.

ConsultasPrenatales	RangoConsultasPrenatales
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10 - 40	10 o más
Cualquier valor que no se encuentre en los listados previamente	Erróneo

Cuadro 5.6 - Dimensión Madres – RangoConsultasPrenatales

Finalmente, se insertan y/o actualizan los campos en la tabla *d\_madres* según corresponda.

## Dimensión Fallecidos

En la Figura 5.11 se presenta un diagrama del flujo de carga/actualización de la dimensión Fallecidos.

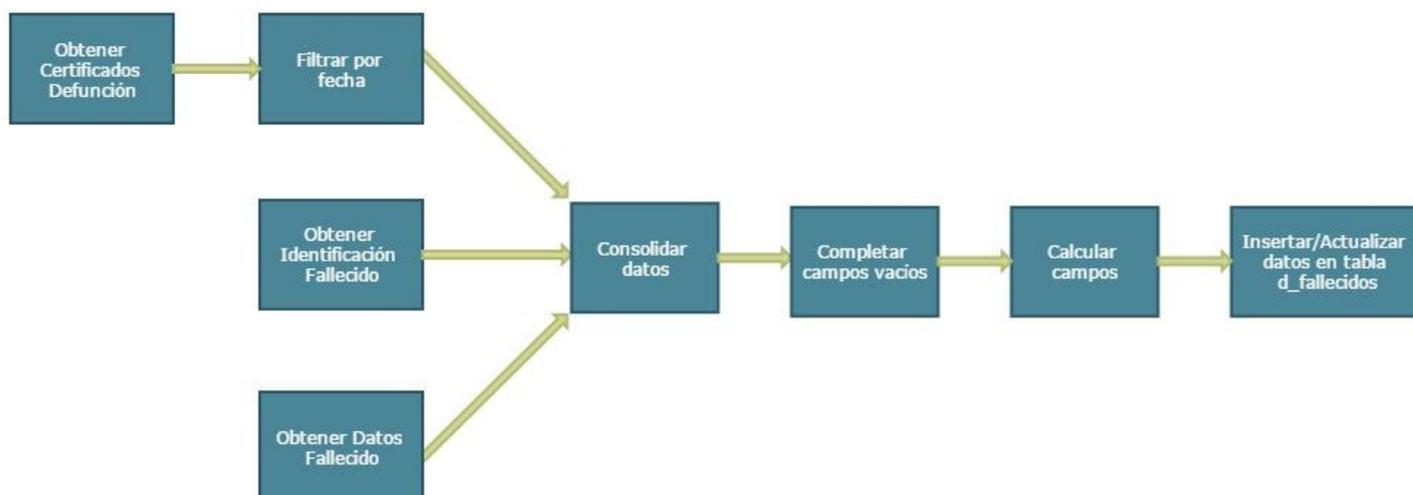


Figura 5.11 - Carga d\_fallecidos

En primer lugar se obtienen los datos de los certificados de defunción no fetales a partir de la tabla *certificados\_defuncion* de las fuentes. Luego, se obtienen las fechas ingresadas por parámetro, que determinan el período seleccionado para la carga y/o actualización de datos, en caso de que no se encuentren vacías, se filtran los certificados cuya fecha de muerte está dentro de dicho período. Concurrentemente, se recuperan los datos de los fallecidos y la identificación de los mismos a partir de las tablas fuentes *cd\_datos\_fallecido* y *cd\_identificacion\_fallecido*.

Posteriormente, se consolidan los datos de forma tal que el resultado contenga únicamente la información de los fallecimientos ocurridos en el período ingresado. Los campos vacíos se reemplazan por tuplas *dummy* con descripción *No Indicado*, de modo que no se generen jerarquías incompletas.

Como identificador del nivel *Fallecido* se utiliza un arreglo de 20 caracteres formado por la concatenación del identificador y la serie del certificado. Se utiliza este identificador ya que no hay ninguna restricción que determine que la cédula del fallecido no pueda ser vacía. Además, se extrae el nombre y datos del documento del fallecido. En caso de que cualquiera de éstos últimos sea desconocido, se completa con el valor *No Indicado*.

A continuación se calculan los demás campos de la dimensión. Para definir el nivel *Embarazada*, se utiliza el atributo *embarazada*, cuyo significado se puede observar en el Cuadro 5.7.

Valor del atributo <i>embarazada</i>	Significado
1	Sí
2	No
3	Se Ignora

4	No Corresponde
5	No Aplica

Cuadro 5.7 - Dimensión Fallecidos – Embarazada

El nivel *Embarazada* se identifica con un arreglo de 20 caracteres que coincide con los significados que figuran en el Cuadro 5.7. Para los casos en que el atributo *embarazada* es vacío o cero, el valor que toma el nivel se define con el siguiente criterio: si el sexo del fallecido es Femenino y la edad está entre 9 y 70 años, el valor que toma el nivel *Embarazada* es *Se Ignora*; si es Femenino y la edad no está dentro del rango previamente mencionado, el valor del nivel es *No Corresponde*; si el sexo del fallecido es Masculino, se le asigna el valor *No Corresponde*; y finalmente si el sexo no es ni Femenino ni Masculino, el nivel toma el valor *Desconocido*. Cabe mencionar que la edad del fallecido es calculada como la diferencia entre la fecha de muerte y la fecha de nacimiento, es decir que en este caso no se utiliza ninguno de los atributos de las fuentes referentes a la edad del fallecido. En los casos en que el atributo *embarazada* tome cualquier otro valor que no sea vacío o cero y no se encuentre en el Cuadro 5.7, el nivel toma el valor *Erróneo*.

En el último paso se persisten los datos en la tabla *d\_fallecidos*.

### Dimensión Fallecidos Menores

La Figura 5.12 ilustra las etapas que forman parte del proceso de carga y actualización de la dimensión Fallecidos Menores.

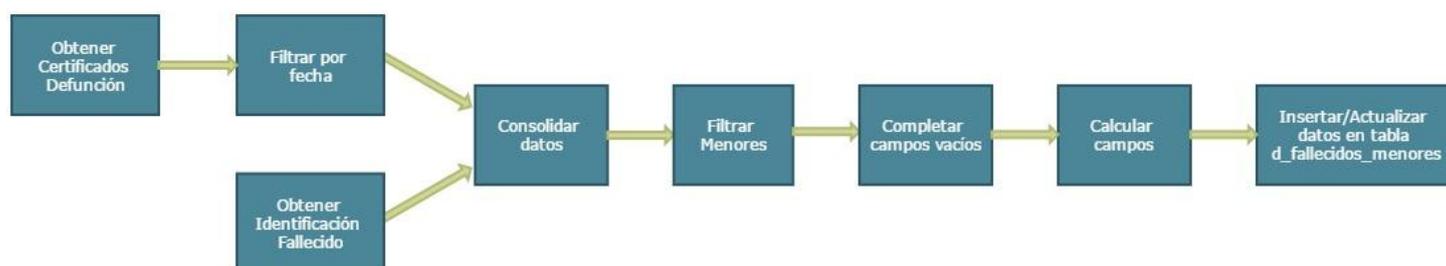


Figura 5.12 - Carga *d\_fallecidos\_menores*

Las primeras etapas de este proceso son análogas a las del proceso de carga/actualización de la dimensión Fallecidos. La diferencia principal es que, una vez asociados los datos, se calcula la edad del fallecido (como la diferencia en años entre la fecha de muerte y la fecha de nacimiento) y se realiza un filtro, de forma de que sólo permanezcan los datos de los fallecidos menores de 5 años.

Posteriormente, durante la etapa de cálculo de campos, se calcula el identificador del nivel *Fallecido Menor* de forma análoga al identificador del nivel *Fallecido* de la dimensión Fallecidos. Luego, se

calcula el nivel *EdadMenor*, como la diferencia en días entre la fecha de muerte y la fecha de nacimiento, y en base a esta se definen los valores del nivel *GrupoEtarioMenor* tal como se muestra en el Cuadro 5.8.

Intervalo de EdadMenor	GrupoEtarioMenor
0-7	Neonatal precoz
8-27	Neonatal tardía
28-364	Postneonatal
365-1826	1 año o más

Cuadro 5.8 - Dimensión Fallecidos Menores - GrupoEtarioMenor

Para finalizar, se insertan y/o actualizan los datos en la tabla *d\_fallecidos\_menores*.

### Tabla de hecho Nacimientos

El diagrama en alto nivel del proceso de carga y actualización de la tabla de hecho Nacimientos se puede observar en la Figura 5.13.

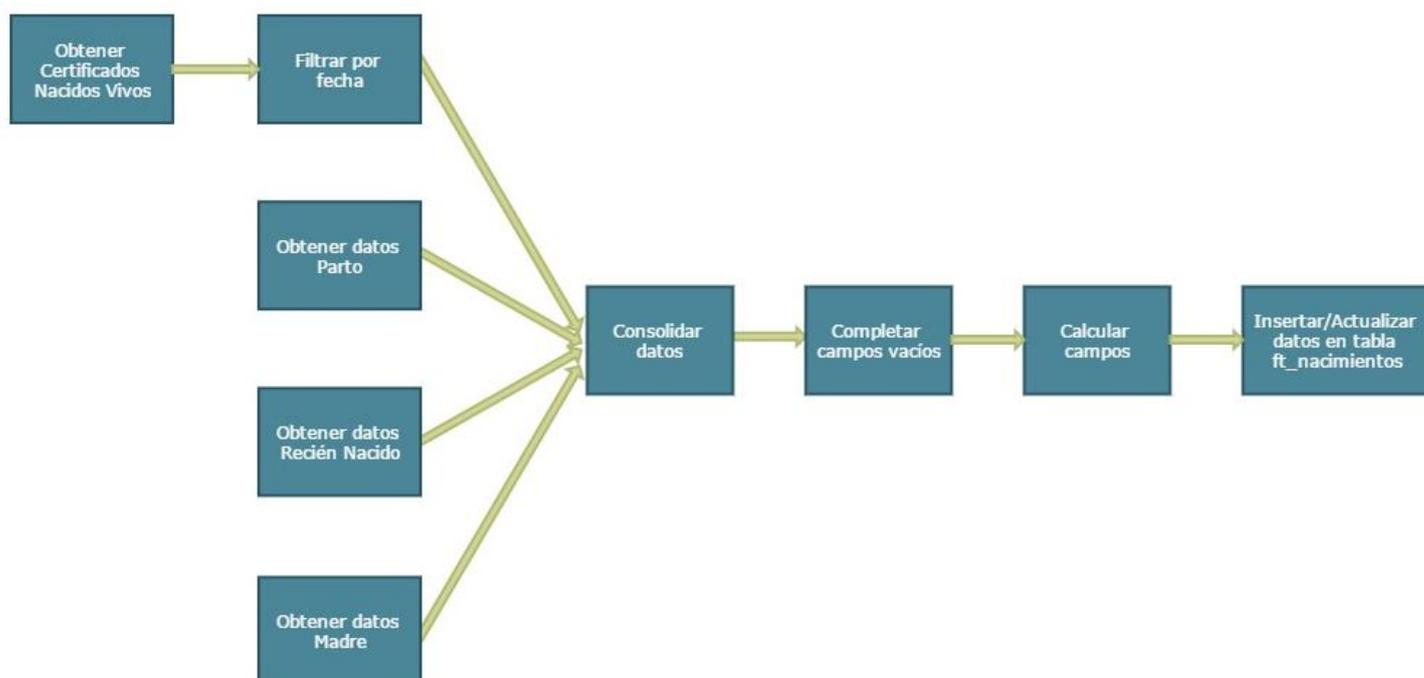


Figura 5.13 - Carga ft\_nacimientos

Al igual que en procesos anteriores, en el primer paso se extraen los datos de los certificados de nacimiento desde la tabla *certificados\_nacidos\_vivos*. A continuación, se obtienen las fechas ingresadas por parámetro que definen el período ingresado para la actualización y/o carga de datos. En caso de que estas no sean vacías, se filtran los nacimientos cuya fecha de ocurrencia está dentro del período seleccionado. En caso contrario, se continúa trabajando con todos los datos extraídos

desde las fuentes que contengan fecha de parto no vacía. En paralelo, se obtienen los datos de los partos, los recién nacidos y las madres a partir de las tablas *cnv\_datos\_parto*, *cnv\_datos\_recien\_nacido* y *cnv\_datos\_madre* respectivamente.

Luego, se consolidan los datos. En caso de que se hayan filtrado los nacimientos por fecha, el resultado de la asociación contiene únicamente los datos de los nacimientos filtrados y los datos de los partos, los recién nacidos y las madres asociadas a ellos. Una vez realizada la asociación se completan los campos vacíos con los identificadores de las tuplas *dummy* de las dimensiones correspondientes.

En el paso siguiente se calculan los campos faltantes de la tabla de hecho. Para calcular el identificador de la madre, se concatena el identificador y la serie del certificado. A continuación se calcula la clave subrogada de la fecha a partir del atributo *fecha\_parto*, y la edad de la madre, como la diferencia entre la fecha de parto y la fecha de nacimiento. Luego, se calcula el identificador del sexo tal como se muestra en el Cuadro 5.9.

Atributo <i>sexo</i> en fuentes	idSexo
F	1
M	2
I	3
D	4
Vacío	5
Cualquier valor que no se encuentre en los listados anteriormente	

Cuadro 5.9 – Tabla de hecho Nacimientos - Sexo

Los demás componentes de la clave de la tabla de hecho Nacimientos no necesitan ser calculados ya que vienen directo de las fuentes. Finalmente se insertan/actualizan, según corresponda, los datos en la tabla *ft\_nacimientos*.

### **Tabla de hecho Mortalidad**

En las Figuras 5.14.1 y 5.14.2 se puede observar el proceso de carga y actualización de la tabla de hecho Mortalidad.

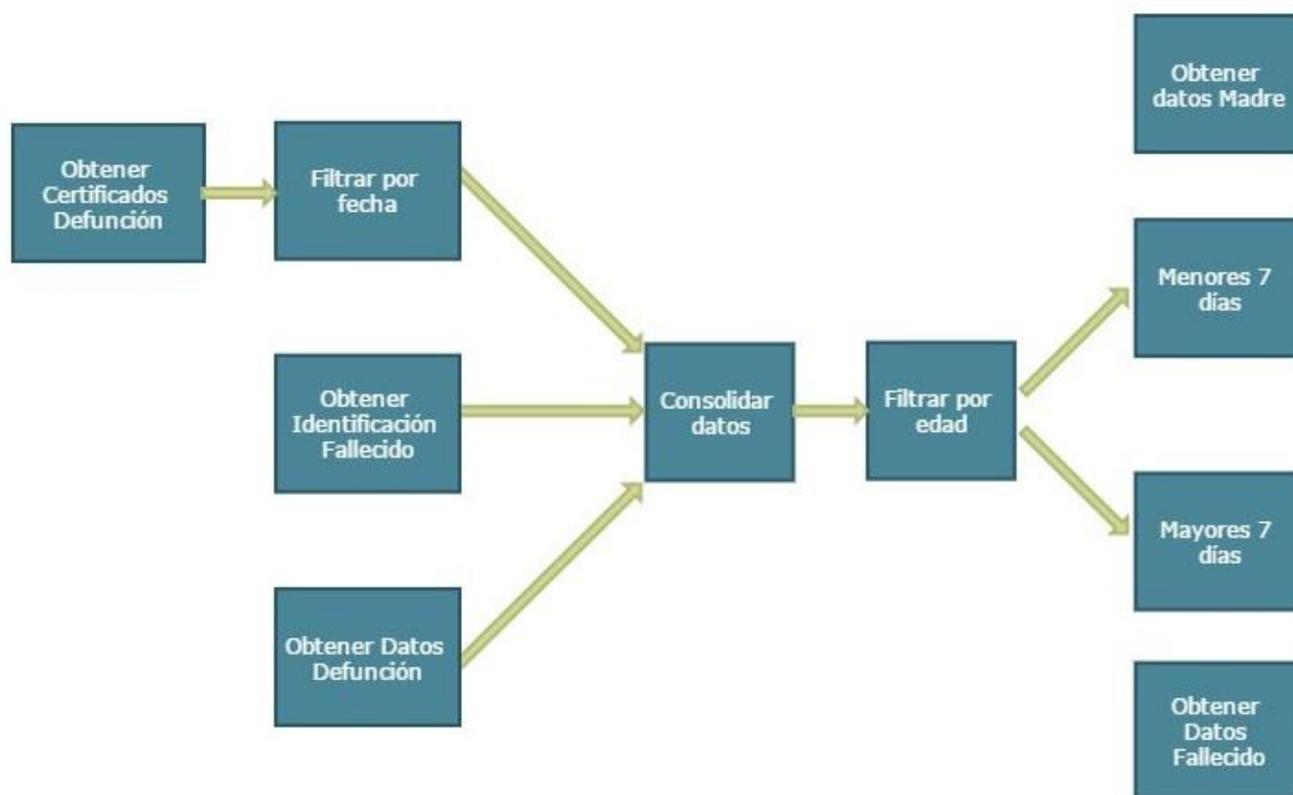


Figura 5.14.1 - Carga ft\_mortalidad

De forma análoga al proceso de carga de la dimensión Fallecidos, en los primeros pasos de este proceso se recuperan los datos de los certificados de defunción no fetales de la tabla *certificados\_defuncion* y se filtran de acuerdo a las fechas ingresadas por parámetro. Esto último se realiza en caso de que las fechas ingresadas no se encuentren vacías. En caso contrario, se continúa trabajando con todos los datos recuperados que tengan fecha de muerte no vacía. Paralelamente, se recuperan los datos de los fallecidos y la identificación de los mismos desde las tablas *cd\_datos\_fallecido* y *cd\_identificacion\_fallecido* de la base de datos fuente. Luego se consolidan los datos de modo de respetar los filtros previamente aplicados en caso que corresponda.

A continuación, se calcula la edad del fallecido en días, como la diferencia entre la fecha de muerte y la fecha de nacimiento. Una vez calculada la edad, se divide el flujo de datos en dos grupos; por un lado los menores de 7 días, y por el otro, los fallecidos de 7 ó más días de edad. Esto se realiza ya que los datos del domicilio y localidad de residencia del fallecido provienen de diferentes tablas fuentes de acuerdo a la edad del mismo. Para los casos en que el fallecido es menor de 7 días, el domicilio y la localidad de residencia se obtienen a partir de los datos de la madre; mientras que para los mayores dichos datos se obtienen a partir de los datos del fallecido.

Concurrentemente se obtienen los datos de las madres y de los fallecidos, a partir de las tablas *cd\_identificacion\_madre* y *cd\_datos\_fallecido* respectivamente. Es pertinente aclarar que la tabla *cd\_identificacion\_madre* sólo contiene datos para los fallecidos menores de 7 días.

Todos los pasos previamente descritos son los que se pueden observar en la Figura 5.14.1.

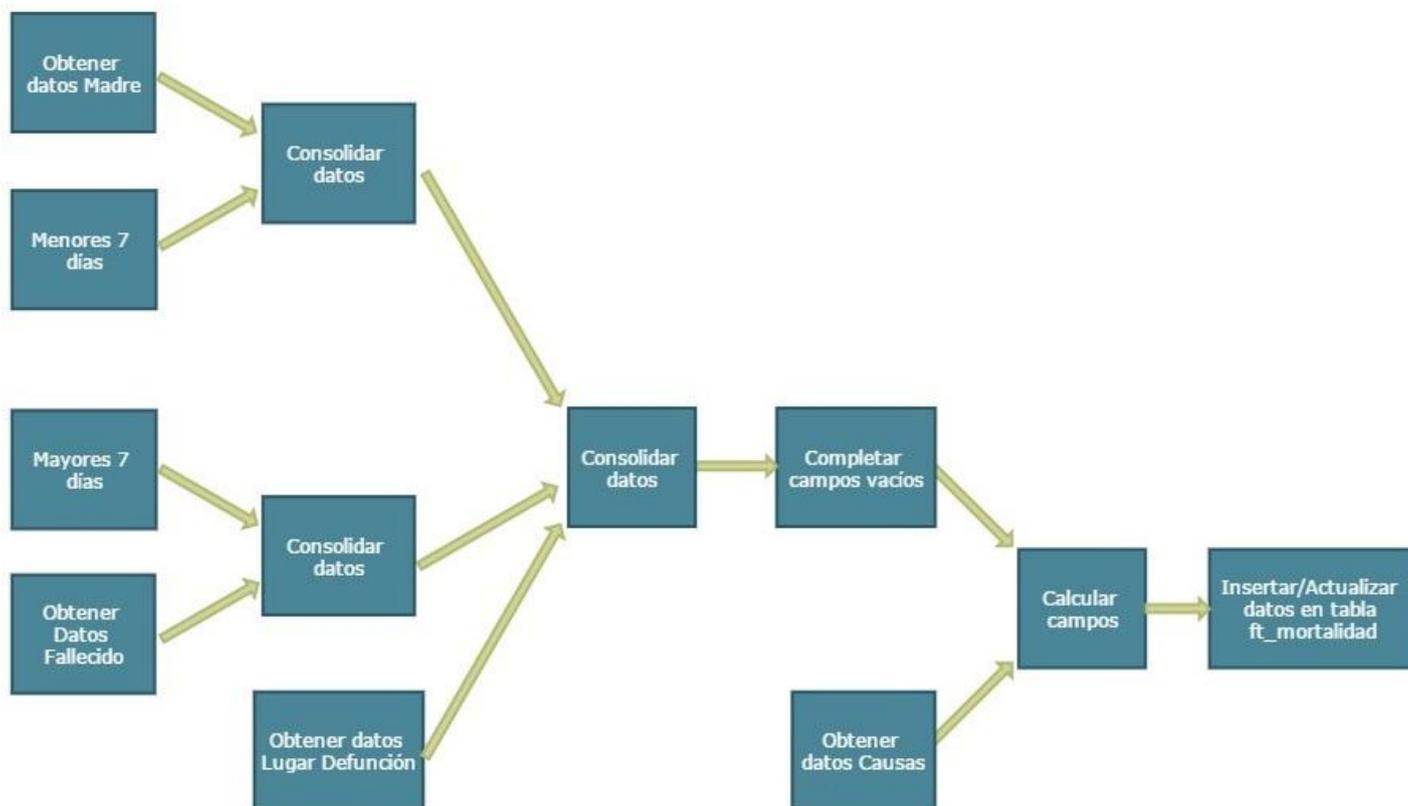


Figura 5.14.2 - Carga ft\_mortalidad

Continuando con el proceso, una vez obtenidos todos los datos previamente mencionados, se procede a realizar las siguientes asociaciones: por un lado se asocian los datos de los fallecidos menores de 7 días con los datos de las madres, y por otro, se asocian los datos de los fallecidos de 7 ó más días de edad con los datos obtenidos a partir de la tabla *cd\_datos\_fallecido*. Además, en paralelo, se obtienen los datos de los lugares de defunción.

En el siguiente paso se consolidan todos los datos previamente obtenidos; es decir, se asocian los datos de los menores de 7 días y sus madres junto con los datos de los fallecidos mayores junto con los datos de los lugares de defunción, obteniendo así todos los datos necesarios de las fuentes para cargar la tabla de hecho Mortalidad. Luego, al igual que en procesos anteriores, se eliminan los campos vacíos completándolos con los identificadores de las tuplas *dummy* correspondientes.

Al mismo tiempo, se extraen los datos de las causas CIE10 a partir de la tabla *d\_causascie10* del Data Warehouse, previamente cargada. Esto se hace con el fin de validar que todos los códigos de las enfermedades obtenidas a partir de los datos de las fuentes, específicamente los valores del atributo

*Enf\_Causante\_Cie10* de la tabla *cd\_datos\_defuncion*, tengan su correspondiente en la dimensión CausasCIE10, ya que el origen de los datos utilizados para cargar dicha dimensión es distinto al origen de los datos utilizados para cargar la tabla de hecho Mortalidad.

Posteriormente, se calculan los campos faltantes de la tabla de hecho. En cuanto al identificador del fallecido, el mismo se calcula como la concatenación del identificador y la serie del certificado. Luego, se calcula la clave subrogada de la fecha a partir del atributo *fecha\_muerte* y el identificador del sexo de manera análoga a como se hizo para la tabla de hecho Nacimientos. Finalmente se insertan y/o actualizan los datos, según corresponda, en la tabla *ft\_mortalidad*.

Las últimas etapas del proceso se pueden observar en la Figura 5.14.2.

### **Tabla de hecho Mortalidad Infantil**

Las Figuras 5.15.1 y 5.15.2 ilustran el proceso de carga y actualización de la tabla de hecho Mortalidad Infantil.

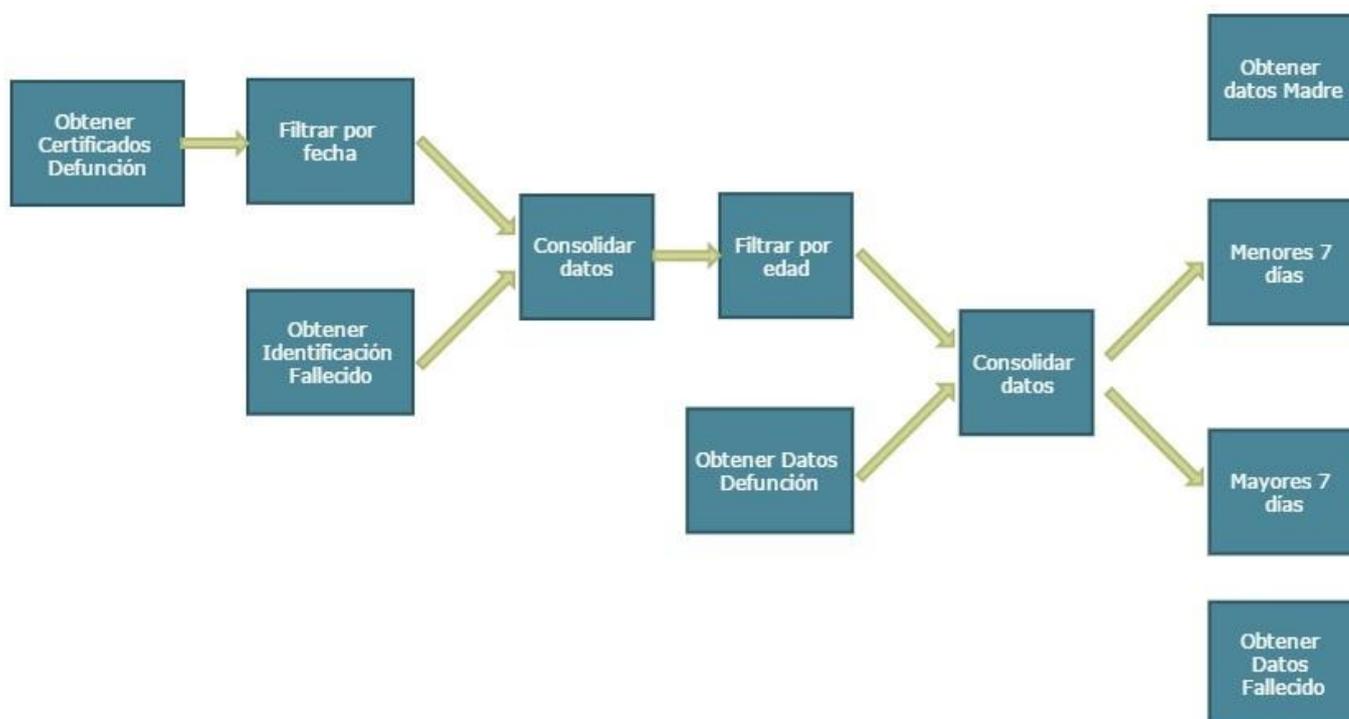


Figura 5.15.1 - Carga *ft\_mortalidad\_infantil*

Las etapas de este proceso son análogas a las del proceso de carga y actualización de la tabla de hecho Mortalidad Infantil. La única diferencia es que, una vez obtenidos y consolidados los datos de los certificados de defunción e identificación de los fallecidos, se calcula la edad del fallecido (como

la diferencia entre la fecha de muerte y la fecha de nacimiento) y realiza un filtro, de forma de que sólo permanezcan los datos de los fallecidos menores de 5 años.

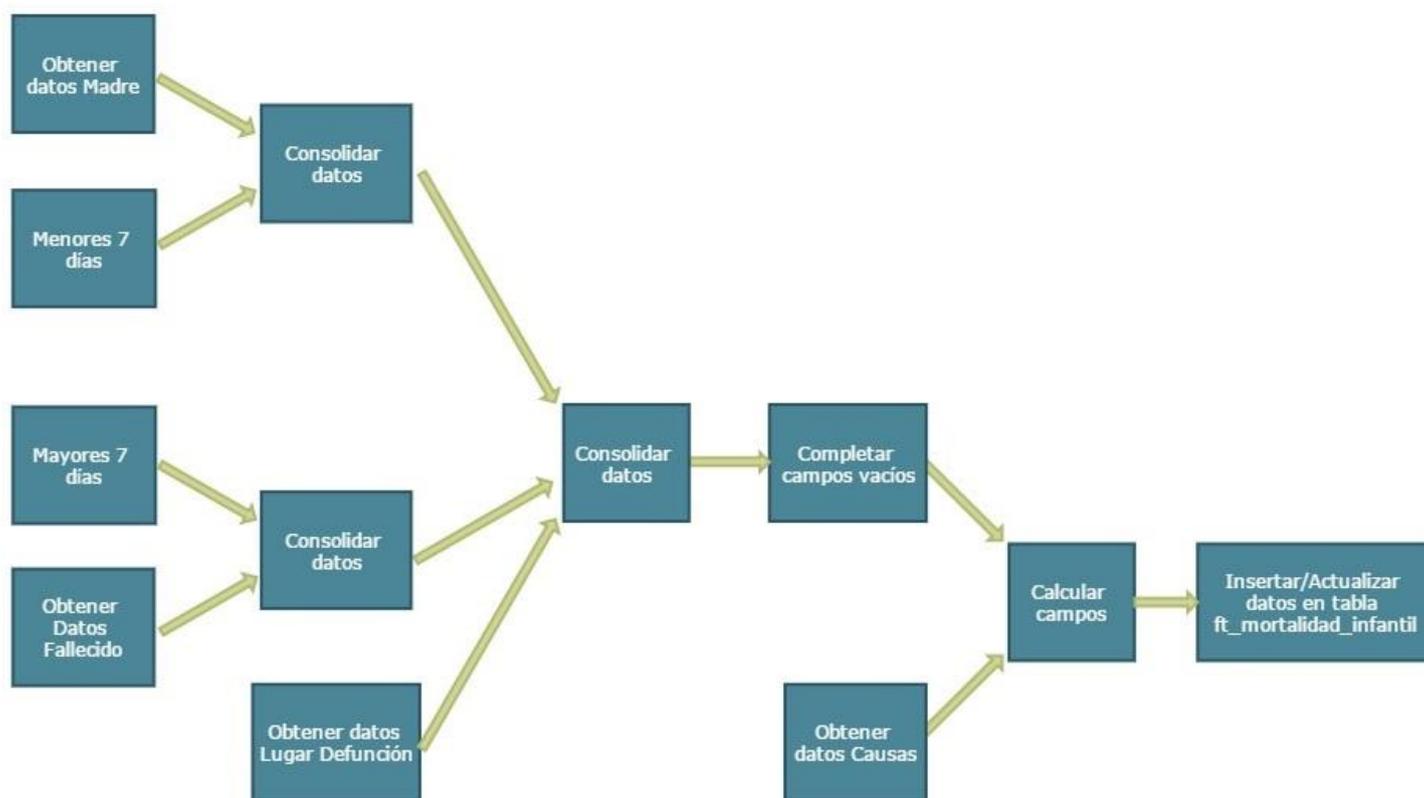


Figura 5.15.2 - Carga ft\_mortalidad\_infantil

### 5.2.3 Verificación de los procesos de ETL

Una vez implementados los procesos de extracción, transformación y carga, se verificó que el resultado de los mismos fuera correcto, es decir, que el Data Warehouse se encontrara completamente cargado y sin inconsistencias con respecto a las fuentes. Para ello, se implementaron consultas SQL sobre las fuentes que retornan los datos que se deberían haber cargado en las tablas de dimensión y de hecho. Luego, sus resultados fueron comparados con el contenido del Data Warehouse, también mediante sentencias SQL. Cabe mencionar que esta verificación se realizó únicamente para las tablas de hecho y de dimensión cuyo contenido proviene directamente de las fuentes, dejando por fuera las tablas de dimensión autogeneradas, como es el caso de Tiempo, Sexo y Edades.

Si bien la realización de esta etapa de verificación requirió un esfuerzo adicional, dada la complejidad de las consultas SQL que se debieron implementar, se considera que tuvo un resultado positivo ya que permitió detectar y corregir errores en etapas tempranas del proyecto. El detalle de dichas consultas SQL se puede encontrar en el Anexo III.

## 5.3 Implementación de cubos

A continuación se presenta una breve introducción al motor OLAP Mondrian y al lenguaje MDX, así como el detalle de los cubos implementados utilizando dicha herramienta y su verificación.

### 5.3.1 Implementación de cubos en Mondrian

Para implementar los cubos se utilizó la herramienta Pentaho Schema Workbench. Como se mencionó en la descripción de la misma, permite trabajar con los archivos XML que definen el esquema de la base de datos multidimensional en Mondrian. Dicho esquema contiene un modelo lógico cuyos componentes principales son cubos, medidas y dimensiones, y los mapeos de este modelo al esquema lógico relacional. Típicamente, como ocurre en el caso de este proyecto, el esquema lógico relacional es del tipo estrella, pero pueden ser utilizados otros diseños como el esquema copo de nieve.

Un cubo es una colección de medidas y dimensiones, que tienen en común una tabla de hecho. Esta última está compuesta por los atributos con los que se calculan las medidas y las referencias a las tablas que corresponden a las dimensiones.

Para definir una dimensión en Mondrian es necesario establecer el atributo *foreignKey*, que pertenece a la tabla de hecho y relaciona a la misma con la dimensión; y las jerarquías que la componen. Estas últimas quedan determinadas mediante una tabla; el atributo *primaryKey*, que pertenece a dicha tabla y se corresponde con el atributo *foreignKey* de la dimensión; y la definición de los niveles que la constituyen. Esto permite que las jerarquías pertenecientes a una misma dimensión tengan distintas tablas asociadas, siempre y cuando estas estén relacionadas a la tabla de hecho del cubo mediante el mismo atributo, específico de la dimensión. En lo que refiere al proyecto, las tablas y los atributos *primaryKey* serán los mismos para todas las jerarquías dentro de una misma dimensión, al estar diseñada la base de datos del Data Warehouse mediante un esquema estrella. Por otra parte, cada nivel tiene asociada una clave dentro de la tabla de la jerarquía a la que pertenece que se especifica en el atributo *column*, y un atributo *nameColumn* que define los nombres de los miembros de dicho nivel.

Cabe destacar que algunas dimensiones pueden ser utilizadas por múltiples cubos y para ello deben estar definidas como dimensiones compartidas. Es importante señalar también que por defecto el nivel más alto de una jerarquía es '(All)' y contiene un único miembro llamado '(All {nombreJerarquía})'. Este miembro es padre de todos los otros miembros de la jerarquía y por lo tanto representa un total general.

Las medidas se pueden calcular utilizando las funciones de agregación provistas por Mondrian, que son *sum*, *count*, *min*, *max*, *avg* y *distinct count*, sobre alguna de las columnas de la tabla de hecho asociada al cubo. Si se requiere utilizar funciones de agregación más complejas, Mondrian permite

utilizar miembros calculados, que se definen en base a una combinación de datos del cubo, operadores aritméticos, números y funciones utilizando fórmulas en el lenguaje MDX.

Mondrian no permite realizar funciones de *drill-across*, pero provee la funcionalidad de crear cubos virtuales, que permiten unir dos cubos existentes pudiendo así analizar dimensiones y medidas de ambos como si fueran uno solo.

## *Expresiones multidimensionales (MDX)*

MDX es un lenguaje propuesto como estándar por Microsoft para realizar consultas sobre bases de datos OLAP [20]. En esta sección se presentará una breve descripción de dicho lenguaje para que al lector le sea más sencillo comprender algunas expresiones en MDX que se encuentran en secciones posteriores.

El lenguaje MDX permite analizar y realizar cálculos sobre estructuras multidimensionales, pero no permite modificarlas ni manipularlas. Entre los conceptos claves de dicho lenguaje se encuentran cubos, medidas, dimensiones, jerarquías y niveles. Cabe destacar que en MDX las medidas forman parte de una dimensión especial denominada *Measures*. Además, MDX maneja el concepto de miembro que representa un dato concreto dentro de una dimensión, siendo este el nivel más bajo de referencia cuando se describe una celda en un cubo. Por ejemplo, el año 2014 es un miembro de la dimensión Tiempo, como también lo es el valor Si del nivel Embarazada de la dimensión Fallecidos.

En una expresión MDX, para hacer referencia de forma inequívoca a dimensiones, jerarquías, niveles o miembros dentro de un cubo se utilizan nombres únicos, que pueden verse como rutas dentro del cubo y están compuestos por el nombre del objeto y de los ancestros de este (dimensión, jerarquía y nivel según corresponda). Aunque no es necesario, excepto en los casos en que el nombre del objeto contiene espacios, se suele delimitar a los mismos con paréntesis rectos ([ ]). Como ejemplo, para acceder al miembro Enero del nivel Mes de la dimensión Tiempo se debe utilizar el nombre único *[Tiempo].[Mes].[Enero]*. En caso de que se quiera trabajar con alguna medida, como por ejemplo *Cant\_Nacimientos*, se debe consultar a la dimensión *Measures* de la forma *[Measures].[Cant\_Nacimientos]*.

Dentro de una expresión MDX, los miembros pueden estar organizados en tuplas o en conjuntos. Una tupla es una colección ordenada de miembros de una o más dimensiones cuya intersección identifica una sección del cubo. Si está compuesta por un miembro de cada dimensión describe completamente una celda del mismo. Cuando una tupla contiene más de un miembro, estos deben aparecer entre paréntesis curvos ((,)) como se muestra en el siguiente ejemplo, en donde se desea obtener la medida *Cant\_Nacimientos* para el mes de Enero:

```
([Measures].[Cant_Nacimientos], [Tiempo].[Mes].[Enero])
```

Es importante mencionar que en este caso el motor OLAP deberá aplicar la función de agregación correspondiente para la medida `Cant_Nacimientos` al hacer un *roll-up* sobre la dimensión `Tiempo`.

Por otra parte, un conjunto es una colección ordenada de cero, una o más tuplas. En la sintaxis MDX, las mismas deben estar entre llaves (`{}`) como se muestra en el ejemplo a continuación, en donde se precisa obtener la medida `Cant_Fallecimientos` para el mes de Febrero y `Razon_MortalidadMaterna` para el mes de Diciembre:

```
{ ([Measures].[Cant_Fallecimientos], [Tiempo].[Mes].[Febrero]),  
  ([Measures].[Razon_MortalidadMaterna], [Tiempo].[Mes].[Diciembre]) }
```

Una consulta básica escrita en lenguaje MDX tiene, a grandes rasgos, la estructura que se muestra en la Figura 5.16.

```
SELECT [<especificacion_eje>  
       [, <especificacion_eje>...]]  
FROM  [<especificacion_cubo>]  
[WHERE [<especificacion_eje_segmentador>]]
```

Figura 5.16 - Estructura básica de una consulta MDX [21]

Esta consulta especifica el subconjunto de datos multidimensionales que deben ser devueltos de un determinado cubo. La cláusula `SELECT` determina la cantidad de ejes y el conjunto de tuplas a incluir en cada uno de ellos. Una consulta MDX soporta hasta 128 ejes, pero lo más común es utilizar únicamente dos, para los cuales se utilizan los alias `COLUMNS` y `ROWS`, obteniendo de esta forma el resultado en un formato matricial. Por otra parte, la cláusula `FROM` determina qué cubo va a ser consultado, y la cláusula `WHERE`, opcional, qué tupla se utilizará para restringir los datos devueltos. Resulta útil destacar que el miembro `All` de cada jerarquía es utilizado por defecto para calcular los valores de las celdas cuando la misma no está incluida en los ejes del `SELECT` o en la especificación del `WHERE`.

Así, si se desea analizar la cantidad de nacimientos de niñas para los años 2014 y 2015 en el cubo `Nacimientos` la consulta a ejecutar debe ser la presentada en la Figura 5.17.

```
SELECT {([Measures].[Cant_Nacimientos]} ON COLUMNS  
       {([Tiempo].[Anio].[2014], [Tiempo].[Anio].[2015]} ON ROWS  
FROM  Nacimientos  
WHERE ([Sexo].[Femenino])
```

Figura 5.17 - Ejemplo de una consulta MDX

Adicionalmente, como parte de la consulta MDX es posible crear miembros y medidas que no estén almacenados en el cubo sino que sean calculados en tiempo de ejecución. Estos son denominados **miembros calculados** y se definen mediante la cláusula `WITH` que se incluye al comienzo de la consulta junto con el nombre único del nuevo miembro y la expresión con la que se debe calcular. El alcance de dicho miembro se limita al ámbito de la consulta, es decir que cada vez que se quiera utilizar debe ser incluido en la misma. Sin embargo, Mondrian permite definir miembros calculados

dentro del esquema del modelo lógico, para poder ser utilizados dentro de cualquier consulta sin necesidad de incluir su definición en esta [20]. Para ello basta con señalar el nombre del nuevo miembro, la dimensión y jerarquía a la cual va a pertenecer, y la fórmula MDX para calcularlo. Esta última puede estar compuesta por identificadores, funciones, constantes y operadores. Supongamos que se desea obtener la medida Proporción\_Niñas que se calcula como la cantidad de niñas nacidas sobre el total de nacimientos. Entonces, para ello se debe incluir en el esquema Mondrian dentro del cubo Nacimientos la definición XML especificada en la Figura 5.18.

```
<CalculatedMember name="Proporcion_Niñas" dimension="Measures">
  <Formula>{ [Measures].[Cant_Nacimientos], [Sexo].[Femenino] } /
  [Measures].[Cant_Nacimientos]
</Formula>
</CalculatedMember>
```

Figura 5.18 - Ejemplo en Mondrian de miembros calculados

Cabe resaltar que al incluir esta medida en una consulta MDX, el valor de la misma se obtiene agregando las medidas que son utilizadas en la fórmula y realizando luego los cálculos pertinentes. Como ejemplo vamos a utilizar la consulta de la Figura 5.19 con la cual se desea obtener la proporción de niñas nacidas en el año 2014.

```
SELECT { [Measures].[Proporcion_Niñas] } ON COLUMNS
      { [Tiempo].[Anio].[2014] } ON ROWS
FROM Nacimientos
```

Figura 5.19 - Ejemplo en Mondrian consulta de miembros calculados

Para calcular la medida Proporción\_Niñas se agregan el numerador y el denominador por separado y luego se realiza la división entre los dos resultados. El numerador se corresponde con el valor de la medida Cant\_Nacimientos para los miembros [Sexo].[Femenino] y [Tiempo].[Anio].[2014], y el miembro All de las jerarquías restantes del cubo Nacimientos ya que estas no aparecen en ningún eje de la consulta. Por otra parte, el denominador se corresponde con el valor de la medida Cant\_Nacimientos para el miembro [Tiempo].[Anio].[2014] y el miembro All de las jerarquías restantes de dicho cubo, incluido el miembro All de la jerarquía Sexo. Luego para obtener el resultado de la consulta se debe dividir el valor obtenido para el numerador sobre el valor obtenido para el denominador. Así, si la cantidad de niñas nacidas en el 2014 es 20.000 y la cantidad de nacimientos totales en ese año es 40.000, entonces el resultado de la consulta será  $\frac{1}{2}$ .

Con esta breve introducción acerca del lenguaje MDX, resultará más sencillo comprender las fórmulas utilizadas para la definición de los miembros calculados que se presentarán en la siguiente sección, la cual detalla las características de los cubos implementados.

## 5.3.2 Cubos implementados

De acuerdo a las relaciones dimensionales detalladas en el Capítulo 4 se crearon tres cubos, Nacimientos, Mortalidad y Mortalidad\_Infantil. Como se mencionó en dicho capítulo, para calcular las medidas Razon\_MortalidadMaterna y Tasa\_MortalidadInfantil es necesario realizar un *drill-across* entre el cubo Nacimientos y los cubos Mortalidad y Mortalidad\_Infantil respectivamente. Sin embargo, Mondrian no permite realizar esta operación, por lo que se debieron implementar dos cubos virtuales denominados Mortalidad\_Virtual y Mortalidad\_Infantil\_Virtual, de forma tal de poder unir los cubos originales con el cubo Nacimientos y trabajar con las dimensiones y medidas de ambos como si fueran uno solo.

Todos los cubos implementados contienen tanto dimensiones propias como compartidas. Estas últimas son Tiempo, Localidad Ocurrencia, Localidad Residencia, Domicilio Residencia, Sexo y Edades. A continuación se presenta una breve descripción de las mismas, así como también, la implementación de cada cubo. En el Anexo IV se puede encontrar el archivo XML con la especificación del esquema.

## Dimensiones compartidas

En la Figura 5.20 se especifica cómo están compuestas las dimensiones compartidas. Dichas dimensiones fueron creadas de esta forma ya que era necesario acceder a las mismas desde todos los cubos implementados.

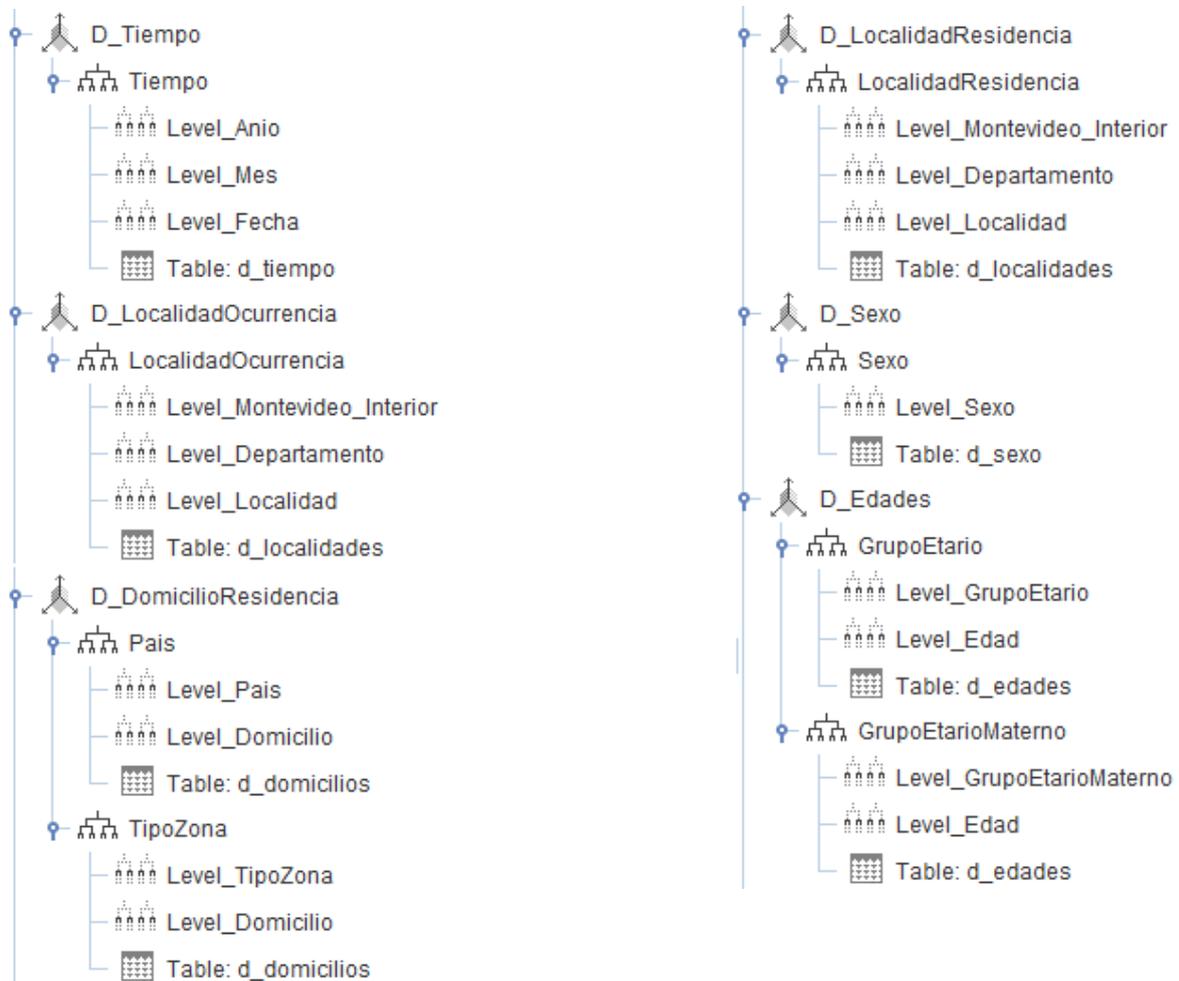


Figura 5.20 - Especificación de dimensiones compartidas

## Cubo Nacimientos

El cubo Nacimientos se definió en base a la relación dimensional Nacimientos presentada en el Capítulo 4, y por lo tanto, la tabla de hecho asociada al mismo es *ft\_nacimientos*. En la Figura 5.21 se pueden observar las dimensiones, tanto las propias como las compartidas, así como las medidas que lo componen.

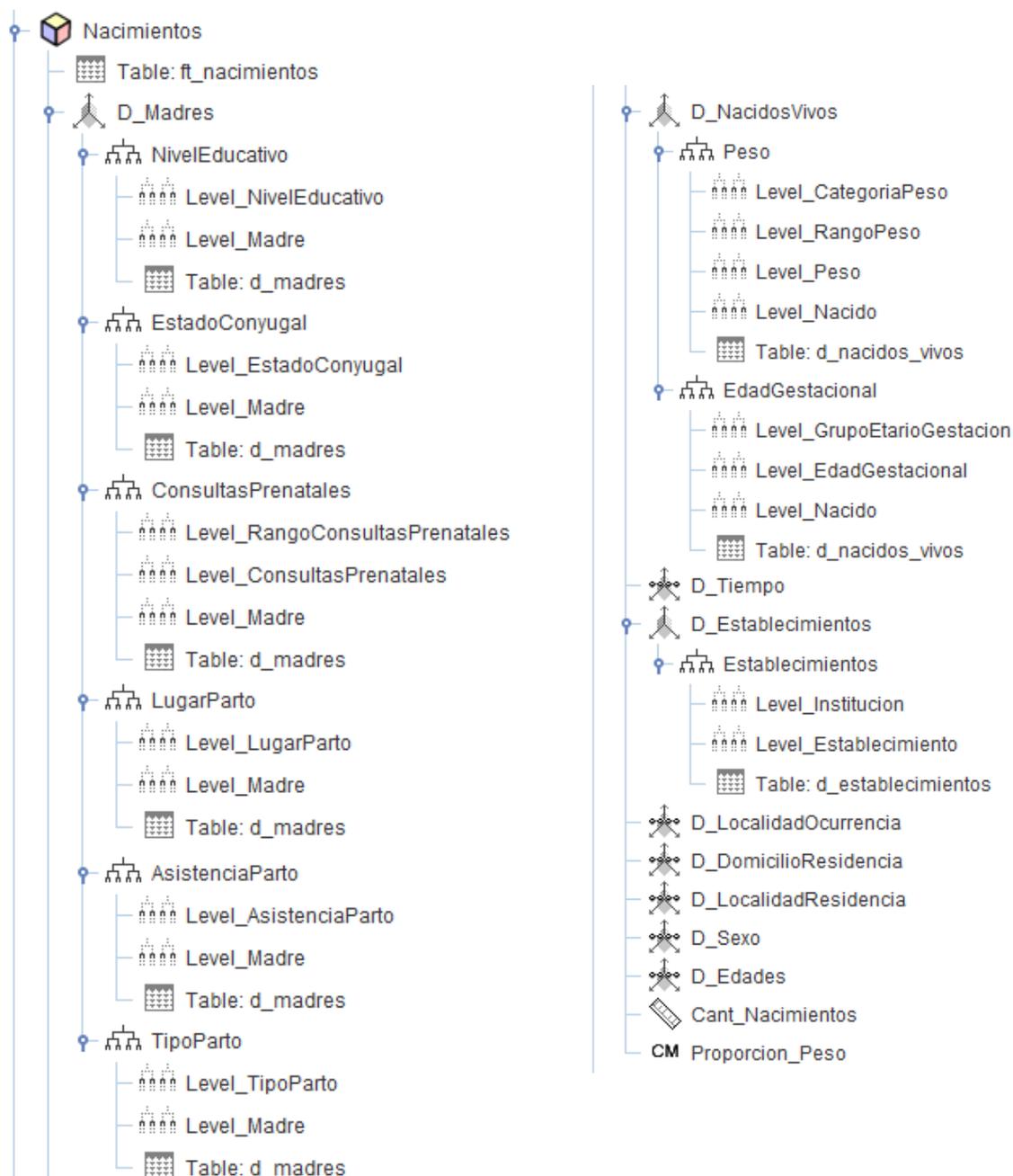


Figura 5.21 - Dimensiones del cubo Nacimientos

La medida `Cant_Nacimientos` se calcula utilizando la función de agregación `count` provista por Mondrian, sobre el atributo `cedulaNacido` de la tabla de hecho `ft_nacimientos`, es decir, contando la cantidad de registros de la tabla de hecho. Por otra parte, para calcular la medida `Proporcion_PorPeso` fue necesario crear un miembro calculado, ya que esto no era posible con las funciones de agregación existentes. Esta última se calcula como:

$$(Cantidad\ de\ nacidos\ vivos\ con\ determinado\ peso / Cantidad\ de\ nacidos\ vivos) \times 100.$$

Por lo que la fórmula del miembro calculado utilizado fue la siguiente:

```
([Measures].[Cant_Nacimientos]) /  
([Measures].[Cant_Nacimientos], [D_NacidosVivos].[Peso].[All]) * 100
```

El numerador de esta fórmula es la medida `Cant_Nacimientos` y el denominador es la medida `Cant_Nacimientos` agrupada por el miembro `All` de la jerarquía `Peso`, es decir, la cantidad total de nacimientos sin importar el peso del recién nacido.

Al seleccionar un nivel de la jerarquía `Peso` de la dimensión `Nacidos Vivos` en una consulta, por ejemplo `CategoriaPeso`, la cantidad de nacimientos del numerador queda agrupada por categoría mientras que la del denominador no, cumpliendo así con la definición de la medida.

Por otra parte, al seleccionar un nivel de una jerarquía distinta de `Peso` en una consulta, los datos se agrupan por este nivel tanto en el denominador como en el numerador, mientras que si no están presentes se toma el nivel más alto por defecto, `All`.

## Cubo Mortalidad

Este cubo implementa la relación dimensional Mortalidad. Por este motivo, la tabla de hecho asociada al mismo es *ft\_mortalidad*. Las dimensiones, tanto propias como compartidas, y las medidas que lo conforman se pueden observar en la Figura 5.22.

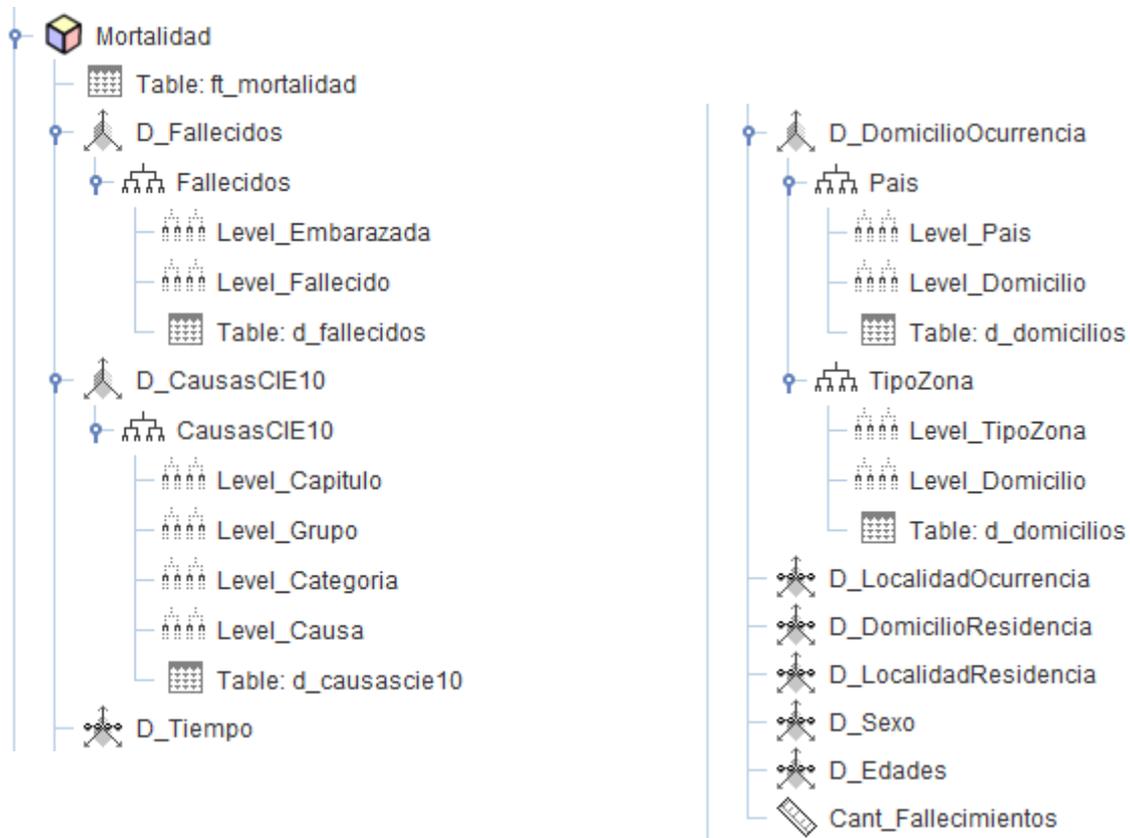


Figura 5.22 - Dimensiones del cubo Mortalidad

La medida Cant\_Fallecimientos se calcula utilizando la función de agregación *count* provista por Mondrian, sobre el atributo *idFallecido* de la tabla de hecho *ft\_mortalidad*.

## Cubo Mortalidad Infantil

El cubo Mortalidad Infantil está definido por la relación dimensional Mortalidad\_Infantil, por lo que la tabla de hecho asociada a dicho cubo es *ft\_mortalidad\_infantil*. En cuanto a las dimensiones, tanto propias como compartidas, y medidas que lo componen, las mismas se pueden observar en la Figura 5.23.

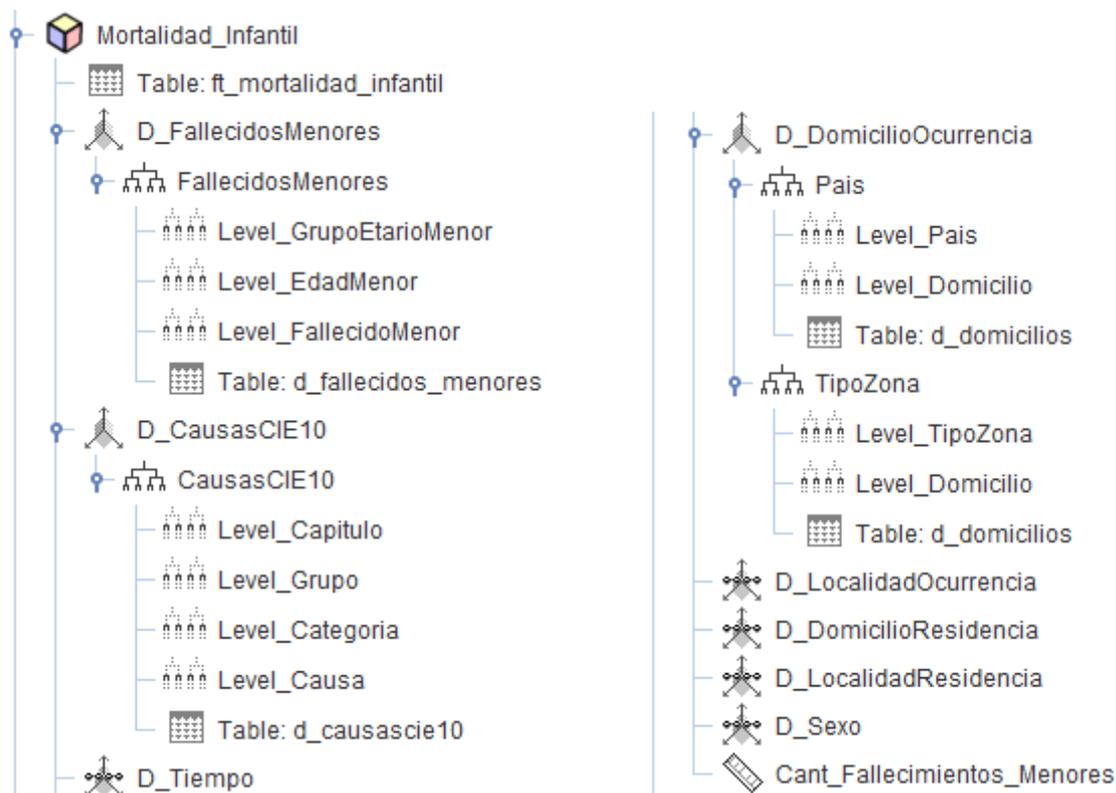


Figura 5.23 - Dimensiones del cubo Mortalidad Infantil

De forma análoga a la medida Cant\_Fallecimientos del cubo Mortalidad, la medida Cant\_Fallecimientos\_Menores se calcula utilizando la función de agregación *count* provista por Mondrian, sobre el atributo *idFallecidoMenor* de la tabla de hecho *ft\_mortalidad\_infantil*.

## ***Cubo virtual Mortalidad***

La necesidad de crear un cubo virtual surgió del hecho de que para calcular la medida Razon\_MortalidadMaterna se precisaba, además de las dimensiones y la medida Cant\_Fallecimientos del cubo Mortalidad, la medida Cant\_Nacimientos del cubo Nacimientos. La fórmula de esta medida es la siguiente:

$$(Cantidad\ de\ defunciones\ maternas / Cantidad\ de\ nacidos\ vivos) \times 100.000$$

Para ello se importaron todas las dimensiones del cubo Mortalidad, junto con la medida Cant\_Fallecimientos, y la medida Cant\_Nacimientos del cubo Nacimientos. Luego, se creó el miembro calculado Razon\_Mortalidad\_Materna, cuya fórmula es:

```
([Measures].[Cant_Fallecimientos], [D_Fallecidos].[Level_Embarazada].[Si]) /  
([Measures].[Cant_Nacimientos], [D_Fallecidos].[Fallecidos].[All],  
[D_DomicilioOcurrencia].[Pais].[All], [D_DomicilioOcurrencia].[TipoZona].[All],  
[D_CausasCIE10].[CausasCIE10].[All]) * 100000
```

En el numerador se obtiene la medida Cant\_Fallecimientos para el miembro *Si* del nivel Embarazada de la dimensión Fallecidos (esto indica que la mujer fallecida estaba embarazada o en parto o puerperio), con la expresión *[D\_Fallecidos].[Level\_Embarazada].[Si]*.

Asimismo, en el denominador se obtiene la medida Cant\_Nacimientos, agrupada por el miembro *All* de las jerarquías que componen las dimensiones Fallecidos, Domicilio Ocurrencia y Causas CIE10, ya que las mismas no existen en el cubo Nacimientos por lo que no es posible agrupar esta medida por dichas dimensiones.

## ***Cubo virtual Mortalidad Infantil***

Este cubo virtual fue creado con el fin de calcular la medida Tasa\_MortalidadInfantil ya que, además de las dimensiones y la medida Cant\_Fallecimientos\_Menores del cubo Mortalidad Infantil, se precisaba la medida Cant\_Nacimientos del cubo Nacimientos. A continuación se presenta la fórmula de esta medida:

$$(Cantidad\ de\ defunciones\ menores\ de\ 5\ años / Cantidad\ de\ nacidos\ vivos) \times 1000$$

Para ello se importaron todas las dimensiones del cubo Mortalidad Infantil, junto con la medida Cant\_Fallecimientos\_Menores, y la medida Cant\_Nacimientos del cubo Nacimientos. Posteriormente, se creó el miembro calculado Tasa\_MortalidadInfantil cuya fórmula se puede representar como:

```
([Measures].[Cant_Fallecimientos_Menores]) /
([Measures].[Cant_Nacimientos], [D_FallecidosMenores].[FallecidosMenores].[All],
[D_DomicilioOcurrencia].[Pais].[All], [D_DomicilioOcurrencia].[TipoZona].[All],
[D_CausasCIE10].[CausasCIE10].[All]) * 1000
```

En el numerador se obtiene la medida Cant\_Fallecimientos\_Menores, mientras que en el denominador se obtiene la medida Cant\_Nacimientos agrupada por el miembro All de las jerarquías que componen las dimensiones Fallecidos Menores, Causas CIE10 y Domicilio Ocurrencia, ya que las mismas no existen en el cubo Nacimientos por lo que no es posible agrupar dicha medida por estas dimensiones.

### 5.3.3 Verificación de la implementación de cubos

Luego de implementados los cubos se llevó a cabo una etapa de verificación de los mismos. Para ello, se realizaron consultas sobre los cubos utilizando la herramienta Saiku y sus resultados se compararon con los datos publicados por el MSP [11]. En el caso de los nacimientos, se constató que los datos coincidían o presentaban desviaciones poco significativas con respecto a los publicados. En los Cuadros 5.10 y 5.11 se presentan algunos ejemplos de las comparaciones realizadas para los nacimientos.

Cantidad de nacidos vivos por año		
	DW	MSP
2013	48.701	48.681
2014	48.387	48.368

Cuadro 5.10 – Verificación de cubo Nacimientos

Cantidad de nacidos vivos por sexo (Masculino / Femenino) y departamento de residencia materna para el año 2014				
	DW		MSP	
	Masculino	Femenino	Masculino	Femenino
Artigas	579	542	578	542
Canelones	3.501	3.303	3.509	3.319
Cerro Largo	667	638	667	637
Colonia	916	797	915	799
Durazno	431	450	430	447
Flores	161	146	161	146
Florida	450	413	451	413
Lavalleja	383	342	389	343

Maldonado	1.335	1.279	1.338	1.281
Montevideo	10.193	9.569	10.274	9.644
Paysandú	878	817	879	819
Río Negro	446	451	446	448
Rivera	827	809	827	811
Rocha	504	458	506	457
Salto	1.102	1.145	1.103	1.145
San José	735	735	741	737
Soriano	587	543	587	543
Tacuarembó	647	635	646	641
Treinta y tres	305	327	304	326

*Cuadro 5.11 – Verificación de cubo Nacimientos (II)*

Sin embargo, al momento de corroborar los datos sobre mortalidad se observaron desviaciones significativas. Dichas diferencias surgen a raíz de las fuentes provistas, cuyo contenido no coincide exactamente con los datos publicados. Por ejemplo, el número total de defunciones para el período 2012 – 2014 publicado por el MSP es 98.271, mientras que en las fuentes, que contienen datos para el período 2012 – 2015, la cantidad total de tuplas de la tabla de defunciones es 96.287. Por este motivo, no fue posible utilizar las publicaciones del MSP para verificar la implementación de los cubos y se decidió elaborar sentencias SQL sobre las fuentes para comparar con los datos obtenidos a partir de los mismos. En el Cuadro 5.12 se muestran algunas de las comparaciones realizadas, que como se puede observar, coinciden en todos los casos.

Indicadores	DW			Fuentes		
	2013	2014	2015	2013	2014	2015
Tasa de mortalidad infantil, por 1.000 nacidos vivos (menores de 1 año)	6,3	3,5	5,9	6,3	3,5	5,9
Número de defunciones infantiles	308	169	212	308	169	212
Tasa de mortalidad infantil, por 1.000 nacidos vivos neonatal (de 0 a 27 días)	3,9	1,6	3,9	3,9	1,6	3,9
Tasa de mortalidad infantil, por 1.000 nacidos vivos postneonatal (de 28 días a <1 año)	2,4	1,9	2	2,4	1,9	2
Mortalidad de menores de 5 años reportada	7,4	4,2	6,7	7,4	4,2	6,7
Razón de mortalidad materna por 100.000 nacidos vivos	0	0	36,1	0	0	36,1
Número de defunciones maternas	0	0	13	0	0	13

*Cuadro 5.12 – Verificación de indicadores*



## Capítulo 6 – Interfaz de usuario

Con el fin de brindar una herramienta que facilite tanto la carga y actualización del Data Warehouse como la realización de consultas y la creación de planillas personalizadas, se implementó una aplicación web para uso exclusivo del personal de la UINS, que permite realizar dichas tareas de forma centralizada. Esta pretende ser lo más amigable posible con el usuario, pero requiere de cierto conocimiento del sistema y de las bases de datos para poder manejarse de forma adecuada en caso de que haya un error.

El *back-end* de la aplicación web fue desarrollado con el software Java Enterprise Edition versión 8 [22]. A fin de habilitar la carga y actualización del Data Warehouse desde la aplicación, se utilizaron las bibliotecas provistas por la herramienta Pentaho Kettle que permiten ejecutar trabajos y transformaciones mediante código Java. Además, se incluyó la biblioteca log4j [23], para la creación de logs, con el objetivo de dejar un registro diario del resultado de estas ejecuciones. Por otra parte, para poder realizar consultas MDX, crear tablas con los resultados y exportarlos en diferentes formatos se utilizó la biblioteca Pivot4j [24], que provee una API Java para servidores OLAP que facilita la creación de interfaces analíticas.

En cuanto al *front-end*, el mismo se desarrolló utilizando el *framework* Bootstrap [25], basado en HTML y CSS, y el lenguaje JavaScript, incluyendo las bibliotecas JQuery [26], para mejorar la interacción con el usuario, y Chart.js [27] para la creación de gráficas.

La aplicación puede ser utilizada luego de ser desplegada en un servidor Apache Tomcat [28], como se especifica en las instrucciones de instalación que se encuentran en el Anexo V. La misma cuenta con un archivo de configuración que permite especificar los parámetros de conexión de las bases de datos fuente y destino, y la ruta donde se encuentra el trabajo Kettle con el que se carga y actualiza el Data Warehouse. Luego de desplegada la aplicación se puede acceder localmente a la misma mediante la URL [http://localhost:8080/interfaz\\_msp/](http://localhost:8080/interfaz_msp/).

La interfaz está compuesta una barra de navegación y tres pantallas, Inicio, Consultas y Carga, cuyas principales características serán presentadas a continuación. En el Anexo V se puede encontrar el manual de usuario de la herramienta que contiene una descripción detallada de las funcionalidades provistas e instrucciones de cómo utilizarlas.

## 6.1 Inicio

Al ingresar a la aplicación se carga por defecto la pantalla de Inicio, cuyo contenido se muestra en la Figura 6.1. El diseño de la misma se asemeja al de una consola de administrador, y su cometido es brindar a los funcionarios de la UINS, mediante tablas y gráficas, una visualización rápida y atractiva de los indicadores, su evolución en los últimos años, y un análisis de los mismos por distintos criterios.

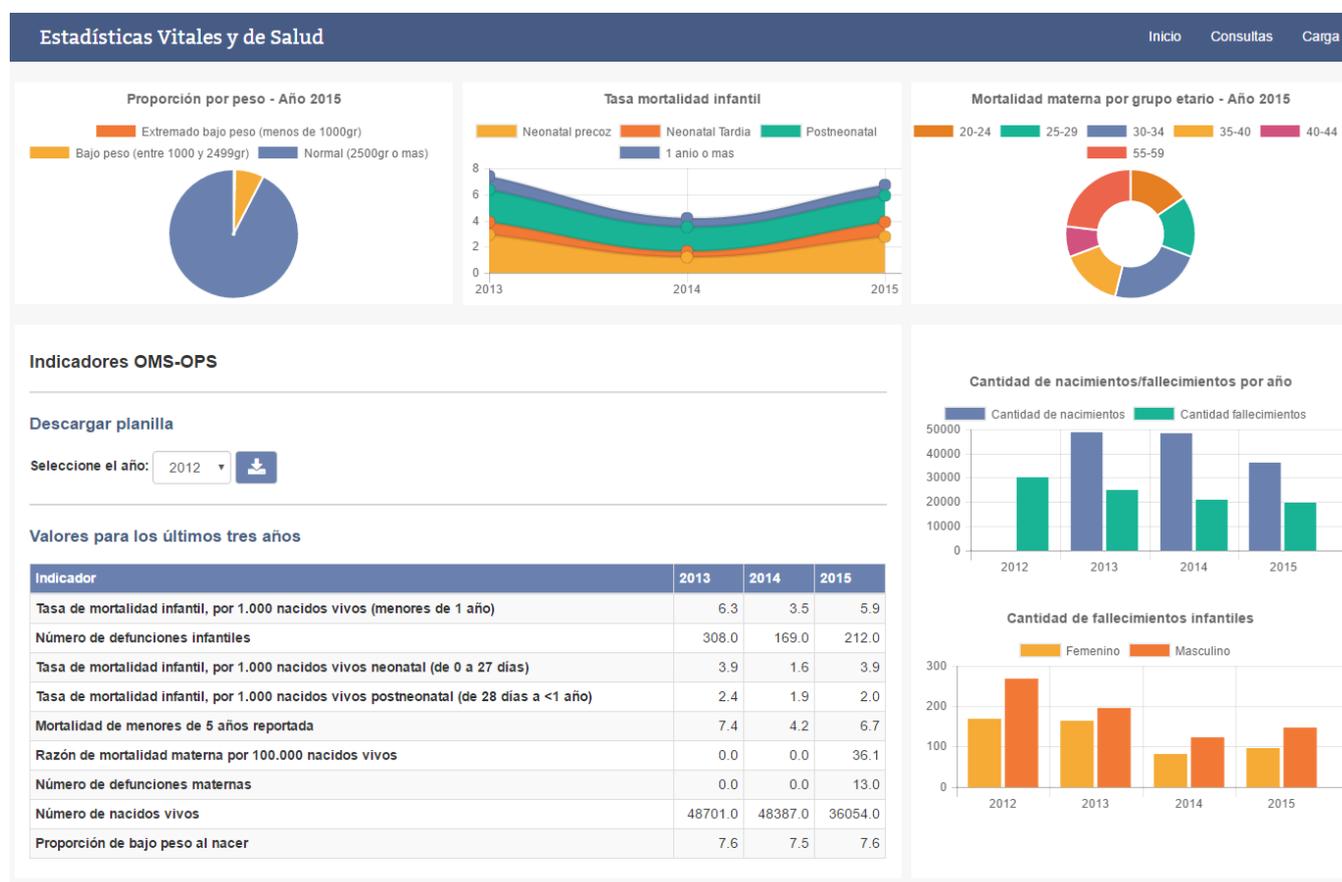


Figura 6.1 – Pantalla Inicio

Al momento de definir la combinación de datos y tipos de gráficas a desplegar en la pantalla principal, se evaluó cuáles aportarían más valor y darían un panorama general de la situación actual al usuario final. La estructura de dichas gráficas es estática, es decir el formato no puede ser modificado sin alterar el código fuente, pero no así su contenido, ya que este varía según los datos que se encuentren cargados en el Data Warehouse.

Por otra parte, la UINS debe entregar anualmente a la OMS una planilla con los valores de los indicadores para el año acaecido. Con el objetivo de facilitar a los usuarios dicha tarea, en la sección inferior izquierda de la pantalla de Inicio se encuentra una lista desplegable con los años disponibles y un botón de descarga, que al seleccionarlo permite descargar una planilla Excel que contiene los

valores de dichos indicadores para el año elegido. Además, en dicha sección se muestra una tabla comparativa de estos valores para los últimos tres años.

## 6.2 Consultas

Esta pantalla permite que usuarios sin conocimientos técnicos realicen consultas personalizadas en el Data Warehouse mediante una interfaz sencilla e intuitiva. Los resultados de dichas consultas se muestran en una tabla que puede ser exportada tanto en formato Excel como en PDF. De esta forma, los funcionarios de la UINS podrán crear de forma rápida tablas a demanda requeridas por autoridades y organizaciones y, en un futuro, esta sección podría ser incluida en una aplicación web destinada al público en general. En la Figura 6.2 se puede observar el resultado de consultar la medida Cant\_Nacimientos por los niveles Montevideo/Interior de la dimensión Localidad Ocurrencia y Año de la dimensión Tiempo.

The screenshot shows the 'Estadísticas Vitales y de Salud' interface. On the left is a sidebar with a tree view of dimensions and measures. The main area has three panels: 'Medidas' (Measures) containing 'Cantidad de nacimientos', 'Columnas' (Columns) containing 'Montevideo / Interior', and 'Filas' (Rows) containing 'Año'. Below these is a 'Filtro' (Filter) section with the text 'Arrastre un criterio de análisis'. A 'Consultar' button is on the right. At the bottom is a table with the following data:

Tiempo	Localidad ocurrencia		
	Montevideo	Interior	No Indicado
2013	25.877	22.546	278
2014	25.976	22.249	162
2015	18.969	17.081	4

Figura 6.2 – Pantalla Consultas

En la sección lateral izquierda de la pantalla se muestra una lista desplegable que permite seleccionar el cubo con el que se desea trabajar. Al elegirlo, se presenta un árbol que contiene las medidas y dimensiones, junto con las jerarquías y niveles que las componen, pertenecientes a dicho cubo. Para armar una consulta se deben arrastrar hacia los paneles que se encuentran en la sección superior derecha, las medidas y los niveles del cubo por los que interesa realizar el análisis. En dichos paneles se pueden especificar las medidas a analizar, los criterios de análisis que se deben incluir en las columnas y filas de la tabla, y los niveles cuyos miembros se utilizarán para aplicar un filtro. A

modo de ejemplo, en la Figura 6.3 se muestra la ventana que emerge al soltar el nivel Mes de la dimensión Tiempo sobre el panel de filtros.



*Figura 6.3 – Ventana filtro*

Luego de armar la consulta y seleccionar el botón Consultar, el resultado de la misma se despliega en la sección inferior derecha de la pantalla en formato tabla junto con dos botones en su extremo superior izquierdo que permiten exportarla en los formatos previamente mencionados.

## 6.3 Carga

Mediante la pantalla de Carga los usuarios de la aplicación pueden cargar y actualizar el Data Warehouse. Tal como se muestra en la Figura 6.4, esta contiene un formulario que ofrece distintas opciones de carga, permitiendo realizar desde una carga inicial, donde se crea el esquema de la base de datos y se cargan todos los datos disponibles, hasta la actualización de una dimensión particular. Si ocurre un error durante cualquiera de estas ejecuciones, se muestra en pantalla un log con los detalles del problema ocurrido. Además, se guarda en el directorio TOMCAT\_HOME/logs un archivo con nombre cargaDW-fechaDelDía con el resultado detallado de todas las ejecuciones realizadas durante ese día. A continuación se presenta una breve descripción de cada una de las opciones de carga provistas.

Estadísticas Vitales y de Salud Inicio Consultas Carga

**Opciones de carga**

Cargar todas las tablas (Se debe realizar la primera vez)

Únicamente fact tables

Personalizada

Nacimientos  Tiempo  Sexo

Mortalidad  Edades  Causas CIE-10

Mortalidad Infantil  Domicilios y Localidades  Establecimientos

**Seleccione las fechas de inicio y de fin para la carga de la tabla Nacimientos**

Fecha de inicio  Fecha de fin

Figura 6.4 – Pantalla Carga

### Inicial

Crea el esquema de la base de datos, o en caso de que ya exista lo elimina y lo crea nuevamente, y carga las tablas de dimensión y de hecho. Es posible indicar, mediante una fecha de inicio y otra de fin, el período de tiempo dentro del cual deben estar comprendidos los datos que se extraerán desde la base de datos fuente. En caso de que no se completen estos campos, se cargarán todos los datos encontrados en dicha base de datos.

## Actualización

Actualiza las tablas de hecho Nacimientos, Mortalidad y Mortalidad Infantil, junto con tablas de dimensión Nacidos Vivos, Madres, Fallecidos Menores y Fallecidos, ya que estas últimas mantienen una correspondencia uno a uno con las tablas de hecho. Es necesario indicar las fechas de inicio y de fin que determinen el período dentro del cual deben estar comprendidos los datos a cargar desde la base de datos fuente. Cabe señalar que al intentar cargar datos ya existentes en el Data Warehouse, los mismos reemplazarán a los anteriores.

## Personalizada

Permite indicar las tablas que deberán ser cargadas o actualizadas en el Data Warehouse. De forma análoga a la opción *Actualización*, la carga de la tabla de hecho Nacimientos implica la carga de las tablas de dimensión Madres y Nacidos Vivos, mientras que la carga de Mortalidad y Mortalidad Infantil implica la carga de Fallecidos y Fallecidos Menores respectivamente. En cualquiera de estos casos se deberán completar los campos de entrada correspondientes a las fechas de inicio y de fin. Resulta útil mencionar que al seleccionar una tabla de dimensión, tanto esta como las tablas de hecho, se vaciarán y se cargarán nuevamente para evitar inconsistencias.

## 6.4 Verificación de la interfaz de usuario

Una vez finalizada la implementación de la interfaz se realizó una verificación de la misma que consistió de dos etapas. En la primera se verificó que proporcionara al usuario todas las funcionalidades necesarias para cumplir con los requerimientos definidos. Los resultados de dicha verificación se detallan en el Cuadro 6.1.

Requerimientos	Resultado
R1, R2	Para obtener la cantidad de nacidos vivos y/o la proporción por peso al nacer, analizadas según distintos criterios, alcanza con acceder a la pantalla Consultas y seleccionar el cubo Nacimientos. En la pantalla que se despliega a continuación, muy similar a la de la Figura 6.2, es sencillo notar que se proporcionan ambas medidas, así como también todos los criterios de análisis establecidos en dichos requerimientos.
R3, R5	Si se desea consultar la cantidad de defunciones y/o la razón de mortalidad materna, analizadas según distintos criterios, basta con acceder a la pantalla Consultas y seleccionar el cubo Mortalidad. Del mismo modo que en el caso anterior, es sencillo notar que se proporcionan ambas medidas, así como todos

	los criterios de análisis establecidos en los requerimientos.
R4	De forma análoga a los casos anteriores, para obtener la tasa de mortalidad infantil según diferentes criterios de análisis, alcanza con acceder a la pantalla Consultas y elegir el cubo Mortalidad Infantil. A continuación se despliega el árbol que contiene todos los criterios solicitados en el requerimiento, así como la medida en cuestión.
Indicadores	Los indicadores pueden ser accedidos a través de la pantalla Inicio, tal como se puede observar en la Figura 6.1.  Además, mediante la pantalla de Consultas es posible obtener los indicadores seleccionando las medidas y los criterios de análisis correspondientes, así como aplicando filtros cuando sea necesario.

*Cuadro 6.1 – Verificación de interfaz de usuario (I)*

Luego, en la segunda etapa, se corroboró que los resultados obtenidos a través de la interfaz coincidieran con los obtenidos al consultar el Data Warehouse desde la herramienta Saiku. En el Cuadro 6.2 se presentan algunos de los valores comparados.

	Interfaz de usuario	Saiku
Proporción bajo peso para el año 2015	7,2	7,2
Tasa de mortalidad infantil, por 1.000 nacidos vivos postneonatal (de 28 días a < 1 año) para el año 2013	2,4	2,4
Tasa de mortalidad infantil, por 1.000 nacidos vivos (menores de 1 año) para el año 2014	3,5	3,5
Cantidad de nacimientos en 2013	48701	48701
Cantidad de nacimientos en 2014	48387	48387
Cantidad de fallecimientos en 2012	30196	30196
Cantidad de fallecimientos en 2015	19582	19582
Cantidad de fallecimientos infantiles de niñas en 2013	164	164
Cantidad de fallecimientos infantiles de niños en 2013	195	195

*Cuadro 6.2 – Verificación de interfaz de usuario (II)*



## Capítulo 7 - Conclusiones

En el proyecto realizado se ha podido cumplir con los objetivos planteados al comienzo del mismo. A continuación se presentan los resultados obtenidos:

- Un análisis de calidad de los datos de las fuentes proporcionadas, que fue tomado como insumo al momento de realizar los procesos de carga y actualización del Data Warehouse.
- Un prototipo de Data Warehouse que permite obtener tanto indicadores de salud requeridos por la OMS, como cuadros estadísticos que comúnmente se publican en el sitio web del MSP, y llevar a cabo consultas personalizadas para satisfacer solicitudes a demanda.
- Una aplicación web que permite realizar consultas y reportes personalizados, así como realizar la carga y actualización del Data Warehouse, de forma sencilla y amigable.
- Documentación detallada del trabajo realizado.
- Guía de instalación y manual de usuario de la aplicación web desarrollada.

Si bien no formaba parte de los requerimientos planteados por los analistas del MSP, se decidió llevar a cabo un análisis de calidad de las fuentes y utilizarlo como insumo principal para realizar un proceso de limpieza y mejora de los datos. Se considera que aunque esto requirió de cierto esfuerzo adicional, ya que se analizaron varios aspectos de calidad sobre numerosos atributos, proporcionó un panorama de los principales aspectos a considerar durante el proceso carga y actualización, permitiendo la automatización de varias tareas de calidad que hoy en día los funcionarios de la UINS realizan manualmente.

Con el fin de cumplir con los objetivos, al momento de implementar el prototipo solicitado, fue necesario resolver problemas complejos de manejo de datos, como heterogeneidades, calidad de datos y cálculos de indicadores con fórmulas no triviales. Los problemas referidos a la calidad tuvieron un impacto considerable durante la elaboración de los procesos de ETL, ya que implicaron pasos adicionales para ser resueltos, como por ejemplo, el agregado de tuplas *dummy*, el cálculo de atributos mediante fórmulas ya que el atributo original contenía información incorrecta, y el manejo de valores nulos. Esto, sumado al hecho de que se contaba con un número significativo de dimensiones, tablas de hecho que requerían de datos de numerosas tablas de las fuentes y mapeos que precisaban de cierta lógica para ser resueltos, resultó en un trabajoso proceso de elaboración de ETLs que demandó una cantidad de tiempo significativa. Por otra parte, durante la implementación de cubos, la construcción de indicadores fue uno de los aspectos más complejos de resolver, ya que sus fórmulas de cálculo requerían de operaciones más complejas que las ofrecidas por la herramienta OLAP por defecto, por lo que fue necesario familiarizarse con el lenguaje MDX y la definición de miembros calculados.

Por otro lado, si bien en un principio no se solicitó que los programas de carga y de consulta se encontraran integrados, la aplicación desarrollada brinda una solución centralizada para el manejo

del sistema de Data Warehouse, ya que se consideró que esto facilitaría su uso. Además, se incluyeron algunas funcionalidades adicionales que se entiende le aportan un valor agregado a la herramienta, como la visualización de gráficas, la posibilidad de exportar los resultados de las consultas en distintos formatos y de realizar cargas personalizadas, entre otras.

Adicionalmente, se realizaron etapas de verificación durante la ejecución del proyecto, tanto del prototipo de Data Warehouse como de la aplicación web, con el objetivo de comprobar la correctitud de los mismos.

En lo personal, creemos que el sistema implementado simplificaría las tareas realizadas por el personal de la UINS, permitiéndoles cumplir con las solicitudes, tanto de autoridades como de terceros, de forma fácil y rápida con el apoyo de una interfaz sencilla y flexible que permite la emisión de reportes personalizados. Además, consideramos que el mantenimiento del sistema no resultaría una tarea compleja, al contar con una interfaz de carga amigable que contempla varios escenarios posibles. Igualmente entendemos que existen varias mejoras, tanto sobre el Data Warehouse como sobre la aplicación, que podrían incluirse en trabajos futuros.

## 7.1 Trabajo a futuro

A continuación se presentan lineamientos generales sobre posibles mejoras a tener en cuenta para trabajos futuros.

### *Validación con el usuario*

En este proyecto no fue posible concretar una etapa de validación con el usuario. Se considera que la misma permitiría no solo corroborar que lo implementado efectivamente cumple con las expectativas de los usuarios, sino también recibir retroalimentación y propuestas de mejoras, principalmente en lo que respecta a la aplicación y su usabilidad.

### *Integración con Sistema de Información Perinatal*

Como se mencionó anteriormente, no fue posible utilizar la base de datos del Sistema de Información Perinatal (SIP) ya que no se contaba con documentación de la misma y los datos proporcionados contenían valores fijos en algunos atributos, lo que imposibilitó la integración con la base mejorada. Esta restricción significó la eliminación de algunos indicadores de interés para el MSP de los requerimientos funcionales. Si se integrara el SIP al sistema implementado se podrían obtener varios indicadores adicionales, así como más información sobre los ya existentes. Además, dicha integración permitiría la incorporación de nuevas mejoras de calidad sobre los datos, posibilitando la realización de chequeos cruzados.

### ***Nuevas funcionalidades en interfaz de usuario***

Con el objetivo de generar reportes más completos, se podría incluir en la sección de Consultas la posibilidad de graficar los resultados, permitiendo al usuario seleccionar el tipo de gráfica y los datos a incluir.

Asimismo, resultaría conveniente que tanto las gráficas como las tablas que se muestran en el panel del administrador fueran personalizables, brindando así una interfaz más flexible y adaptable a los intereses particulares de cada usuario.

En la actualidad, el sistema no ofrece la posibilidad de guardar consultas previamente armadas, lo que facilitaría la construcción de consultas futuras. Se considera por lo tanto, que este aspecto podría ser mejorado en una nueva versión de la aplicación.

### ***Extensión de la interfaz de usuario***

Actualmente la interfaz de usuario está enfocada al uso interno del MSP por parte del personal de la UINS. En el contexto de las políticas de datos abiertos que han sido impulsadas en este último tiempo, sería interesante ampliar dicha interfaz para que pueda ser utilizada también por el público en general, permitiendo de esta forma que cualquier usuario pueda realizar consultas cuando así lo desee. Por ejemplo, agregando manejo de sesión, permisos y perfiles.



## Bibliografía

- [1] E. Malinowski y E. Zimányi, *Advanced Data Warehouse Design*, Springer, 2008.
- [2] W. Inmon, *Building the Data Warehouse*, Wiley, 2005.
- [3] A. Marotta, *Diapositivas del Curso de Diseño y Construcción de Data Warehouse*, Montevideo, 2013.
- [4] F. Carpani, «CMDM: Un Modelo Conceptual para la Especificación de Bases Multidimensionales,» Montevideo, 2000.
- [5] M. Golfarelli, D. Maio y S. Rizzi, «Conceptual Design of Data Warehouses from E/R Schemes,» de *International Conference on Systems Science*, Hawaii, 1998.
- [6] M. Golfarelli y S. Rizzi, *Data Warehouse Design: Modern Principles and Methodologies*, McGraw-Hill, 2009.
- [7] Pentaho, «Pentaho Community Edition,» [En línea]. Available: <http://community.pentaho.com>. [Último acceso: 12 Febrero 2016].
- [8] A. Marotta, *Diapositivas del curso de Calidad de Datos*, Montevideo, 2013.
- [9] Fundación Julio Ricaldoni, «Convenio MSP - FJR,» Montevideo, 2012.
- [10] Organización Panamericana de la Salud, Unidad de Información y Análisis de Salud (HA), «Glosario de indicadores,» [En línea]. Available: <http://www.paho.org/hq/index.php>. [Último acceso: 6 Junio 2016].
- [11] Ministerio de Salud Pública, «Estadísticas,» [En línea]. Available: <http://www.msp.gub.uy/noticia/estad%C3%ADsticas>. [Último acceso: 15 Mayo 2016].
- [12] Organización Panamericana de la Salud, «CIE-10 Volumen 1 Español,» [En línea]. Available: <http://ais.paho.org/classifications/Chapters/index.htm>. [Último acceso: 7 Abril 2016].
- [13] Poder Judicial de la Provincia del Chaco, Instituto Médico Forense, «CIE-10,» [En línea]. Available: [www.justiciachaco.gov.ar/IMCIF/Contenido/cie10.xls](http://www.justiciachaco.gov.ar/IMCIF/Contenido/cie10.xls). [Último acceso: 7 Abril 2016].
- [14] MySQL Documentation, «Optimizing Database Structure,» [En línea]. Available: <http://dev.mysql.com/doc/refman/5.7/en/data-size.html>. [Último acceso: 16 Abril 2016].
- [15] Pentaho, Documentación versión 5.3, *Getting Acquainted with Pentaho Components*, 2015.
- [16] Mondrian, «Mondrian Documentation,» [En línea]. Available: <http://mondrian.pentaho.com/documentation/>. [Último acceso: 10 Abril 2016].
- [17] Pentaho, «Pentaho Data Integration Wiki,» [En línea]. Available: [http://wiki.pentaho.com/display/EAI/Latest+Pentaho+Data+Integration+\(aka+Kettle\)+Documentation](http://wiki.pentaho.com/display/EAI/Latest+Pentaho+Data+Integration+(aka+Kettle)+Documentation). [Último acceso: 19 Febrero 2016].
- [18] Pentaho, «Pentaho Schema Workbench,» [En línea]. Available:

- <http://mondrian.pentaho.com/documentation/workbench.php>. [Último acceso: 12 Diciembre 2015].
- [19] Meteorite, «Saiku,» [En línea]. Available: <http://www.meteorite.bi/products/saiku>. [Último acceso: 9 Diciembre 2015].
- [20] Mondrian, «Mondrian Documentation MDX,» [En línea]. Available: <http://mondrian.pentaho.com/documentation/mdx.php>. [Último acceso: 10 Abril 2016].
- [21] Microsoft, «Documentación MDX,» [En línea]. Available: <https://msdn.microsoft.com/en-us/library/ms145506.aspx>. [Último acceso: 16 Marzo 2016].
- [22] Oracle, «Java,» [En línea]. Available: <http://www.oracle.com/technetwork/java/javase/overview/index.html>. [Último acceso: 25 Noviembre 2015].
- [23] Apache, «Log4j,» [En línea]. Available: <http://logging.apache.org/log4j/2.x/>. [Último acceso: 11 Febrero 2016].
- [24] Pivot4J Team, «Pivot4J,» [En línea]. Available: <http://www.pivot4j.org/>. [Último acceso: 16 Marzo 2016].
- [25] Bootstrap, «Bootstrap,» [En línea]. Available: <http://getbootstrap.com/getting-started/>. [Último acceso: 29 Enero 2016].
- [26] The jQuery Foundation, «jQuery,» [En línea]. Available: <https://jquery.com/>. [Último acceso: 22 Marzo 2016].
- [27] Chart.js, «Chart.js,» [En línea]. Available: <http://www.chartjs.org/docs/>. [Último acceso: 10 Abril 2016].
- [28] Apache, «Apache Tomcat,» [En línea]. Available: <http://tomcat.apache.org/>. [Último acceso: 21 Setiembre 2015].
- [29] Ministerio de Salud Pública, «Certificado De Nacido Vivo Electrónico – Manual De Usuario,» Montevideo.
- [30] Oracle, «Downloads,» [En línea]. Available: <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>. [Último acceso: 4 Enero 2016].

## Anexo I - Calidad de datos

En el presente Anexo se expone el modelo de calidad planteado sobre la base mejorada, obtenido a partir de un análisis previo del perfil de los datos en lo que refiere tanto a valores faltantes como erróneos y a dependencias funcionales difusas. Asimismo, se detalla el análisis realizado posteriormente sobre dicho modelo, junto con los procedimientos utilizados para manejar los errores encontrados y prevenir posibles errores que pudieran surgir en futuras actualizaciones de los datos. Estos procedimientos fueron implementados durante la etapa de extracción, transformación y carga.

A continuación se presenta en forma de tabla la siguiente información: en la primera sección se detalla el modelo de calidad sobre cada atributo, o conjunto de atributos que interesa estudiar; luego, en la segunda sección, se encuentran los resultados del análisis y las consultas SQL ejecutadas para realizarlo, así como las soluciones implementadas durante la carga de las tablas correspondientes del Data Warehouse. Cabe destacar que se estudiaron únicamente los atributos que resultaban relevantes para la carga de las tablas de hecho y de dimensión del proyecto. Por otra parte, es importante mencionar que en cuanto a la corrección de errores que refieren a dominios o rangos, se tomó como criterio no corregir valores atípicos en el nivel más bajo con el fin de facilitar su detección, ya que podrían causar desviaciones en los estudios estadísticos. Sin embargo, en niveles superiores estos se agregan en miembros que agrupan valores erróneos (*Erróneo*) o faltantes (*No Indicado*).

## Certificados Nacidos Vivos

### cnv\_datos\_recien\_nacido

**Tablas:**

*cnv\_datos\_recien\_nacido*

**Atributos:**

- a. *cnv\_datos\_recien\_nacido.primer\_nombre*
- b. *cnv\_datos\_recien\_nacido.primer\_apellido*
- c. *cnv\_datos\_recien\_nacido.numero\_documento\_reserva*
- d. *cnv\_datos\_recien\_nacido.semanas\_gestacion\_nacimiento*
- e. *cnv\_datos\_recien\_nacido.peso*
- f. *cnv\_datos\_recien\_nacido.sexo*

#### Modelo de calidad

	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Complejidad	Densidad	<u>Atributo nulo:</u> Verifica si la celda tiene valor nulo  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo no es NULL devuelve 1, sino 0	Porcentaje de valores en 1 en toda la columna.
2	Consistencia	Integridad de dominio	<u>Rango Edad Gestacional:</u> Verifica que el valor esté comprendido entre 20 y 50  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo <i>semanas_gestacion_nacimiento</i> está comprendido entre 20 y 50 devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna
3		Integridad de dominio	<u>Rango Peso:</u> Verifica que el valor esté comprendido entre 500 y 8000.  Tipo de resultado: (0,1)	Si el valor del atributo está comprendido entre 500 y 8000 devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna

			Granularidad: Celda		
4	Exactitud	Correctitud sintáctica	<p><u>Formato Sexo:</u>                  Verifica que el valor sea 'F', 'M', 'D' o 'I'</p> <p>Tipo de resultado: (0,1)                  Granularidad: Celda</p>	Si el valor del atributo <i>sexo</i> es 'F','M','D' o 'I' devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna
<b>Análisis y solución</b>					
	<b>Consulta</b>	<b>Resultado</b>	<b>Solución</b>		
1.a	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE primer_nombre IS NOT NULL;	100%	Se le asigna el valor <i>No Indicado</i> .		
1.b	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE primer_apellido IS NOT NULL;	100%			
1.c	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE numero_documento_reserv a IS NOT NULL;	100%	Según el manual del sistema CNV [29]: <i>“Cuando el médico firma electrónicamente un certificado de nacido vivo, el sistema se comunica con la Dirección Nacional de Identificación Civil para solicitar un número de reserva de su cédula de identidad ...                  El número de reserva se almacena en el certificado de nacido vivo electrónico y se incluye en el comprobante que se emite para la Dirección General de Registro de Estado Civil.”</i> Por lo tanto concluimos que este atributo no puede ser nulo.		

1.d	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE semanas_gestacion_nacimie nto IS NOT NULL;	100%	-
1.e	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE peso IS NOT NULL;	100%	-
1.f	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE sexo IS NOT NULL;	100%	Se le asigna el valor <i>Desconocido</i> .
2	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE semanas_gestacion_nacimie nto BETWEEN 20 AND 50;	99,95%	Se muestran todos los valores aunque sean incorrectos. Pero si la cantidad de semanas de gestación al nacer excede el valor 50 o es menor a 20, se le asigna el rango con id 11 y descripción <i>Erróneo</i> .
3	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE peso BETWEEN 500 AND 8000;	99,79%	Se muestran todos los valores aunque sean incorrectos. Pero si el peso excede el valor 8000 o es menor a 500, se le asigna el rango con id 9 y descripción <i>Erróneo</i> .
4	SELECT COUNT(*) FROM cnv_datos_recien_nacido WHERE sexo IN ( 'F', 'M', 'D', 'I' );	99,98%	Se le asigna el valor <i>Erróneo</i> .

[cnv\\_datos\\_parto](#)

**Tablas:**

*cnv\_datos\_parto*

*tipos\_terminacion\_parto*

**Atributos:**

- a. *cnv\_datos\_parto.tipo\_parto*
- b. *tipos\_terminacion\_parto.id\_terminacion*
- c. *cnv\_datos\_atendio\_parto.id\_tipo\_atendio\_parto*
- d. *cnv\_datos\_parto.lugar\_parto*
- e. *cnv\_datos\_parto.establecimiento\_ocurrencia*
- f. *cnv\_datos\_parto.localidad\_ocurrencia*

**Modelo de calidad**

	<b>Dimensión</b>	<b>Factor</b>	<b>Métrica</b>	<b>Procedimiento de Medición</b>	<b>Agregación</b>
1	Complejidad	Densidad	<p><u>Atributo nulo:</u>                      Verifica si la celda tiene valor nulo</p> <p>Tipo de resultado: (0,1)                      Granularidad: Celda</p>	Si el atributo no es <i>NULL</i> devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna.
2	Consistencia	Integridad inter-relación	<p><u>Integridad Tipo Parto:</u>                      Verifica que se cumpla la restricción de integridad entre <i>cnv_datos_parto</i> y <i>tipos_terminacion_parto</i></p> <p>Tipo de resultado: (0,1)                      Granularidad: Celda</p>	Si el atributo <i>tipo_parto</i> coincide con algún valor del atributo <i>id_terminacion</i> devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna

**Análisis y solución**

	<b>Consulta</b>	<b>Resultado</b>	<b>Solución</b>
1.a	SELECT COUNT(*)	100%	Se le asigna el id 21 con descripción

	FROM cnv_datos_parto WHERE tipo_parto IS NOT NULL;		<i>No Indicado</i> existente en la tabla <i>tipos_terminacion_parto</i> .
1.c	SELECT COUNT(*) FROM cnv_datos_atendio_parto WHERE id_tipo_atendio_parto IS NOT NULL;	99,01%	Se le asigna el id 0 con descripción <i>No Indicado</i> .
1.d	SELECT COUNT(*) FROM cnv_datos_parto WHERE lugar_parto IS NOT NULL;	99,98%	Se le asigna el id 0 con descripción <i>No Indicado</i> .
1.e	SELECT COUNT(*) FROM cnv_datos_parto WHERE establecimiento_ocurrenci a IS NOT NULL;	99,43%	Se le asigna el id 0 que se corresponde con el establecimiento <i>No Indicado</i> .
1.f	SELECT COUNT(*) FROM cnv_datos_parto WHERE localidad_ocurrencia IS NOT NULL;	99,67%	Se le asigna el id 0 que se corresponde con la localidad <i>No Indicado</i> .
2	SELECT COUNT(*) FROM cnv_datos_parto INNER JOIN tipos_terminacion_parto ON tipo_parto = id_terminacion;	100%	Se le asigna el id 22 con descripción <i>Erróneo</i> .

*cnv\_datos\_madre*

**Tablas:**

*cnv\_datos\_madre*

**Atributos:**

- a. *cnv\_datos\_madre.primer\_nombre*
- b. *cnv\_datos\_madre.primer\_apellido*
- c. *cnv\_datos\_madre.numero\_documento*
- d. *cnv\_datos\_madre.tipo\_documento*
- e. *cnv\_datos\_madre.pais\_emisor\_documento*
- f. *cnv\_datos\_madre.total\_consultas\_prenatales*
- g. *cnv\_datos\_madre.mayor\_nivel\_estudios*
- h. *cnv\_datos\_madre.estado\_civil*
- i. *cnv\_datos\_madre.edad*
- j. *cnv\_datos\_madre.id\_domicilio*

**Modelo de calidad**

	<b>Dimensión</b>	<b>Factor</b>	<b>Métrica</b>	<b>Procedimientos de Medición</b>	<b>Agregación</b>
1	Complejitud	Densidad	<u>Atributo nulo:</u> Verifica si la celda tiene valor nulo  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo no es NULL devuelve 1, sino 0	Porcentaje de valores en 1 en toda la columna.
2	Consistencia	Integridad de dominio	<u>Rango Cantidad Consultas Prenatales:</u> Verifica que el valor esté comprendido entre 0 y 41  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo <i>total_consultas_prenatales</i> está comprendido entre 0 y 41 devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna
3		Integridad de dominio	<u>Rango Edad Madre:</u> Verifica que el valor esté comprendido entre 9 y 70	Si el atributo <i>edad</i> está comprendido entre 9 y 70 devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna

			Tipo de resultado: (0,1) Granularidad: Celda		
<b>Análisis y Solución</b>					
	<b>Consulta</b>	<b>Resultado</b>	<b>Solución</b>		
1.a	SELECT COUNT(*) FROM cnv_datos_madre WHERE primer_nombre IS NOT NULL;	100%	Se le asigna el valor <i>No Indicado</i> .		
1.b	SELECT COUNT(*) FROM cnv_datos_madre WHERE primer_apellido IS NOT NULL;	100%			
1.c	SELECT COUNT(*) FROM cnv_datos_madre WHERE numero_documento IS NOT NULL;	100%			
1.d	SELECT COUNT(*) FROM cnv_datos_madre WHERE tipo_documento IS NOT NULL;	99,94%	Se le asigna el id 0 con descripción <i>No Indicado</i> .		
1.e	SELECT COUNT(*) FROM cnv_datos_madre WHERE pais_emisor_documento IS NOT NULL;	99,94%			
1.f	SELECT COUNT(*) FROM cnv_datos_madre WHERE total_consultas_prenatales IS NOT NULL;	100%	-		

1.g	SELECT COUNT(*) FROM cnv_datos_madre WHERE mayor_nivel_estudios IS NOT NULL;	83,26%	Se le asigna el id 42 con descripción <i>No Indicado.</i>
1.h	SELECT COUNT(*) FROM cnv_datos_madre WHERE estado_civil IS NOT NULL;	100%	Se le asigna el id 21 con descripción <i>No Indicado.</i>
1.i	SELECT COUNT(*) FROM cnv_datos_madre WHERE edad IS NOT NULL;	99,99%	Se calcula la edad como la diferencia entre la fecha de parto y la fecha de nacimiento de la madre.
1.j	SELECT COUNT(*) FROM cnv_datos_madre WHERE id_domicilio IS NOT NULL;	99,48%	Se le asigna el id 0 que se corresponde con el domicilio <i>No Indicado.</i>
2	SELECT COUNT(*) FROM cnv_datos_madre WHERE total_consultas_prenatales BETWEEN 0 AND 40;	99,92%	Se muestran todos los valores aunque sean incorrectos. Pero si la cantidad de consultas excede el valor 40 o es menor a 0, se le asigna el rango con id 11 y descripción <i>Erróneo.</i>
3	SELECT COUNT(*) FROM cnv_datos_madre WHERE edad BETWEEN 9 AND 70;	99,82%	Se muestran todos los valores aunque sean incorrectos. Pero si la edad excede los 70 años o es menor a 9, se le asigna el grupo etario materno con id 9 y descripción <i>Erróneo.</i>

[certificados nacidos vivos](#)

<p><b>Tablas:</b>  <i>certificados_nacidos_vivos</i></p> <p><b>Atributo:</b>  <i>certificados_nacidos_vivos.fecha_parto</i></p>					
Modelo de calidad					
	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Compleitud	Densidad	<p><u>Fecha Parto nulo:</u>                      Verifica si la celda tiene valor nulo</p> <p>Tipo de resultado: (0,1)                      Granularidad: Celda</p>	<p>Si el atributo <i>fecha_parto</i> no es NULL devuelve 1, sino 0.</p>	<p>Porcentaje de valores en 1 en toda la columna.</p>
Análisis y solución					
	Consulta	Resultado	Solución		
1	<pre>SELECT COUNT(*) FROM certificados_nacidos_vivos WHERE fecha_parto IS NOT NULL;</pre>	100%	<p>Se descartan las tuplas cuya fecha de parto tenga valor NULL.</p>		

## Certificados Defunción

### cd\_identificacion\_fallecido

<b>Tablas:</b> <i>cd_identificacion_fallecido</i>					
<b>Atributos:</b> <ol style="list-style-type: none"> <li><i>cd_identificacion_fallecido.Primer_Nombre</i></li> <li><i>cd_identificacion_fallecido.Primer_Apellido</i></li> <li><i>cd_identificacion_fallecido.N_Doc</i></li> <li><i>cd_identificacion_fallecido.Tipo_Doc_Id</i></li> <li><i>cd_identificacion_fallecido.Pais_Doc_Id</i></li> <li><i>cd_identificacion_fallecido.Sexo</i></li> </ol>					
Modelo de calidad					
	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Compleitud	Densidad	<u>Atributo nulo:</u> Verifica si la celda tiene valor nulo  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo no es NULL devuelve 1, sino 0	Porcentaje de valores en 1 en toda la columna.
2	Exactitud	Correctitud sintáctica	<u>Formato Sexo:</u> Verifica que el valor sea 'F', 'M', 'D' o 'I'  Tipo de resultado: (0,1) Granularidad: Celda	Si el valor del atributo <i>sexo</i> es 'F','M','D' o 'I' devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna
Análisis y solución					
	Consulta	Resultado	Solución		
1.a	SELECT COUNT(*) FROM cd_identificacion_fallecido WHERE Primer_Nombre IS	100%	Se le asigna el valor <i>No Indicado</i> .		

	NOT NULL;		
1.b	SELECT COUNT(*) FROM cd_identificacion_fallecido WHERE Primer_Apellido IS NOT NULL;	100%	
1.c	SELECT COUNT(*) FROM cd_identificacion_fallecido WHERE N_Doc IS NOT NULL;	100%	
1.d	SELECT COUNT(*) FROM cd_identificacion_fallecido WHERE Tipo_Doc_Id IS NOT NULL;	99,12%	Se le asigna el id 0 con descripción <i>No Indicado.</i>
1.e	SELECT COUNT(*) FROM cd_identificacion_fallecido WHERE Pais_Doc_Id IS NOT NULL;	99,17%	
1.f	SELECT COUNT(*) FROM cd_identificacion_fallecido WHERE Sexo IS NOT NULL;	99,2%	Se le asigna el valor <i>Desconocido.</i>
2	SELECT COUNT(*) FROM cd_identificacion_fallecido WHERE Sexo IN ('F','M','D','I');	99,16%	Se le asigna el valor <i>Erróneo.</i>

cd datos fallecido

**Tablas:**

cd\_datos\_fallecido

**Atributos:**

- a. cd\_datos\_fallecido.Embarazada
- b. cd\_datos\_fallecido.id\_domicilio

Modelo de calidad					
	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Compleitud	Densidad	<u>Embarazada o Domicilio nulo:</u> Verifica si la celda tiene valor nulo  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo no es NULL devuelve 1, sino 0	Porcentaje de valores en 1 en toda la columna.
2	Exactitud	Correctitud sintáctica	<u>Formato Embarazada:</u> Verifica que el valor se encuentre comprendido entre 1 y 5.  Tipo de resultado: (0,1) Granularidad: Celda	Si el valor del atributo <i>Embarazada</i> se encuentra entre 1 y 5 devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna.
Análisis y solución					
	Consulta		Resultado	Solución	
1.a	SELECT COUNT(*) FROM cd_datos_fallecido WHERE Embarazada IS NOT NULL;		99,95%	Se estudia el sexo y si es Femenino se estudia la edad, y se asignan los siguientes valores: <ul style="list-style-type: none"> <li>• Femenino con edad entre 9 y 70 años: id 3 con descripción <i>Se Ignora</i></li> <li>• Femenino menor de 9 o mayor de 70 años: id 4 con</li> </ul>	

			descripción <i>No Corresponde</i> <ul style="list-style-type: none"> <li>• Masculino: id 4 con descripción <i>No Corresponde</i></li> <li>• Cualquier sexo que no sea Femenino ni Masculino: id 6 con descripción <i>Desconocido</i></li> </ul>
1.b	SELECT COUNT(*) FROM cd_datos_fallecido WHERE id_domicilio IS NOT NULL;	99,9%	Se le asigna el id 0 que se corresponde con el domicilio <i>No Indicado</i> .
2	SELECT COUNT(*) FROM cd_datos_fallecido WHERE Embarazada BETWEEN 1 AND 5;	44,02%	Se le asigna el valor <i>Erróneo</i> .

[cd datos defuncion](#)

<b>Tablas:</b> <i>cd_datos_defuncion</i>					
<b>Atributos:</b> <ul style="list-style-type: none"> <li>a. <i>cd_datos_defuncion.enf_causante_cie10</i></li> <li>b. <i>cd_datos_defuncion.edad_anios_def_calculada</i></li> <li>c. <i>cd_datos_defuncion.edad_anios_def_digit</i></li> </ul>					
Modelo de calidad					
	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Compleitud	Densidad	<u>Causa CIE-10 nulo:</u> Verifica si la celda tiene valor nulo  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo <i>enf_causante_cie10</i> no es NULL devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna.

2	Consistencia	Integridad de dominio	<u>Rango Edad Calculada:</u> Verifica que el valor esté comprendido entre 0 y 150  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo <i>edad_anios_def_digitada</i> está comprendido entre 0 y 150 devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna
3		Integridad de dominio	<u>Rango Edad Digitada:</u> Verifica que el valor esté comprendido entre 0 y 150  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo <i>edad_anios_def_dgit</i> está comprendido entre 0 y 150 devuelve 1, sino 0.	Porcentaje de valores en 1 en toda la columna
<b>Análisis y solución</b>					
	<b>Consulta</b>	<b>Resultado</b>	<b>Solución</b>		
1	SELECT COUNT(*) FROM cd_datos_defuncion WHERE enf_causante_cie10 IS NOT NULL;	89,66%	Se le asigna el id 9999 que se corresponde con la causa <i>No Indicado</i> .		
2	SELECT COUNT(*) FROM cd_datos_defuncion WHERE edad_anios_def_calculada BETWEEN 0 AND 150;	98,04%	Se estudia el atributo <i>edad_anios_def_dgit</i> .		
3	SELECT COUNT(*) FROM cd_datos_defuncion WHERE edad_anios_def_dgit BETWEEN 0 AND 150;	89,65%	Se le asigna el id 151 que se corresponde con la edad <i>Erróneo</i> .		

cd lugar defuncion

<p><b>Tablas:</b>  <i>cd_lugar_defuncion</i></p> <p><b>Atributo:</b>  <i>cd_lugar_defuncion.Lugar_Def_Id</i></p>					
Modelo de calidad					
	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Compleitud	Densidad	<p><u>Domicilio Ocurrencia nulo:</u>                      Verifica si la celda tiene valor nulo</p> <p>Tipo de resultado: (0,1)                      Granularidad: Celda</p>	<p>Si el atributo <i>Lugar_Def_Id</i> no es NULL devuelve 1, sino 0.</p>	<p>Porcentaje de valores en 1 en toda la columna.</p>
Análisis y solución					
	Consulta	Resultado	Solución		
1	<pre>SELECT COUNT(*) FROM cd_lugar_defuncion WHERE Lugar_Def_Id IS NOT NULL;</pre>	100%	<p>Se le asigna el id 0 que se corresponde con el domicilio <i>No Indicado</i>.</p>		

[certificados defuncion](#)

<p><b>Tablas:</b>  <i>certificados_defuncion</i>  <i>cd_identificacion_fallecido</i></p> <p><b>Atributo:</b>                  a. <i>certificados_defuncion.fecha_muerte</i>                  b. <i>cd_identificacion_fallecido.Fecha_Nac</i></p>					
Modelo de calidad					
	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Complejidad	Densidad	<p><u>Fecha Muerte nulo:</u>                      Verifica si la celda tiene valor nulo</p> <p>Tipo de resultado: (0,1)                      Granularidad: Celda</p>	<p>Si el atributo <i>fecha_muerte</i> no es NULL devuelve 1, sino 0.</p>	<p>Porcentaje de valores en 1 en toda la columna.</p>
2	Consistencia	Integridad inter-relación	<p><u>Integridad relacional entre Fecha Muerte y Fecha Nacimiento:</u>                      Verifica que la diferencia en años entre Fecha Muerte y Fecha Nacimiento esté comprendida entre 0 y 150</p> <p>Tipo de resultado: (0,1)                      Granularidad: Celda</p>	<p>Si la diferencia en años entre <i>fecha_muerte</i> y <i>Fecha_Nac</i> está comprendida entre 0 y 150 devuelve 1, sino 0.</p>	<p>Porcentaje de valores en 1 en toda la columna</p>
Análisis y solución					
	Consulta		Resultado	Solución	
1	<pre>SELECT COUNT(*) FROM certificados_defuncion</pre>		100%	<p>Se descartan las tuplas cuya fecha de muerte tenga valor NULL.</p>	

	WHERE fecha_muerte IS NOT NULL;		
2	SELECT COUNT(*) FROM certificados_defuncion NATURAL JOIN cd_identificacion_fallecido WHERE TIMESTAMPDIFF(YEAR, Fecha_Nac, fecha_muerte) BETWEEN 0 AND 150;	98,04%	Se estudia el atributo <i>edad_anios_def_calculada</i> de la tabla <i>cd_datos_defuncion</i> .

## Codigueras

### Establecimientos

<b>Tablas:</b> <i>Establecimientos</i>					
<b>Atributo:</b> <i>establecimientos.institucion</i>					
<b>Modelo de calidad</b>					
	<b>Dimensión</b>	<b>Factor</b>	<b>Métrica</b>	<b>Procedimientos de Medición</b>	<b>Agregación</b>
1	Compleitud	Densidad	<u>Institución nulo:</u> Verifica si la celda tiene valor nulo  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo <i>institucion</i> no es NULL devuelve 1, sino 0	Porcentaje de valores en 1 en toda la columna.
<b>Análisis y solución</b>					
	<b>Consulta</b>		<b>Resultado</b>	<b>Solución</b>	
1	SELECT COUNT(*) FROM establecimientos		100%	Se le asigna el valor <i>No Indicado</i> .	

	WHERE institucion IS NOT NULL;		
--	--------------------------------	--	--

### Localidades

<b>Tablas:</b> <i>localidades</i>					
<b>Atributos:</b> <i>localidades.departamento</i>					
Modelo de calidad					
	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Complejidad	Densidad	<u>Departamento nulo:</u> Verifica si la celda tiene valor nulo  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo <i>departamento</i> no es NULL devuelve 1, sino 0	Porcentaje de valores en 1 en toda la columna.
Análisis y solución					
	Consulta	Resultado	Solución		
1	SELECT COUNT(*) FROM localidades WHERE departamento IS NOT NULL;	100%	-		

### Domicilios

<b>Tablas:</b> <i>domicilios</i>					
<b>Atributos:</b> a. <i>domicilios.tipo_zona</i> b. <i>domicilios.id_pais</i> c. <i>domicilios.id_calle</i> d. <i>domicilios.id_localidad</i>					
Modelo de calidad					
	Dimensión	Factor	Métrica	Procedimientos de Medición	Agregación
1	Compleitud	Densidad	<u>Tipo Zona, País, Calle o Localidad nulo:</u> Verifica si la celda tiene valor nulo  Tipo de resultado: (0,1) Granularidad: Celda	Si el atributo no es NULL devuelve 1, sino 0	Porcentaje de valores en 1 en toda la columna.
Análisis y solución					
	Consulta	Resultado	Solución		
1.a	SELECT COUNT(*) FROM domicilios WHERE tipo_zona IS NOT NULL;	43,81%	Se le asigna el id 0 con descripción <i>No Indicado</i> .		
1.b	SELECT COUNT(*) FROM domicilios WHERE id_pais IS NOT NULL;	99,99%			
1.c	SELECT COUNT(*) FROM domicilios WHERE id_calle IS NOT NULL;	71,58%	Se le asigna el valor del atributo <i>domicilio_otro</i> de la tabla <i>domicilios</i> , y en caso de que este también sea vacío se le asigna el id 0 con descripción <i>No Indicado</i> .		

1.d	SELECT COUNT(*) FROM domicilios WHERE id_localidad IS NOT NULL;	98,61%	Se le asigna el id 0 con descripción <i>No Indicado.</i>
-----	--	--------	---



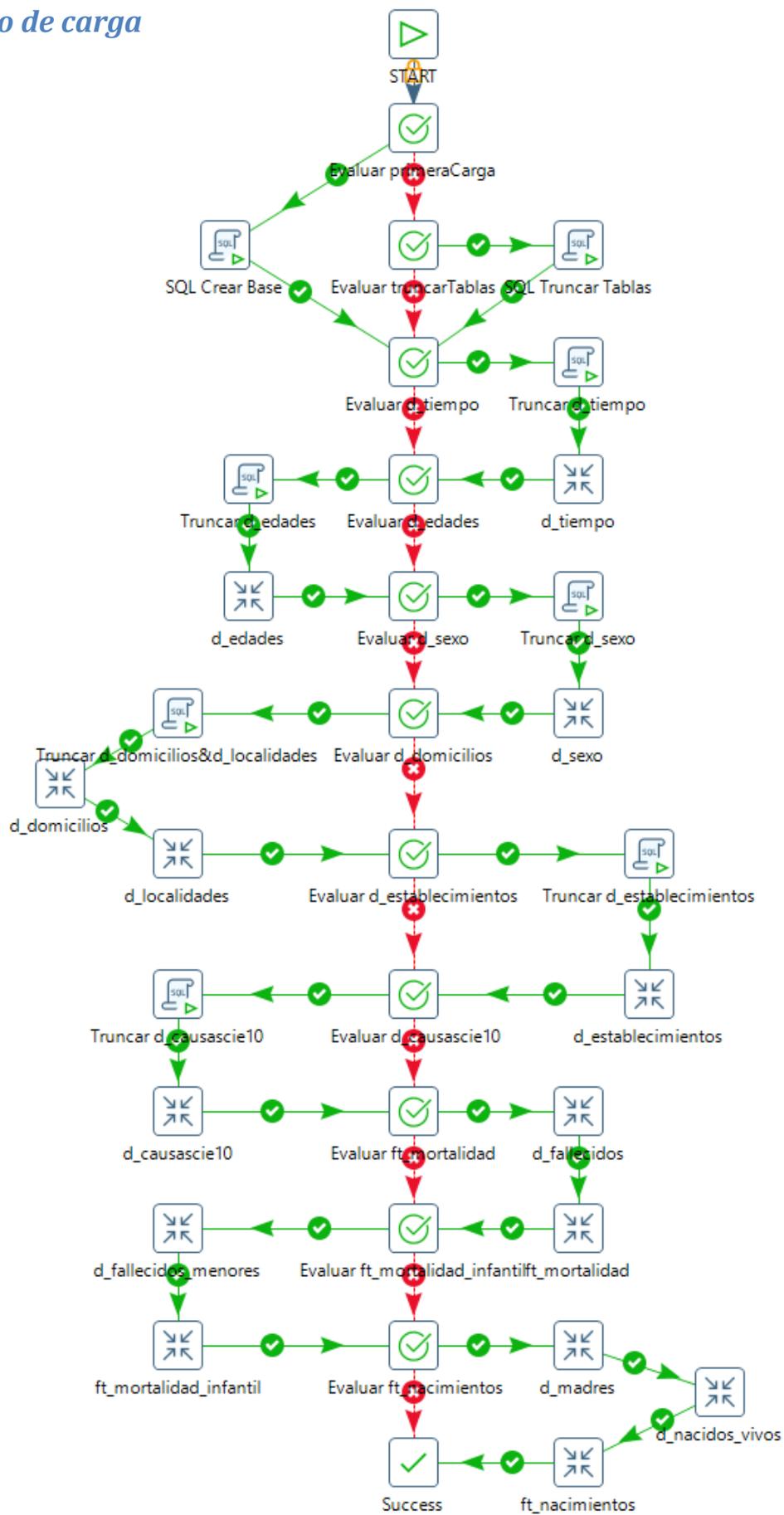
## Anexo II – Carga y actualización

A continuación se detallan los procesos de extracción, transformación y carga del Data Warehouse. Además, en el cuadro III.1 se presentan los valores por defecto de cada uno de los parámetros utilizados por los procesos ETL, que fueron descritos en el Capítulo 5 en la sección *Carga y actualización*.

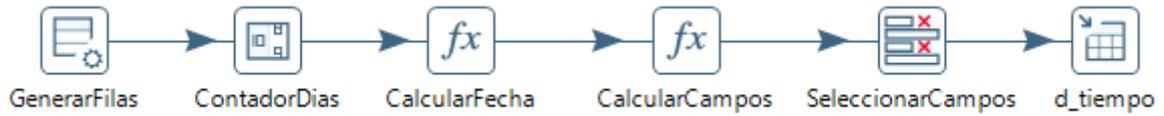
Parámetro	Valor por defecto
<i>primerCarga</i>	<i>false</i>
<i>d_causascie10</i>	<i>false</i>
<i>d_domicilios</i>	<i>false</i>
<i>d_edades</i>	<i>false</i>
<i>dsexo</i>	<i>false</i>
<i>d_establecimientos</i>	<i>false</i>
<i>d_tiempo</i>	<i>false</i>
<i>ft_nacimientos</i>	<i>false</i>
<i>ft_mortalidad</i>	<i>false</i>
<i>ft_mortalidad_infantil</i>	<i>false</i>
<i>truncarTablas</i>	<i>false</i>
<i>fechaInicio</i>	<i>0001-01-01</i>
<i>fechaFin</i>	<i>2200-01-01</i>
<i>host</i>	<i>localhost</i>
<i>port</i>	<i>3306</i>
<i>user</i>	<i>root</i>
<i>password</i>	<i>password</i>
<i>esquemaFuentes</i>	<i>(vacío)</i>

Cuadro III.1 - Carga y actualización del Data Warehouse - Parámetros por defecto

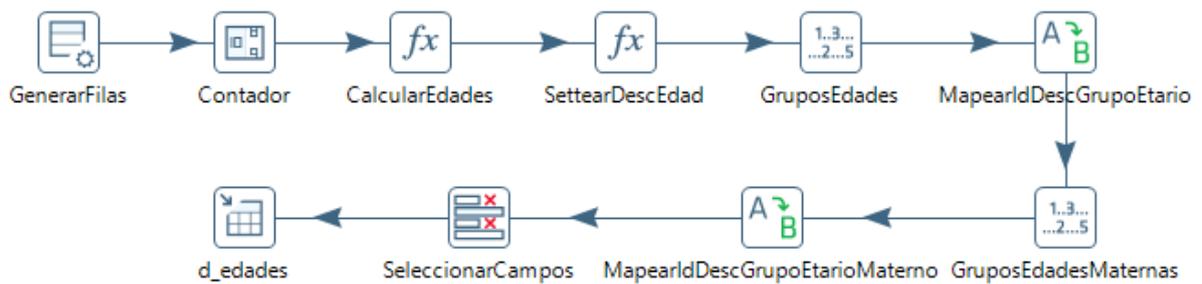
### Proceso de carga



### Dimensión Tiempo



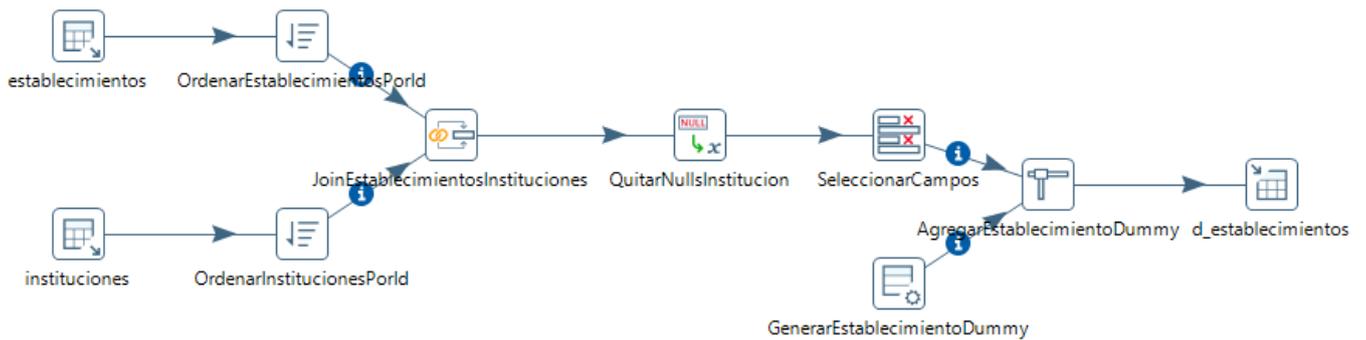
### Dimensión Edades



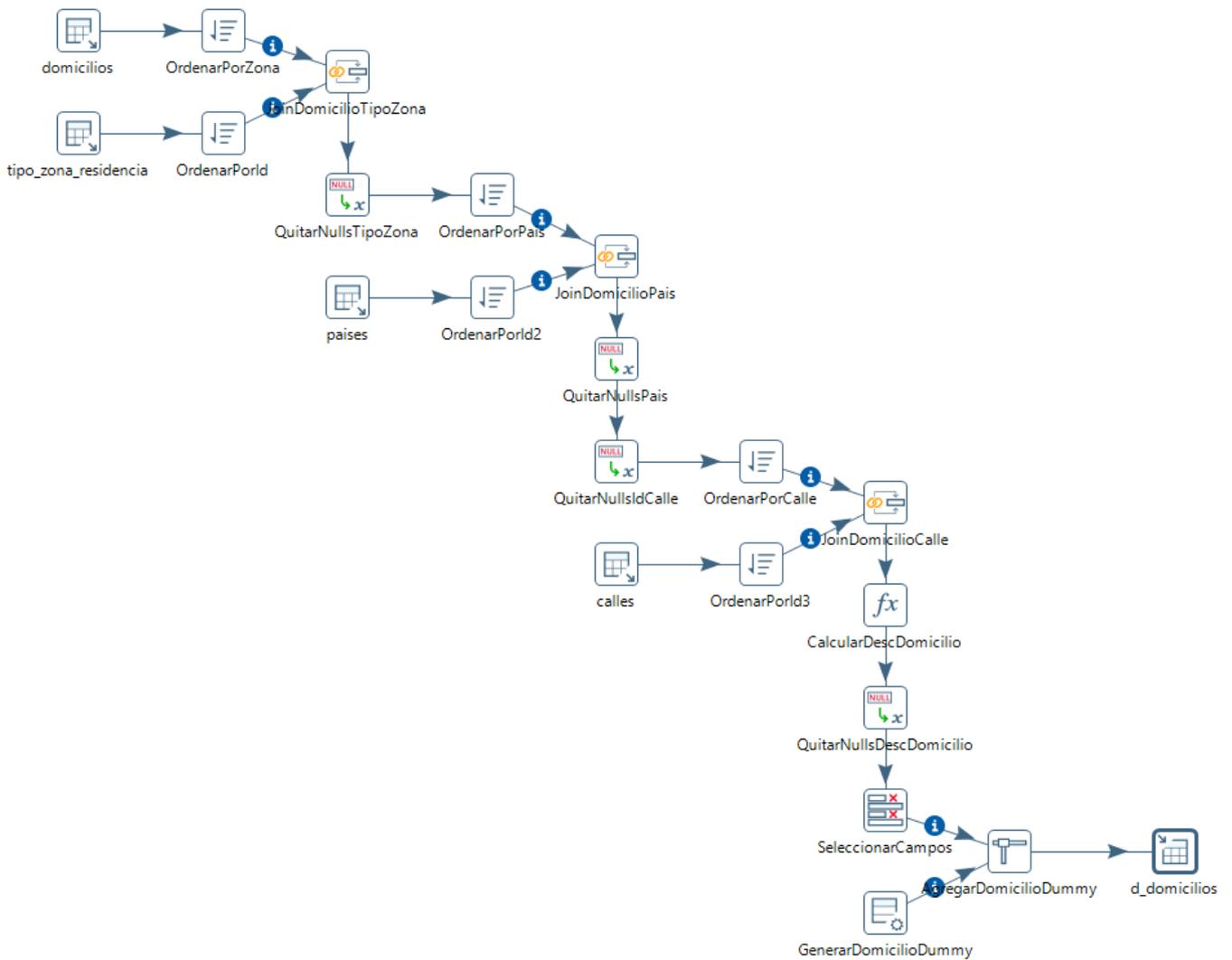
### Dimensión Sexo



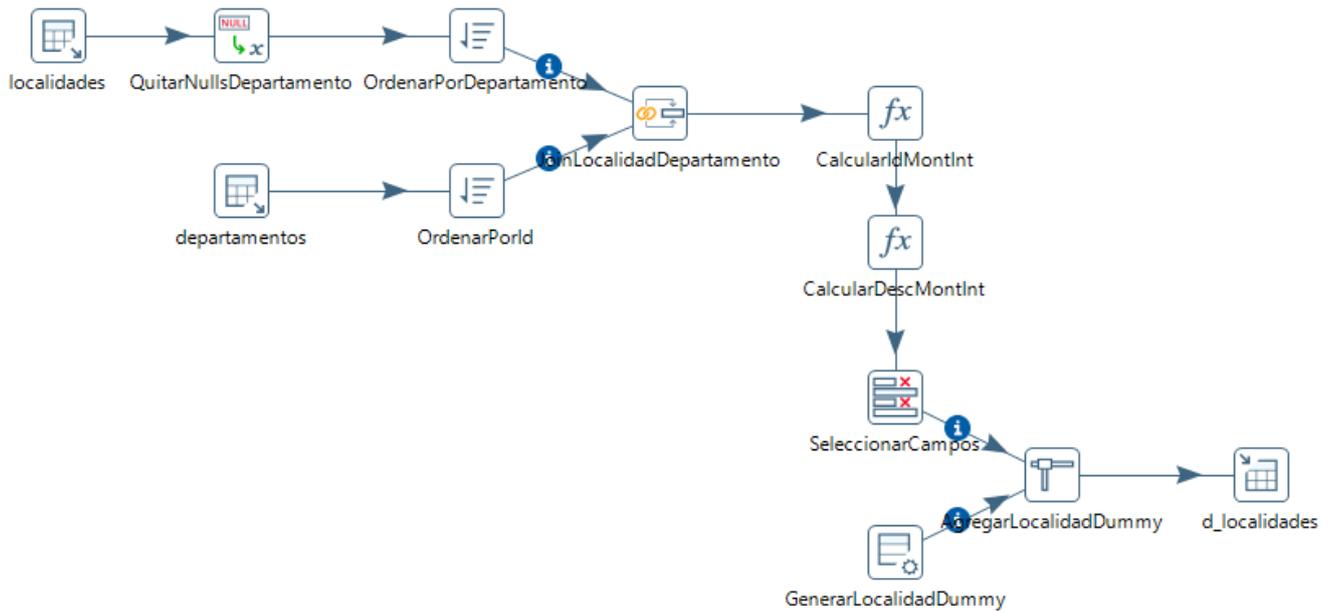
### Dimensión Establecimientos



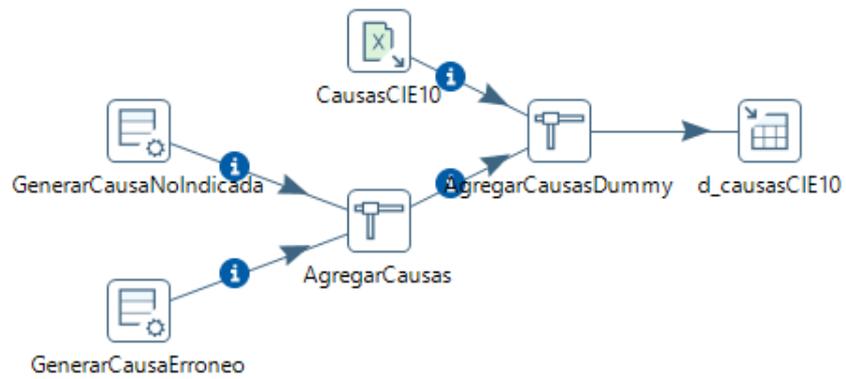
### Dimensión Domicilios



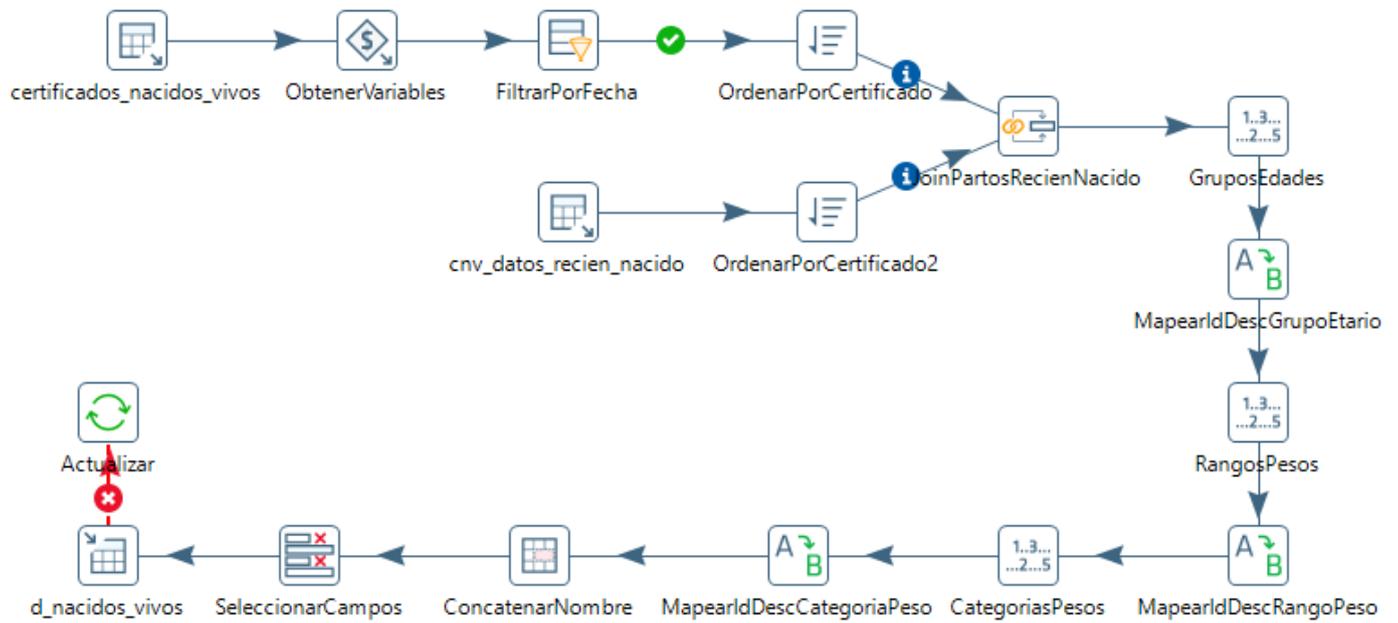
### Dimensión Localidades



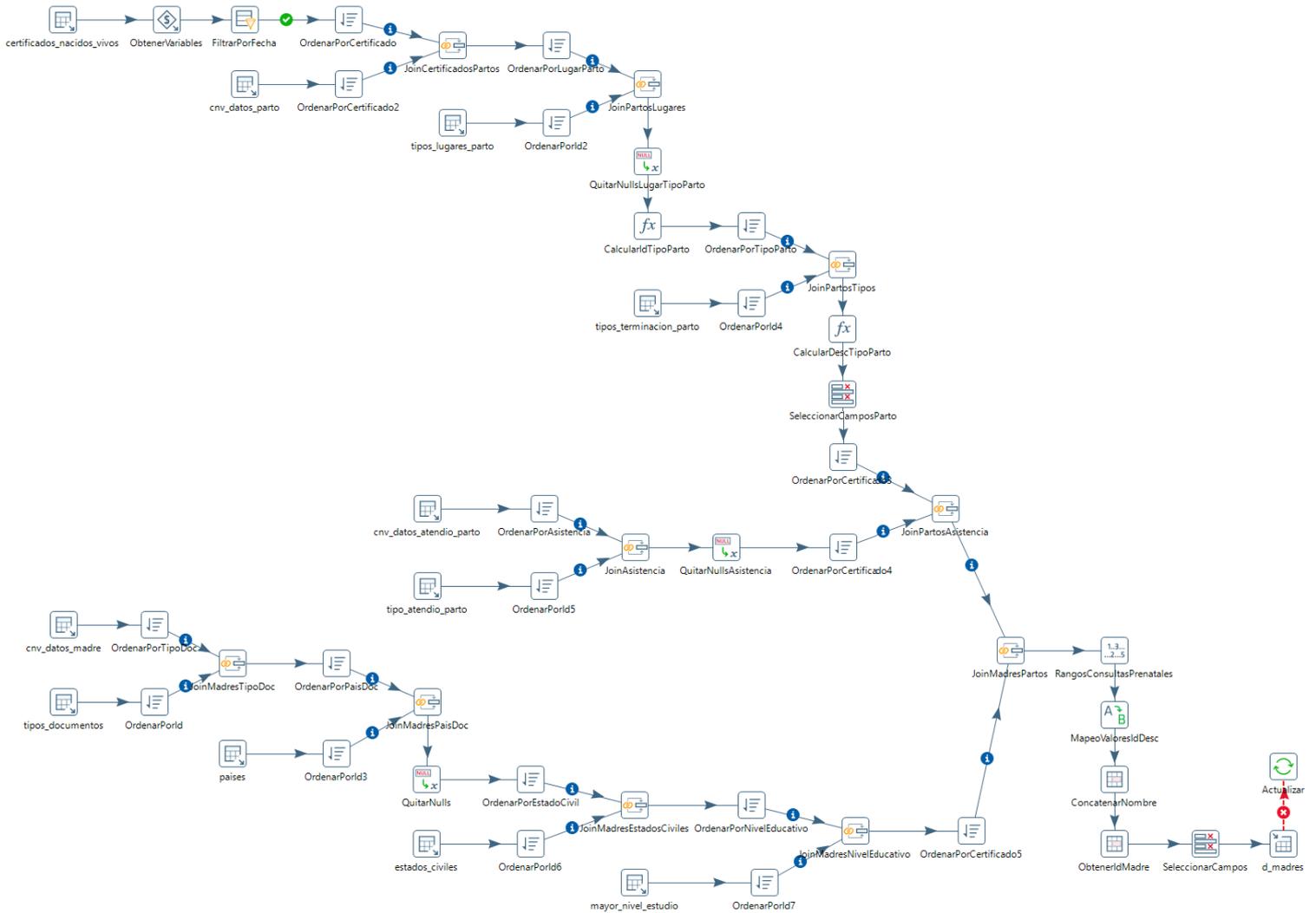
### Dimensión CausasCIE10



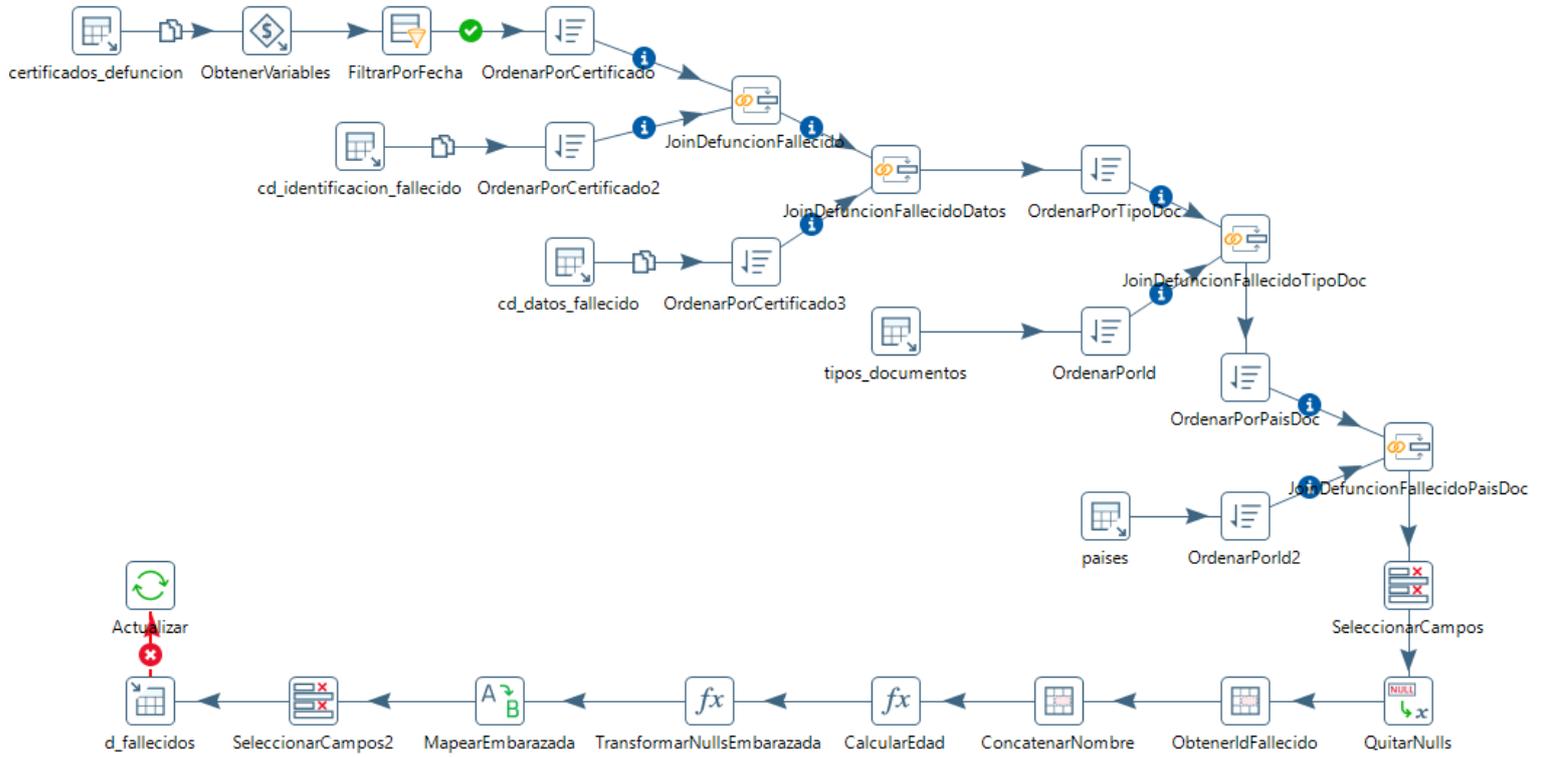
### Dimensión Nacidos Vivos



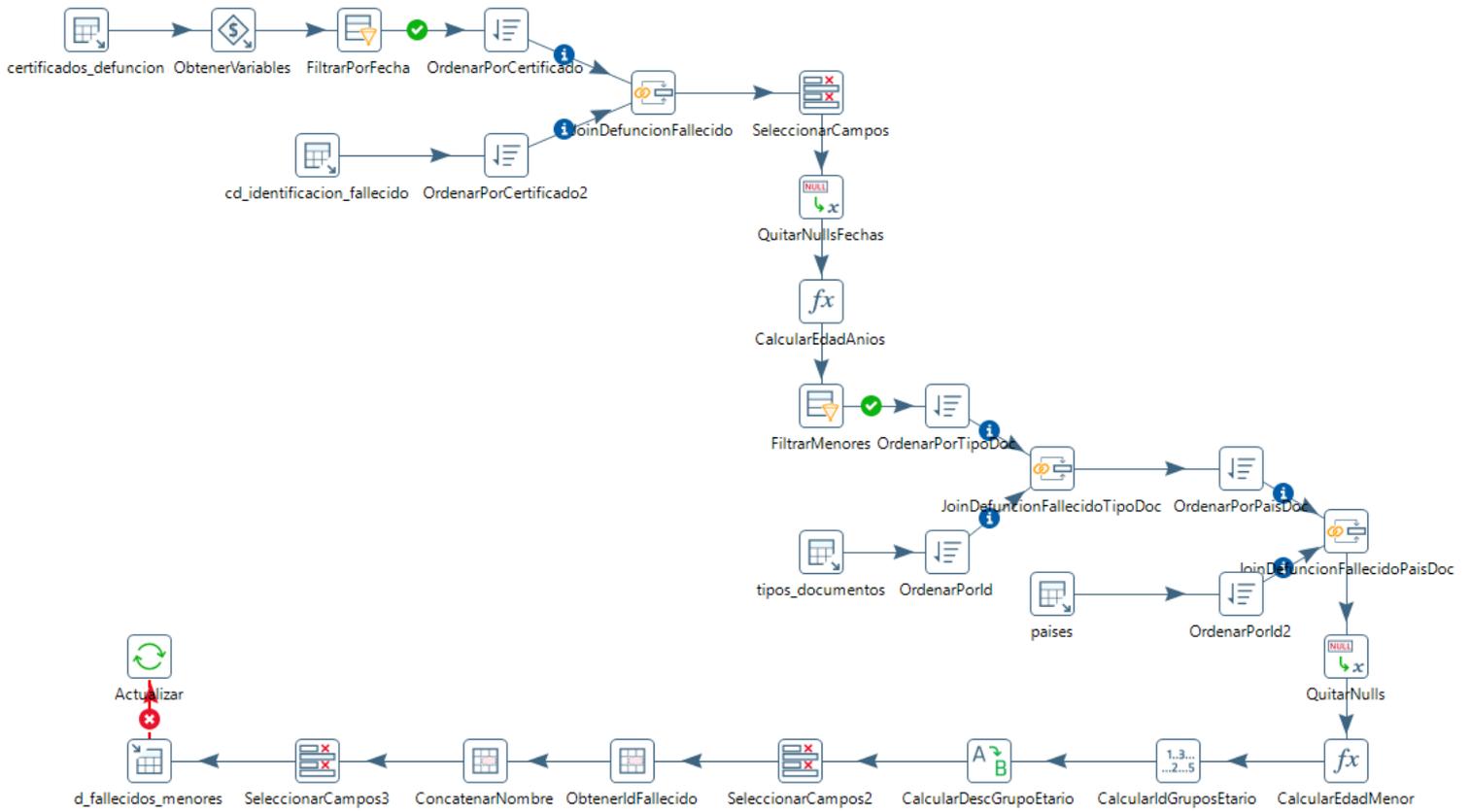
## Dimensión Madres



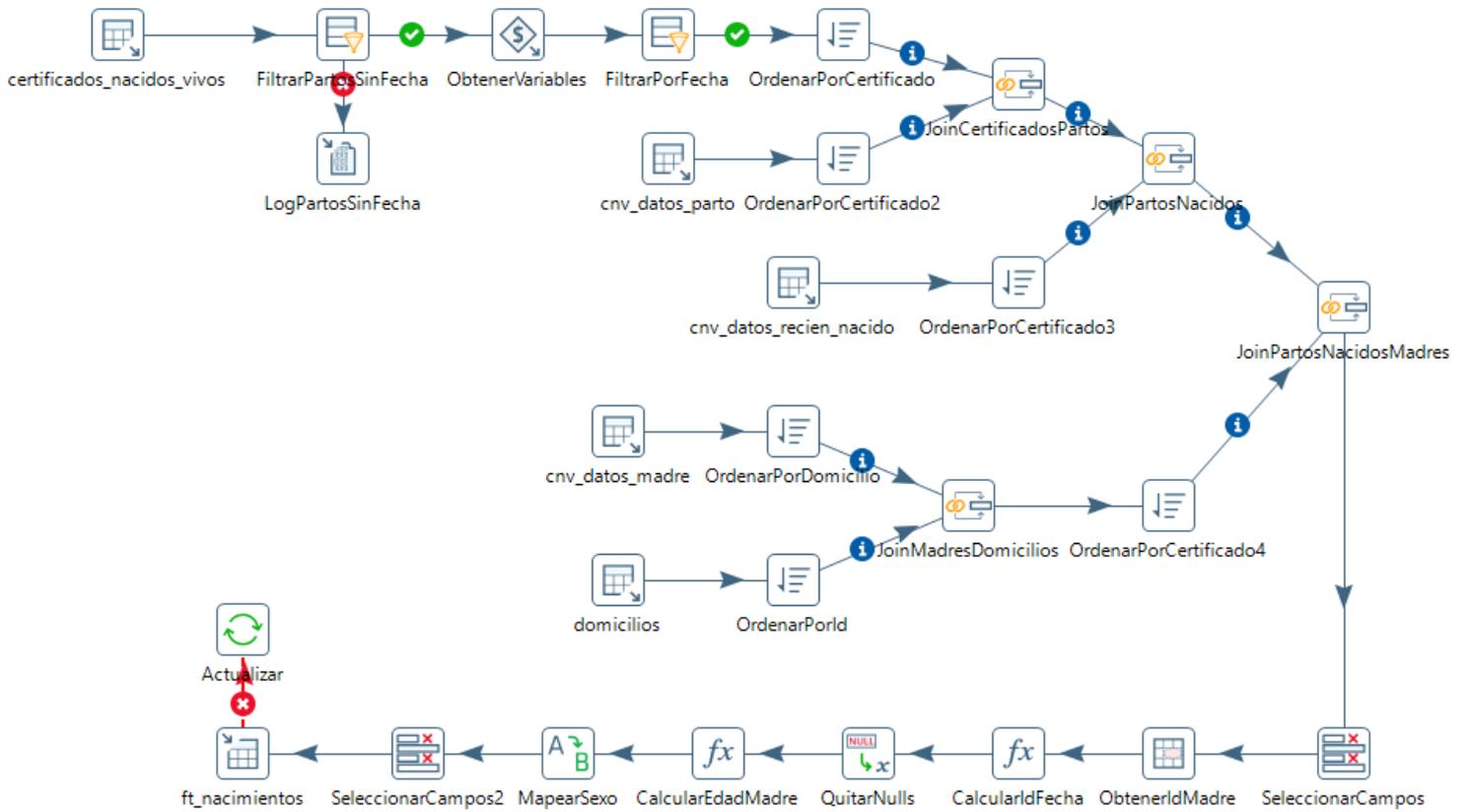
## Dimensión Fallecidos



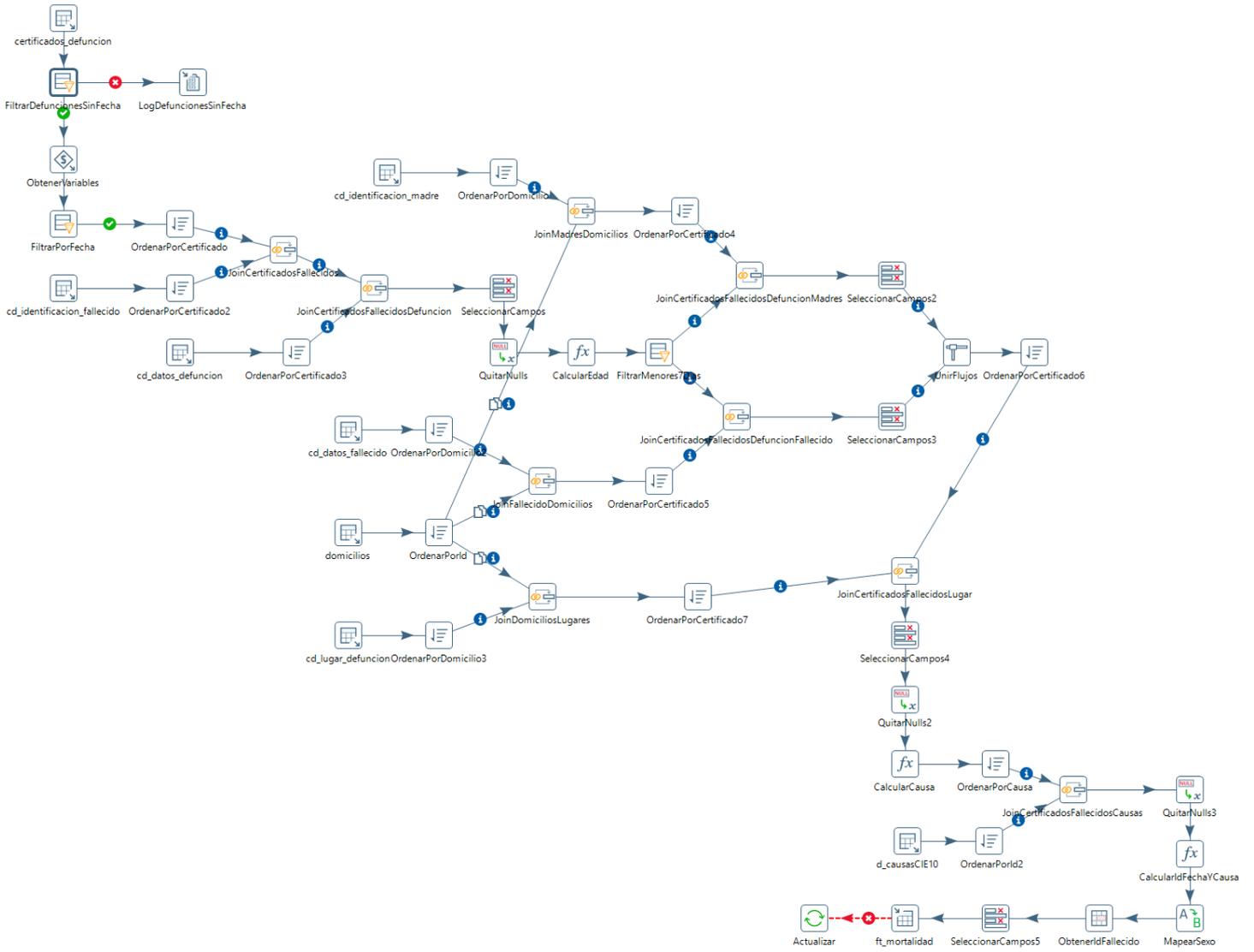
## Dimensión Fallecidos Menores



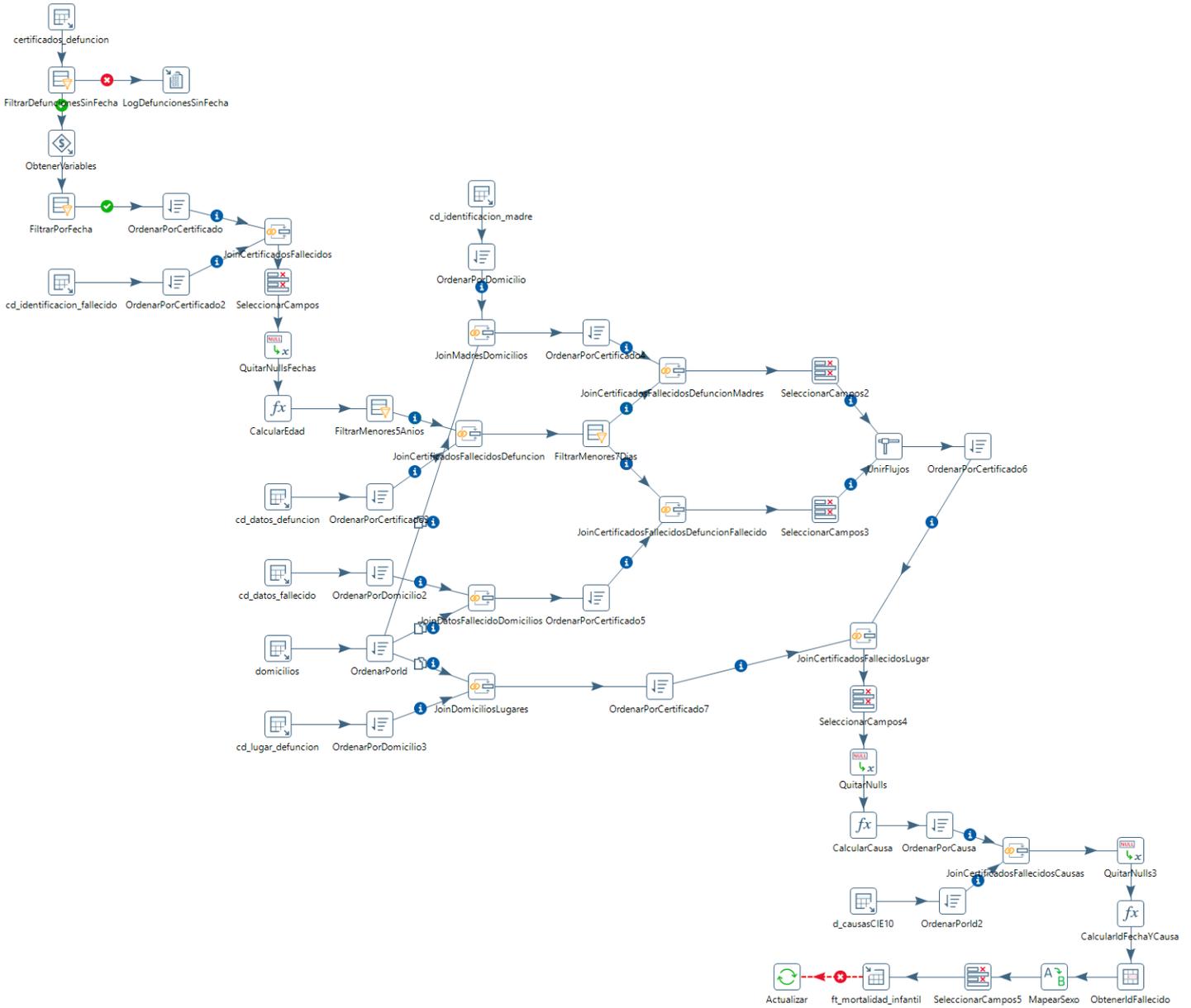
### Tabla de hecho Nacimientos



## Tabla de hecho Mortalidad



## Tabla de hecho Mortalidad Infantil



## Anexo III – Consultas para verificación de ETL

En este Anexo se detallan las consultas SQL elaboradas para la verificación del proceso de extracción, transformación y carga del Data Warehouse.

```
/* ESTABLECIMIENTOS */
SELECT idEstablecimientos AS idEstablecimiento,
       e.Nombre AS descEstablecimiento,
       COALESCE(institucion, 0) AS idInstitucion,
       COALESCE(i.Nombre, 'No Indicado') AS descInstitucion
FROM establecimientos e LEFT JOIN instituciones i ON institucion = id_institucion
UNION
SELECT 0 AS idEstablecimiento,
       'No Indicado' AS descEstablecimiento,
       0 AS idInstitucion,
       'No Indicado' AS descInstitucion;

/* DOMICILIOS */
SELECT id_domicilio AS idDomicilio,
       CASE COALESCE(d.id_Calle, 0)
         WHEN 0 THEN IF(domicilio_otro IS NULL, 'No Indicado', CONCAT_WS(' ', domicilio_otro,
                               numero_puerta))
         ELSE CONCAT_WS(' ', c.Nombre, numero_puerta)
       END AS descDomicilio,
       COALESCE(tipo_zona, 0) AS idTipoZona,
       COALESCE(tz.Nombre, 'No Indicado') AS descTipoZona,
       COALESCE(d.id_pais, 0) AS idPais,
       COALESCE(p.Nombre, 'No Indicado') AS descPais
FROM domicilios d
LEFT JOIN tipo_zona_residencia tz ON tipo_zona = idZona_residencia
LEFT JOIN paises p ON d.id_pais = p.id_pais
LEFT JOIN calles c ON d.id_Calle = c.id_Calle
UNION
SELECT 0 AS idDomicilio,
       'No Indicado' AS descDomicilio,
       0 AS idTipoZona,
       'No Indicado' AS descTipoZona,
       0 AS idPais,
       'No Indicado' AS descPais;

/* LOCALIDADES */
SELECT idLocalidad,
       l.nombre as descLocalidad,
       COALESCE(idDepartamento, 22) AS idDepartamento,
       COALESCE(d.nombre, (' NO INDICADO')) AS descDepartamento,
       CASE COALESCE(idDepartamento, 22)
         WHEN 10 THEN 1
         WHEN 20 THEN 3
         WHEN 22 THEN 4
         ELSE 2
       END AS idMontInt,
       CASE COALESCE(idDepartamento, 22)
         WHEN 10 THEN 'Montevideo'
         WHEN 20 THEN 'Desconocido'
         WHEN 22 THEN 'No Indicado'
         ELSE 'Interior'
       END AS descMontInt
FROM localidades l
LEFT JOIN departamentos d ON departamento = idDepartamento
UNION
SELECT 0 AS idLocalidad,
       'No Indicado' AS descLocalidad,
       22 AS idDepartamento,
       ' NO INDICADO' AS descDepartamento,
       4 AS idMontInt,
       'No Indicado' AS descMontInt;
```

```

/* NACIDOS VIVOS */
SELECT numero_documento_reserva AS cedulaNacido,
    COALESCE(CONCAT_WS(' ', TRIM(primer_nombre), TRIM(primer_apellido)), 'No Indicado') AS
    nombreNacido,
    peso,
    CASE
        WHEN peso = 0 THEN 0
        WHEN peso BETWEEN 500 and 999 THEN 1
        WHEN peso BETWEEN 1000 and 1499 THEN 2
        WHEN peso BETWEEN 1500 and 1999 THEN 3
        WHEN peso BETWEEN 2000 and 2499 THEN 4
        WHEN peso BETWEEN 2500 and 2999 THEN 5
        WHEN peso BETWEEN 3000 and 3499 THEN 6
        WHEN peso BETWEEN 3500 and 3999 THEN 7
        WHEN peso BETWEEN 4000 and 8000 THEN 8
        ELSE 9
    END AS idRangoPeso,
    CASE
        WHEN peso = 0 THEN 'No Indicado'
        WHEN peso BETWEEN 500 and 999 THEN 'Menos de 1000'
        WHEN peso BETWEEN 1000 and 1499 THEN 'De 1000 a 1499'
        WHEN peso BETWEEN 1500 and 1999 THEN 'De 1500 a 1999'
        WHEN peso BETWEEN 2000 and 2499 THEN 'De 2000 a 2499'
        WHEN peso BETWEEN 2500 and 2999 THEN 'De 2500 a 2999'
        WHEN peso BETWEEN 3000 and 3499 THEN 'De 3000 a 3499'
        WHEN peso BETWEEN 3500 and 3999 THEN 'De 3500 a 3999'
        WHEN peso BETWEEN 4000 and 8000 THEN 'De 4000 a 8000'
        ELSE 'Erroneo'
    END AS descRangoPeso,
    CASE
        WHEN peso = 0 THEN 0
        WHEN peso BETWEEN 500 and 999 THEN 1
        WHEN peso BETWEEN 1000 and 2499 THEN 2
        WHEN peso BETWEEN 2500 and 8000 THEN 3
        ELSE 4
    END AS idCategoriaPeso,
    CASE
        WHEN peso = 0 THEN 'No Indicado'
        WHEN peso BETWEEN 500 and 999 THEN 'Extremado bajo peso (menos de 1000gr)'
        WHEN peso BETWEEN 1000 and 2499 THEN 'Bajo peso (entre 1000 y 2499gr)'
        WHEN peso BETWEEN 2500 and 8000 THEN 'Normal (2500gr o mas)'
        ELSE 'Erroneo'
    END AS descCategoriaPeso,
    semanas_gestacion_nacimiento AS semanaGestacion,
    CASE
        WHEN semanas_gestacion_nacimiento = 0 THEN 0
        WHEN semanas_gestacion_nacimiento BETWEEN 20 AND 31 THEN 1
        WHEN semanas_gestacion_nacimiento BETWEEN 32 AND 33 THEN 2
        WHEN semanas_gestacion_nacimiento BETWEEN 34 AND 35 THEN 3
        WHEN semanas_gestacion_nacimiento = 36 THEN 4
        WHEN semanas_gestacion_nacimiento = 37 THEN 5
        WHEN semanas_gestacion_nacimiento = 38 THEN 6
        WHEN semanas_gestacion_nacimiento = 39 THEN 7
        WHEN semanas_gestacion_nacimiento = 40 THEN 8
        WHEN semanas_gestacion_nacimiento = 41 THEN 9
        WHEN semanas_gestacion_nacimiento = 42 THEN 10
        ELSE 11
    END AS idGrupoEtarioGestacion,
    CASE
        WHEN semanas_gestacion_nacimiento = 0 THEN 'No Indicado'
        WHEN semanas_gestacion_nacimiento BETWEEN 20 AND 31 THEN 'Entre 20 y 31'
        WHEN semanas_gestacion_nacimiento BETWEEN 32 AND 33 THEN 'Entre 32 y 33'
        WHEN semanas_gestacion_nacimiento BETWEEN 34 AND 35 THEN 'Entre 34 y 35'
        WHEN semanas_gestacion_nacimiento = 36 THEN '36'
        WHEN semanas_gestacion_nacimiento = 37 THEN '37'
        WHEN semanas_gestacion_nacimiento = 38 THEN '38'
        WHEN semanas_gestacion_nacimiento = 39 THEN '39'
        WHEN semanas_gestacion_nacimiento = 40 THEN '40'
        WHEN semanas_gestacion_nacimiento = 41 THEN '41'
        WHEN semanas_gestacion_nacimiento = 42 THEN 'Mas de 41'
        ELSE 'Erroneo'
    END AS descGrupoEtarioGestacion
FROM certificados_nacidos_vivos cnv
LEFT JOIN cnv_datos_recien_nacido drn
ON (cnv.id_certificado = drn.id_certificado AND cnv.serie = drn.serie);
    
```

```

/* MADRES */
SELECT CONCAT(TRIM(cnvp.id_certificado), TRIM(cnvp.serie)) AS idMadre,
COALESCE(tipo_documento, 0) AS idDocumentoTipoMadre,
COALESCE(td.Nombre, 'No Indicado') AS descDocumentoTipoMadre,
COALESCE(pais_emisor_documento, 0) AS idDocumentoPaisMadre,
COALESCE(p.Nombre, 'No Indicado') AS descDocumentoPaisMadre,
COALESCE(numero_documento, 'No Indicado') AS documentoNroMadre,
COALESCE(CONCAT_WS(' ', TRIM(primer_nombre), TRIM(primer_apellido)), 'No Indicado') AS nombreMadre,
idTipoParto,
IF(idTipoParto = 22, 'Erroneo', ttp.Nombre) AS descTipoParto,
COALESCE(id_tipo_atendio_parto, 0) AS idAsistenciaParto,
COALESCE(tap.nombre, 'No Indicado') AS descAsistenciaParto,
COALESCE(lugar_parto, 0) AS idLugarParto,
COALESCE(tlp.Nombre, 'No Indicado') AS descLugarParto,
total_consultas_prenatales AS cantConsultasPrenatales,
CASE
    WHEN total_consultas_prenatales = 0 THEN 0
    WHEN total_consultas_prenatales = 1 THEN 1
    WHEN total_consultas_prenatales = 2 THEN 2
    WHEN total_consultas_prenatales = 3 THEN 3
    WHEN total_consultas_prenatales = 4 THEN 4
    WHEN total_consultas_prenatales = 5 THEN 5
    WHEN total_consultas_prenatales = 6 THEN 6
    WHEN total_consultas_prenatales = 7 THEN 7
    WHEN total_consultas_prenatales = 8 THEN 8
    WHEN total_consultas_prenatales = 9 THEN 9
    WHEN total_consultas_prenatales BETWEEN 10 AND 40 THEN 10
    ELSE 11
END AS idRangoConsultasPrenatales,
CASE
    WHEN total_consultas_prenatales = 0 THEN '0'
    WHEN total_consultas_prenatales = 1 THEN '1'
    WHEN total_consultas_prenatales = 2 THEN '2'
    WHEN total_consultas_prenatales = 3 THEN '3'
    WHEN total_consultas_prenatales = 4 THEN '4'
    WHEN total_consultas_prenatales = 5 THEN '5'
    WHEN total_consultas_prenatales = 6 THEN '6'
    WHEN total_consultas_prenatales = 7 THEN '7'
    WHEN total_consultas_prenatales = 8 THEN '8'
    WHEN total_consultas_prenatales = 9 THEN '9'
    WHEN total_consultas_prenatales BETWEEN 10 AND 40 THEN '10 o mas'
    ELSE 'Erroneo'
END AS descRangoConsultasPrenatales,
COALESCE(mayor_nivel_estudios, 41) AS idNivelEducativo,
COALESCE(mne.Descripcion, 'NO INDICADO') AS descNivelEducativo,
COALESCE(estado_civil, 21) AS idEstadoConyugal,
COALESCE(ec.descripcion, 'NO INDICADO') AS descEstadoConyugal
FROM (SELECT cnv.id_certificado, cnv.serie, cnv.fecha_parto, tipo_certificado,
CASE tipo_certificado
    WHEN 1 THEN (CASE COALESCE(tipo_parto, 21)
        WHEN 0 THEN 1
        WHEN 1 THEN 2
        ELSE 22
    END)
    ELSE COALESCE(tipo_parto, 21)
END AS idTipoParto, lugar_parto
FROM certificados_nacidos_vivos cnv
LEFT JOIN cnv_datos_parto dp
ON (cnv.id_certificado = dp.id_certificado AND cnv.serie = dp.serie)) cnvp
LEFT JOIN tipos_lugares_parto tlp
ON (lugar_parto = id_lugar)
LEFT JOIN tipos_terminacion_parto ttp
ON (idTipoParto = id_terminacion)
LEFT JOIN (cnv_datos_atendio_parto ap
LEFT JOIN tipo_atendio_parto tap
ON (id_tipo_atendio_parto = id_tipo))
ON (cnvp.id_certificado = ap.id_certificado AND cnvp.serie = ap.serie)
LEFT JOIN (cnv_datos_madre dm
LEFT JOIN tipos_documentos td
ON (tipo_documento = idTipo_Documento)
LEFT JOIN paises p
ON (pais_emisor_documento = id_pais)
LEFT JOIN estados_civiles ec
ON (estado_civil = idEstados_Civiles)
LEFT JOIN mayor_nivel_estudio mne
ON (mayor_nivel_estudios = id))
ON (cnvp.id_certificado = dm.id_certificado AND cnvp.serie = dm.serie);
    
```

```

/* FALLECIDOS */
SELECT CONCAT(TRIM(cd.ID_Certificado), TRIM(cd.Serie)) AS idFallecido,
COALESCE(Tipo_Doc_Id, 0) AS idDocumentoTipo,
COALESCE(td.Nombre, 'No Indicado') AS descDocumentoTipo,
COALESCE(Pais_Doc_Id, 0) AS idDocumentoPais,
COALESCE(p.Nombre, 'No Indicado') AS descDocumentoPais,
N_Doc AS documentoNro,
COALESCE(CONCAT_WS(' ', TRIM(primer_nombre), TRIM(primer_apellido)), 'No Indicado') AS
nombreFallecido,
CASE COALESCE(Embarazada, 0)
    WHEN 0 THEN (CASE Sexo
        WHEN 'F' THEN IF(TIMESTAMPDIFF(YEAR, Fecha_Nac, fecha_muerte) BETWEEN 9 AND 70,
            'Se Ignora', 'No Corresponde')
        WHEN 'M' THEN 'No Corresponde'
        ELSE 'Desconocido'
    END)
    WHEN 1 THEN 'Si'
    WHEN 2 THEN 'No'
    WHEN 3 THEN 'Se Ignora'
    WHEN 4 THEN 'No Corresponde'
    WHEN 5 THEN 'No Aplica'
    WHEN 6 THEN 'Desconocido'
    ELSE 'Erroneo'
END AS embarazada
FROM certificados_defuncion cd
LEFT JOIN cd_identificacion_fallecido idf
ON (cd.ID_Certificado = idf.ID_Certificado AND cd.Serie = idf.Serie)
LEFT JOIN cd_datos_fallecido df
ON (cd.ID_Certificado = df.ID_Certificado AND cd.Serie = df.Serie)
LEFT JOIN tipos_documentos td
ON (Tipo_Doc_Id = idTipo_Documento)
LEFT JOIN paises p
ON (Pais_Doc_Id = id_pais)
WHERE grupo_etario <> 1;

/* FALLECIDOS MENORES */
SELECT CONCAT(TRIM(cd.ID_Certificado), TRIM(cd.Serie)) AS idFallecidoMenor,
COALESCE(Tipo_Doc_Id, 0) AS idDocumentoTipo,
COALESCE(td.Nombre, 'No Indicado') AS descDocumentoTipo,
COALESCE(Pais_Doc_Id, 0) AS idDocumentoPais,
COALESCE(p.Nombre, 'No Indicado') AS descDocumentoPais,
N_Doc AS documentoNro,
COALESCE(CONCAT_WS(' ', TRIM(primer_nombre), TRIM(primer_apellido)), 'No Indicado') AS
nombreFallecidoMenor,
TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01/01/0001')), COALESCE(fecha_muerte, '01/01/0001'))
AS edadMenor,
CASE
    WHEN TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte,
'01-01-0001')) BETWEEN 0 AND 7 THEN 1
    WHEN TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte,
'01-01-0001')) BETWEEN 8 AND 27 THEN 2
    WHEN TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte,
'01-01-0001')) BETWEEN 28 AND 364 THEN 3
    WHEN TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte,
'01-01-0001')) BETWEEN 365 AND 1827 THEN 4
END AS idGrupoEtarioMenor,
CASE
    WHEN TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte,
'01-01-0001')) BETWEEN 0 AND 7 THEN 'Neonatal precoz'
    WHEN TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte,
'01-01-0001')) BETWEEN 8 AND 27 THEN 'Neonatal tardia'
    WHEN TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte,
'01-01-0001')) BETWEEN 28 AND 364 THEN 'Postneonatal'
    WHEN TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte,
'01-01-0001')) BETWEEN 365 AND 1827 THEN '1 año o mas'
END AS descGrupoEtarioMenor
FROM certificados_defuncion cd
LEFT JOIN cd_identificacion_fallecido
ON (cd.ID_Certificado = cd_identificacion_fallecido.ID_Certificado AND cd.Serie =
cd_identificacion_fallecido.Serie)
LEFT JOIN cd_datos_fallecido
ON (cd.ID_Certificado = cd_datos_fallecido.ID_Certificado AND cd.Serie = cd_datos_fallecido.Serie)
LEFT JOIN tipos_documentos td
ON (Tipo_Doc_Id = idTipo_Documento)
LEFT JOIN paises p
ON (Pais_Doc_Id = id_pais)
WHERE grupo_etario <> 1
AND TIMESTAMPDIFF(YEAR, DATE(COALESCE(Fecha_Nac, '01-01-0001')), COALESCE(fecha_muerte, '01-01-0001'))
< 5;
    
```

```

/* NACIMIENTOS */
SELECT numero_documento_reserva AS cedulaNacido,
    CONCAT(TRIM(cnv.id_certificado), TRIM(cnv.serie)) AS idMadre,
    CASE
        WHEN ((COALESCE(edad, 999) = 999 AND TIMESTAMPDIFF(YEAR, DATE(COALESCE(fecha_nacimiento,
            '01-01-0001')), DATE(cnv.fecha_parto)) > 150) OR edad > 150) THEN 151
        ELSE edad
    END AS edadMadre,
    YEAR(cnv.fecha_parto)*10000 + MONTH(cnv.fecha_parto)*100 + DAY(cnv.fecha_parto) AS idFecha,
    CASE
        WHEN COALESCE(sexo, 'D') = 'F' THEN 1
        WHEN COALESCE(sexo, 'D') = 'M' THEN 2
        WHEN COALESCE(sexo, 'D') = 'I' THEN 3
        WHEN COALESCE(sexo, 'D') = 'D' OR (LENGTH(sexo) = 0) THEN 4
        ELSE 5
    END AS idSexo,
    COALESCE(establecimiento_ocurrencia, 0) AS idEstablecimiento,
    COALESCE(localidad_ocurrencia, 0) AS idLocalidadOcurrencia,
    COALESCE(dm.id_domicilio, 0) AS idDomicilioResidenciaMadre,
    COALESCE(d.id_localidad, 0) AS idLocalidadResidenciaMadre
FROM certificados_nacidos_vivos cnv
LEFT JOIN cnv_datos_parto dp
    ON (cnv.id_certificado = dp.id_certificado AND cnv.serie = dp.serie)
LEFT JOIN cnv_datos_recien_nacido drn
    ON (cnv.id_certificado = drn.id_certificado AND cnv.serie = drn.serie)
LEFT JOIN (cnv_datos_madre dm
            LEFT JOIN domicilios d
                ON (dm.id_domicilio = d.id_domicilio))
    ON (cnv.id_certificado = dm.id_certificado AND cnv.serie = dm.serie)
WHERE cnv.fecha_parto IS NOT NULL;

/* MORTALIDAD */
SELECT CONCAT(TRIM(cd.ID_Certificado), TRIM(cd.Serie)) AS idFallecido,
    COALESCE(idCausaCIE, 9999) AS idCausa,
    YEAR(cd.fecha_muerte)*10000 + MONTH(cd.fecha_muerte)*100 + DAY(cd.fecha_muerte) AS idFecha,
    IF(TIMESTAMPDIFF(YEAR, DATE(COALESCE(Fecha_Nac, STR_TO_DATE('01,01,0001','%d,%m,%Y'))), DATE(cd.
    fecha_muerte)) BETWEEN 0 AND 150,
    TIMESTAMPDIFF(YEAR, DATE(COALESCE(Fecha_Nac, STR_TO_DATE('01,01,0001','%d,%m,%Y'))), DATE(cd.
    fecha_muerte)),
    IF(edad_anios_def_calculada BETWEEN 0 AND 150,
        edad_anios_def_calculada,
        IF(edad_anios_def_digit BETWEEN 0 AND 150, edad_anios_def_digit, 151))) AS idEdad,
    CASE
        WHEN COALESCE(sexo, 'D') = 'F' THEN 1
        WHEN COALESCE(sexo, 'D') = 'M' THEN 2
        WHEN COALESCE(sexo, 'D') = 'I' THEN 3
        WHEN COALESCE(sexo, 'D') = 'D' OR (LENGTH(sexo) = 0) THEN 4
        ELSE 5
    END AS idSexo,
    COALESCE(Lugar_Def_Id, 0) AS idDomicilioOcurrencia,
    COALESCE(domd.id_localidad, 0) AS idLocalidadOcurrencia,
    IF(TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, STR_TO_DATE('01,01,0001','%d,%m,%Y'))), DATE(cd.
    fecha_muerte)) BETWEEN 0 AND 6,
    COALESCE(idm.id_domicilio, 0),
    COALESCE(df.id_domicilio, 0)) AS idDomicilioResidencia,
    IF(TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, STR_TO_DATE('01,01,0001','%d,%m,%Y'))), DATE(cd.
    fecha_muerte)) BETWEEN 0 AND 6,
    COALESCE(domm.id_localidad, 0),
    COALESCE(domf.id_localidad, 0)) AS idLocalidadResidencia
FROM certificados_defuncion cd
LEFT JOIN cd_identificacion_fallecido idf
    ON (cd.ID_Certificado = idf.ID_Certificado AND cd.Serie = idf.Serie)
LEFT JOIN (SELECT ddef.ID_Certificado, ddef.Serie, idCausa AS idCausaCIE,
    edad_anios_def_calculada, edad_anios_def_digit
    FROM cd_datos_defuncion ddef
    LEFT JOIN dw_msp.d_causasCIE10
        ON (idCausa = IF(SUBSTRING(COALESCE(enf_causante_cie10, '0000'), CHAR_LENGTH(COALESCE(
        enf_causante_cie10, '0000'))) = 'X',
        SUBSTRING(COALESCE(enf_causante_cie10, '0000') FROM 1 FOR CHAR_LENGTH(
        COALESCE(enf_causante_cie10, '0000'))-1),
        COALESCE(enf_causante_cie10, '0000')))) dd
    ON (cd.ID_Certificado = dd.ID_Certificado AND cd.Serie = dd.Serie)
LEFT JOIN (cd_datos_fallecido df
            LEFT JOIN domicilios domf
                ON (df.id_domicilio = domf.id_domicilio))
    ON (cd.ID_Certificado = df.ID_Certificado AND cd.Serie = df.Serie)
LEFT JOIN (cd_identificacion_madre idm
            LEFT JOIN domicilios domm
                ON (idm.id_domicilio = domm.id_domicilio))
    ON (cd.ID_Certificado = idm.ID_Certificado AND cd.Serie = idm.Serie)
LEFT JOIN (cd_lugar_defuncion ld
            LEFT JOIN domicilios domld
                ON (Lugar_Def_Id = domld.id_domicilio))
    ON (cd.ID_Certificado = ld.ID_Certificado AND cd.Serie = ld.Serie)
WHERE grupo_etario <> 1
    AND cd.fecha_muerte IS NOT NULL;

```

```

/* MORTALIDAD INFANTIL */
SELECT CONCAT(TRIM(cd.ID_Certificado), TRIM(cd.Serie)) AS idFallecidoMenor,
    COALESCE(idCausaCIE, 9999) AS idCausa,
    YEAR(cd.fecha_muerte)*10000 + MONTH(cd.fecha_muerte)*100 + DAY(cd.fecha_muerte) AS idFecha,
    CASE
        WHEN COALESCE(sexo, 'D') = 'F' THEN 1
        WHEN COALESCE(sexo, 'D') = 'M' THEN 2
        WHEN COALESCE(sexo, 'D') = 'I' THEN 3
        WHEN COALESCE(sexo, 'D') = 'D' OR (LENGTH(sexo) = 0) THEN 4
        ELSE 5
    END AS idSexo,
    COALESCE(Lugar_Def_Id, 0) AS idDomicilioOcurrencia,
    COALESCE(domld.id_localidad, 0) AS idLocalidadOcurrencia,
    IF(TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, STR_TO_DATE('01,01,0001','%d,%m,%Y'))), DATE(cd.
    fecha_muerte)) BETWEEN 0 AND 6,
        COALESCE(idm.id_domicilio, 0),
        COALESCE(df.id_domicilio, 0)) AS idDomicilioResidencia,
    IF(TIMESTAMPDIFF(DAY, DATE(COALESCE(Fecha_Nac, STR_TO_DATE('01,01,0001','%d,%m,%Y'))), DATE(cd.
    fecha_muerte)) BETWEEN 0 AND 6,
        COALESCE(domm.id_localidad, 0),
        COALESCE(domf.id_localidad, 0)) AS idLocalidadResidencia
FROM certificados_defuncion cd
LEFT JOIN cd_identificacion_fallecido idf
ON (cd.ID_Certificado = idf.ID_Certificado AND cd.Serie = idf.Serie)
LEFT JOIN (SELECT ddef.ID_Certificado, ddef.Serie, idCausa AS idCausaCIE,
    edad_anios_def_calculada, edad_anios_def_digit
    FROM cd_datos_defuncion ddef
    LEFT JOIN dw_msp.d_causasCIE10
    ON (idCausa = IF(SUBSTRING(COALESCE(enf_causante_cie10, '0000'), CHAR_LENGTH(COALESCE(
    enf_causante_cie10, '0000')))) = 'X',
        SUBSTRING(COALESCE(enf_causante_cie10, '0000') FROM 1 FOR CHAR_LENGTH(
        COALESCE(enf_causante_cie10, '0000'))-1),
        COALESCE(enf_causante_cie10, '0000')))) dd
ON (cd.ID_Certificado = dd.ID_Certificado AND cd.Serie = dd.Serie)
LEFT JOIN (cd_datos_fallecido df
    LEFT JOIN domicilios domf
    ON (df.id_domicilio = domf.id_domicilio))
ON (cd.ID_Certificado = df.ID_Certificado AND cd.Serie = df.Serie)
LEFT JOIN (cd_identificacion_madre idm
    LEFT JOIN domicilios domm
    ON (idm.id_domicilio = domm.id_domicilio))
ON (cd.ID_Certificado = idm.ID_Certificado AND cd.Serie = idm.Serie)
LEFT JOIN (cd_lugar_defuncion ld
    LEFT JOIN domicilios domld
    ON (Lugar_Def_Id = domld.id_domicilio))
ON (cd.ID_Certificado = ld.ID_Certificado AND cd.Serie = ld.Serie)
WHERE grupo_etario <> 1
    AND cd.fecha_muerte IS NOT NULL
    AND TIMESTAMPDIFF(YEAR, DATE(COALESCE(Fecha_Nac, STR_TO_DATE('01,01,0001','%d,%m,%Y'))), DATE(cd.
    fecha_muerte)) < 5;
    
```

## Anexo IV – Implementación de cubos

A continuación se expone en detalle el archivo XML que implementa los cubos presentados durante el informe.

```
<Schema name="MSP" description="Data Warehouse de estadísticas vitales y de salud del MSP">
  <Dimension type="StandardDimension" visible="true" highCardinality="false" name="D_Tiempo" caption="
  Tiempo" description="Tiempo">
    <Hierarchy name="Tiempo" visible="true" hasAll="true" allMemberName="All" primaryKey="idFecha"
    caption="Tiempo" description="Tiempo">
      <Table name="d_tiempo">
      </Table>
      <Level name="Level_Anio" visible="true" table="d_tiempo" column="anio" nameColumn="anio" type=
      "Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Anio" description
      ="Anio">
      </Level>
      <Level name="Level_Mes" visible="true" table="d_tiempo" column="idMes" nameColumn="mes" type=
      "String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Mes" description=
      "Mes">
      </Level>
      <Level name="Level_Fecha" visible="true" table="d_tiempo" column="idFecha" nameColumn="fecha" type=
      "Date" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Fecha" description=
      "Fecha">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" visible="true" highCardinality="false" name="D_LocalidadOcurrencia"
  caption="Localidad ocurrencia" description="Localidad de ocurrencia">
    <Hierarchy name="LocalidadOcurrencia" visible="true" hasAll="true" allMemberName="All" primaryKey=
    "idLocalidad" caption="Localidad ocurrencia" description="Localidad ocurrencia">
      <Table name="d_localidades">
      </Table>
      <Level name="Level_Montevideo_Interior" visible="true" table="d_localidades" column="idMontInt"
      nameColumn="descMontInt" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf=
      "Never" caption="Montevideo / Interior" description="Montevideo / Interior">
      </Level>
      <Level name="Level_Departamento" visible="true" table="d_localidades" column="idDepartamento"
      nameColumn="descDepartamento" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf
      ="Never" caption="Departamento ocurrencia" description="Departamento de ocurrencia">
      </Level>
      <Level name="Level_Localidad" visible="true" table="d_localidades" column="idLocalidad" nameColumn=
      "descLocalidad" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
      caption="Localidad ocurrencia" description="Localidad de ocurrencia">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" visible="true" highCardinality="false" name="D_DomicilioResidencia"
  caption="Domicilio residencia" description="Domicilio residencia">
    <Hierarchy name="Pais" visible="true" hasAll="true" allMemberName="All" primaryKey="idDomicilio"
    caption="Pais residencia" description="Pais residencia">
      <Table name="d_domicilios">
      </Table>
      <Level name="Level_Pais" visible="true" table="d_domicilios" column="idPais" nameColumn="descPais"
      type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Pais
      residencia" description="Pais residencia">
      </Level>
      <Level name="Level_Domicilio" visible="true" table="d_domicilios" column="idDomicilio" nameColumn=
      "descDomicilio" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
      caption="Domicilio residencia" description="Domicilio residencia">
      </Level>
    </Hierarchy>
  </Dimension>
</Schema>
```

```

<Hierarchy name="TipoZona" visible="true" hasAll="true" allMemberName="All" primaryKey="idDomicilio"
caption="Tipo zona residencia" description="Tipo zona residencia">
  <Table name="d_domicilios">
  </Table>
  <Level name="Level_TipoZona" visible="true" table="d_domicilios" column="idTipoZona" nameColumn=
"descTipoZona" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Tipo zona residencia" description="Tipo zona residencia">
  </Level>
  <Level name="Level_Domicilio" visible="true" table="d_domicilios" column="idDomicilio" nameColumn=
"descDomicilio" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Domicilio residencia" description="Domicilio residencia">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" highCardinality="false" name="D_LocalidadResidencia"
caption="Localidad residencia" description="Localidad residencia">
  <Hierarchy name="LocalidadResidencia" visible="true" hasAll="true" allMemberName="All"
allMemberCaption="Todas las localidades" primaryKey="idLocalidad" caption="Localidad residencia"
description="Localidad residencia">
  <Table name="d_localidades">
  </Table>
  <Level name="Level_Montevideo_Interior" visible="true" table="d_localidades" column="idMontInt"
nameColumn="descMontInt" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Montevideo / Interior" description="Montevideo / Interior">
  </Level>
  <Level name="Level_Departamento" visible="true" table="d_localidades" column="idDepartamento"
nameColumn="descDepartamento" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Departamento residencia" description="Departamento residencia">
  </Level>
  <Level name="Level_Localidad" visible="true" table="d_localidades" column="idLocalidad" nameColumn=
"descLocalidad" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Localidad residencia" description="Localidad residencia">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" highCardinality="false" name="D_Sexo" caption="Sexo"
description="Sexo">
  <Hierarchy name="Sexo" visible="true" hasAll="true" allMemberName="All" primaryKey="idSexo" caption=
"Sexo" description="Sexo">
  <Table name="d_sexo">
  </Table>
  <Level name="Level_Sexo" visible="true" table="d_sexo" column="idSexo" nameColumn="descSexo" type=
"Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Sexo" description=
"Sexo">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" highCardinality="false" name="D_Edades" caption=
Edades" description="Edades">
  <Hierarchy name="GrupoEtario" visible="true" hasAll="true" allMemberName="All" primaryKey="idEdad"
caption="Grupo etario" description="Grupo etario">
  <Table name="d_edades">
  </Table>
  <Level name="Level_GrupoEtario" visible="true" table="d_edades" column="idGrupoEtario" nameColumn=
"descGrupoEtario" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Grupo etario" description="Grupo etario">
  </Level>
  <Level name="Level_Edad" visible="true" table="d_edades" column="idEdad" nameColumn="descEdad" type=
"Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Edad"
description="Edad">
  </Level>
</Hierarchy>
<Hierarchy name="GrupoEtarioMaterno" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idEdad" caption="Grupo etario materno" description="Grupo etario materno">
  <Table name="d_edades">
  </Table>
  <Level name="Level_GrupoEtarioMaterno" visible="true" table="d_edades" column=
"idGrupoEtarioMaterno" nameColumn="descGrupoEtarioMaterno" type="Numeric" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never" caption="Grupo etario materno" description="Grupo etario
materno">
  </Level>
  <Level name="Level_Edad" visible="true" table="d_edades" column="idEdad" nameColumn="descEdad" type=
"Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Edad"
description="Edad">
  </Level>
</Hierarchy>
</Dimension>

```

```

<Cube name="Nacimientos" caption="Nacimientos" visible="true" description="Nacimientos" cache="true"
enabled="true">
  <Table name="ft_nacimientos">
  </Table>
  <Dimension type="StandardDimension" visible="true" foreignKey="idMadre" highCardinality="false" name="
D_Madres" caption="Madres" description="Madres">
  <Hierarchy name="NivelEducativo" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idMadre" caption="Nivel educativo" description="Nivel educativo de la madre">
    <Table name="d_madres">
    </Table>
    <Level name="Level_NivelEducativo" visible="true" table="d_madres" column="idNivelEducativo"
nameColumn="descNivelEducativo" type="Numeric" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never" caption="Nivel educativo" description="Nivel educativo">
    </Level>
    <Level name="Level_Madre" visible="true" table="d_madres" column="idMadre" nameColumn=
"nombreMadre" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Madre" description="Madre">
    </Level>
  </Hierarchy>
  <Hierarchy name="EstadoConyugal" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idMadre" caption="Estado conyugal" description="Estado conyugal de la madre">
    <Table name="d_madres">
    </Table>
    <Level name="Level_EstadoConyugal" visible="true" table="d_madres" column="idEstadoConyugal"
nameColumn="descEstadoConyugal" type="Numeric" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never" caption="Estado conyugal" description="Estado conyugal de la madre">
    </Level>
    <Level name="Level_Madre" visible="true" table="d_madres" column="idMadre" nameColumn=
"nombreMadre" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Madre" description="Madre">
    </Level>
  </Hierarchy>
  <Hierarchy name="ConsultasPrenatales" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idMadre" caption="Consultas prenatales" description="Cantidad de consultas prenatales">
    <Table name="d_madres">
    </Table>
    <Level name="Level_RangoConsultasPrenatales" visible="true" table="d_madres" column=
"idRangoConsultasPrenatales" nameColumn="descRangoConsultasPrenatales" type="Numeric"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Rango consultas
prenatales" description="Rango de consultas prenatales">
    </Level>
    <Level name="Level_ConsultasPrenatales" visible="true" table="d_madres" column=
"cantConsultasPrenatales" nameColumn="cantConsultasPrenatales" type="Numeric" uniqueMembers=
"false" levelType="Regular" hideMemberIf="Never" caption="Consultas prenatales" description=
"Consultas prenatales">
    </Level>
    <Level name="Level_Madre" visible="true" table="d_madres" column="idMadre" nameColumn=
"nombreMadre" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Madre" description="Madre">
    </Level>
  </Hierarchy>
  <Hierarchy name="LugarParto" visible="true" hasAll="true" allMemberName="All" primaryKey="idMadre"
caption="Lugar parto" description="Lugar de parto">
    <Table name="d_madres">
    </Table>
    <Level name="Level_LugarParto" visible="true" table="d_madres" column="idLugarParto" nameColumn=
"descLugarParto" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Lugar parto" description="Lugar de parto">
    </Level>
    <Level name="Level_Madre" visible="true" table="d_madres" column="idMadre" nameColumn=
"nombreMadre" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Madre" description="Madre">
    </Level>
  </Hierarchy>
  <Hierarchy name="AsistenciaParto" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idMadre" caption="Asistencia parto" description="Asistencia parto">
    <Table name="d_madres">
    </Table>
    <Level name="Level_AsistenciaParto" visible="true" table="d_madres" column="idAsistenciaParto"
nameColumn="descAsistenciaParto" type="Numeric" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never" caption="Asistencia parto" description="Asistencia parto">
    </Level>
    <Level name="Level_Madre" visible="true" table="d_madres" column="idMadre" nameColumn=
"nombreMadre" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Madre" description="Madre">
    </Level>
  </Hierarchy>

```

```

<Hierarchy name="TipoParto" visible="true" hasAll="true" allMemberName="All" primaryKey="idMadre"
caption="Tipo parto" description="Tipo de parto">
  <Table name="d_madres">
  </Table>
  <Level name="Level_TipoParto" visible="true" table="d_madres" column="idTipoParto" nameColumn=
"descTipoParto" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Tipo parto" description="Tipo de parto">
  </Level>
  <Level name="Level_Madre" visible="true" table="d_madres" column="idMadre" nameColumn=
"nombreMadre" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Madre" description="Madre">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="cedulaNacido" highCardinality="false"
name="D_NacidosVivos" caption="Nacidos vivos" description="Nacidos vivos">
  <Hierarchy name="Peso" visible="true" hasAll="true" allMemberName="All" primaryKey="cedulaNacido"
caption="Peso" description="Peso">
  <Table name="d_nacidos_vivos">
  </Table>
  <Level name="Level_CategoriaPeso" visible="true" table="d_nacidos_vivos" column="idCategoriaPeso"
nameColumn="descCategoriaPeso" type="Numeric" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never" caption="Categoria peso" description="Categoria peso">
  </Level>
  <Level name="Level_RangoPeso" visible="true" table="d_nacidos_vivos" column="idRangoPeso"
nameColumn="descRangoPeso" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Rango peso" description="Rango peso">
  </Level>
  <Level name="Level_Peso" visible="true" table="d_nacidos_vivos" column="peso" nameColumn="peso"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption="Peso"
description="Peso">
  </Level>
  <Level name="Level_Nacido" visible="true" table="d_nacidos_vivos" column="cedulaNacido"
nameColumn="nombreNacido" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Nacido vivo" description="Nacido vivo">
  </Level>
</Hierarchy>
<Hierarchy name="EdadGestacional" visible="true" hasAll="true" allMemberName="All" primaryKey=
"cedulaNacido" caption="Edad gestacional" description="Edad gestacional">
  <Table name="d_nacidos_vivos">
  </Table>
  <Level name="Level_GrupoEtarioGestacion" visible="true" table="d_nacidos_vivos" column=
"idGrupoEtarioGestacion" nameColumn="descGrupoEtarioGestacion" type="Numeric" uniqueMembers=
"false" levelType="Regular" hideMemberIf="Never" caption="Grupo etario" description="Grupo
etario">
  </Level>
  <Level name="Level_EdadGestacional" visible="true" table="d_nacidos_vivos" column=
"semanaGestacion" nameColumn="semanaGestacion" type="Numeric" uniqueMembers="false" levelType=
"Regular" hideMemberIf="Never" caption="Edad gestacional" description="Edad gestacional">
  </Level>
  <Level name="Level_Nacido" visible="true" table="d_nacidos_vivos" column="cedulaNacido"
nameColumn="nombreNacido" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Nacido vivo" description="Nacido vivo">
  </Level>
</Hierarchy>
</Dimension>
<DimensionUsage source="D_Tiempo" name="D_Tiempo" caption="Tiempo" visible="true" foreignKey=
"idFecha" highCardinality="false">
</DimensionUsage>
<Dimension type="StandardDimension" visible="true" foreignKey="idEstablecimiento" highCardinality=
"false" name="D_Establecimientos" caption="Establecimientos" description="Establecimientos">
  <Hierarchy name="Establecimientos" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idEstablecimiento" caption="Establecimientos" description="Establecimientos">
  <Table name="d_establecimientos">
  </Table>
  <Level name="Level_Institucion" visible="true" table="d_establecimientos" column="idInstitucion"
nameColumn="descInstitucion" type="Numeric" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never" caption="Institucion" description="Institucion">
  </Level>
  <Level name="Level_Establecimiento" visible="true" table="d_establecimientos" column=
"idEstablecimiento" nameColumn="descEstablecimiento" type="Numeric" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never" caption="Establecimiento" description="Establecimiento">
  </Level>
</Hierarchy>
</Dimension>

```

```

<DimensionUsage source="D_LocalidadOcurrencia" name="D_LocalidadOcurrencia" caption="Localidad
ocurrencia" visible="true" foreignKey="idLocalidadOcurrencia" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_DomicilioResidencia" name="D_DomicilioResidencia" caption="Domicilio
residencia materna" visible="true" foreignKey="idDomicilioResidenciaMadre" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_LocalidadResidencia" name="D_LocalidadResidencia" caption="Localidad
residencia materna" visible="true" foreignKey="idLocalidadResidenciaMadre" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_Sexo" name="D_Sexo" caption="Sexo" visible="true" foreignKey="idSexo"
highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_Edades" name="D_Edades" caption="Edad materna" visible="true" foreignKey=
"edadMadre" highCardinality="false">
</DimensionUsage>
<Measure name="Cant_Nacimientos" column="cedulaNacido" aggregator="count" caption="Cantidad de
nacimientos" description="Cantidad de nacimientos" visible="true">
</Measure>
<CalculatedMember name="Proporcion_Peso" caption="Proporcion por peso" formula=
"(([Measures].Cant_Nacimientos)/([Measures].Cant_Nacimientos,[D_NacidosVivos.Peso].[All]))*100"
dimension="Measures" visible="true">
</CalculatedMember>
</Cube>
<Cube name="Mortalidad_Infantil" caption="Mortalidad infantil" visible="false" description="Mortalidad
infantil" cache="true" enabled="true">
<Table name="ft_mortalidad_infantil">
</Table>
<Dimension type="StandardDimension" visible="true" foreignKey="idFallecidoMenor" highCardinality=
"false" name="D_FallecidosMenores" caption="Fallecidos menores" description="Fallecidos menores">
<Hierarchy name="FallecidosMenores" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idFallecidoMenor" caption="Fallecidos menores" description="FallecidosMenores">
<Table name="d_fallecidos_menores">
</Table>
<Level name="Level_GrupoEtarioMenor" visible="true" table="d_fallecidos_menores" column=
"idGrupoEtarioMenor" nameColumn="descGrupoEtarioMenor" type="Numeric" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never" caption="Grupo etario menor" description="Grupo etario
menor">
</Level>
<Level name="Level_EdadMenor" visible="true" table="d_fallecidos_menores" column="edadMenor"
nameColumn="edadMenor" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Edad menor" description="Edad menor">
</Level>
<Level name="Level_FallecidoMenor" visible="true" column="idFallecidoMenor" nameColumn=
"nombreFallecidoMenor" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Fallecido menor" description="Fallecido">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="idCausa" highCardinality="false" name=
"D_CausasCIE10" caption="Causas CIE10" description="Causas CIE10">
<Hierarchy name="CausasCIE10" visible="true" hasAll="true" allMemberName="All" primaryKey="idCausa"
caption="Causas CIE10" description="Causas CIE10">
<Table name="d_causascie10">
</Table>
<Level name="Level_Capitulo" visible="true" table="d_causascie10" column="idCapitulo" nameColumn=
"descCapitulo" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Capitulo" description="Capitulo">
</Level>
<Level name="Level_Grupo" visible="true" table="d_causascie10" column="idGrupo" nameColumn=
"descGrupo" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption=
"Grupo" description="Grupo">
</Level>
<Level name="Level_Categoria" visible="true" table="d_causascie10" column="idCategoria"
nameColumn="descCategoria" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Categoria" description="Categoria">
</Level>
<Level name="Level_Causa" visible="true" table="d_causascie10" column="idCausa" nameColumn=
"descCausa" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption=
"Causa CIE10" description="Causa CIE10">
</Level>
</Hierarchy>
</Dimension>

```

```
<DimensionUsage source="D_Tiempo" name="D_Tiempo" caption="Tiempo" visible="true" foreignKey=
"idFecha" highCardinality="false">
</DimensionUsage>
<Dimension type="StandardDimension" visible="true" foreignKey="idDomicilioOcurrencia" highCardinality
="false" name="D_DomicilioOcurrencia" caption="Domicilio ocurrencia" description="Domicilio
ocurrencia">
  <Hierarchy name="Pais" visible="true" hasAll="true" allMemberName="All" primaryKey="idDomicilio"
caption="Pais ocurrencia" description="Pais ocurrencia">
    <Table name="d_domicilios">
    </Table>
    <Level name="Level_Pais" visible="true" table="d_domicilios" column="idPais" nameColumn=
"descPais" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption=
"Pais ocurrencia" description="Pais ocurrencia">
    </Level>
    <Level name="Level_Domicilio" visible="true" table="d_domicilios" column="idDomicilio" nameColumn
="descDomicilio" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Domicilio ocurrencia" description="Domicilio ocurrencia">
    </Level>
  </Hierarchy>
  <Hierarchy name="TipoZona" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idDomicilio" caption="Tipo zona ocurrencia" description="Tipo zona ocurrencia">
    <Table name="d_domicilios">
    </Table>
    <Level name="Level_TipoZona" visible="true" table="d_domicilios" column="idTipoZona" nameColumn=
"descTipoZona" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Tipo zona ocurrencia" description="Tipo zona ocurrencia">
    </Level>
    <Level name="Level_Domicilio" visible="true" table="d_domicilios" column="idDomicilio" nameColumn
="descDomicilio" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Domicilio ocurrencia" description="Domicilio ocurrencia">
    </Level>
  </Hierarchy>
</Dimension>
<DimensionUsage source="D_LocalidadOcurrencia" name="D_LocalidadOcurrencia" caption="Localidad
ocurrencia" visible="true" foreignKey="idLocalidadOcurrencia" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_DomicilioResidencia" name="D_DomicilioResidencia" caption="Domicilio
residencia materna" visible="true" foreignKey="idDomicilioResidenciaMadre" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_LocalidadResidencia" name="D_LocalidadResidencia" caption="Localidad
residencia materna" visible="true" foreignKey="idLocalidadResidenciaMadre" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_Sexo" name="D_Sexo" caption="Sexo" visible="true" foreignKey="idSexo"
highCardinality="false">
</DimensionUsage>
<Measure name="Cant_Fallecimientos_Menores" column="idFallecidoMenor" aggregator="count" caption=
"Cantidad fallecimientos menores" description="Cantidad fallecimientos menores" visible="true">
</Measure>
</Cube>
<Cube name="Mortalidad" caption="Mortalidad" visible="false" description="Mortalidad" cache="true"
enabled="true">
  <Table name="ft_mortalidad">
  </Table>
  <Dimension type="StandardDimension" visible="true" foreignKey="idFallecido" highCardinality="false"
name="D_Fallecidos" caption="Fallecidos" description="Fallecidos">
    <Hierarchy name="Fallecidos" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idFallecido" caption="Fallecidos" description="Fallecidos">
      <Table name="d_fallecidos">
      </Table>
      <Level name="Level_Embarazada" visible="true" table="d_fallecidos" column="embarazada" nameColumn
="embarazada" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Embarazada" description="Embarazada">
      </Level>
      <Level name="Level_Fallecido" visible="true" table="d_fallecidos" column="idFallecido" nameColumn
="nombreFallecido" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Fallecido" description="Fallecido">
      </Level>
    </Hierarchy>
  </Dimension>
```

```

<Dimension type="StandardDimension" visible="true" foreignKey="idCausa" highCardinality="false" name="
D_CausasCIE10" caption="Causas CIE10" description="Causas CIE10">
  <Hierarchy name="CausasCIE10" visible="true" hasAll="true" allMemberName="All" primaryKey="idCausa"
caption="Causas CIE10" description="Causas CIE10">
    <Table name="d_causascie10">
    </Table>
    <Level name="Level_Capitulo" visible="true" table="d_causascie10" column="idCapitulo" nameColumn=
"descCapitulo" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Capitulo" description="Capitulo">
    </Level>
    <Level name="Level_Grupo" visible="true" table="d_causascie10" column="idGrupo" nameColumn=
"descGrupo" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption=
"Grupo" description="Grupo">
    </Level>
    <Level name="Level_Categoria" visible="true" table="d_causascie10" column="idCategoria"
nameColumn="descCategoria" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf=
"Never" caption="Categoria" description="Categoria">
    </Level>
    <Level name="Level_Causa" visible="true" table="d_causascie10" column="idCausa" nameColumn=
"descCausa" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption=
"Causa CIE10" description="Causa CIE10">
    </Level>
  </Hierarchy>
</Dimension>
<DimensionUsage source="D_Tiempo" name="D_Tiempo" caption="Tiempo" visible="true" foreignKey=
"idFecha" highCardinality="false">
</DimensionUsage>
<Dimension type="StandardDimension" visible="true" foreignKey="idDomicilioOcurrencia" highCardinality
="false" name="D_DomicilioOcurrencia" caption="Domicilio ocurrencia" description="Domicilio
ocurrencia">
  <Hierarchy name="Pais" visible="true" hasAll="true" allMemberName="All" primaryKey="idDomicilio"
caption="Pais residencia" description="Pais residencia">
    <Table name="d_domicilios">
    </Table>
    <Level name="Level_Pais" visible="true" table="d_domicilios" column="idPais" nameColumn=
"descPais" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never" caption=
"Pais residencia" description="Pais residencia">
    </Level>
    <Level name="Level_Domicilio" visible="true" table="d_domicilios" column="idDomicilio" nameColumn
="descDomicilio" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Domicilio residencia" description="Domicilio residencia">
    </Level>
  </Hierarchy>
  <Hierarchy name="TipoZona" visible="true" hasAll="true" allMemberName="All" primaryKey=
"idDomicilio" caption="Tipo zona residencia" description="Tipo zona residencia">
    <Table name="d_domicilios">
    </Table>
    <Level name="Level_TipoZona" visible="true" table="d_domicilios" column="idTipoZona" nameColumn=
"descTipoZona" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Tipo zona residencia" description="Tipo zona residencia">
    </Level>
    <Level name="Level_Domicilio" visible="true" table="d_domicilios" column="idDomicilio" nameColumn
="descDomicilio" type="Numeric" uniqueMembers="false" levelType="Regular" hideMemberIf="Never"
caption="Domicilio residencia" description="Domicilio residencia">
    </Level>
  </Hierarchy>
</Dimension>
<DimensionUsage source="D_LocalidadOcurrencia" name="D_LocalidadOcurrencia" caption="Localidad
ocurrencia" visible="true" foreignKey="idLocalidadOcurrencia" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_DomicilioResidencia" name="D_DomicilioResidencia" caption="Domicilio
residencia" visible="true" foreignKey="idDomicilioResidencia" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_LocalidadResidencia" name="D_LocalidadResidencia" caption="Localidad
residencia" visible="true" foreignKey="idLocalidadResidencia" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_Sexo" name="D_Sexo" caption="Sexo" visible="true" foreignKey="idSexo"
highCardinality="false">
</DimensionUsage>
<DimensionUsage source="D_Edades" name="D_Edades" caption="Edades" visible="true" foreignKey="idEdad"
highCardinality="false">
</DimensionUsage>
<Measure name="Cant_Fallecimientos" column="idFallecido" aggregator="count" caption="Cantidad
fallecimientos" description="Cantidad fallecimientos" visible="true">
</Measure>

```

```

<VirtualCube enabled="true" name="Mortalidad_Infantil_Virtual" caption="Mortalidad infantil" visible=
"true" description="Cubo virtual de mortalidad infantil">
  <CubeUsages>
    <CubeUsage cubeName="Nacimientos" ignoreUnrelatedDimensions="true">
    </CubeUsage>
    <CubeUsage cubeName="Mortalidad_Infantil" ignoreUnrelatedDimensions="false">
    </CubeUsage>
  </CubeUsages>
  <VirtualCubeDimension caption="Tiempo" visible="true" highCardinality="false" name="D_Tiempo">
  </VirtualCubeDimension>
  <VirtualCubeDimension cubeName="Mortalidad_Infantil" caption="Fallecidos menores" visible="true"
highCardinality="false" name="D_FallecidosMenores">
  </VirtualCubeDimension>
  <VirtualCubeDimension cubeName="Mortalidad_Infantil" caption="Causas CIE10" visible="true"
highCardinality="false" name="D_CausasCIE10">
  </VirtualCubeDimension>
  <VirtualCubeDimension cubeName="Mortalidad_Infantil" caption="Domicilio ocurrencia" visible="true"
highCardinality="false" name="D_DomicilioOcurrencia">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Localidad ocurrencia" visible="true" highCardinality="false" name=
"D_LocalidadOcurrencia">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Domicilio residencia materna" visible="true" highCardinality="false"
name="D_DomicilioResidencia">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Localidad residencia materna" visible="true" highCardinality="false"
name="D_LocalidadResidencia">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Sexo" visible="true" highCardinality="false" name="D_Sexo">
  </VirtualCubeDimension>
  <VirtualCubeMeasure cubeName="Nacimientos" name="[Measures].[Cant_Nacimientos]" visible="false">
  </VirtualCubeMeasure>
  <VirtualCubeMeasure cubeName="Mortalidad_Infantil" name="[Measures].[Cant_Fallecimientos_Menores]"
visible="true">
  </VirtualCubeMeasure>
  <CalculatedMember name="Tasa_Mortalidad_Infantil" caption="Tasa mortalidad infantil" formula=
"((([Measures].[Cant_Fallecimientos_Menores])/([Measures].[Cant_Nacimientos],[D_FallecidosMenores.Falle
cidosMenores].[All],[D_CausasCIE10.CausasCIE10].[All],[D_DomicilioOcurrencia.Pais].[All],[D_DomicilioO
currencia.TipoZona].[All]))*1000" dimension="Measures" visible="true">
  </CalculatedMember>
</VirtualCube>
<VirtualCube enabled="true" name="Mortalidad_Virtual" caption="Mortalidad" visible="true" description=
"Cubo virtual de mortalidad">
  <VirtualCubeDimension cubeName="Mortalidad" caption="Fallecidos" visible="true" highCardinality=
"false" name="D_Fallecidos">
  </VirtualCubeDimension>
  <VirtualCubeDimension cubeName="Mortalidad" caption="Causas CIE10" visible="true" highCardinality=
"false" name="D_CausasCIE10">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Tiempo" visible="true" highCardinality="false" name="D_Tiempo">
  </VirtualCubeDimension>
  <VirtualCubeDimension cubeName="Mortalidad" caption="Domicilio ocurrencia" visible="true"
highCardinality="false" name="D_DomicilioOcurrencia">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Localidad ocurrencia" visible="true" highCardinality="false" name=
"D_LocalidadOcurrencia">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Domicilio residencia" visible="true" highCardinality="false" name=
"D_DomicilioResidencia">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Localidad residencia" visible="true" highCardinality="false" name=
"D_LocalidadResidencia">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Sexo" visible="true" highCardinality="false" name="D_Sexo">
  </VirtualCubeDimension>
  <VirtualCubeDimension caption="Edades" visible="true" highCardinality="false" name="D_Edades">
  </VirtualCubeDimension>
  <VirtualCubeMeasure cubeName="Nacimientos" name="[Measures].[Cant_Nacimientos]" visible="false">
  </VirtualCubeMeasure>
  <VirtualCubeMeasure cubeName="Mortalidad" name="[Measures].[Cant_Fallecimientos]" visible="true">
  </VirtualCubeMeasure>
  <CalculatedMember name="Razon_Mortalidad_Materna" caption="Razon mortalidad materna" formula=
"((([Measures].[Cant_Fallecimientos],[D_Fallecidos],[Level_Embarazada].[Si])/([Measures].[Cant_Nacimien
tos],[D_Fallecidos.Fallecidos].[All],[D_DomicilioOcurrencia.Pais].[All],[D_DomicilioOcurrencia.TipoZon
a].[All],[D_CausasCIE10.CausasCIE10].[All]))*100000" dimension="Measures" visible="true">
  </CalculatedMember>
</VirtualCube>
</Schema>

```

## Anexo V – Manual de usuario

En el presente anexo se encuentra la documentación detallada de la interfaz de usuario implementada. La misma incluye requerimientos de software para correr la aplicación, instrucciones de instalación y explicaciones paso a paso sobre el uso de las funcionalidades provistas.

### Instalación

Esta sección constituye una guía de instalación de la aplicación.

#### *Software requerido*

A continuación se presentan los requisitos mínimos de software para un correcto funcionamiento de la aplicación.

##### *JDK7+*

Se debe contar con una versión mayor o igual a 7 del software Java Development Kit (JDK) instalada en el sistema. Dicho software puede ser descargado de la página de descargas de Oracle SE [30]. En caso de existir una versión previamente instalada, se puede verificar que sea mayor o igual a la versión requerida corriendo el comando `java -version` en la línea de comandos.

##### *Apache Tomcat*

Es necesario disponer de alguna versión del servidor Apache Tomcat instalada, ya que el mismo será utilizado para el despliegue de la aplicación web. Si no se cuenta con dicho software, se recomienda descargar la última versión estable, que puede ser obtenida accediendo a la sección de descargas de la página de Apache Tomcat [28].

#### *Instrucciones*

Los pasos necesarios para poder acceder a la aplicación, luego de contar con el software requerido, son los siguientes:

1. Copiar el archivo *interfaz\_msp.war*, que se encuentra dentro de la carpeta *Instalación*, en el directorio `TOMCAT_HOME/webapps`.
2. Iniciar Tomcat.

3. Editar el archivo *config.properties* que se encuentra en el directorio TOMCAT\_HOME\webapps\interfaz\_msp\WEB-INF\classes especificando las siguientes propiedades:

- **Host:** nombre del host donde corren las bases de datos fuente y destino.
- **Port:** puerto de conexión de las bases de datos fuente y destino. El puerto por defecto de MySQL es 3306, si este no ha sido modificado, dicho valor debe ser el establecido.
- **User:** usuario con permisos de administrador sobre las bases de datos fuente y destino.
- **Password:** contraseña del usuario especificado.
- **esquemaFuentes:** nombre del esquema de la base de datos fuente.
- **MondrianSchema:** ruta del archivo *Cubos\_MSP.xml*. En caso de no haber modificado su ubicación, el mismo se encuentra dentro de la carpeta *Instalación*.
- **Job:** ruta del archivo *DW\_MSP.kjb*. Es importante mencionar que los archivos de las transformaciones también deben encontrarse en dicho directorio. En caso de no haber modificado su ubicación, los mismos se encuentran dentro de la carpeta *Instalación*.

4. Reiniciar Tomcat.

A continuación es posible acceder a la aplicación ingresando en un navegador, preferentemente Google Chrome, a la dirección [http://localhost:8080/interfaz\\_msp/](http://localhost:8080/interfaz_msp/), siendo localhost y 8080, el host y el puerto por defecto de Tomcat.

Cabe destacar que al cambiar las propiedades de las bases de datos fuente y destino y/o la ubicación o nombre de cualquiera de los archivos mencionados, estos deben modificarse en el archivo de configuración para garantizar el correcto funcionamiento de la aplicación.

## Documentación

A lo largo de esta sección se presentan las funcionalidades provistas por la herramienta y se brindan instrucciones de cómo utilizarlas. Dichas funcionalidades se encuentran disponibles dentro de las tres pantallas principales de la aplicación, Inicio, Consultas y Carga.

Es posible acceder a dichas pantallas mediante la barra de navegación que se encuentra en la sección superior de la aplicación. Dependiendo de la pantalla a la que se desea acceder, el link de la barra de navegación que se debe seleccionar.

Figura V.1 – Barra de navegación

La pantalla de Inicio es cargada por defecto al ingresar a la aplicación. El diseño de la misma se asemeja al de una consola de administrador permitiendo visualizar mediante tablas y gráficas los valores de los indicadores durante los últimos años. Además permite generar planillas Excel, muy similares a las requeridas por la OMS, conteniendo los valores de los indicadores para un año determinado.

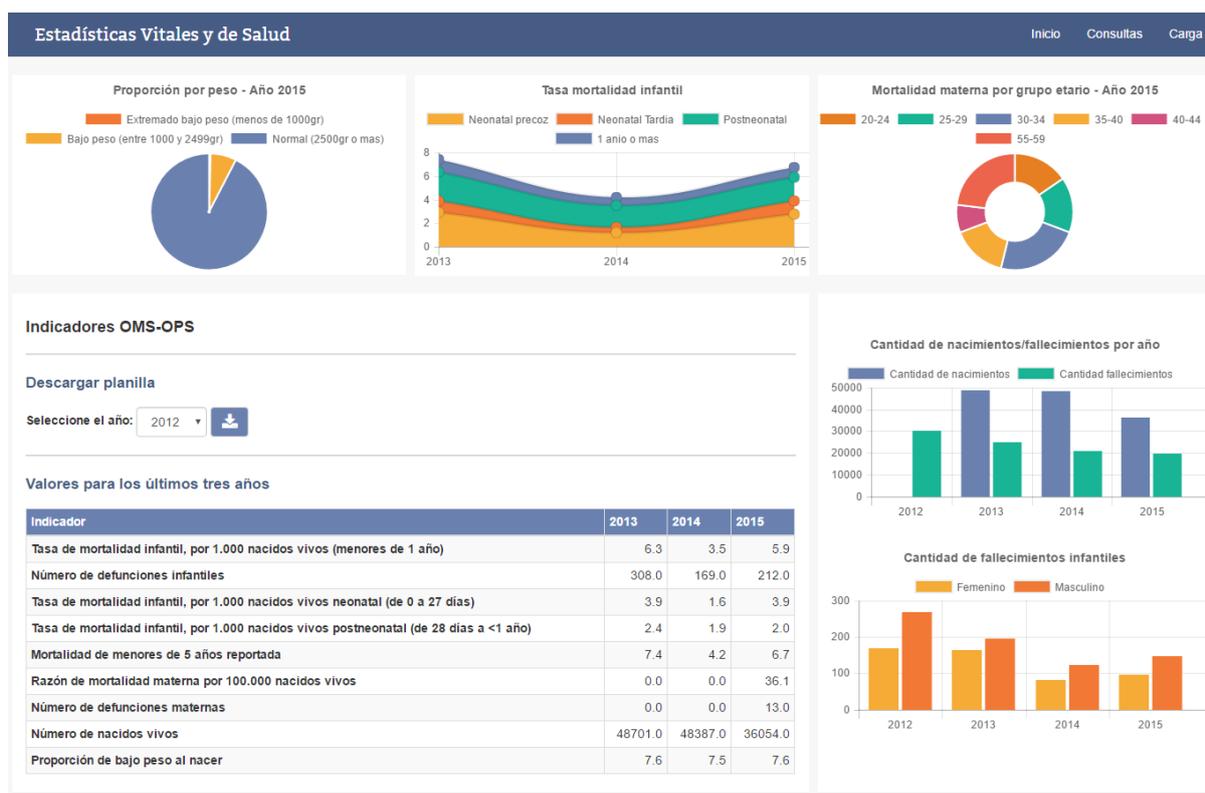


Figura V.2 – Pantalla de Inicio

La pantalla de Consultas dispone de una interfaz para realizar consultas personalizadas sobre el Data Warehouse, mostrando los resultados en forma de tabla. En dichas consultas es posible seleccionar las medidas a analizar, los criterios a incluir en las filas y las columnas de la tabla, y los filtros a aplicar. Adicionalmente, la tabla generada puede ser exportada en formato Excel y PDF.

The screenshot shows the 'Estadísticas Vitales y de Salud' application interface. At the top, there are navigation links for 'Inicio', 'Consultas', and 'Carga'. The main area is divided into several sections:

- Measures:** A panel containing 'Cantidad de nacimientos'.
- Columns:** A panel containing 'Montevideo / Interior'.
- Rows:** A panel containing 'Año'.
- Filter:** A section with the instruction 'Arrastre un criterio de análisis' and a 'Consultar' button.
- Data Table:** A table showing birth statistics by year and locality.

Tiempo	Localidad ocurrencia		
	Montevideo	Interior	No Indicado
	<b>Cantidad de nacimientos</b>	<b>Cantidad de nacimientos</b>	<b>Cantidad de nacimientos</b>
2013	25.877	22.546	278
2014	25.976	22.249	162
2015	18.969	17.081	4

Figura V.3 – Pantalla de Consultas

Por último, la pantalla de Carga permite realizar tanto la carga inicial del Data Warehouse como las sucesivas actualizaciones. Además dispone de una opción para realizar cargas o actualizaciones personalizadas, seleccionando las tablas con las que se desea trabajar.

The screenshot shows the 'Estadísticas Vitales y de Salud' application interface for the 'Carga' (Load) screen. It includes the following elements:

- Opciones de carga:** Radio buttons for 'Inicial', 'Actualización', and 'Personalizada'.
- Selección de tablas:** A grid of checkboxes for selecting data tables:
  - Nacimientos
  - Mortalidad
  - Mortalidad Infantil
  - Tiempo
  - Edades
  - Domicilios y Localidades
  - Sexo
  - Causas CIE-10
  - Establecimientos
- Fecha de inicio y de fin:** Two input fields for 'Fecha de inicio' and 'Fecha de fin' with a placeholder 'dd/mm/aaaa'.
- Cargar:** A button to execute the load operation.

Figura V.4 – Pantalla de Carga

## Funcionalidades

Para utilizar correctamente todas las funcionalidades provistas por la herramienta, la base de datos del Data Warehouse, cuyas propiedades se especificaron en el archivo de configuración al realizar la instalación, debe encontrarse levantada. En caso contrario, se mostrará una advertencia al acceder a la aplicación, tal como se muestra en la Figura V.5. Si la base se encuentra levantada y se sigue mostrando dicho error, se debe verificar que las propiedades especificadas sean las correctas y, en caso de que no lo sean, modificar el archivo de configuración, reiniciar Tomcat y volver a ingresar a la aplicación.



Figura V.5 – Error de conexión

## Visualizar indicadores mediante gráficas y tablas

En la pantalla de Inicio se encuentran disponibles una serie de gráficas y tablas que comparan y analizan algunos de los indicadores requeridos por la OMS según distintos criterios. Con el objetivo de brindar un panorama general de la situación actual, las gráficas presentadas son las siguientes:

- Gráfica de torta que muestra la proporción por peso al nacer analizada por categoría peso para el último año del que se disponen datos.



Figura V.6 – Gráfica Proporción por peso

- Gráfica de líneas que analiza la tasa de mortalidad infantil por grupo etario para los últimos tres años.

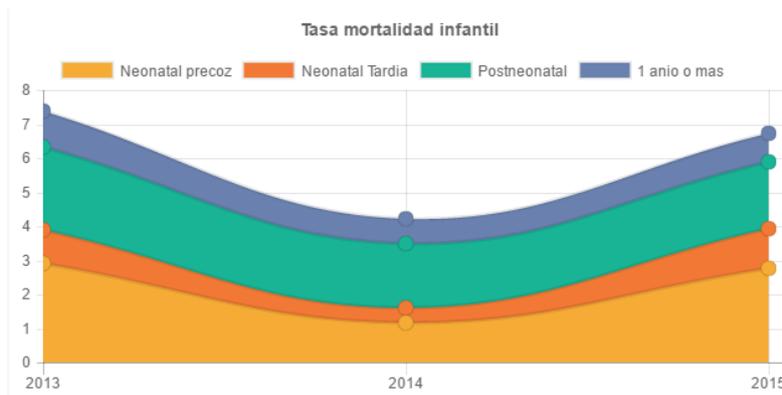


Figura V.7 – Gráfica Tasa mortalidad infantil

- Gráfica de anillos que presenta las muertes maternas analizadas por grupo etario para el último año del que se disponen datos.

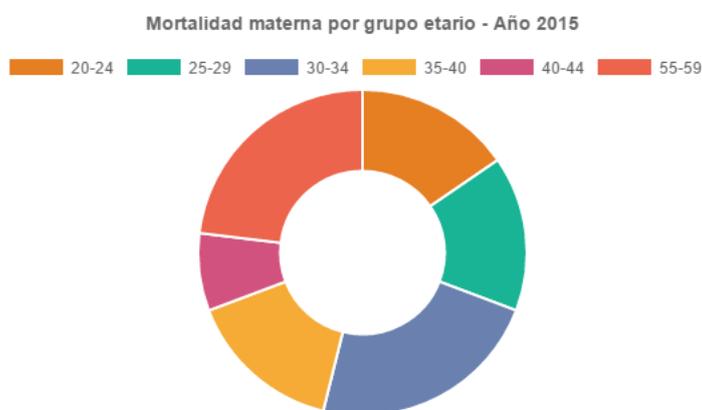


Figura V.8 – Gráfica Mortalidad materna

- Gráfica de barras que compara la cantidad de nacimientos con la cantidad de fallecimientos para los últimos cuatro años.

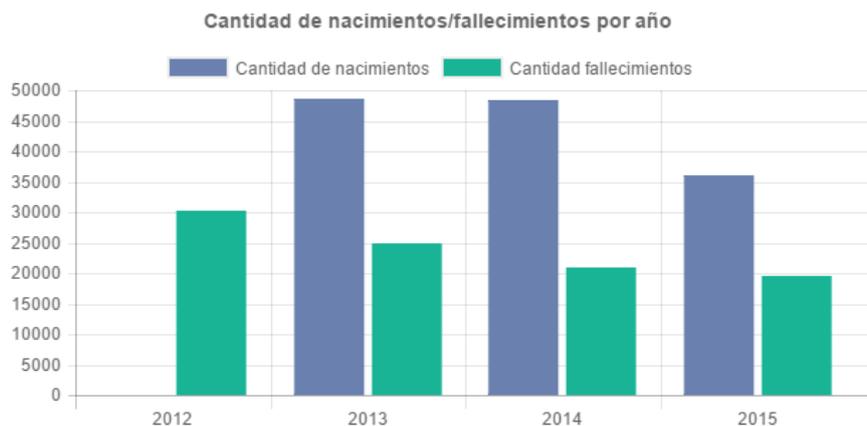


Figura V.9 – Gráfica comparativa nacimientos/fallecimientos

- Gráfica de barras que analiza la cantidad de fallecimientos infantiles por sexo para los últimos cuatro años.

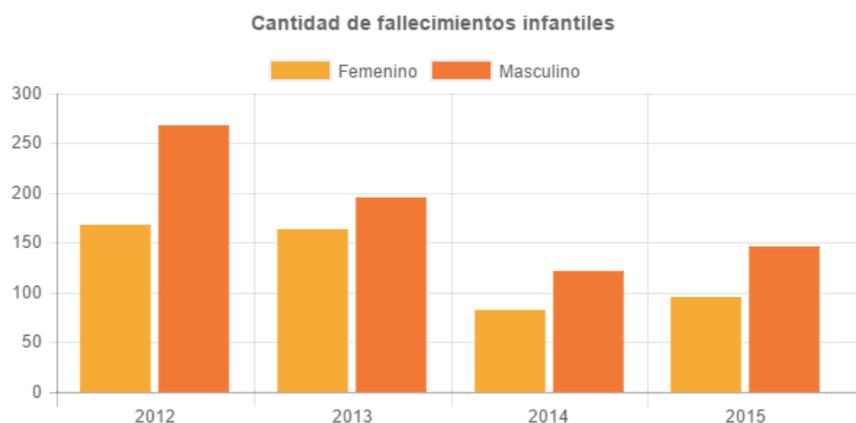


Figura V.10 – Gráfica fallecimientos infantiles

Además, en la sección inferior izquierda de la pantalla, se encuentra disponible una tabla comparativa de los indicadores requeridos por la OMS para los últimos tres años.

#### Valores para los últimos tres años

Indicador	2013	2014	2015
Tasa de mortalidad infantil, por 1.000 nacidos vivos (menores de 1 año)	6.3	3.5	5.9
Número de defunciones infantiles	308.0	169.0	212.0
Tasa de mortalidad infantil, por 1.000 nacidos vivos neonatal (de 0 a 27 días)	3.9	1.6	3.9
Tasa de mortalidad infantil, por 1.000 nacidos vivos postneonatal (de 28 días a <1 año)	2.4	1.9	2.0
Mortalidad de menores de 5 años reportada	7.4	4.2	6.7
Razón de mortalidad materna por 100.000 nacidos vivos	0.0	0.0	36.1
Número de defunciones maternas	0.0	0.0	13.0
Número de nacidos vivos	48701.0	48387.0	36054.0
Proporción de bajo peso al nacer	7.6	7.5	7.6

Figura V.11 – Tabla comparativa indicadores OMS

## Descargar planilla con indicadores OMS para determinado año

1. Acceder a la pantalla de Inicio
2. En la sección inferior izquierda de la pantalla, titulada *INDICADORES OMS-OPS*, se encuentra un sector de descarga de planillas. Allí se muestra una lista desplegable que permite seleccionar el año para el cual se desea generar la planilla.

### Indicadores OMS-OPS

#### Descargar planilla

Seleccione el año:  

Valores para los  años

Indicador	2013	2014	2015
Tasa de mortalidad infantil, por 1.000 nacidos vivos (menores de 1 año)	6.3	3.5	5.9
Número de defunciones infantiles	308.0	169.0	212.0
Tasa de mortalidad infantil, por 1.000 nacidos vivos neonatal (de 0 a 27 días)	3.9	1.6	3.9
Tasa de mortalidad infantil, por 1.000 nacidos vivos postneonatal (de 28 días a <1 año)	2.4	1.9	2.0
Mortalidad de menores de 5 años reportada	7.4	4.2	6.7
Razón de mortalidad materna por 100.000 nacidos vivos	0.0	0.0	36.1
Número de defunciones maternas	0.0	0.0	13.0
Número de nacidos vivos	48701.0	48387.0	36054.0
Proporción de bajo peso al nacer	7.6	7.5	7.6

Figura V.12 – Descarga planilla indicadores OMS – Selección año

3. Luego de elegido el año, seleccionar el botón que se encuentra a la derecha de la lista desplegable , que descarga el archivo generado en la carpeta *Descargas* del sistema. En la Figura V.13 se muestra la planilla para el año 2014.

	A	B
1	<b>INDICADOR</b>	<b>VALOR</b>
2	Tasa de mortalidad infantil, por 1.000 nacidos vivos (menores de 1 año)	3,5
3	Número de defunciones infantiles	169
4	Tasa de mortalidad infantil, por 1.000 nacidos vivos neonatal (de 0 a 27 días)	1,6
5	Tasa de mortalidad infantil, por 1.000 nacidos vivos postneonatal (de 28 días a <1 año)	1,9
6	Mortalidad de menores de 5 años reportada	4,2
7	Razón de mortalidad materna por 100.000 nacidos vivos	0
8	Número de defunciones maternas	0
9	Número de nacidos vivos	48387
10	Proporción de bajo peso al nacer	7,5

Figura V.13 – Planilla indicadores OMS – Año 2014

## Crear consultas personalizadas

Para realizar consultas personalizadas sobre el Data Warehouse se debe acceder a la pantalla de Consultas mediante la barra de navegación. Para comenzar a armar la consulta, es necesario seleccionar el tema de análisis. Luego, para el tema elegido, se deben seleccionar las medidas a analizar y los criterios de análisis a aplicar. A continuación se presentan instrucciones de cómo realizar cada una de estas tareas.

### Seleccionar el tema de análisis

Al acceder a la pantalla de Consultas, en la sección superior, se encuentra una lista desplegable que permite seleccionar entre las opciones Nacimientos, Mortalidad y Mortalidad Infantil el tema a analizar.



Figura V.14 – Selección tema de análisis

Luego de seleccionar una opción, se despliega un árbol que contiene las medidas y los criterios de análisis para el tema elegido, junto con los paneles de Medidas, Columnas, Filas y Filtro.



Figura V.15 – Despliegue de paneles y árbol

### Seleccionar medidas a analizar

Para realizar una consulta es necesario incluir al menos una medida. Las mismas se seleccionan siguiendo los siguientes pasos:

1. En el árbol desplegado al seleccionar un tema de análisis, hacer clic en la opción Medidas. A continuación se mostrarán las medidas disponibles para el tema elegido, identificadas mediante el ícono .

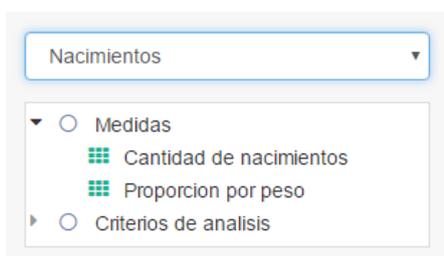


Figura V.16 – Medidas

2. Seleccionar la medida que se desea analizar y arrastrarla hacia el panel de Medidas que se encuentra en la sección superior de la página. En el ejemplo se analizarán la cantidad de nacimientos.

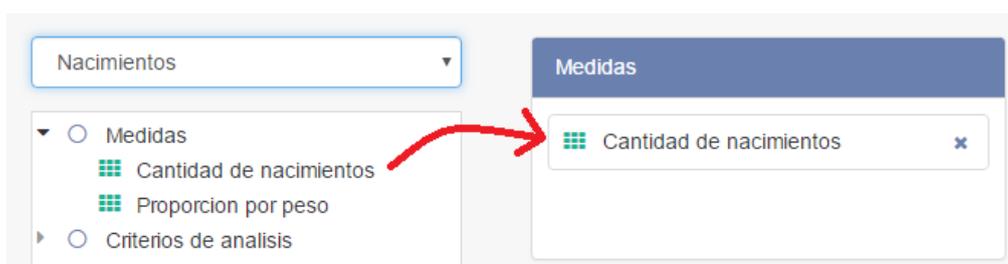


Figura V.17 – Panel de medidas

### Seleccionar criterios de análisis

Los criterios de análisis pueden ser incluidos en las filas o columnas de la tabla de resultados, o como parte de un filtro. A continuación se detalla cómo indicar qué criterios se desean incluir en la consulta. Es importante mencionar que para realizar una consulta es necesario seleccionar al menos un criterio de análisis para incluir en las filas de la tabla de resultados.

1. En el árbol desplegado al seleccionar un tema de análisis, hacer clic en la opción Criterios de análisis. A continuación se mostrarán los grupos de criterios relacionados por los cuales se pueden analizar las medidas para el tema de análisis elegido.

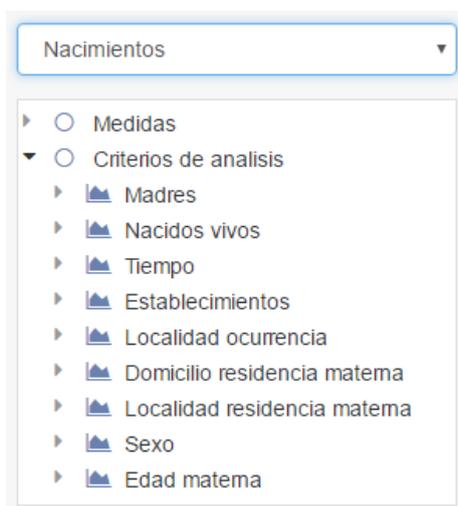


Figura V.18 – Criterios de análisis (I)

2. Al hacer clic en un grupo se encuentran sub-grupos, señalados con el ícono . Dentro de estos se despliegan también los criterios de análisis que los integran, identificados con el ícono . Estos últimos deben ser arrastrados hacia los paneles Columnas, Filas y Filtro según corresponda, como se explica a continuación. Cabe destacar que no podrá ser seleccionado más de un criterio de un mismo sub-grupo.

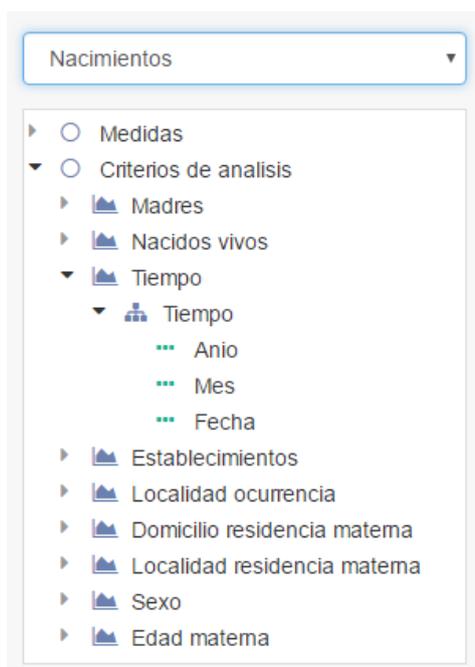


Figura V.19 – Criterios de análisis (II)

### Columnas

Es posible seleccionar qué criterios de análisis aparecerán en las columnas de la tabla de resultados arrastrándolos desde el árbol de criterios hacia el panel de Columnas que se encuentra en la sección superior de la página. A modo de ejemplo se analizarán la cantidad de nacimientos por año, ubicando este último criterio en las columnas de la tabla.

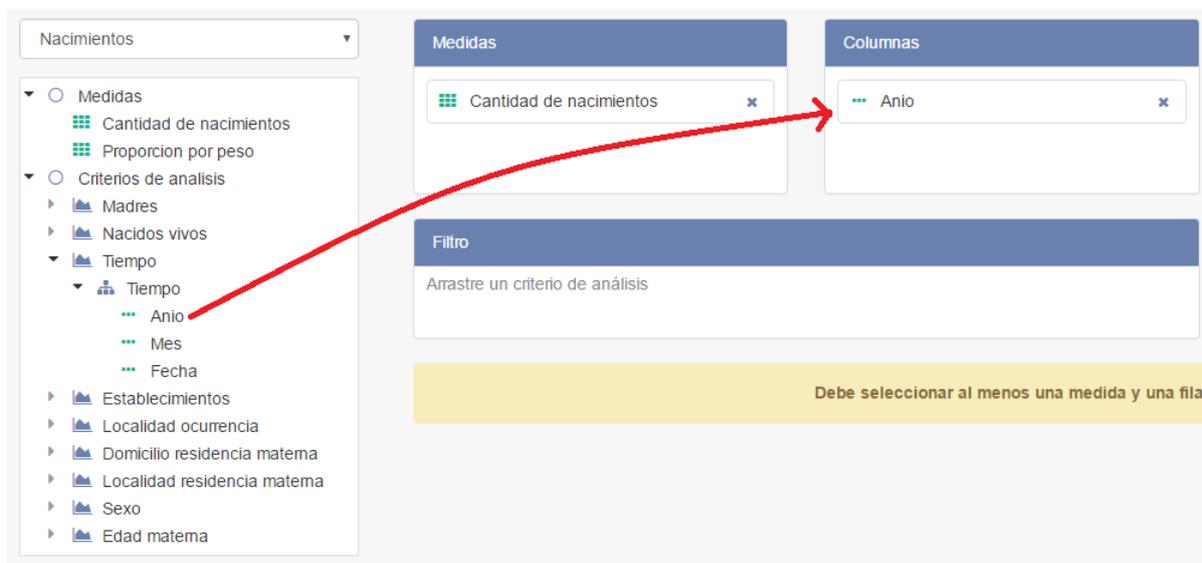


Figura V.20 – Panel Columnas

### Filas

De forma análoga a las columnas es posible seleccionar qué criterios de análisis aparecerán en las filas de la tabla, arrastrándolos desde el árbol de criterios hacia el panel de Filas que se encuentra en la sección superior de la página. Se debe recordar que es necesario seleccionar al menos una fila para poder realizar la consulta. Continuando con el ejemplo, se analizarán la cantidad de nacimientos por año en las columnas y por sexo del nacido en las filas.

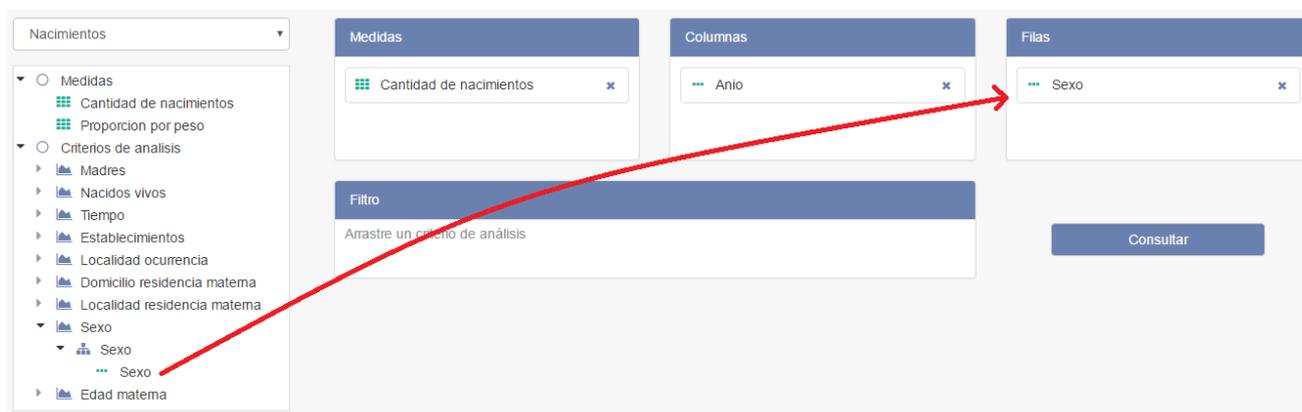


Figura V.21 – Panel Filas

## Filtros

La aplicación permite definir filtros basados en los criterios de análisis. Para ello se debe arrastrar el criterio elegido desde el árbol hacia el panel de Filtros.

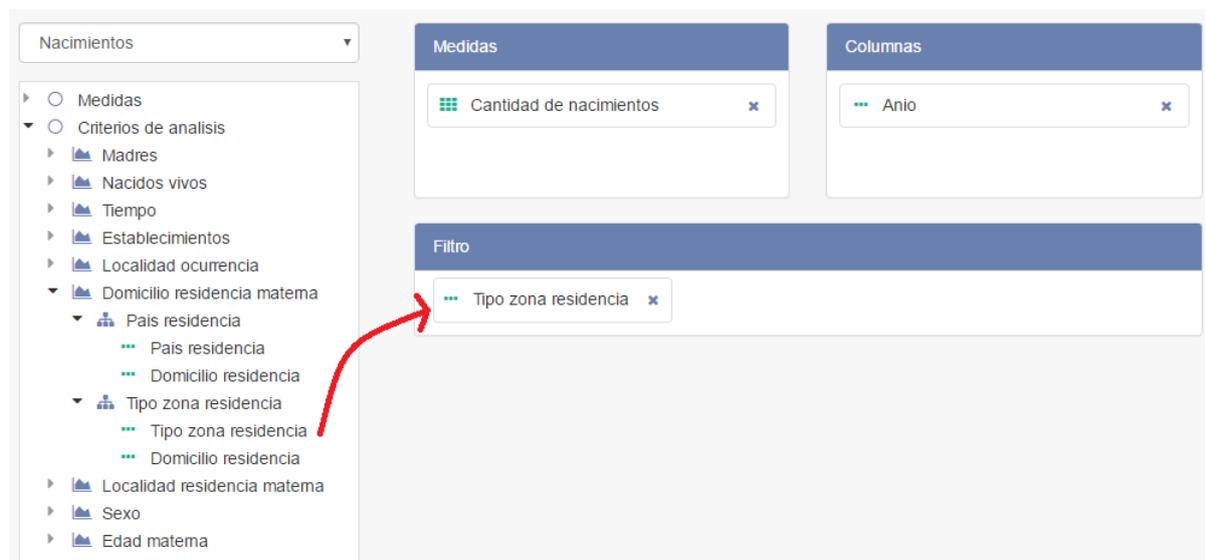


Figura V.22 – Panel Filtro

Al soltar el criterio sobre el panel, o al hacer clic sobre este, se muestra una ventana emergente que permite seleccionar los valores de dicho criterio por los cuales se desea acotar los datos a incluir en el resultado de la consulta.



Figura V.23 – Ventana Filtro

Luego de elegidos los valores se debe seleccionar el botón Guardar. En este ejemplo, se tendrán en cuenta para el análisis únicamente los nacidos vivos cuya madre, al momento del parto, reside en una zona urbana o rural. Como se puede observar en la Figura V.23, en la parte superior de la ventana se encuentra disponible la opción Seleccionar Todos, que marca todos los valores posibles del criterio seleccionado.

### Realizar consulta

Luego de armada la consulta se debe hacer clic sobre el botón Consultar que se encuentra a la derecha del panel de Filtros. Esta debe contener al menos una medida a analizar y un criterio de análisis en las filas.



Figura V.24 – Botón Consultar

La tabla resultante de ejecutar la consulta del ejemplo, que se muestra debajo de la sección de paneles, se puede observar en la Figura V.25. Las medidas elegidas siempre se muestran en las columnas de la tabla.

	Tiempo		
	2013	2014	2015
Sexo	Cantidad de nacimientos	Cantidad de nacimientos	Cantidad de nacimientos
Femenino	3.339	3.374	2.510
Masculino	3.411	3.589	2.663
Indefinido		1	
Desconocido	2		1

Figura V.25 –Tabla resultado

### **Exportar tablas personalizadas**

Al realizar una consulta, en el borde superior izquierdo de la tabla resultante se encuentran dos botones, uno verde que permite exportar la tabla en formato Excel, y uno rojo que permite exportarla en formato PDF.

Al seleccionar el botón de descarga en formato PDF , el archivo se guarda en la carpeta *Descargas* de la computadora desde la cual se accedió a la aplicación, con el nombre del tema de análisis elegido. El archivo generado para la consulta de ejemplo, cuyo nombre por defecto es *Nacimientos.pdf*, se muestra en la Figura V.26.

	Tiempo		
	2013	2014	2015
Sexo	Cantidad de nacimientos	Cantidad de nacimientos	Cantidad de nacimientos
Femenino	3.339	3.374	2.510
Masculino	3.411	3.589	2.663
Indefinido		1	
Desconocido	2		1

Figura V.26 –Resultados en formato PDF

Asimismo, al seleccionar el botón de descarga en formato Excel  , la planilla generada se guarda en la carpeta *Descargas* con el nombre del tema de análisis elegido. El resultado de exportar la tabla del ejemplo se muestra en la Figura V.27.

	A	B	C	D
1		Tiempo		
2		2013	2014	2015
3				
4	Sexo	Cantidad de nacimientos	Cantidad de nacimientos	Cantidad de nacimientos
5	Femenino	3.339,00	3.374,00	2.510,00
6	Masculino	3.411,00	3.589,00	2.663,00
7	Indefinido		1,00	
8	Desconocido	2,00		1,00

Figura V.27 –Resultados en formato Excel

## Modificar consultas

La herramienta provee las siguientes tres funcionalidades que permiten modificar las consultas previamente armadas. Las mismas se presentan a continuación.

### Eliminar criterios y medidas

Las medidas y criterios de análisis que se encuentran dentro de los paneles contienen un botón en su extremo derecho, como se muestra en la Figura V.28, que al seleccionarlo elimina la medida o el criterio del panel correspondiente.



Figura V.28 –Eliminar medida/criterio

### Modificar el orden de los criterios/medidas dentro de un panel

La disposición de las medidas y criterios dentro de los paneles Medidas, Filas y Columnas determinan el orden en que aparecerán en los cabezales de la tabla resultado. Es posible modificar la posición de una medida o criterio dentro de un panel seleccionándolo y arrastrándolo a la ubicación deseada dentro del mismo.

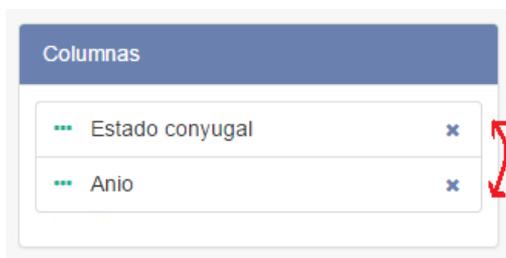


Figura V.29 –Modificar el orden de medidas/criterios

### Mover criterios entre paneles

La herramienta permite mover criterios de análisis entre los paneles Columnas, Filas y Filtro. Para ello se debe seleccionar y arrastrar el criterio desde el panel donde se encuentra hacia el panel deseado.

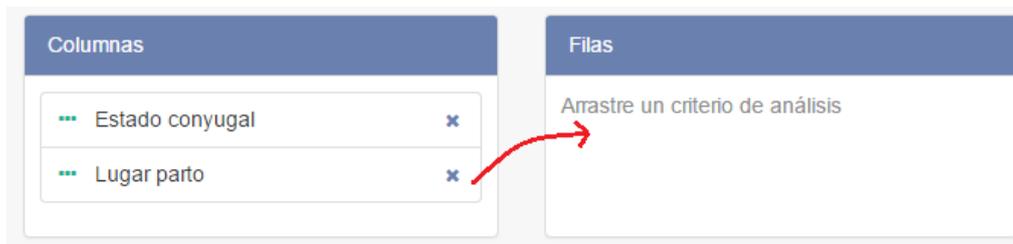


Figura V.30 –Intercambiar criterios entre paneles

## ***Cargar y actualizar el Data Warehouse***

Las funcionalidades de carga y actualización del Data Warehouse se encuentran disponibles dentro de la pantalla de Carga. A continuación se presenta una descripción de dichas funcionalidades, no sin antes destacar algunas características que son comunes para todas ellas.

- Es indispensable que las propiedades de la base de datos y la ruta del archivo que se utilizará tanto para realizar la carga como la actualización, especificados en el archivo de configuración, sean los correctos para permitir el correcto funcionamiento de dichas tareas.
- Mientras se estén procesando las operaciones, se mostrará un icono de espera  debajo del botón de carga. Luego, si la ejecución resulta exitosa se despliega el mensaje “Se han cargado los datos exitosamente”. En caso contrario, se presenta en la sección inferior de la

página un *log* que contiene los errores ocurridos, tal como se muestra en la Figura V.31. Cabe mencionar que puede ser necesario contar con cierto conocimiento técnico para comprender dichos mensajes.

```
Ha ocurrido un error. A continuación se presenta el log con los detalles:

2016/06/21 20:57:31 - DW_MSP - Start of job execution
2016/06/21 20:57:31 - DW_MSP - exec(0, 0, START.0)
2016/06/21 20:57:31 - START - Starting job entry
2016/06/21 20:57:31 - DW_MSP - Starting entry [Evaluar primeraCarga]
2016/06/21 20:57:31 - DW_MSP - exec(1, 0, Evaluar primeraCarga.0)
2016/06/21 20:57:31 - Evaluar primeraCarga - Starting job entry
2016/06/21 20:57:31 - Evaluar primeraCarga - Value to evaluate is true
2016/06/21 20:57:31 - DW_MSP - Starting entry [SQL Crear Base]
```

Figura V.31 –Log en pantalla

- En el directorio TOMCAT\_HOME/logs se guardan *logs* diarios, con nombre *cargaDW-fechaDelDía*, conteniendo el resultado detallado de todas las ejecuciones realizadas durante ese día.

### Carga inicial del Data Warehouse

Al realizar la primera carga del Data Warehouse, se crea el esquema de la base de datos y se cargan las tablas de dimensión y de hecho. Los pasos a seguir para utilizar esta funcionalidad son:

1. Seleccionar la opción de carga Inicial.

Opciones de carga

Inicial

Actualización

Personalizada

Seleccione las fechas de inicio y de fin para la carga de las tablas Nacimientos, Mortalidad y Mortalidad Infantil

Fecha de inicio: dd/mm/aaaa

Fecha de fin: dd/mm/aaaa

Cargar

Figura V.32 –Carga inicial

2. En el formulario se muestran dos campos opcionales, una fecha de inicio y otra de fin. Estos determinan el período de tiempo dentro del cual deben estar comprendidos los datos que se extraerán desde la base de datos fuente para cargar las tablas de hecho. En caso de que no

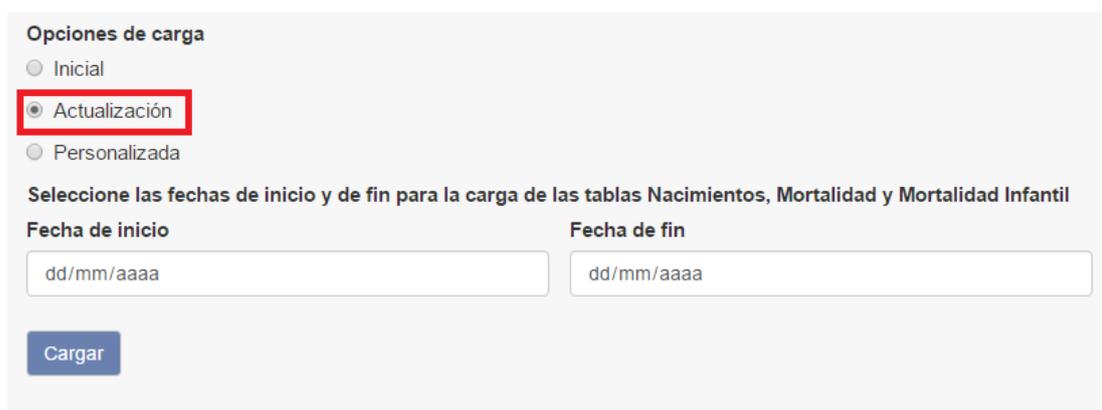
se completen estos campos, se cargarán todos los datos encontrados en dicha base de datos.

3. Hacer clic en el botón Cargar.

### Actualización del Data Warehouse

Al seleccionar esta opción se actualizarán las tablas de hecho Nacimientos, Mortalidad y Mortalidad Infantil, junto con tablas de dimensión Nacidos Vivos, Madres, Fallecidos Menores y Fallecidos. Cabe señalar que al intentar cargar datos ya existentes en el Data Warehouse, los mismos reemplazarán a los anteriores. Los pasos a seguir para ejecutar esta tarea son:

1. Seleccionar la opción Actualización.



Opciones de carga

Inicial

Actualización

Personalizada

Seleccione las fechas de inicio y de fin para la carga de las tablas Nacimientos, Mortalidad y Mortalidad Infantil

Fecha de inicio

Fecha de fin

Figura V.33 –Actualización

2. En este caso, se deben completar obligatoriamente las fechas de inicio y de fin para determinar los datos a actualizar.
3. Hacer clic en el botón Cargar.

### Carga/Actualización personalizada

Es posible indicar, mediante esta opción, las tablas que deberán ser cargadas o actualizadas en el Data Warehouse. De forma análoga a la opción *Actualización*, la carga de la tabla de hecho Nacimientos implica la carga de las tablas de dimensión Madres y Nacidos Vivos, mientras que la carga de Mortalidad y Mortalidad Infantil implica la carga de Fallecidos y Fallecidos Menores respectivamente. Resulta útil mencionar que al seleccionar una tabla de dimensión, tanto esta como las tablas de hecho, se vaciarán y se cargarán nuevamente para evitar inconsistencias. Para utilizar esta funcionalidad se deberá:

1. Seleccionar la opción Personalizada.

The screenshot shows a web interface for data loading. At the top, under 'Opciones de carga', there are three radio buttons: 'Inicial', 'Actualización', and 'Personalizada'. The 'Personalizada' option is selected and highlighted with a red box. Below this, there are nine checkboxes arranged in three columns: 'Nacimientos' (checked), 'Mortalidad', 'Mortalidad Infantil', 'Tiempo', 'Edades', 'Domicilios y Localidades', 'Sexo' (checked), 'Causas CIE-10', and 'Establecimientos'. Below the checkboxes, there is a section titled 'Seleccione las fechas de inicio y de fin para la carga de la tabla Nacimientos'. It contains two text input fields: 'Fecha de inicio' and 'Fecha de fin', both with placeholder text 'dd/mm/aaaa'. Below each field is a red error message: 'Debe seleccionar una fecha de inicio' and 'Debe seleccionar una fecha de fin'. At the bottom left, there is a blue button labeled 'Cargar'.

*Figura V.34 –Carga personalizada*

2. Indicar las tablas que se desean cargar o actualizar. En caso de elegir una tabla de hecho, se deberán completar obligatoriamente las fechas de inicio y de fin.
3. Hacer clic en el botón Cargar.