

Proyecto de Grado  
Ingeniería en Computación

**Sistema de Business Intelligence para análisis de  
datos de calidad del aire de la Intendencia de  
Montevideo**

Gonzalo Melo  
gonzalo.melo@fing.edu.uy

Tutoras:  
Dra. Adriana Marotta  
amarotta@fing.edu.uy  
MSc. Flavia Serra  
fserra@fing.edu.uy

Facultad de Ingeniería  
Universidad de la República  
20 de diciembre de 2019



## **Resumen**

Hoy en día existen muchas organizaciones que promueven el mantenimiento de una buena calidad del Medio Ambiente. Es por esto que cada vez son más las autoridades que se sienten comprometidas en realizar distintos controles que permitan el monitoreo ambiental. La Intendencia de Montevideo (IM) ha implementado un programa de monitoreo de calidad del aire, que incluye una Red de Monitoreo que comenzó a operar en 2005 [5].

El objetivo de la red es conocer la calidad del aire de base de Montevideo. Para este objetivo, la Red de Monitoreo opera con estaciones que permiten evaluar la concentración de material particulado en diversas fracciones y de gases que habitualmente se encuentran en ambientes urbanos (dióxido de azufre y dióxido de nitrógeno). Las estaciones de monitoreo están ubicadas en sitios fijos que permiten caracterizar la calidad de base de la ciudad.

A partir de todas estas mediciones se obtiene una gran cantidad de datos, que actualmente están siendo manipulados y analizados en forma manual por los especialistas de la intendencia.

El presente proyecto tiene como objetivo principal la disponibilización de una plataforma que resuelva la recolección, integración y depuración de los datos de distintas fuentes, permitiendo al Servicio de Calidad de Aire el acceso inmediato a los mismos en forma adecuada e intuitiva para el posterior análisis y generación de reportes para la toma de decisiones.

Como resultado principal del proyecto se implanta y disponibiliza, en un ambiente productivo, un sistema de Business Intelligence (BI) con datos de mediciones actuales e históricas de las estaciones de la red, del volumen vehicular en las principales arterias de Montevideo y de la actividad de distintas industrias, almacenados en un Data Warehouse (DW), con el fin de ser analizados en conjunto según distintos indicadores y reportes predefinidos.

Priorizando la adopción, utilización y correcto funcionamiento de la herramienta, objetivo primordial del proyecto, se logran dos resultados relevantes. Primero, se obtiene un conjunto de herramientas que facilitan la interacción con la plataforma: un conjunto predefinido de indicadores, gráficas, consultas y reportes para el análisis de BI, una aplicación web desarrollada a medida para la carga y actualización del DW, la integración de herramientas estadísticas conocidas y utilizadas por los especialistas del servicio, y un manual de usuario web con vídeos y tutoriales sobre el uso de la plataforma. Segundo, se logran objetivos técnicos destacables como una correcta separación en distintos componentes arquitectónicos y el despliegue en contenedores, haciendo foco en la facilidad de instalación, actualización y soporte del sistema en producción.

## **Palabras clave**

Red de Monitoreo, Business Intelligence, Data Warehouse, Calidad de Aire, Análisis Multidimensional



# Índice general

<b>Resumen</b>	<b>1</b>
<b>1. Introducción</b>	<b>5</b>
<b>2. Red de Monitoreo de Calidad de Aire de Montevideo</b>	<b>7</b>
2.1. Estaciones . . . . .	7
2.2. Contaminantes y Métodos . . . . .	9
2.2.1. Monóxido de carbono (CO) . . . . .	9
2.2.2. Óxidos de Nitrógeno (NOx) . . . . .	10
2.2.3. Dióxido de Azufre (SO2) . . . . .	10
2.2.4. Material Particulado (PM) . . . . .	10
2.3. Índice de Calidad de Aire (ICAire) . . . . .	11
2.4. Informes semanales y anuales . . . . .	13
<b>3. Conceptos básicos de sistemas de Business Intelligence (BI)</b>	<b>15</b>
3.1. Introducción . . . . .	15
3.2. Proceso de desarrollo . . . . .	17
3.2.1. Planificación y Arquitectura . . . . .	18
3.2.2. Modelado de Datos e Integración . . . . .	21
3.2.3. Reportes y Analítica . . . . .	26
3.2.4. Implantación, Mantenimiento y Crecimiento . . . . .	27
<b>4. Análisis</b>	<b>29</b>
4.1. Planificación . . . . .	29
4.2. Análisis . . . . .	31
4.2.1. Enfoque orientado a las fuentes . . . . .	31
4.2.2. Enfoque orientado al usuario . . . . .	40
4.2.3. Estudio de factibilidad . . . . .	40
4.2.4. Especificación de requerimientos funcionales . . . . .	41
4.2.5. Especificación de requerimientos no funcionales . . . . .	50
<b>5. Diseño</b>	<b>51</b>
5.1. Arquitectura . . . . .	51
5.1.1. Fuentes de datos . . . . .	52
5.1.2. Back-end . . . . .	52
5.1.3. DW . . . . .	52
5.1.4. OLAP . . . . .	53
5.1.5. Front-end . . . . .	54

5.1.6. Operaciones del sistema . . . . .	54
5.2. Diseño . . . . .	56
5.2.1. Diseño Conceptual . . . . .	56
5.2.2. Diseño Lógico . . . . .	82
5.2.3. Diseño Físico . . . . .	87
<b>6. Implementación</b>	<b>89</b>
6.1. Extracción, Transformación y Carga (ETL) . . . . .	89
6.2. Implementación de cubos . . . . .	109
6.2.1. Dimensiones compartidas . . . . .	109
6.2.2. Cubo Vehículos Contaminación . . . . .	110
6.2.3. Cubo Industrias Contaminación . . . . .	111
6.2.4. Cubo Días Importantes Contaminación . . . . .	112
6.3. Generación de reporte de BI de sección del Informe Anual . . . . .	113
6.4. Herramientas utilizadas . . . . .	119
6.5. Interfaz de usuario . . . . .	122
6.5.1. Aplicación web de carga, actualización y reportes . . . . .	122
6.5.2. Análisis . . . . .	125
<b>7. Conclusiones y trabajo futuro</b>	<b>129</b>
7.1. Conclusiones . . . . .	129
7.2. Trabajo futuro . . . . .	131
<b>Bibliografía</b>	<b>135</b>
<b>Anexos</b>	<b>139</b>
<b>A. Monitoreo</b>	<b>141</b>
<b>B. Métricas</b>	<b>145</b>
<b>C. Carga y actualización</b>	<b>149</b>
<b>D. Implementación de cubos</b>	<b>155</b>

# Capítulo 1

## Introducción

Montevideo tiene –en líneas generales– una buena calidad de aire. La predominancia de vientos, el suave relieve donde no destacan accidentes topográficos de importancia y la cercanía del Río de la Plata proveen una situación favorable para la dispersión natural de posibles contaminantes.

La calidad del aire de una ciudad se ve presionada constantemente por las numerosas actividades que en ella se desarrollan. Estas incluyen la quema de combustible para el transporte y la industria; el uso de leña tanto en el área comercial como residencial para la calefacción y cocción de alimentos, la generación de energía eléctrica a partir de centrales térmicas y quema de residuos a cielo abierto, entre otras.

Es evidente que todas estas actividades conviven en Montevideo y son responsables de que se emitan continuamente gases y partículas con potencial impacto en la calidad del aire. Por tanto, es posible que, en algunas oportunidades, la capacidad depuradora natural no sea suficiente y en algunas zonas de la ciudad se observe mayor concentración de contaminantes, lo que eventualmente podría hacer que la calidad de aire fuera inadecuada.

Por esta situación resulta imprescindible que la IM mantenga el monitoreo y evaluación de la calidad de aire en forma continua, una tarea que realiza a través de la Unidad Calidad de Aire del Servicio de Evaluación de la Calidad y Control Ambiental (ECCA).

Para desarrollar este cometido se ha implementado un programa de monitoreo de calidad del aire, que incluye una Red de Monitoreo que comenzó a operar en 2005.

Las distintas estaciones de monitoreo realizan mediciones periódicas de las cuales se obtiene una gran cantidad de datos, que actualmente están siendo manipulados y analizados en forma manual por los especialistas de la intendencia, Ing. Quím. Andrea De Nigris e Ing. Quím. Pablo Franco, y su equipo.

De lo antedicho se desprende la necesidad de contar con un sistema que permita la realización de las tareas de manera efectiva, simplificando el trabajo efectuado manualmente al automatizar el proceso de recolección y procesamiento de los datos y la generación de herramientas de análisis para su posterior utilización.

Para lograr este propósito se propone el uso de un sistema de BI que sirve para brindar una plataforma unificada para extraer, procesar, almacenar y disponibilizar los datos que se poseen para convertirlos en información relevante mediante el uso herramientas de análisis. Para este

cometido se utiliza una plataforma de DW que sirve como repositorio de datos del sistema de BI.

Este trabajo se plantea como objetivo desarrollar un sistema de BI que permita la automatización de la extracción, integración, limpieza y carga de los datos provistos por las estaciones, y brinde una interfaz a medida para que los usuarios puedan analizar toda la información de interés.

El objetivo general de este proyecto es desarrollar e implantar una herramienta de BI para el análisis de información relativa a la calidad del aire en Montevideo.

Los objetivos específicos del proyecto son los siguientes:

1. Lograr integrar datos de distintas fuentes de interés para los especialistas de la IM.
2. Proveer una interfaz sencilla que permita las cargas sucesivas de los datos al DW.
3. Proveer diferentes herramientas de BI, implementadas a medida de los requerimientos de los usuarios, para el análisis de los datos, incluyendo herramientas Online Analytical Processing (OLAP) [27], tableros de control, reportes, visualización de los datos desplegados en mapas, etc.
4. Lograr satisfacer los requerimientos de análisis de los usuarios y proveerles un sistema que sea amigable en su utilización.

Los resultados esperados son:

1. Prototipo de la herramienta de BI.
2. Documentación del proyecto y manual de la herramienta.
3. Adquisición por parte de los estudiantes de conocimiento a fondo de los problemas y soluciones en la implementación de este tipo de sistemas.

El resto del informe tiene la siguiente estructura:

El segundo capítulo consta de una introducción a la Red de Monitoreo de Calidad de Aire de Montevideo donde se brinda un acercamiento al contexto del proyecto.

El tercer capítulo brinda un abordaje teórico de los sistemas de BI. Además se presentan las herramientas y tecnologías utilizadas en la plataforma.

En el cuarto capítulo se presenta la planificación del proyecto y la etapa de análisis del sistema de BI que incluye la descripción de datos fuente y la especificación de requerimientos.

El quinto capítulo engloba los aspectos relacionados a la arquitectura de software y al modelado conceptual, lógico y físico que componen el diseño del sistema de BI.

El sexto capítulo contiene la implementación del sistema de BI. En el mismo se describen las herramientas de software utilizadas, la carga del DW, la implementación de los cubos, la generación del reporte de la sección del Informe Anual y las herramientas de interfaz de usuario que serán utilizadas por los especialistas para realizar los análisis.

El séptimo capítulo consta de las conclusiones logradas del trabajo y una visión de los próximos puntos de acción.

Al final se detalla la bibliografía utilizada y se anexa el contenido que sirve para brindar la completitud de la documentación del sistema.

# Capítulo 2

## Red de Monitoreo de Calidad de Aire de Montevideo

En Montevideo se evalúa la calidad del aire desde 1968, habiéndose constituido propiamente en una red de monitoreo a partir del año 2005. Los datos provenientes de las mediciones del período anterior (1968-2002) fueron utilizados como antecedentes para la realización de una campaña de diagnóstico de la calidad del aire en el período 2003-2004 que dio lugar a la red tal como está configurada actualmente.

### 2.1. Estaciones

En el año 2003, la Dirección Nacional de Medio Ambiente (DINAMA) del Ministerio de Vivienda, Ordenamiento Territorial y Medio Ambiente (MVOTMA) junto con los especialistas del Servicio de ECCA, dependiente del Departamento de Desarrollo Ambiental (DDA), de la IM, llevaron adelante la campaña de diagnóstico de la calidad del aire de Montevideo, que ofició como línea de base para el diseño de la Red y el Programa de Monitoreo para permitir la evaluación y vigilancia de la calidad del aire en el departamento.

En el año 2005, la red de monitoreo comenzó utilizando equipamiento de la DINAMA y de la IM, siendo gestionada por personal de la IM. Esta Red comenzó operando mayoritariamente en forma manual recabando información cada seis días. En el año 2018 DINAMA instaló una estación automática en Avenida Italia y Barradas cuyos datos pasaron a integrar los informes de la Red.

En el año 2009, a partir de un acuerdo con la Administración Nacional de Usinas y Trasmisiones Eléctricas (UTE), se sumaron los datos correspondientes a una estación automática propiedad de este organismo, ubicada en las cercanías de la planta de La Tablada (Colón) y posteriormente una segunda estación automática en el año 2011, ubicada en las cercanías del Palacio Legislativo. Por último, en el año 2014 comenzaron a incluirse en los informes de la Red de Monitoreo datos de la estación Bella Vista, ubicada en las inmediaciones de la central Termoeléctrica José Batlle y Ordoñez.

En el año 2012 la Administración Nacional de Combustibles Alcohol y Portland (ANCAP) se incorporó a la Red de Monitoreo, aportando datos provenientes de una estación automática

localizada en los alrededores de la refinería de La Teja.

La IM cuenta a la fecha con cuatro estaciones de calidad de aire con equipamiento automático y manual ubicadas en Ciudad Vieja, Curva de Maroñas, Tres Cruces y Colón. Los equipos de operación manual evalúan la concentración promedio de algunos contaminantes durante 24 horas, permiten verificar las tendencias de más largo plazo y se utilizan como datos de control para los equipos automáticos.

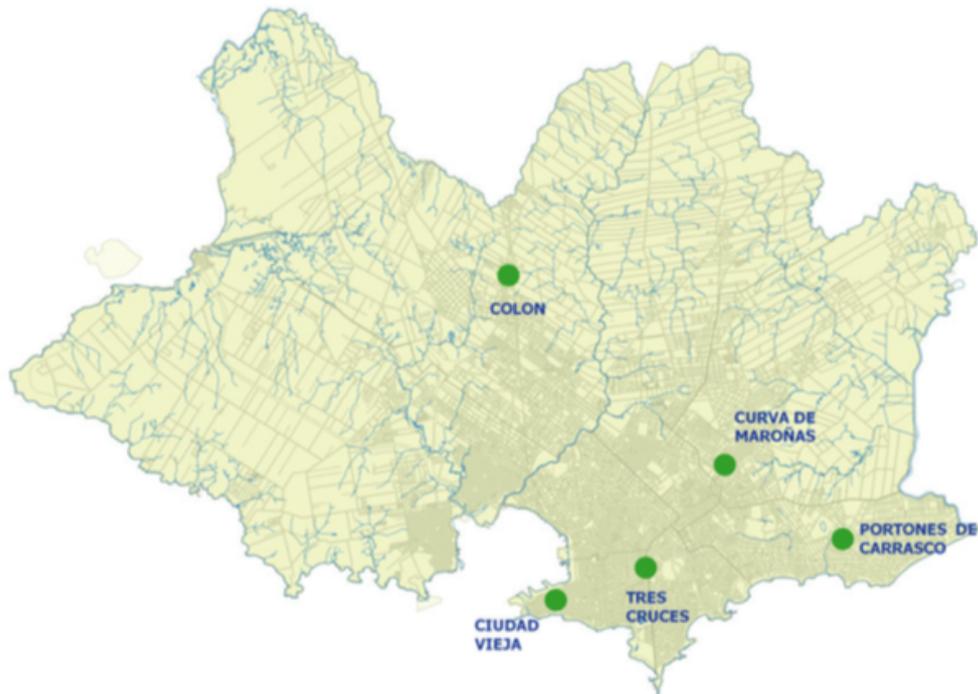


Figura 2.1: Red de Monitoreo de Base [5]

Nombre	Referencia	Dirección	Material Particulado					Tren de Monitoreo
			PTS	PM10	PM2,5	NO2	O3	
Ciudad Vieja	Academia Uruguay Asociación de Bancarios del Uruguay.	Rincón y JC Gómez Camacurá y Reconquista			X			X
Tres Cruces	Shopping Tres Cruces	Br Artigas y V Haedo		X		X	X	X
Curva de Maroñas	Centro Comunal Zonal 9	Av. 8 de Octubre y Marcos Sastre		X	X	X	X	X
Portones de Carrasco	Policlínica Lugo – Casa de Galicia	Av Italia y Av Bolivia		X				X
Colón	Servicio de Máquinas MTOP	Garzón y Camino Colman	X		X	X	X	

Figura 2.2: Red de Monitoreo de Base - Contaminantes [5]

Las estaciones automáticas tienen la ventaja de recabar información en forma horaria, funcionando todos los días del año y son mantenidas por sus respectivos propietarios (UTE, ANCAP y DINAMA). La IM recibe los datos, realiza el análisis, procesamiento, y los publica.

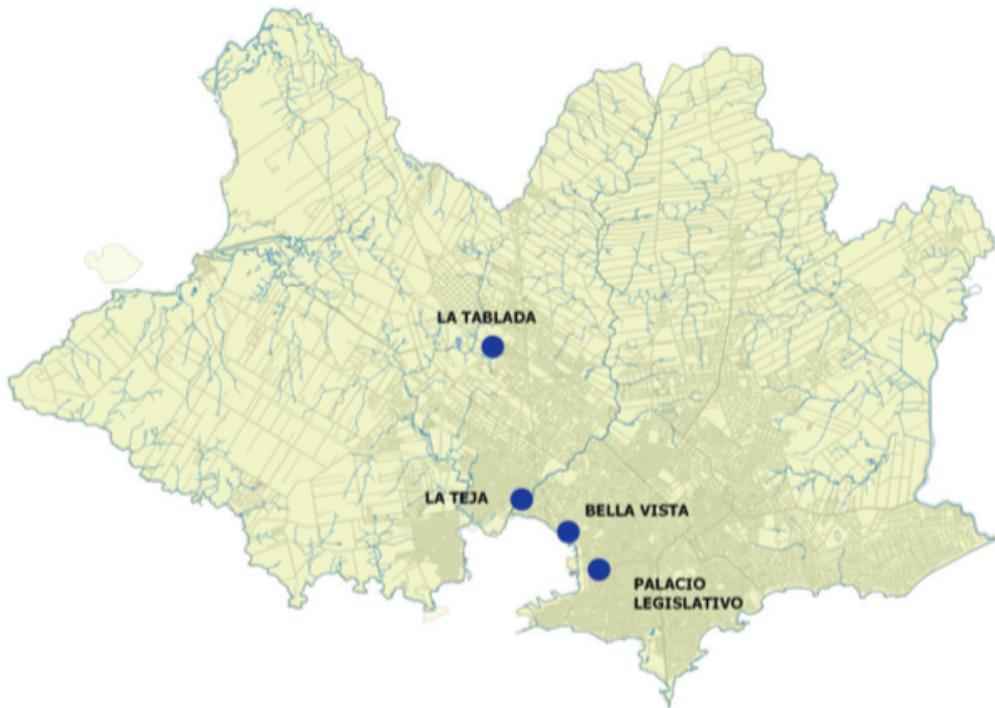


Figura 2.3: Red de Monitoreo Vigilancia de Fuentes Significativas [5]

Nombre	Propietario	Dirección	PM10	PM2.5	SO2	NO2	CO	TRS	Parametros Meteorológicos
La Teja	ANCAP	Del Cid y Yañez Pinzón		X		X	X	X	X
Palacio Legislativo	UTE	Guatemala y Acuña de Figueroa	X		X	X	X		
La Tablada	UTE	Teofilo Díaz y Máximo Santos	X		X	X	X		X
Bella Vista	UTE	Río Grande y Dionisio Carbajal	X		X	X	X		X

Figura 2.4: Red de Monitoreo Vigilancia de Fuentes Significativas - Contaminantes [5]

## 2.2. Contaminantes y Métodos

### 2.2.1. Monóxido de carbono (CO)

El CO es un gas que se produce a partir de la combustión a bajas concentraciones de oxígeno, lo que se denomina combustión incompleta. La bibliografía indica que 86% de las emisiones proviene del transporte, seguida con 6% por quema de combustible en la industria y 3% por procesos industriales, el 4% restante se origina en quemas y otros procesos no identificados. En forma natural se genera a partir de la oxidación de metano, comúnmente producida por la descomposición de materia orgánica.

El CO puede causar efectos adversos en la salud ya que compite con el Oxígeno (O<sub>2</sub>) en el

torrente sanguíneo, lo que reduce la capacidad de la sangre de transportar el oxígeno a los diferentes órganos. Las personas sensibles, particularmente aquellas con problemas cardíacos, pueden ver disminuida su capacidad de oxigenación.

Sin embargo, las concentraciones de CO raras veces exceden los límites establecidos para la preservación de la salud, incluso en grandes centros urbanos [4].

### **2.2.2. Óxidos de Nitrógeno (NO<sub>x</sub>)**

Los óxidos de nitrógeno describen una mezcla de dos gases: óxido nítrico (NO) y dióxido de nitrógeno (NO<sub>2</sub>). Son gases inorgánicos formados por la combinación de oxígeno con el nitrógeno del aire. El NO es producido en cantidades mucho mayores pero se oxida rápidamente a NO<sub>2</sub> en la atmósfera. La emisión de este gas se debe fundamentalmente a transporte (62 %), combustión para generación de energía, mecánica y eléctrica (30 %) y procesos industriales (7 %). Son originados naturalmente por descomposición bacteriana, incendios forestales y actividad volcánica.

Actualmente la evidencia científica relaciona la exposición de corto plazo con efectos respiratorios. Se ha encontrado que la concentración de NO<sub>2</sub> en las cercanías de vías de tránsito importantes puede ser considerable, por lo que es importante considerar el efecto en individuos sensibles.

El NO<sub>2</sub> causa efectos perjudiciales en los bronquios, puede irritar los pulmones y bajar la resistencia a infecciones respiratorias. Contribuyen a la formación de la lluvia ácida, aumentan la concentración de nitratos en suelos y aguas superficiales. Estos óxidos reaccionan con otras partículas en el ambiente y se integran al material particulado; en presencia de compuestos orgánicos volátiles y radiación solar reaccionan generando Ozono (O<sub>3</sub>) que también puede tener efectos adversos sobre el sistema respiratorio de la población sensible [4].

### **2.2.3. Dióxido de Azufre (SO<sub>2</sub>)**

El SO<sub>2</sub> es un gas muy reactivo en la atmósfera. Casi todos los combustibles fósiles tienen rastros de azufre en su composición, por lo que el SO<sub>2</sub> se emite fundamentalmente a través de la quema de combustibles fósiles (93 %) tanto en la industria como en la generación de energía eléctrica, seguido por las emisiones asociadas a los procesos de producción y el transporte.

Actualmente la Organización Mundial de la Salud (OMS) ha establecido una correlación directa entre los efectos negativos que afectan al aparato respiratorio y la concentración de SO<sub>2</sub> en el aire, mencionándose especialmente la aparición de broncoespasmos y efectos en asmáticos.

El SO<sub>2</sub> se utiliza como indicador de la familia de los óxidos azufrados que se denominan genéricamente SO<sub>x</sub>. La presencia de SO<sub>2</sub> en la atmósfera es responsable directamente de la acidez de las lluvias [4].

### **2.2.4. Material Particulado (PM)**

El término Material Particulado incluye partículas sólidas o líquidas que, por su pequeño tamaño, permanecen suspendidas en el aire. La caracterización de las partículas suspendidas en

el aire se realiza de acuerdo a su tamaño. Este se indica en el nombre PM<sub>n</sub>, correspondiendo la n al diámetro aerodinámico de las partículas retenidas (usualmente expresado en  $\mu\text{m}$ , o sea, micrómetros).

En general, se considera que cuando el diámetro aerodinámico de estas partículas es mayor a 100  $\mu\text{m}$  sedimentan rápidamente y no permanecen en el aire un tiempo suficiente para ser trasladadas más allá de algunos metros. Las partículas cuyo diámetro aerodinámico es menor de 100 y mayor de 0,01  $\mu\text{m}$  son las que tienen comparativamente una alta estabilidad en la atmósfera.

Estas partículas incluyen polen, material biológico microscópico, polvo, recirculación de suelo, hollín y otros pequeños sólidos. En el ambiente urbano se originan fundamentalmente a partir de resuspensión de suelo, quema de combustibles en fuentes móviles e industrias, incineración no controlada, algunos procesos industriales y quema de leña. Por otra parte y adicionalmente, en cercanías del mar se producen aerosoles con altas concentraciones de cloruro de sodio de origen natural.

El material particulado total se denomina Partículas Totales en Suspensión (PTS) y es la suma de todas las partículas suspendidas hasta 100  $\mu\text{m}$  de diámetro aerodinámico. El particulado grueso o PM<sub>10</sub> corresponde a la fracción de partículas cuyo diámetro es menor de 10  $\mu\text{m}$ ; este grupo es particularmente relevante para la salud, porque pueden ser inhaladas y penetrar las vías respiratorias más allá de la laringe. El material particulado fino corresponde a la fracción menor de 2,5  $\mu\text{m}$  de diámetro, denominada también fracción respirable porque eventualmente puede llegar al torrente sanguíneo.

La porción de material particulado proveniente de la combustión incompleta de combustible fósil contiene una alta concentración de carbono orgánico en su composición. Este parámetro se puede evaluar a través de diferentes metodologías que permiten estimar su concentración en el aire. Usualmente se lo conoce como humo negro y es uno de los mayores contribuyentes del particulado fino; es suficientemente pequeño como para ser inhalado [4].

## 2.3. Índice de Calidad de Aire (ICAire)

Los resultados obtenidos semanalmente en la Red de Monitoreo de Calidad de Aire se informan utilizando el ICAire [1]. El protocolo para su cálculo se generó en 2007 por el Laboratorio de Calidad Ambiental del Departamento de Desarrollo Ambiental de forma similar al Air Quality Index (AQI) [1], utilizado en Agencia de Protección Ambiental de Estados Unidos (EPA). En 2010 se modificó y se agregó una categoría y otros parámetros.

Para establecer el índice se tomaron en cuenta diversas experiencias exitosas de otros países, en particular la efectividad de la comunicación con esta herramienta.

El índice transforma las concentraciones de los diferentes contaminantes a una escala arbitraria de 0 a 400, similar para todos, distribuida en seis categorías como se muestra en la figura 2.5.

Las categorías consideradas son:

1. Muy buena: Si el ICAire está entre 0 y 25, se considera que la calidad del aire es muy buena.

Nivel 1	0 a 25	Muy buena		<b>Calidad de aire óptima.</b>
Nivel 2	26 a 50	Buena		<b>Situación favorable a la realización de todo tipo de actividades.</b>
Nivel 3	50 a 100	Aceptable		<b>Calidad aceptable para la mayoría de las personas.</b>
Nivel 4	101 a 200	Inadecuada		<b>Aparición de molestias en personas sensibles.</b>
Nivel 5	201 a 300	Mala		<b>Molestias e intolerancia en personas con padecimiento respiratorio.</b>
Nivel 6	301 a 400	Muy mala		<b>Aparición de síntomas e intolerancia en la población.</b>

Figura 2.5: ICAire [1]

2. Buena: Si el ICAire está entre 26 y 50, se considera que la calidad del aire es buena.
3. Aceptable: Si el ICAire se encuentra entre 51 y 100, la calidad del aire es aceptable. A pesar de que existe algún contaminante, sus niveles no presentan riesgos para la mayoría de las personas, aunque es posible que algunas personas especialmente sensibles se vean afectadas.
4. Inadecuada: Aparición de molestias en personas sensibles cuando el ICAire está entre 101 y 200. En general, la mayoría no se ve afectada.
5. Mala: Los valores de ICAire entre 201 y 300 puede provocar que cualquier persona sienta sus efectos, los que pueden ser severos en individuos sensibles.
6. Muy mala: Con valores de ICAire mayores de 301, el contaminante puede provocar efectos en la salud de cualquier persona.

La categoría de cada estación se define a partir del ICAire de cada uno de los parámetros y se establece que la categoría que caracteriza a la estación en estudio corresponde al mayor valor obtenido en ese día de muestreo.

En las estaciones de la Red Manual se informa el ICAire para los días de muestreo, es decir, cada seis días. En cambio, para las estaciones de la Red Automática se informa el ICAire para cada día.

## 2.4. Informes semanales y anuales

Semanalmente se emiten públicamente, a modo de información y monitoreo de la calidad del aire actual, dos informes de la Red de Monitoreo orientados a los resultados de las distintas estaciones según su objetivo [3]:

1. Informe semanal sobre Red Base: Orientado a la evaluación de la calidad base de la Ciudad, donde se incluyen los registros de estaciones en Ciudad Vieja, Tres Cruces, Curva de Maroñas, Portones de Carrasco y Colón.
2. Informe semanal sobre Red Fuentes Significativas: Orientado a evaluar el impacto generado por algunas de las fuentes más significativas del departamento como las cercanías de ruta 5 y camino Lecocq, del Palacio Legislativo, la central termoeléctrica José Batlle y Ordóñez, y la refinería de La Teja. Esta red incluye las estaciones en La Teja, Bella Vista, Palacio Legislativo y La Tablada.

Anualmente se emite públicamente y se presenta un informe elaborado [2] a los distintos jerarcas de la IM donde se indican todos los resultados obtenidos desde enero a diciembre, incluyendo un análisis histórico comparativo con los años anteriores. Utilizando dicho informe se puede realizar un análisis de la calidad de aire del año transcurrido.



# Capítulo 3

## Conceptos básicos de sistemas de Business Intelligence (BI)

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

### 3.1. Introducción

El contexto de la sociedad de la información ha propiciado la necesidad de tener mejores, más rápidos y más eficientes métodos para extraer y transformar los datos de una organización en información y distribuirla a lo largo de la cadena de valor.

El Business Intelligence responde a dicha necesidad. Sin embargo, este concepto, que actualmente se considera crítico en la gran mayoría de empresas, no es nuevo. Desde hace décadas el concepto ha evolucionado aunando diferentes tecnologías, metodologías y términos bajo su paraguas.

Actualmente se entiende por Business Intelligence al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización [23].

En los últimos años, el mercado Business Intelligence se ha visto marcado por una clara evolución que lo destaca como un mercado maduro:

1. Se ha producido una consolidación del mercado mediante la compra de empresas pequeñas por parte de los principales agentes del mercado (SAP, IBM, Microsoft y Oracle).
2. Pentaho, una de las plataformas open source más completas y en la cual se basa la solución de este proyecto, fue adquirida por Hitachi en 2015.
3. Se ha enriquecido con soluciones open source que cubren el espectro de necesidades de una organización para la explotación de la información.
4. Han aparecido nuevas empresas con foco en la innovación cubriendo nuevos nichos en el mercado de la inteligencia de negocio como la visualización, el análisis predictivo, los contenedores livianos y/o el Business Intelligence en tiempo real.

5. A pesar de los momentos de la crisis económica el mercado de inteligencia de negocio sigue en una fase de crecimiento estable al posicionarse como una necesidad crítica para toda organización.

Como se puede visualizar en el cuadrante mágico de Gartner, en la figura 3.1, para herramientas de integración de datos en Agosto 2019, Hitachi Vantara (Pentaho) se posiciona como un una empresas de nicho. En la misma se diferencian del resto en que normalmente están especializados en segmentos muy específicos de su mercado o trabajan exclusivamente en sectores verticalizados.



Figura 3.1: Cuadrante mágico de Gartner [6]

La implantación de estos sistemas de información proporciona diversos beneficios, entre los que podemos destacar:

1. Crear un círculo virtuoso de la información (en la figura 3.2 se visualiza que los datos se transforman en información que genera un conocimiento que permite tomar mejores decisiones que se traducen en mejores resultados y que generan nuevos datos).
2. Permitir una visión única, conformada, histórica, persistente y de calidad de toda la información.
3. Crear, manejar y mantener métricas, Key Performance Indicator (KPI) y Key Goal Indicator (KGI) fundamentales para la empresa.
4. Aportar información actualizada tanto a nivel agregado como en detalle.
5. Reducir el diferencial de orientación de negocio entre el departamento TI y la organización.

6. Mejorar comprensión y documentación de los sistemas de información en el contexto de una organización.
7. Mejorar de la competitividad de la organización como resultado de ser capaces de:
  - a) Diferenciar lo relevante sobre lo superfluo.
  - b) Acceder más rápido a información.
  - c) Tener mayor agilidad en la toma de las decisiones.

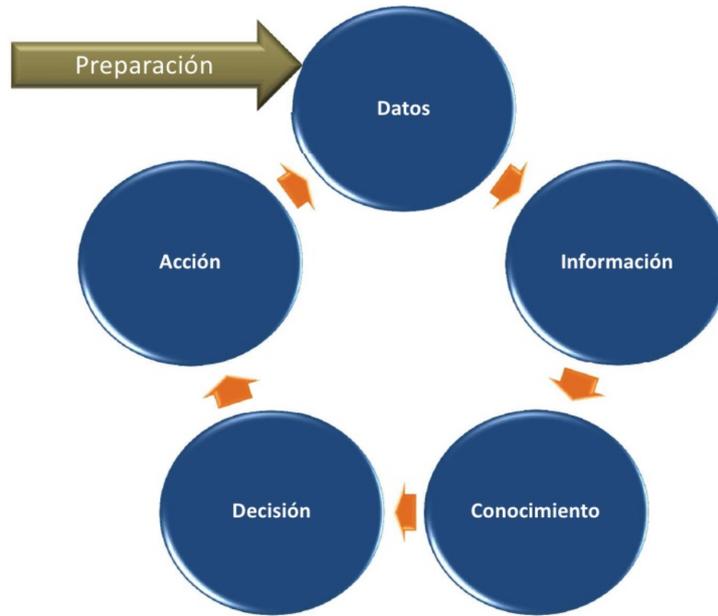


Figura 3.2: Círculo Virtuoso - Análisis de Datos

## 3.2. Proceso de desarrollo

La Metodología Kimball, empleada para la construcción de un sistema de BI y DW, se basa en el Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle) [25].

Este ciclo de vida del proyecto de BI, está basado en cuatro principios básicos:

1. Centrarse en el negocio.
2. Construir una infraestructura de información adecuada.
3. Realizar entregas en incrementos significativos (este principio consiste en crear el DW en incrementos entregables en plazos de 6 a 12 meses, en este punto, la metodología se parece a las metodologías ágiles de construcción de software).
4. Ofrecer la solución completa con todos los elementos necesarios para entregar valor a los usuarios de negocios:
  - a) Centrarse en el negocio.
  - b) DW bien diseñado.

- c) Entregar herramientas de consulta ad hoc.
- d) Aplicaciones para informes y análisis avanzado.
- e) Capacitación.
- f) Soporte.
- g) Sitio web.
- h) Documentación.

La construcción de una solución de BI y DW es sumamente compleja, y Kimball propone una metodología que ayuda a simplificar esa complejidad. Las tareas de esta metodología (ciclo de vida) se presenta en la figura 3.3 y se describen a continuación.

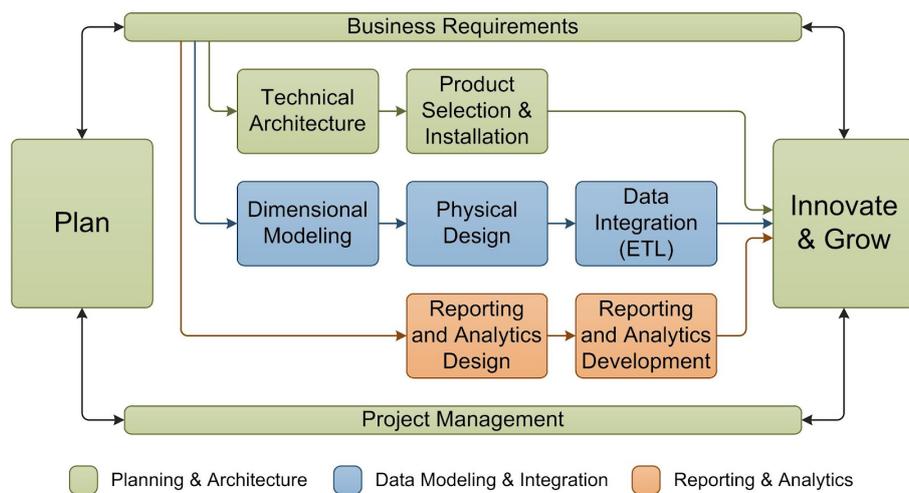


Figura 3.3: Ciclo de Vida Dimensional del Negocio [25]

### 3.2.1. Planificación y Arquitectura

El primer recuadro en el Ciclo de Vida Dimensional del Negocio se centra en conseguir lanzar el proyecto, incluyendo la determinación del alcance, la justificación y la dotación de personal. A lo largo del ciclo de vida, la planificación y la dirección de las tareas del proyecto mantiene las actividades en marcha.

Los entornos BI/DW se encargan de la integración de numerosas tecnologías, almacenes de datos y metadatos asociados. La ruta de la tecnología en el Ciclo de Vida Dimensional del Negocio empieza con el diseño del sistema de arquitectura para establecer una lista de la obtención de las capacidades necesarias, seguidas de la selección e instalación de productos que satisfagan esas necesidades de la arquitectura.

#### Planificación del Proyecto

En este proceso se determina el propósito del proyecto de BI/DW, sus objetivos específicos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información.

Esta tarea incluye las siguientes acciones típicas de un plan de proyecto:

1. Definir el alcance (entender los requerimientos del negocio).
2. Identificar las tareas.
3. Programar las tareas.
4. Planificar el uso de los recursos.
5. Asignar la carga de trabajo a los recursos.
6. Elaboración de un documento final que representa un plan del proyecto.

Además en esta parte definimos cómo realizar la administración o gestión con las siguientes actividades:

1. Monitoreo del estado de los procesos y actividades.
2. Rastreo de problemas.
3. Desarrollo de un plan de comunicación comprensiva que dirija la empresa y las áreas de TI.

### **Definición de Requerimientos del Negocio**

Identificar las necesidades del negocio es una tarea clave en el ciclo de vida Kimball ya que estos descubrimientos dirigen la mayoría de decisiones ascendentes y descendentes. Las necesidades se recogen para determinar los factores clave que repercuten en el negocio centrándose en lo que los usuarios de negocio hacen hoy (o lo que quieren hacer en el futuro) en lugar de preguntar “¿Qué quieres almacenar en el almacén de datos?”, se identifican las oportunidades más significativas en la empresa, basándose en el valor del negocio y su viabilidad. Posteriormente las necesidades detalladas son recopiladas en la primera iteración del sistema de desarrollo BI/DW.

En la etapa de relevamiento de requerimientos se pueden utilizar distintos enfoques para el análisis según las capacidades y posibilidades de los datos, del equipo y del proyecto en si.

#### Enfoque desde Requerimientos

En el enfoque desde requerimientos se descubren junto a los usuarios del proyecto los requerimientos de negocio que necesitan. Dicho enfoque tiene las siguientes características:

1. Los requerimientos son el universo de información.
2. Las bases fuente se relacionarán luego.
3. Aplicable cuando se tienen Bases Fuentes complejas. (Se analizan con los requerimientos en mente).

#### Enfoque desde Datos

En el enfoque desde los datos se analizan los datos fuentes para visualizar qué respuestas pueden brindar para el negocio. Dicho enfoque tiene las siguientes características:

1. Datos fuentes son el universo de información.
2. El DW se obtiene transformando las fuentes.
3. Aplicable cuando los requerimientos están poco claros.

### **Gestión de Proyecto**

Esta fase tiene como objetivo hacer un seguimiento constante a las actividades, los progresos obtenidos y los riesgos que se pueden presentar durante todo el proyecto. En la metodología de Kimball se incluye una fase de gestión del proyecto, en la cual se realizan las siguientes actividades:

1. Reunión de información del estado del proyecto.
2. Revisar el plan de proyecto.
3. Administrar el alcance del proyecto.
4. Controlar los cambios.
5. Desarrollar un plan de comunicaciones.

### Diseño de la Arquitectura Técnica

El área de arquitectura técnica cubre los procesos y herramientas que se aplican a los datos. En el área técnica existen distintos conjuntos que tienen distintos requerimientos, brindan sus propios servicios y componentes de almacenaje de datos.

En la figura 3.4 se puede observar la arquitectura típica de un Data Warehouse según Malinowski [26], que consta de las capas Fuentes de datos, Back-end, DW, OLAP, Front-end explicadas a continuación.

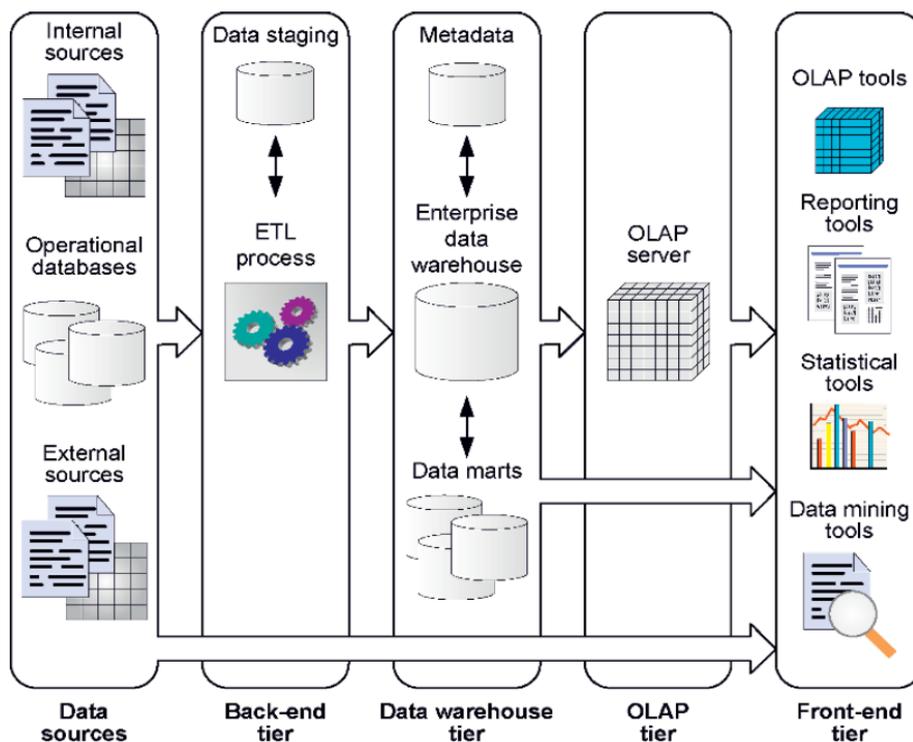


Figura 3.4: Arquitectura [27]

#### Fuentes de datos

Esta representa las diferentes fuentes de datos que alimentan los datos del DW. La fuente de datos puede estar en cualquier formato: archivo de texto plano, base de datos relacional, otros

tipos de base de datos, archivo Excel, etc. Las mismas pueden ser internas o externas a la organización.

### Capa back-end

Esta compuesta por herramientas de ETL de datos, utilizadas para alimentar el DW. A continuación se detallan los tres pasos de este proceso:

1. Extracción: recopilación de datos de diversas fuentes. Las mismas pueden ser bases de datos relacionales u otras fuentes con distintos formatos.
2. Transformación: conversión de los datos del formato de la fuente al formato del DW. Incluye tareas de depuración, filtrado, integración y agregación de los mismos.
3. Carga: carga del DW 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 de datos, antes de realizar la carga del DW. Esto es lo que se conoce como Data Staging.

### Capa DW

Esta compuesta por un DW corporativo y/o varios Data Marts, y un repositorio de metadata que almacena información sobre el DW y su contenido. Un DW corporativo es un DW centralizado que abarca todas las áreas funcionales y departamentos de una organización. Por otra parte, un Data Mart es un DW especializado, orientado hacia un área funcional particular o un grupo de usuarios en una organización.

### Capa OLAP

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:

1. Relational OLAP (ROLAP): almacena los datos en un motor relacional.
2. Multidimensional OLAP (MOLAP): trabaja sobre almacenamiento especializado.
3. Hybrid OLAP (HOLAP): combina ambas estrategias.

### Capa front-end

Esta compuesta por herramientas clientes, como por ejemplo herramientas OLAP, de estadísticas y generación de reportes, que permiten visualizar y analizar interactivamente el contenido del DW.

## **Selección de Productos e Instalación**

Durante la fase de selección de productos, Kimball propone hacer una investigación exhaustiva acerca de los productos que se encuentran disponibles en el mercado, y opcionalmente desarrollar prototipos para conocer la forma en la que productos satisfacen sus necesidades.

### **3.2.2. Modelado de Datos e Integración**

La ruta de datos en el Ciclo de Vida Dimensional del Negocio empieza con el diseño de un modelo dimensional como objetivo para enfrentarse a los requisitos del negocio, mientras consideramos

las realidades de datos subyacentes. El mundo Kimball es sinónimo del modelo dimensional donde los datos se dividen en medidas o dimensiones descriptivas. Los modelos dimensionales pueden estar instanciados en bases de datos relacionales, referidas como esquemas de estrella, o bases de datos multidimensionales, conocidas como los cubos OLAP. Independientemente de la plataforma, los modelos dimensionales intentan dirigirse a dos objetivos simultáneos: facilidad de uso desde la perspectiva de los usuarios y rapidez en la realización de la consulta.

El modelo dimensional se transforma en diseño físico. De esta manera, se aborda el diseño del sistema ETL y los desafíos para el desarrollo. El ciclo de vida describe subsistemas del proceso de extracción transformación. Estos subsistemas se agrupan en 4 operaciones mayores:

1. Extraer los datos de la fuente.
2. Llevar a cabo la limpieza y el ajuste de las transformaciones
3. Proporcionar los datos en la capa de presentación.
4. Dirigir el proceso ETL de mantenimiento y entorno.

## **Modelado Dimensional**

El modelado dimensional es una técnica de diseño que busca presentar la información en un marco estándar e intuitivo que permita un acceso de alto rendimiento. Este modelado se vale de los principios de la disciplina que emplea el modelo relacional con algunas importantes restricciones. El modelado dimensional es esencialmente útil para resumir y organizar los datos y la presentación de información para soportar el análisis de la misma. Existen algunos conceptos básicos para comprender la filosofía de este tipo de modelado: Temas, Medidas, Dimensiones y Jerarquías.

### Temas

Un tema como es una cuestión de interés de una función empresarial. Los temas en conjunto constituyen el ámbito de implementación del DW.

### Medidas

Para especificar los temas se deben identificar las medidas. Una medida o indicador es un cuantificador del desempeño de un ítem o una actividad del negocio. La información que brinda una medida es usada por los usuarios en sus consultas para evaluar el desempeño de un tema.

### Dimensiones

El DW organiza un gran conjunto de datos operacionales mediante múltiples dimensiones. Una dimensión es una colección de miembros o entidades del mismo tipo y constituye un calificador conceptual que provee el contexto o significado para una medida. Los miembros de una dimensión pueden estar organizados en una o más jerarquías. Una jerarquía es un conjunto de miembros de una dimensión, los cuales se definen por su posición relativa con respecto a los otros miembros de la misma dimensión, y forman en su totalidad una estructura de árbol. Partiendo de la raíz del árbol, los miembros son progresivamente más detallados hasta llegar a las hojas, donde obtenemos el mayor nivel de detalle.

### Jerarquías

Los miembros de una dimensión pueden estar organizados en una o más jerarquías. Una jerarquía es un conjunto de miembros de una dimensión, los cuales se definen por su posición relativa con respecto a los otros miembros de la misma dimensión, y forman en su totalidad

una estructura de árbol. Partiendo de la raíz del árbol, los miembros son progresivamente más detallados hasta llegar a las hojas, donde obtenemos el mayor nivel de detalle.

### Cubos

La realidad se modela como un conjunto de cubos. Cada cubo, esta formado por:

1. Un conjunto de Dimensiones organizadas en jerarquías.
2. Un conjunto de Medidas asociadas a cada Coordenada.

### Operaciones OLAP

Es posible moverse en las jerarquías de las dimensiones y observar de esa forma, diferentes visiones de las medidas. Para esto último se dispone de un conjunto de operaciones sobre MD:

1. *Slice* (Seleccionar dimensiones): Se define un subconjunto del hipercubo especificando sobre qué dimensiones interesa analizar qué medida.
2. *Dice* (Filtrar): Se fijan valores para algunas dimensiones.
3. *Pivot* (Rotación): Selecciona el orden de visualización de las dimensiones.
4. *Drill-up* y *drill-down*: Movimientos en la jerarquía de una dimensión. Son agrupamientos y des-agrupamientos.
5. *Roll-up* (Consolidación): Calcula las medidas en función de agrupamientos.
6. *Drill-across*: Relaciona dos cubos.
7. *Drill-through*: Accede desde un dato del cubo a datos en el DW de donde se deriva el cubo.

El modelado dimensional es un proceso dinámico y altamente iterativo. Comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados y descritos en la tarea anterior. El proceso iterativo consiste en cuatro pasos:

1. Elegir el proceso de negocio: que consiste en, elegir el área a modelizar. Esta es una decisión de la dirección, y depende fundamentalmente del análisis de requerimientos y de los temas analíticos anotados en la etapa anterior.
2. Establecer el nivel de granularidad: La granularidad significa especificar el nivel de detalle. La elección de la granularidad depende de los requerimientos del negocio y lo que es posible a partir de los datos actuales. La sugerencia general es comenzar a diseñar el DW al mayor nivel de detalle posible, ya que se podrían realizar agrupamientos posteriores, al nivel deseado.
3. Elegir las dimensiones: Las dimensiones surgen naturalmente de las discusiones del equipo, y facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensiones. Las tablas de dimensiones tienen un conjunto de atributos (generalmente textuales) que brindan una perspectiva o forma de análisis sobre una medida en una tabla hechos. Una forma de identificar las tablas de dimensiones es que sus atributos son posibles candidatos para ser encabezado en los informes, tablas pivote, cubos, o cualquier forma de visualización, unidimensional o multidimensional.

4. Identificar medidas y las tablas de hechos: Este paso, consiste en identificar las medidas que surgen de los procesos de negocios. Una medida es un atributo (campo) de una tabla que se desea analizar, sumando o agrupando sus datos y usando los criterios de corte conocidos como dimensiones. Las medidas habitualmente se vinculan con el nivel de granularidad del punto 2, y se encuentran en tablas que denominamos tablas de hechos. Cada tabla de hechos tiene como atributos una o más medidas de un proceso organizacional, de acuerdo a los requerimientos. Un registro contiene una medida expresada en números, como ser cantidad, tiempo, dinero, etc., sobre la cual se desea realizar una operación de agregación (promedio, conteo, suma, etc.) en función de una o más dimensiones. La granularidad, en este punto, es el nivel de detalle que posee cada registro de una tabla de hechos.

## Diseño Conceptual

El propósito de esta fase es representar los requerimientos especificados en la etapa de Definición de Requerimientos del Negocio 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.

El modelo *Conceptual MultiDimensional Model* (CMDM) [9], propuesto como tesis de Maestría por el grupo CSI del InCo es utilizado para la generación de un modelo conceptual y se basa en el Modelo Dimensional presentado.

El mismo define un modelo que permite la especificación detallada de una base de datos multidimensional. Esta especificación se construye mediante un lenguaje gráfico que permite describir las estructuras de datos y algunas restricciones de integridad, y un lenguaje de restricciones de integridad que permite dar una descripción precisa de las relaciones entre los datos.

Se presentan las 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.

*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. Cabe señalar que las medidas también se representan como dimensiones.

*Relaciones dimensionales*: representan cruzamientos entre dimensiones. Se tiene un elemento en el conjunto relación si y solo 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.

## Diseño Lógico

La etapa de diseño lógico toma como entrada un modelo conceptual y genera un modelo lógico relacional o multidimensional.

La dificultad principal es encontrar un modelo lógico que satisfaga no sólo los requerimientos funcionales de información, sino también requerimientos de performance en la realización de consultas complejas de análisis de datos.

Existen distintos tipos de almacenamiento según la base de datos que se utilizará:

*Sistemas MOLAP*

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Por el contrario, la arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales. Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis.

El sistema MOLAP utiliza una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos precompilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan precalculados y almacenados previamente.

### Sistemas ROLAP

La arquitectura ROLAP, accede a los datos almacenados en un DW para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

La arquitectura ROLAP es capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del DW, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples.

### Sistemas HOLAP

Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

La herramienta OLAP utilizada para la implementación de este proyecto, PBI, requiere que el DW este 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 DW ROLAP se utiliza el modelo multidimensional, que esta 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, almacenan las medidas numéricas del negocio y contienen las claves foráneas a las tablas dimensionales

relacionadas, resultantes de la intersección de las dimensiones; y las tablas de dimensión, donde se guardan las descripciones de las dimensiones.

A la hora de modelar el DW, hay que decidir cuál es el esquema más apropiado para obtener los resultados que queremos conseguir.

*Star Schema* En el Star Schema hay una única tabla central, la tabla de hechos, que contiene todas las medidas y una tabla adicional por cada una de las perspectivas desde las que queremos analizar dicha información, es decir por cada una de las dimensiones.

*Snowflake Schema* La otra alternativa de modelado es la utilización del Snowflake Schema. Esta es una estructura más compleja que el esquema en estrella. La diferencia es que algunas de las dimensiones no están relacionadas directamente con la tabla de hechos, sino que se relacionan con ella a través de otras dimensiones. En este caso también tenemos una tabla de hechos, situada en el centro, que contiene todas las medidas y una o varias tablas adicionales, con un mayor nivel de normalización.

El Star Schema, aunque ocupa más espacio en disco (dato cada vez menos significativo), es más simple de entender por el usuario y ofrece un mejor rendimiento a la hora de ser consultado.

*Star Cluster* Para lograr un mejor balance entre rendimiento, almacenamiento y mantenimiento se puede utilizar un modelo híbrido llamado Star Cluster. El mismo, selectivamente, separa los fragmentos de jerarquías compartidos entre diferentes dimensiones, mientras que el resto de las jerarquías se desnormalizan. Con este modelo se logra el mínimo número de tablas y a la vez se evita el solapamiento entre dimensiones.

## **Diseño Físico**

En el diseño físico se implementa el modelo lógico en el manejador de bases de datos elegido, teniendo en cuenta técnicas de comprensión, índices, selección de vistas e índices, optimización de consultas, carga y mantenimiento, los cuales se hace necesario para acelerar los tiempos de respuesta de las consultas complejas.

## **Diseño e Implementación de ETL**

El subsistema de ETL es la base sobre la cual se alimenta el sistema de BI. Si se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar (grabar) la información en el DW en un formato acorde para la utilización por parte de las herramientas de análisis.

### **3.2.3. Reportes y Analítica**

En esta tarea se proporciona, a una gran comunidad de usuarios una forma más estructurada y por lo tanto, más fácil, de acceder al DW. Se proporciona este acceso estructurado a través de lo que llamamos, aplicaciones de BI. Las aplicaciones de BI son la cara visible de la inteligencia de negocios: los informes y aplicaciones de análisis proporcionan información útil a los usuarios. Las aplicaciones de BI incluyen un amplio espectro de tipos de informes y herramientas de análisis, que van desde informes simples de formato fijo, a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio. Kimball divide a estas aplicaciones en dos categorías basadas en el nivel de sofisticación, y les llama:

1. Informes estándar: son informes relativamente simples, de formato predefinido, y parámetros de consulta fijos, proporcionan a los usuarios un conjunto básico de información acerca de lo que está sucediendo en un área determinada de la empresa y se utilizan día a día.
2. Aplicaciones analíticas: Son más complejas que los informes estándar. Estas aplicaciones pueden incluir algoritmos y modelos de minería de datos, que ayudan a identificar oportunidades o cuestiones subyacentes en los datos, y el usuario puede pedir cambios en los sistemas transaccionales basándose en los conocimientos obtenidos del uso de la aplicación de BI. Algunas aplicaciones analíticas comunes incluyen:
  - a) Análisis de la eficacia de la promociones.
  - b) Análisis de rutas de acceso en un sitio Web.
  - c) Análisis de afinidad de programas.
  - d) Planificación del espacio en espacios comerciales.
  - e) Detección de fraudes.
  - f) Administración y manejo de categorías de productos.

### **Diseño y Desarrollo de Herramientas de BI**

Mientras que algunos miembros de proyectos están inmersos en la tecnología y los datos, otros se centran en identificar y construir un amplio rango de aplicaciones BI, incluyendo informes estandarizados, consultas parametrizadas, tableros, cuadros de mando, modelos analíticos, aplicaciones de extracción de datos junto con las interfaces de navegación asociadas.

#### **3.2.4. Implantación, Mantenimiento y Crecimiento**

Las tres rutas en el Ciclo de Vida Dimensional del Negocio convergen en la implementación, reuniendo la tecnología, los datos y las aplicaciones BI. La implementación iterada entra en fase de mantenimiento, mientras que el crecimiento vuelve atrás hacia la planificación del proyecto para la próxima iteración del sistema BI/DW.

Para administrar el entorno del sistema de BI existente es importante enfocarse en los usuarios de negocio, los cuales son el motivo de su existencia, además de gestionar adecuadamente las operaciones del sistema de BI, medir y proyectar su éxito y comunicarse constantemente con los usuarios para establecer un flujo de retroalimentación, en esto consiste el Mantenimiento. Finalmente, es importante sentar las bases para el crecimiento y evolución del sistema de BI en donde el aspecto clave es manejar el crecimiento y evolución de forma iterativa utilizando el Ciclo de Vida propuesto, y establecer las oportunidades de crecimiento y evolución en orden por nivel prioridad.



# Capítulo 4

## Análisis

En este capítulo se engloban todos los aspectos relacionados a la planificación del proyecto y la etapa de análisis del sistema de BI.

### 4.1. Planificación

El plan de trabajo se enmarca dentro del enfoque presentado por Kimball en The Data Warehouse Toolkit [25] llamado Ciclo de Vida Dimensional del Negocio.

Se listan a continuación las etapas e hitos más relevantes del proceso:

1. Análisis y Especificación de Requerimientos.
2. Diseño Conceptual.
3. Diseño Lógico.
4. Diseño Físico.
5. Implementación.

Es importante destacar que, aunque se hayan seguido los lineamientos presentados, en la fase de *Análisis y Especificación de Requerimientos* se agrega una etapa de *Prototipado*.

Un prototipo es un primer modelo que sirve como representación o simulación del producto final y que nos permite verificar el diseño y confirmar que cuenta con las características específicas planteadas logrando los siguientes objetivos:

1. Aprender rápido y poder incluir las modificaciones necesarias en la fase de diseño. Esto nos permite aportar sobre nuevas funcionalidades e incluso hacer descartes cuando las propuestas no son viables.
2. Realizar pruebas que verifiquen y determinen como reaccionan los usuarios ante diferentes situaciones como la carga inicial, la actualización o las distintas consultas de las herramientas de análisis.
3. Testear la herramienta antes de comenzar el proceso de desarrollo. Los prototipos se analizan y se evalúan para detectar errores y posibles mejoras, confirmar que cuenta con

las características deseadas y asegurar que se puede adaptar los requerimientos de análisis.

En dicha etapa se presentan a los especialistas de la IM distintos prototipos de análisis y uso con las siguientes características:

1. Se utilizó un conjunto reducido de datos teniendo en cuenta que las distintas fuentes pudieran integrarse entre si según su dimensión temporal.
2. No se diseñaron ETLs complejos sino que se manipularon manualmente las fuentes previo a su carga en el DW.
3. Se generaron cubos sencillos, sin jerarquías ni un estudio detallado sobre la aditividad, para poder hacer análisis sobre las relaciones de las distintas fuentes de datos.
4. Se presentaron distintas gráficas y tablas en Saiku para poder visualizar la posible relación entre los datos.

## 4.2. Análisis

La primera etapa del proceso es el Análisis. En esta se trabaja con los especialistas del Servicio de Calidad de Aire para poder comprender y plasmar formalmente las funcionalidades y requisitos que se espera del sistema de BI.

La etapa de Especificación de Requerimientos se realiza mediante un enfoque combinado, utilizando paralelamente el enfoque orientado a las fuentes y el enfoque orientado al usuario, y luego integrando los resultados de ambas etapas para lograr requerimientos que sean de utilidad para los usuarios y que puedan ser realizados con los datos disponibles.

Como base para la identificación de requerimientos, se visualizan las principales actividades contaminantes, como la quema de combustible para el transporte y la industria; el uso de leña tanto en el área comercial como residencial para la calefacción y cocción de alimentos, la generación de energía eléctrica a partir de centrales térmicas y quema de residuos a cielo abierto, de las cuales se necesita realizar un análisis, y las fuentes de datos que se disponen.

### 4.2.1. Enfoque orientado a las fuentes

Como primer paso, se investigan todas las fuentes de datos internas, abiertas al público o no, con las que se dispone desde el Servicio de Calidad de Aire, desde bases de datos disponibles hasta archivos con datos que puedan ser de utilidad.

Luego, se investigan fuentes de datos externas que puedan ser cruzadas con las obtenidas del Servicio de Calidad de Aire para darle un mayor valor agregado al análisis y lograr visualizar las mediciones de contaminación junto con otros indicadores, para identificar relaciones entre los distintos tipos de datos. Para dicha búsqueda se utiliza el Catálogo de Datos Abiertos de la Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento (AGESIC) [8] donde se encuentra una gran variedad de datos y tipos de datos.

A continuación se describe cada fuente de datos analizada:

#### **Barrios de Montevideo**

Shapefile <sup>1</sup> de polígonos con los barrios de Montevideo. En la figura 4.1 se presenta un mapa gráfico de la ciudad de Montevideo diferenciando los distintos polígonos de los barrios en colores y en la figura 4.2 se puede visualizar la información tabular del archivo Shapefile donde se muestra el área ocupada, nombre, número y código.

Los datos de los barrios de Montevideo son mantenidos en el Servicio de Geomática. Los mismos son provistos por el Instituto Nacional de Estadística (INE).

#### **Centros Comunales Zonales (CCZ) de Montevideo**

Shapefile de polígonos con los CCZ de Montevideo. En la figura 4.3 se presenta un mapa gráfico de la ciudad de Montevideo diferenciando los distintos polígonos de los CCZ en colores y en la figura 4.4 se puede visualizar la información tabular del archivo Shapefile donde se muestra un identificador autogenerado y el nombre de cada CCZ.

---

<sup>1</sup>Shapefile: Formato de almacenamiento de datos vectorial que guarda la localización, la forma y los atributos de entidades geográficas.

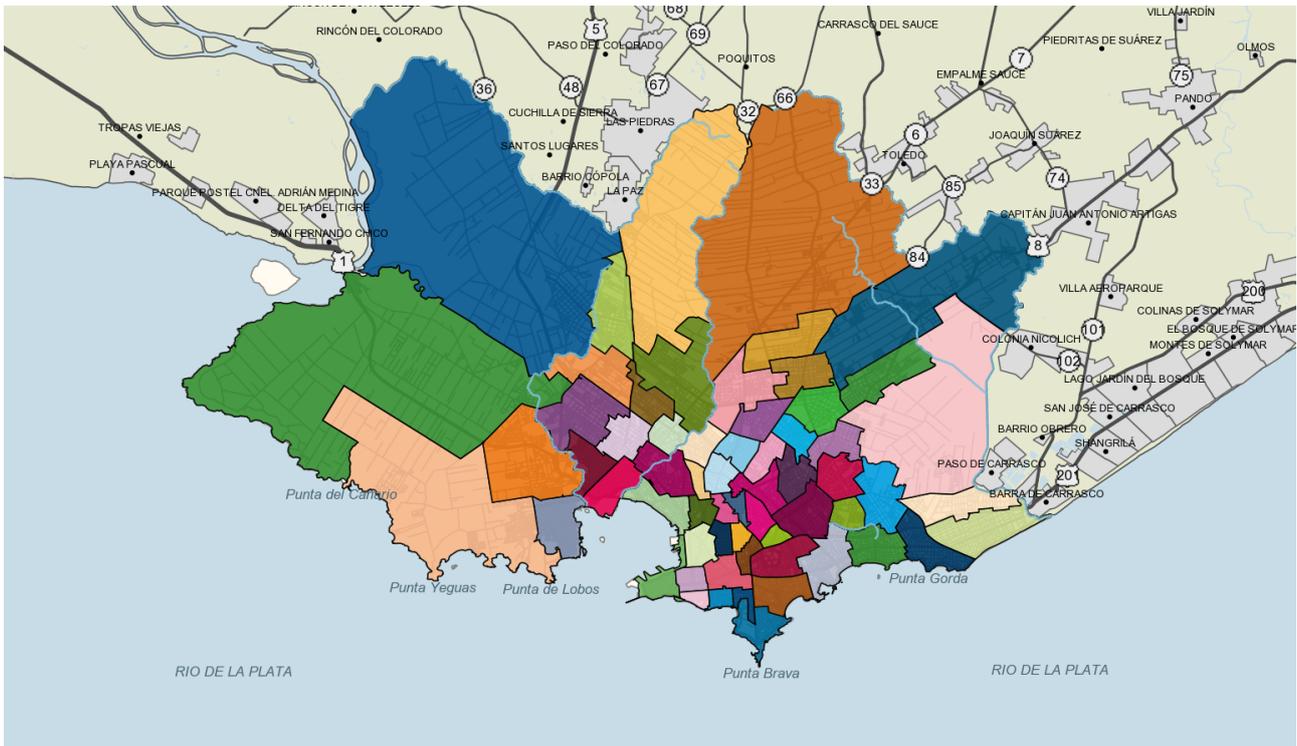


Figura 4.1: Mapa Barrios de Montevideo

AREA_KM,N,1	NOMBARR,C,25	NROB	CODB
2.111	CIUDAD VIEJA	1	CJ
1.297	CENTRO	2	CE
0.694	BARRIO SUR	3	SU
2.279	CORDON	4	CD
0.798	PALERMO	5	PL
0.766	PARQUE RODO	6	PR
2.73	PUNTA CARRETAS	7	PC
4.155	BUCEO	9	BU
3.375	PQUE. BATLLE, V. DOLORES	10	PB
3.501	MALVIN	11	MA
1.714	MALVIN NORTE	12	MN
3.816	PUNTA GORDA	13	PG
6.168	CARRASCO	14	CR
4.643	CARRASCO NORTE	15	CN
32.162	BARRIADOS DE CARRASCO	16	BC
3.126	MARO DE LAS, PARQUE GUARANI	17	MO
2.69	FLOR DE MARO DE LAS	18	FL

Figura 4.2: Datos Barrios de Montevideo

Los datos de los CCZ de Montevideo son mantenidos en el Servicio de Geomática de la IM.

### Descripción de campos de medidas históricas

Archivo de formato texto que sirve de meta-información [13] para comprender los archivos de medidas históricas de calidad de aire. En la figura 4.5 se puede visualizar el formato y significado de cada campo de los archivos de medidas históricas calidad de aire.

Los datos son generados por el Servicio de Calidad de Aire de la IM.

En la misma fuente de datos se especifican los archivos de medidas históricas que se disponibi-

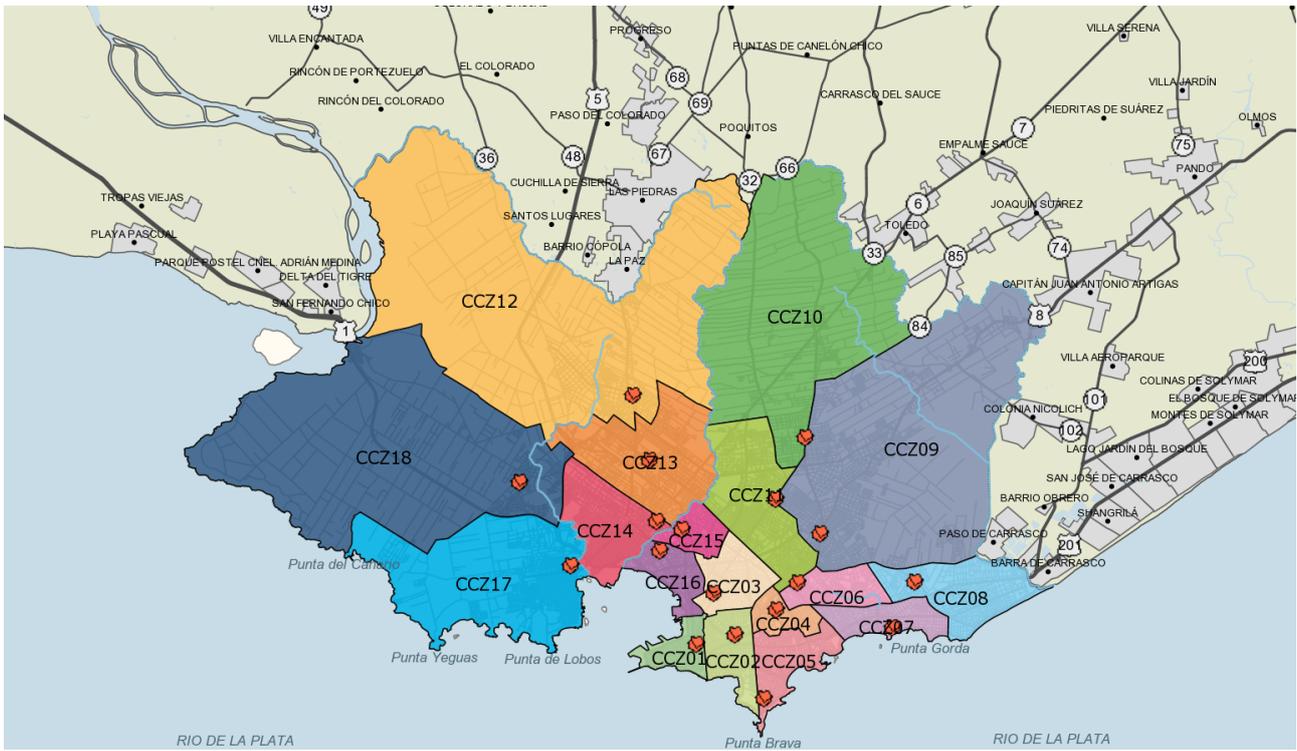


Figura 4.3: Mapa CCZ de Montevideo

GID,N,19,0	ZONA_LEGAL,C,5
8647694	CCZ12
8647695	CCZ10
8647696	CCZ17
8647697	CCZ14
8647698	CCZ09
8647699	CCZ08
8647700	CCZ07
8647701	CCZ13
8647702	CCZ11
8647703	CCZ15
8647704	CCZ02
8647705	CCZ03
8647706	CCZ04
8647707	CCZ05
8647708	CCZ06
8647709	CCZ18
8647710	CCZ01
8647711	CCZ16

Figura 4.4: Datos CCZ de Montevideo

lizan, diferenciando las estaciones automáticas de las manuales, como se puede visualizar en la figura 4.6 y la figura 4.7. Para cada fuente se indica su nombre de archivo, estación de medición y período de cada conjunto de mediciones.

### Medidas históricas calidad de aire

Archivos en formato *Comma-separated values* (CSV) con medidas que van desde el 2005 hasta

Descripcion de campos de los archivos historicos	
POLLUTANT_ID:	Identificacion de contaminante. Estos identificadores se describen en el archivo contaminantes.csv
POLLUTANT_AVERAGING:	Periodo de toma de muestra en horas
DATE:	Fecha y hora que fue tomado el dato
POLLUTANT_VALUE:	Valor del parametro
POLLUTANT_UNIT:	Unidades en que esta expresado el contaminante
STATION_ID:	Denominacion de la zona que caracteriza la estacion.
X:	Localizacion. Coordenada X de la estacion (UTM- Zona 21)
Y:	Localizacion. Coordenada Y de la estacion (UTM- Zona 21)
ID_ESTACIONES:	Identificacion de la estacion segun consta en el archivo estaciones.csv
METHOD_ID:	Identifica el metodo con el cual se midio el Parametro. Los metodos se describen en el archivo metodos.csv

Figura 4.5: Descripción medidas históricas calidad de aire [13]

Las estaciones automaticas con datos horarios se presentan los datos anuales.		
CV_pm10_14.csv	Ciudad vieja	2014
CV_pm2_15.csv	Ciudad vieja	2015
CV_pm2_16.csv	Ciudad vieja	2016
CV_pm2_17.csv	Ciudad vieja	2017
CV_pm2_18.csv	Ciudad vieja	2018
TRC_pm10_15.csv	Tres Cruces F	2014-2015
TRC_pm10_16.csv	Tres Cruces F	2016
TRC_pm10_17.csv	Tres Cruces F	2017
TRC_pm10_18.csv	Tres Cruces F	2018
TRC_NO2_15.csv	Tres Cruces M	2015
TRC_NO2_16.csv	Tres Cruces M	2016
TRC_NO2_17.csv	Tres Cruces M	2017
TRC_NO2_18.csv	Tres Cruces M	2018
CM_pm2_17.csv	Curva de Ma	2017
CM_pm2_18.csv	Curva de Ma	2018
CM_NO2_17.csv	Curva de Ma	2017
CM_NO2_18.csv	Curva de Ma	2018
CM_O3_17.csv	Curva de Ma	2017
CM_O3_18.csv	Curva de Ma	2018
CO_pm2_18.csv	Colon PM2.5	2018
CO_NO2_18.csv	Colon NO2	2018
CO_O3_18.csv	Colon O3	2018

Figura 4.6: Archivos de medidas históricas calidad de aire [13]

Las Estaciones manuales con datos cada 24/48 horas presentan en archivos por parametro		
hn.csv	Humo Negro	2004-2018
PM2.csv	Material part	2007-2013
PM10.csv	Material part	2007-2018
pts.csv	Particulado t	2004-2018
SO2.csv	SO2 manual	2012-2018

Figura 4.7: Archivos de medidas históricas calidad de aire [13]

el 2018 [18] y para cada una se detalla a qué contaminante se mide, las unidades, el valor y el método de medición. Así como otros datos de interés como se explica en el archivo de descripción de la metadata. En la figura 4.5 se pueden visualizar los campos y los valores descritos.

Los datos son generados por el Servicio de Calidad de Aire de la IM de las distintas estaciones de la red.

A	B	C	D	E	F	G	H	I	J	K
	pollutant_id	pollutant_av	date	pollutant_va	pollutant_un	station_id	X	Y	ID_estacion	method_id
1378	PM2	1	1/1/18 0:00	5	ug/m3	UYMVD_E8	570970	6149046	Colon	UYMVD_NO2
1379	PM2	1	1/1/18 1:00	5	ug/m3	UYMVD_E8	570970	6149046	Colon	UYMVD_NO2
1380	PM2	1	1/1/18 2:00	5	ug/m3	UYMVD_E8	570970	6149046	Colon	UYMVD_NO2
1381	PM2	1	1/1/18 3:00	5	ug/m3	UYMVD_E8	570970	6149046	Colon	UYMVD_NO2
1382	PM2	1	1/1/18 4:00	5	ug/m3	UYMVD_E8	570970	6149046	Colon	UYMVD_NO2
1383	PM2	1	1/1/18 5:00	5	ug/m3	UYMVD_E8	570970	6149046	Colon	UYMVD_NO2
1384	PM2	1	1/1/18 6:00	11	ug/m3	UYMVD_E8	570970	6149046	Colon	UYMVD_NO2

Figura 4.8: Medidas históricas calidad de aire [18]

### Medidas históricas de sensores de la plataforma IoT - PM10

Archivo en formato CSV que contiene las medidas históricas de los sensores de calidad del aire de la plataforma de Internet of Things (IoT) [16], esta plataforma cuenta con una estación de medición automática en tiempo real conectada a internet donde envía periódicamente los datos de medición. Los valores en tiempo real de estos sensores se pueden acceder a través de los servicios abiertos de la plataforma. En la figura 4.9 se puede visualizar las columnas identificador autogenerated, identificador de estación o nodo, contaminante medido, instante de tiempo de la medición y valor correspondiente.

Los datos son generados por el Servicio de Calidad de Aire de la IM de la estación automática que se encuentra en Tres Cruces.

_id	id_node	sensor	timestamp	valor_v
1	AQM60	PM10	2017-03...	59138
2	AQM60	PM10	2017-03...	57787
3	AQM60	PM10	2017-03...	54746
4	AQM60	PM10	2017-03...	55875
5	AQM60	PM10	2017-03...	57302

Figura 4.9: Medidas históricas de sensores de la plataforma IoT - PM10 [16]

### Lista de identificación de contaminantes

Archivo en formato CSV con una descripción de los tipos de contaminantes que se miden [17]. En la figura 4.10 se puede visualizar las columnas identificador autogenerated, código de contaminante y una descripción.

Los datos son generados por el Servicio de Calidad de Aire de la IM.

### Categorías de valores de medición por contaminante

Datos en formato tabular que clasifica las concentraciones de los diferentes contaminantes en seis categorías como se muestra en la figura 4.11.

Los datos son generados por el Servicio de Calidad de Aire de la IM.

### Métodos de medición

_id	CONTA...	EXPLICACION
1	PM2	Material particulado menor de 2.5 micras de diam...
2	PM10	Material particulado menor de 10 micras de diam...
3	HN	Humo Negro (black smoke)
4	PTS	Material particulado total
5	SO2	Dioxido de azufre
6	NO2	Dioxido de nitrogeno
7	O3	Ozono

Figura 4.10: Lista de Identificación de contaminantes [17]

Categorías de ICAire	VALOR ICAIRE	PTS ug/m2 *	PM10 ug/m3 *	PM2.5 ug/m3	Humo Negro ug/m3 *	SO2 ug/m3 *	NO2 ug/m3 **	O3 ug/m3 ***	CO mg/m3 ***	TRS ug/m3 *
MUY BUENA	Hasta 25	60	50	25	50	20	40	80	4.5	2.5
BUENA	26 a 50	100	75	32	75	50	75	100	7	5
ACEPTABLE	51 a 100	150	100	50	100	125	200	160	10	10
INADECUADA	101 a 200	375	150	75	150	365	500	240	15	20
MALA	201 a 300	563	225	100	225	550	1130	500	22	30
MUY MALA	> 300	> 564	> 226	> 100	> 225	> 550	>1130	> 500	>22	>30

Figura 4.11: Categorías de Contaminantes [1]

Archivo en formato CSV con una descripción de los métodos de medición utilizados [19]. En la figura 4.12 se puede visualizar las columnas identificador autogenerado, código del método y una descripción.

Los datos son generados por el Servicio de Calidad de Aire de la IM.

_id	METODO	EXPLICACION
1	UYMVD_PM2_b	Medida de material particulado menor de 2.5 micras de diametro (PM2.5) utilizando Nefelometri...
2	UYMVD_PM2_D	Medida de material particulado menor de 2.5 micras de diametro (PM2.5) utilizando Un equipo ...
3	UYMVD_PM10_2	Medida de material particulado menor de 10 micras de diametro (PM10) utilizando Nefelometria...
4	UYMVD_PM10_D	Medida de material particulado menor de 10 micras de diametro (PM10) utilizando equipo dicot...
5	UYMVD_PM10_H	Medida de PM10 utilizando un equipo de alto volumen para recoleccion de muestra
6	UYMVD_PTS	Medida de material particulado total (PTS) utilizando un equipo de alto volumen para recoleccio...
7	UYMVD_SO2_T	Medida de dioxido de azufre (SO2) en tren de monitoreo y posterior analisis con cromatografia i...
8	UYMVD_NO2_2	Medida de dioxido de nitrogeno (NO2), semiconductor sensible al gas
9	UYMVD_HN	Medida de Humo Negro (black smoke) utilizando reflectometria como principio de deteccion
10	UYMVD_O3	Medida de ozono (O3) semiconductor sensible al gas

Figura 4.12: Métodos de medición [19]

## Estaciones de medición

Archivo en formato CSV con una lista las estaciones de medición [15] de la Red de Monitoreo de la IM. En la figura 4.13 se puede visualizar las columnas identificador autogenerado, nombre de la estación, ubicación geográfica mediante latitud y longitud, y el código de la estación.

Los datos son generados por el Servicio de Calidad de Aire de la IM.

_id	ESTACION	X	Y	EXPLICACION
1	Ciudad Vieja 3	572796	6137122	E1
2	Ciudad Vieja 1	572831	6136774	E1
3	Ciudad Vieja 2	572452	6137044	E1
4	Centro	574385	6137072	E2
5	Tres Cruces	576888	6138953	E5

Figura 4.13: Estaciones de medición [15]

## ANCAP

Archivo en formato planilla con datos de mediciones y meteorológicos históricos de la estación de ANCAP. En la figura 4.14 se pueden visualizar las columnas fecha de medición y los valores medidos de CO, NO, NOx, NO2, PM2.5, SO2, TRS, radiación solar global, humedad relativa, temperatura externa, temperatura interna, dirección del viento y velocidad del viento.

Estos datos fueron provistos por ANCAP para el uso exclusivo que se enmarca dentro del Proyecto de Grado.

	ANCAP1 CO Station 1 mg/m3	ANCAP1 NO Station 1 microg/m3	ANCAP1 NOx Station 1 microg/m3	ANCAP1 NO2 Station 1 microg/m3	ANCAP1 PM 2.5 Station 1 microg/m3	ANCAP1 SO2 Station 1 microg/m3	ANCAP1 TRS Station 1 microg/m3	ANCAP1 Global Sun rad 1 W/m2	ANCAP1 Relative Humidity 1 %	ANCAP1 Temperature Extern 1 °C	ANCAP1 Temp Internal 1 °C	ANCAP1 Wind Direction 1 degres	ANCAP1 Wind Speed 1 m/s
12/31/16 22:00	2.00					0.00	1.43	0	64	26.1	19.6	80	2.0
12/31/16 23:00	1.92					0.01	1.26	0	66	25.6	19.6	62	1.4
1/1/17 0:00	1.77					0.01	1.00	0	69	25.3	19.4	69	1.0
1/1/17 1:00	1.40					0.61	1.77	0	73	25.6	19.4	46	1.3
1/1/17 2:00	0.76					0.09	1.01	0	78	25.6	19.4	36	1.6
1/1/17 3:00	0.67					0.12	0.98	0	80	25.6	19.3	13	1.1

Figura 4.14: ANCAP

## Volumen vehicular

Archivos en formato CSV del sistema de conteo vehicular que utiliza el Centro de Gestión de Movilidad (CGM) para la medición de los volúmenes vehiculares que circulan por las arterias que se encuentran centralizadas, utiliza analíticas de vídeo para la generación de los datos necesarios para la elaboración de planes y coordinaciones semaforicas. Los detectores miden el total de vehículos y su clasificación por largo de los mismos, por sentido de circulación, pudiendo discriminarse por carril dentro de cada sentido [11] [12].

El sistema genera un registro cada 5 minutos, las medidas que se explican a continuación corresponden a la suma de todos los vehículos de los 5 minutos anteriores a la fecha y hora del registro. También se publican datos particulares del sensor que permiten conocer su ubicación y entender el sentido de circulación que se está midiendo.

En la figura 4.15 se visualizan los siguientes datos:

1. volumen: Cantidad de vehículos que se contabilizó en los 5 minutos previos en un carril dado.
2. volumen hora: Estimación de la cantidad de vehículos/hora que representan los 5 minutos previos medidos.

id_detector	id_carril	fecha	hora	dsc_avenida	dsc_int_ante	dsc_int_sigui	latitud	longitud	volumen	volumen_hora
1	3	11/1/18	0:00:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	1	12
1	3	11/1/18	0:05:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	1	12
1	3	11/1/18	0:10:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	0:15:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	0:20:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	1	12
1	3	11/1/18	0:25:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	0:30:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	0:35:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	0:40:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	0:45:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	0:50:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	0:55:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0
1	3	11/1/18	1:00:00	Sarmiento	Requena	Rambla	-34.915369	-56.168358	0	0

Figura 4.15: Volumen Vehicular [11]

## Vías de tránsito

Shapefile de líneas con las vías de tránsito de Montevideo [20]. En la figura 4.16 se puede visualizar que contiene los polígonos de las vías de Montevideo.

La información de vías de tránsito es mantenida en el Servicio de Geomática, por la Unidad de Nomenclatura y Numeración.



Figura 4.16: Vías de tránsito [20]

## Encuesta continua de hogares

Microdatos, cuestionarios, diccionario de variables y notas de la Encuesta Continua de Hogares [14]. Dentro del conjunto de datos disponible se define analizar las condiciones de vida de las personas mediante los indicadores socioeconómicos y demográficos, desagregados por CCZ. En la figura 4.17 se puede visualizar un ejemplo de la fuente de datos donde se presentan las distintas variables que se utilizan para las mediciones.

Los datos de la encuesta continua de hogares son mantenidos por el INE.

_id	DESCRIPCIÓN DE LA VARIABLE	ECH 2017
4	HOGAR. IDENTIFICACIÓN	
5	AÑO DE LA ENCUESTA	ANIO
6	IDENTIFICACIÓN DEL CUESTIONARIO	NUMERO
7	DEPARTAMENTO	DPTO
8	NOMBRE DEL DEPARTAMENTO	NOMDPTO
9	SECCIÓN CENSAL	SECC
10	SEGMENTO CENSAL	SEGM
11	LOCALIDAD AGRUPADA	LOCAGR
12	NOMBRE DE LA LOCALIDAD AGRUPADA	NOM_LOCAGR
13	CENTRO COMUNAL ZONAL	CCZ
14	BARRIO	BARRIO
15	NOMBRE DEL BARRIO	NOMBARRIO
16	MES DE LA ENCUESTA	MES
17	ESTRATO	ESTRED13

Figura 4.17: Encuesta continua de hogares [14]

### Esquema de las fuentes de datos

En la figura 4.18 se presenta un modelo conceptual de las fuentes de datos donde se especifican las distintas entidades y sus atributos. A su vez, como se puede reflejar, se muestran las relaciones entre ellos indicando los distintos tipos de asociación como uno a uno, uno a muchos o mucho a muchos.

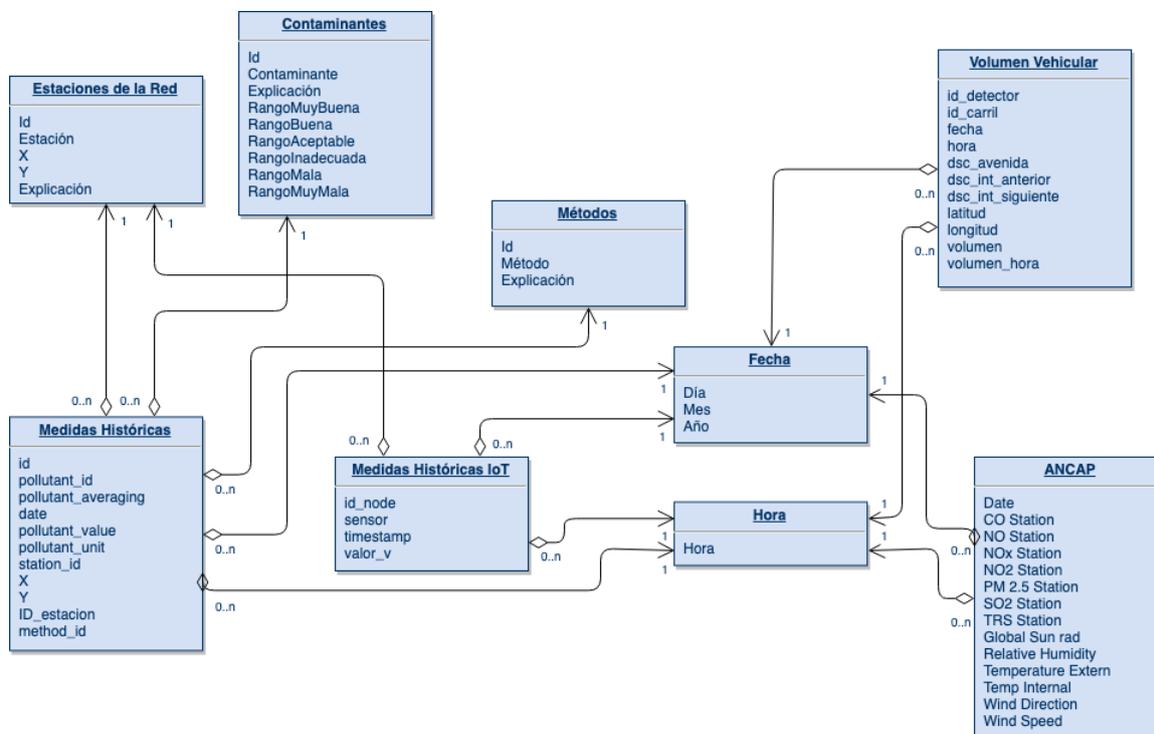


Figura 4.18: Esquema conceptual de las fuentes

## 4.2.2. Enfoque orientado al usuario

Para identificar las necesidades que los usuarios del sistema de BI quieren o pueden atacar, se realizan sucesivas reuniones presenciales donde se discuten y ponen en la mesa ideas o solicitudes de análisis de datos o situaciones puntuales que han surgido en el Servicio de Calidad de Aire en los últimos años.

Dentro de los puntos que se lograron identificar se encuentran:

1. Automatizar la realización de la sección Resultados de Red de Monitoreo del Informe Anual.
2. Relacionar la combustión realizada en la quema de leña o carbón al hacer asados con los días festivos, por ejemplo, un 1ro de mayo.
3. Investigar en qué medida varía la emisión de contaminantes en los días que la Central Batlle se encuentra encendida generando energía.
4. Investigar la relación entre distintos contaminantes, esto es, cuánto influye en la emisión de un contaminante el aumento en la emisión de otro. Por ejemplo, cuánto influye un aumento de emisión de PM10 en la presencia de Humo Negro.

## 4.2.3. Estudio de factibilidad

Con las fuentes externas e internas disponibles se realizan distintos prototipos con cruzamientos entre ellas para identificar, junto a los especialistas de la IM, cuáles son los relacionamientos más relevantes que pueden convertirse en requerimientos para un posterior análisis y desarrollo de los mismos.

Los prototipos fueron plasmados en las herramientas de análisis de datos con distintos ejemplos reales de cómo se visualizan los datos ingresados en el sistema.

Esta forma de trabajo permite obtener feedback de los usuarios en etapas tempranas del desarrollo del sistema, lo que brinda la posibilidad de realizar cambios o ajustar tempranamente los requerimientos.

A continuación, en el cuadro 4.1, se describen los distintos cruzamientos que se realizaron pero que no fueron tomados en cuenta para un posterior desarrollo.

Tomando en cuenta los cruzamientos potenciales descartados, las fuentes de datos descartadas son:

1. Vías de tránsito
2. Encuesta continua de hogares

Cruzamiento	Motivo
Evolución en el tiempo de la contaminación del aire relacionándolo con la densidad de vehículos en las calles y los tipos de caminería.	Aunque el estudio de en qué vías se presenta una mayor contaminación según un mayor volumen vehicular es muy prometedor, sucede que alrededor de las cámaras vehiculares y de las estaciones de monitoreo de la red, los tipos de caminería son iguales, imposibilitando un análisis por tipo de caminería, por ejemplo, logrando una comparación entre pavimento y caminos de tierra.
Evolución en el tiempo de la contaminación del aire relacionándolo con la afectación a las personas que residen en los hogares de los distintos barrios de Montevideo.	Teniendo claro que el objetivo último y más importante de la Red de Monitoreo es poder generar conocimiento que sea utilizado para brindar a los ciudadanos de Montevideo una mejor calidad de aire, se entiende que el análisis de distintas dimensiones relacionadas a la información brindada por la encuesta excede el alcance del proyecto y se prefiere priorizar otros requerimientos.

Cuadro 4.1: Cruzamientos potenciales descartados

#### 4.2.4. Especificación de requerimientos funcionales

Basados en ambos enfoques y en el estudio de factibilidad se presentan en el cuadro 4.2 los requerimientos funcionales acordados con los especialistas de la IM según las fuentes de datos disponibles y las necesidades identificadas. Primeramente se especifican los requerimientos funcionales de análisis.

Luego, en cuadro 4.3, se describen los requerimientos funcionales de uso.

ID	Requerimiento Funcional	Descripción
RFA1	RFA1 - Relación Volumen Vehicular y Contaminación	Analizar la evolución en el tiempo de la contaminación del aire relacionándolo con la densidad de vehículos en las calles.
RFA2	RFA2 - Relación Actividad Industrial y Contaminación	Analizar la evolución en el tiempo de la contaminación del aire que generan las industrias a su alrededor teniendo en cuenta su nivel de actividad.
RFA3	RFA3 - Relación Días Especiales y Contaminación	Analizar la evolución en el tiempo de la contaminación del aire generados por la combustión, específicamente la quema de leña generada por los asados, en días especiales como los festivos, feriados, fechas patria y días con eventos destacables.

Cuadro 4.2: Requerimientos Funcionales de Análisis

ID	Requerimiento Funcional	Descripción
RFU1	RFU1 - Aplicación web de carga y actualización	Permitir a usuarios finales la realización de la carga y actualización del DW mediante un aplicativo que resulte amigable y no requiera de amplio conocimiento técnico.
RFU2	RFU2 - Generación de reporte de BI de sección del Informe Anual	Automatizar la realización de un reporte de BI de la sección Resultados de Red de Monitoreo del Informe Anual, donde detallan las características de cada estación y se muestra un resumen de los resultados expresados utilizando el ICAire.
RFU3	RFU3 - Generación de un dashboard para análisis	Generar un dashboard con gráficas e indicadores pre-definidos para un posterior análisis

Cuadro 4.3: Requerimientos Funcionales de Uso

A continuación se describen en detalle los requerimientos funcionales de análisis.

## **RFA1 - Relación Volumen Vehicular y Contaminación**

*Analizar la evolución en el tiempo de la contaminación del aire relacionándolo con la densidad de vehículos en las calles.*

Este requerimiento intenta realizar un análisis sobre uno de los principales focos de contaminación de aire como es el transporte. El objetivo del mismo es poder visualizar si a mayor densidad de autos existe una mayor contaminación.

El mismo surge del cruzamiento de la fuente de datos de medidas históricas de los sensores de calidad del aire de la plataforma de IoT [16] interna al Servicio de Calidad de Aire y una fuente de datos externa obtenida del catálogo de datos abiertos de AGESIC [8].

Para relacionar el volumen vehicular en las calles de Montevideo con la contaminación, es necesario investigar en paralelo, con datos horarios, la emisión de los contaminantes PM<sub>2</sub>, PM<sub>10</sub> o PTS, junto con el volumen de vehículos que se encuentran alrededor de las estaciones de la red en las que se realizan las mediciones de dichos contaminantes.

Para tomar datos de casos reales para realizar la investigación, se define junto a los especialistas del Servicio de Calidad de Aire un rango de distancia entre una estación de la red y las cámaras de conteo vehicular. El valor definido es *150m* de radio pero es modificable al momento de la carga.

En la etapa de carga de datos de contaminación y volumen vehicular se filtrarán y no se relacionarán los datos obtenidos de estaciones de monitoreo y cámaras de conteo vehicular que se encuentren a un radio de más de *150m*.

Analizando las fuentes de datos con las que se cuenta en el momento de la elaboración del informe, se puede observar que:

Para el conteo vehicular:

1. Se obtuvieron datos del conteo vehicular del 2017 y 2018, por lo tanto se podrá relacionar con las mediciones de contaminación para ese período.
2. Aunque se obtuvo un número considerable de cámaras de conteo vehicular, se podrán relacionar sólo las que se encuentren en un radio de *150m* con alguna estación de la red.

Para las mediciones de contaminación:

1. Existen datos de PM<sub>2</sub>, PM<sub>10</sub> y PTS semanales en varias estaciones de monitoreo, como Ciudad Vieja y Portones, pero no horarios o por minutos como es requerido.
2. Existen datos horarios de PM<sub>10</sub> en Ciudad Vieja y Tres Cruces, pero no se encuentran en el mismo período de medición que el conteo vehicular.

Por lo tanto, al momento de la elaboración de este informe será posible relacionar:

1. Mediciones de contaminación de PM<sub>10</sub>, obtenidas en la estación de la red que se encuentra en Tres Cruces para el período 2017.
2. Datos de conteo vehicular de las cámaras que se encuentran a menos de *150m* de la estación de Tres Cruces para el período 2017. En la figura 4.19 se pueden visualizar las cámaras vehiculares en color verde y las estaciones de la red en color azul para analizar la cercanía entre ellas.

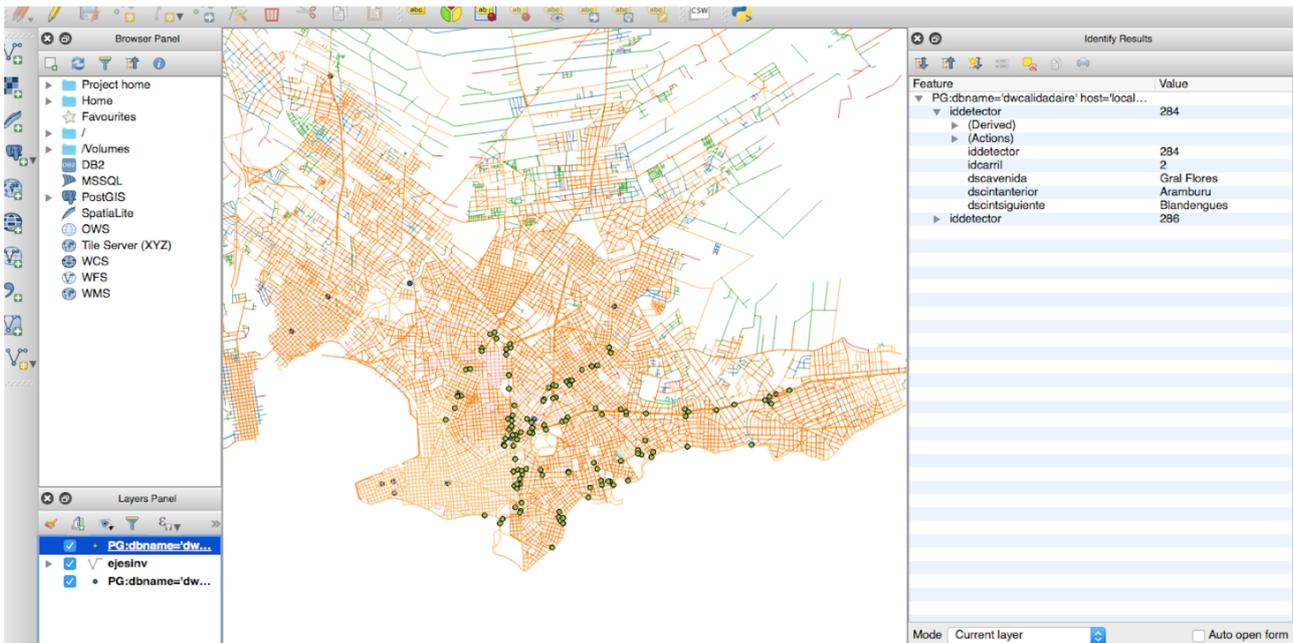


Figura 4.19: Radio de 150m alrededor de las estaciones de la red y cámaras vehiculares

En el prototipo desarrollado para mostrar la relación entre ambos datos, se realizó una carga inicial manual y se generó un cubo de ejemplo para ser visualizado en Saiku, como se puede ver en la figura 4.20, que lograra mostrar básicamente la relación entre ambos datos.

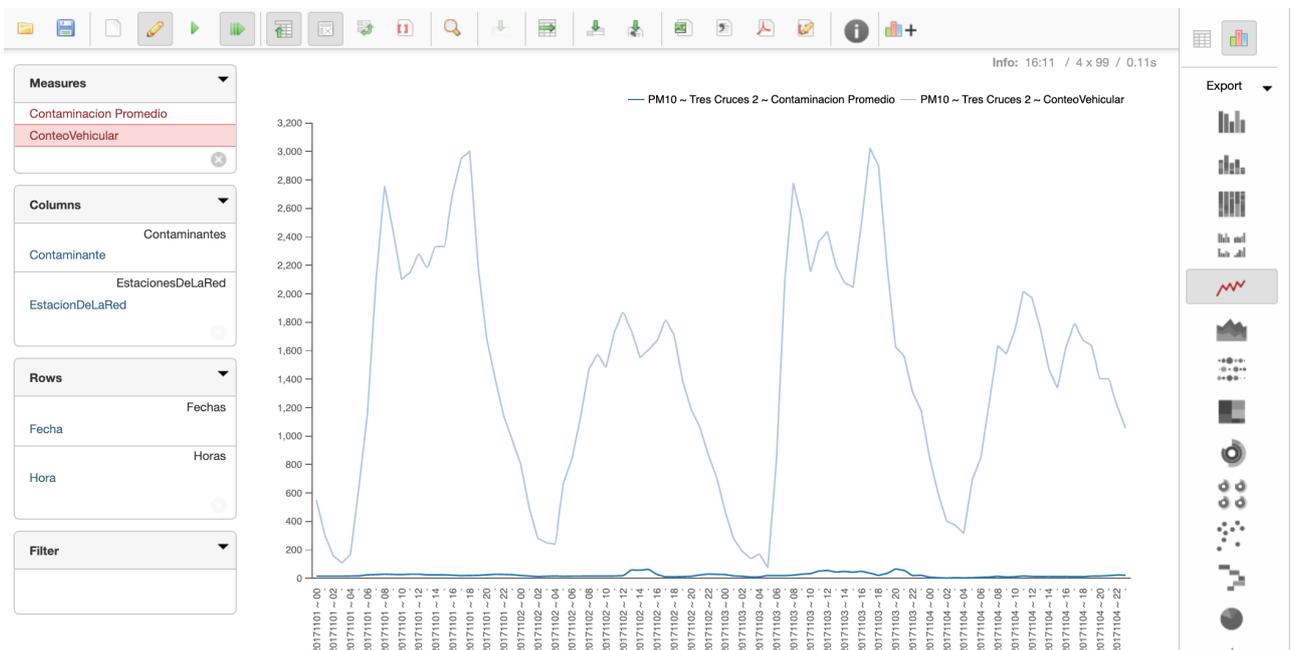


Figura 4.20: Prototipo - Relación Contaminación - Volumen Vehicular

Algunos otros puntos interesantes para estudiar serían las estaciones de monitoreo en el Palacio Legislativo y en Portones de Carrasco, pero del conteo vehicular se tienen datos sólo de 2017 y necesitaríamos datos horarios de los contaminantes en dichas estaciones de monitoreo. UTE tiene estos datos horarios, pero no se logró obtenerlos al momento de realizar el informe. En la estación de ANCAP se tienen datos horarios pero no se tiene conteo vehicular alrededor.

## RFA2 - Relación Actividad Industrial y Contaminación

*Analizar la evolución en el tiempo de la contaminación del aire que generan las industrias a su alrededor teniendo en cuenta su nivel de actividad.*

El requerimiento pretende realizar un análisis sobre otro de los principales focos de contaminación de aire como son los procesos industriales. El objetivo es poder entender si a mayor actividad se encuentra una mayor contaminación en el aire.

El mismo surge del cruzamiento de la fuente de datos de medidas históricas de la Red de Monitoreo [18] interna al Servicio de Calidad de Aire e industrias identificadas con los especialistas que son ingresadas de forma manual.

Para relacionar la actividad industrial con la contaminación, es necesario investigar en paralelo, con datos diarios, la emisión de los contaminantes PM<sub>2</sub>, PM<sub>10</sub> o PTS, junto con el nivel de actividad de las industrias que se encuentran alrededor de las estaciones de la red en las que se realizan las mediciones de dichos contaminantes.

Por lo tanto, para tomar los datos para realizar la investigación, se define junto a los especialistas del Servicio de Calidad de Aire un rango de distancia entre una estación de la red y una industria. El valor definido es  $3km$  de radio pero es modificable al momento de la carga.

En la etapa de carga de datos de contaminación y nivel de actividad se filtrarán y no se relacionarán los datos obtenidos de estaciones de monitoreo e industrias que se encuentren a un radio de más de  $3km$ .

Analizando las fuentes de datos con las que se cuenta en el momento de la elaboración del informe, se puede observar que:

Para el nivel de actividad:

1. Dentro de las industrias a investigar, de la cual se logró conseguir datos actuales, se encuentra la Central Batlle mediante la información provista por Administración del Mercado Eléctrico (ADME) [7].
2. Se obtuvieron datos de nivel de actividad desde 2015 por lo tanto se podrá relacionar con las mediciones de contaminación para ese período.
3. Aunque existen un número considerable de estaciones de la red, se podrán relacionar sólo las que se encuentren en un radio de  $3km$  con la Central Batlle. Para el momento que se realizó el informe, la estación de la red en Ciudad Vieja es la única estación en el rango de la cual se disponían datos para cruzar.

Para las mediciones de contaminación:

1. Existen datos de múltiples contaminantes con periodicidad diaria y semanal.

Por lo tanto, al momento de la elaboración de este informe será posible relacionar:

1. Mediciones de múltiples contaminantes obtenidas en la estación de la red que se encuentra en Ciudad Vieja para el período desde 2005 hasta la fecha.
2. Datos de la actividad industrial de la Central Batlle desde 2015. En la figura 4.21 se puede visualizar la ubicación de la Central Batlle en color amarillo y las estaciones de la red en color rojo para analizar la cercanía entre ellas.

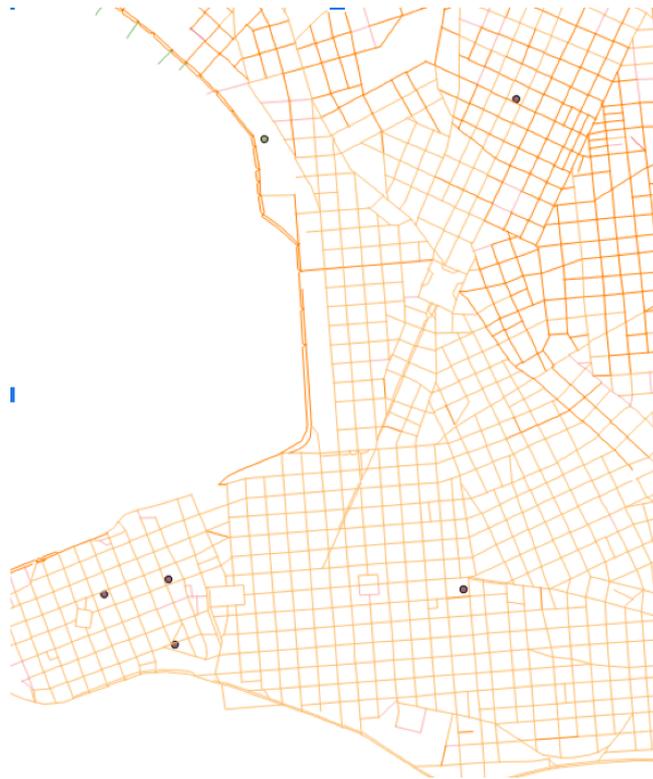


Figura 4.21: Radio de 3km alrededor de la Central Batlle

En el prototipo desarrollado para mostrar la relación entre ambos datos, se realizó una carga inicial manual y se generó un cubo de ejemplo para ser visualizado en Saiku, como se puede ver en la figura 4.22, en la figura 4.23 y en la figura 4.24, que logra mostrar básicamente la relación entre ambos datos.

## PM2\_5

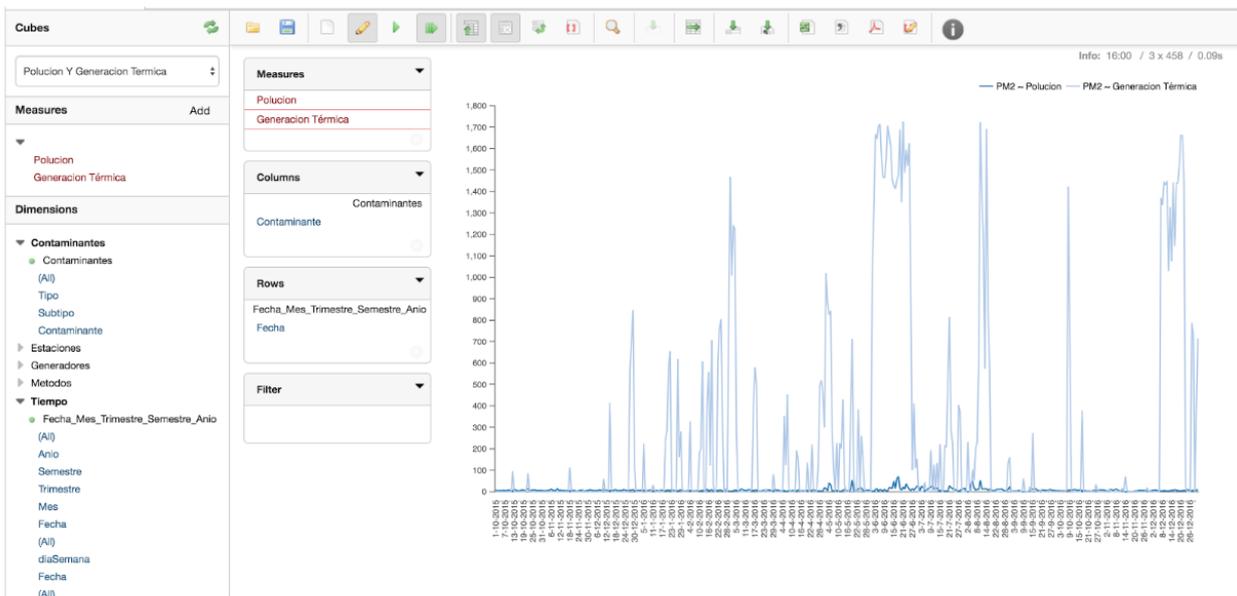


Figura 4.22: Prototipo - Relación Contaminación en PM2.5 - Actividad Industrial

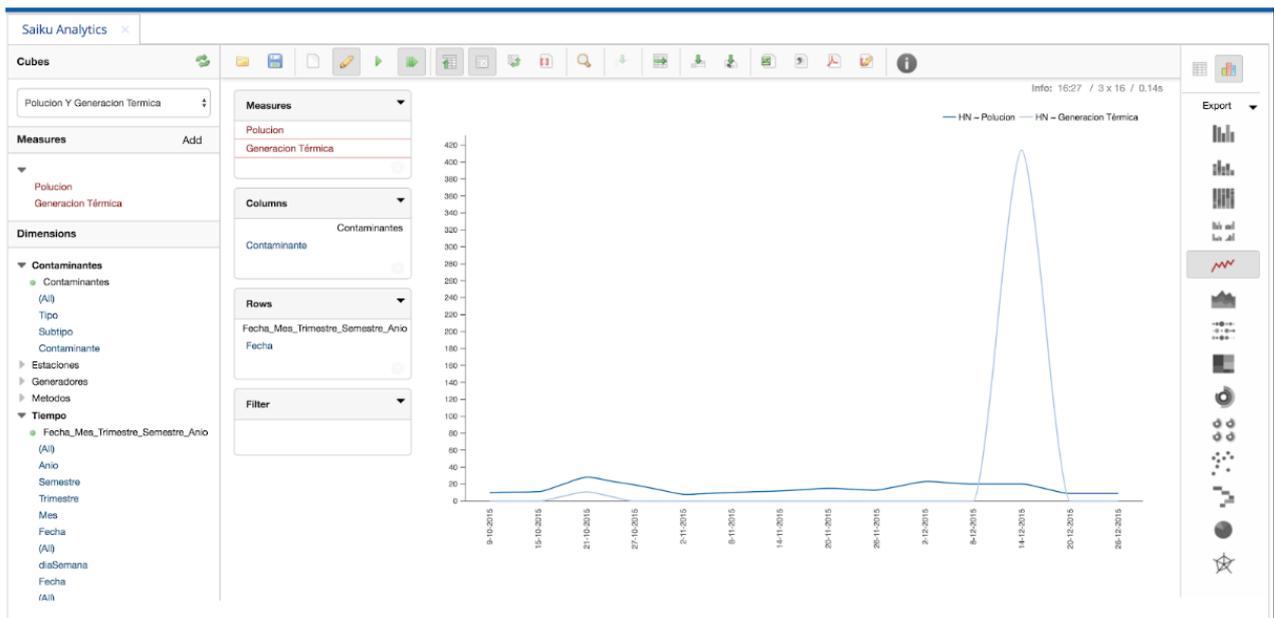


Figura 4.23: Prototipo - Relación Contaminación en Humo Negro - Actividad Industrial

## SO2

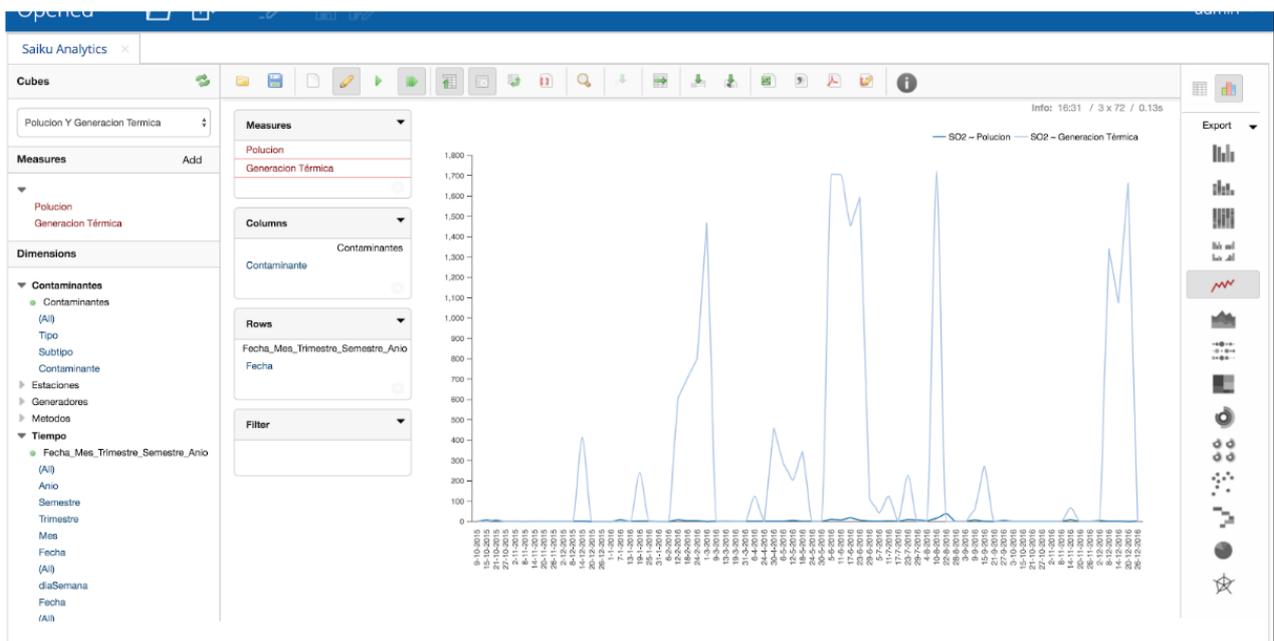


Figura 4.24: Prototipo - Relación Contaminación en SO2 - Actividad Industrial

## RFA3 - Relación Días Especiales y Contaminación

*Analizar la evolución en el tiempo de la contaminación del aire generados por la combustión, específicamente la quema de leña generada por los asados, en días especiales como los festivos, feriados, fechas patria y días con eventos destacables.*

El requerimiento pretende realizar un análisis sobre otro de los principales focos de contaminación de aire como son los procesos de quema de leña por distintos motivos, en especial, los asados. El objetivo es poder entender si en los días que se produce una mayor quema de leña se encuentra una mayor contaminación en el aire.

El mismo surge del cruzamiento de la fuente de datos de medidas históricas de la Red de Monitoreo [18] interna al Servicio de Calidad de Aire con fuentes de datos externas para obtener los días festivos, feriados y fechas patria [21], y del ingreso manual de fechas con eventos destacables.

Para relacionar los días especiales con la contaminación, es necesario investigar en paralelo, con datos horarios y diarios, la emisión de los contaminantes PM<sub>2</sub>, PM<sub>10</sub> o PTS, junto con los días especiales. Cabe destacar que son relevantes los datos obtenidos de todas las estaciones de la red.

Por lo tanto, para tomar los datos para realizar la investigación, se define junto a los especialistas del Servicio de Calidad de Aire los días que serán tomados en cuenta para la realización del análisis.

Analizando las fuentes de datos con las que se cuenta en el momento de la elaboración del informe, se puede observar que:

Para los días especiales:

1. Se obtuvieron datos de los días festivos del Ministerio de Turismo (MINTUR) [21].
2. Se obtienen e ingresan los días que hay partidos de fútbol relevantes (Selección Uruguay de Fútbol, partidos de los equipos grandes).

Para las mediciones de contaminación:

1. Existen datos de múltiples contaminantes con periodicidad diaria y semanal.

Por lo tanto, al momento de la elaboración de este informe será posible relacionar:

1. Mediciones de múltiples contaminantes obtenidos en las estaciones de la red de monitoreo para el período desde 2005 hasta la fecha.
2. Días especiales tomados de las fuentes externas y de las ingresadas manualmente.

En el prototipo desarrollado para mostrar la relación entre ambos datos, se realizó una carga inicial manual y se generó un cubo de ejemplo para ser visualizado en JPivot, como se puede ver en la figura 4.25, que logra mostrar básicamente la relación entre ambos datos.

Fechas	Contaminantes	EstacionesDeLaRed	DiasSemana	EstacionesDelAnio	Horas	Metodos	TemperaturaExterna	DiasImportantess	Measures
All Fecchass	All Contaminantess	All EstacionesDeLaReds	All DiasSemanas	All EstacionesDelAnios	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	8.133
				verano	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	5.346
				otono	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	9.797
				invierno	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	11.943
				primavera	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	5.505
	Contaminante secundario	All EstacionesDeLaReds	All DiasSemanas	All EstacionesDelAnios	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	8.133
				verano	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	5.346
				otono	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	9.797
				invierno	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	11.943
				primavera	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	5.505
	PM	All EstacionesDeLaReds	All DiasSemanas	All EstacionesDelAnios	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	8.133
				verano	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	5.346
				otono	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	9.797
				invierno	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	11.943
				primavera	All Horass	All Metodoss	All TemperaturaExternas	All DiasImportantess	5.505

Figura 4.25: Prototipo - Relación Contaminación en PM2.5 - Días Importantes

A continuación se describen en detalle los requerimientos funcionales de uso.

### RFU1 - Aplicación web de carga y actualización

*Permitir a usuarios finales la realización de la carga y actualización del DW mediante un aplicativo que resulte amigable y no requiera de amplio conocimiento técnico.*

Para promover el uso del sistema de BI es necesario brindar herramientas simples e intuitivas para las tareas que son indispensables de realizar periódicamente para su gestión, como lo son la carga y actualización de los datos desde las fuentes. Para cumplir dicho motivo se implementa y disponibiliza una aplicación web fácilmente accesible.

### RFU2 - Generación de reporte de BI de sección del Informe Anual

*Automatizar la realización de un reporte de BI de la sección Resultados de Red de Monitoreo del Informe Anual, donde detallan las características de cada estación y se muestra un resumen de los resultados expresados utilizando el ICAire.*

Actualmente, los especialistas del servicio ECCA, generan manualmente la información, estadísticas y reportes de la sección Resultados de Red de Monitoreo del Informe Anual como se visualiza en la figura 4.26. Se espera la posibilidad de generar dicho reporte de BI de la sección a demanda, desde aplicativo web a implementar, con los datos que se encuentren en el DW para el año indicado.

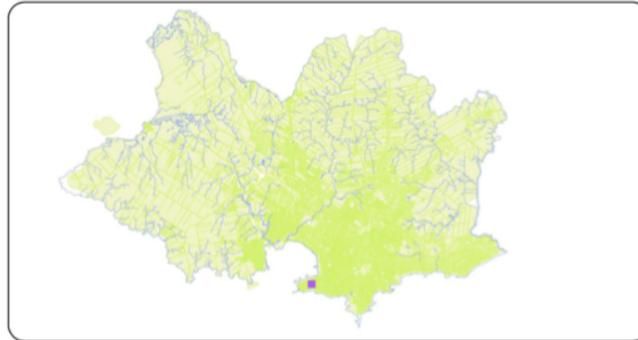
### RFU3 - Generación de un dashboard para análisis

*Generar un dashboard con gráficas e indicadores pre-definidos para un posterior análisis*

Con el objetivo de brindar herramientas de BI visuales para descubrir *insights* que logren retratar fielmente el estado de la calidad de aire en Montevideo mediante un acceso eficaz y eficiente a la información, se genera un Dashboard con gráficas e indicadores preestablecidos para poder ser consultado de manera rápida e intuitiva.

**2.2 Estación 1: Ciudad Vieja**  
Estación de Base

Academia Uruguay				
UTM21 S	X=572795	Altura sobre Nivel del mar 20m		
SIRGAS 2000	Y=6137106	Elevación desde la calzada 9 m		
Sede de Asociación de Empleados Bancarios de Uruguay				
UTM21 S	X=572831	Altura sobre Nivel del mar 14m		
SIRGAS 2000	Y=6136758	Elevación desde la calzada 3 m		
Parámetro	Unidades	Método de medida	Periodo evaluado	
Material Particulado PM2.5	PM10	ug/m3	Light Scattering	Horaria
Humo Negro	HN	ug/m3	Reflectometría	24 horas cada 6 días
Dióxido de Azufre	SO2	ug/m3	Cromatografía Iónica	24 horas cada 6 días



Mapa 2.2 Ciudad Vieja

	Promedio (ug/m3)	IC (ug/m3)	Número de Muestras	Máximo (ug/m3)	Mínimo (ug/m3)	Percentil 95 (ug/m3)
PM2.5	13	(12-15)	348	67	3	26
SO2	5	(3-6)	15	21	1	20
Humo Negro	14	(11-17)	30	49	1	24

Tabla 2.2 Resultados Ciudad Vieja

Figura 4.26: Sección Resultados de Red de Monitoreo - Informe Anual

#### 4.2.5. Especificación de requerimientos no funcionales

Basados en ambos enfoques y en el estudio de factibilidad se presentan los requerimientos no funcionales que deberá cumplir el sistema, acordados con los especialistas de la IM.

1. Implementar la solución utilizando la suite de Pentaho Community Edition (CE).
2. Funcionar en arquitecturas de varios niveles.
3. Mantener una relación adecuada con Bases de Datos (BDs) Fuentes:
  - a) Acceso a BDs heterogéneas y multiplataforma como bases de datos relacionales y geográficas.
  - b) Independiente de los Sistemas de Producción.
4. Prover documentación técnica detallada del sistema, incluyendo su instalación, configuración, inicialización y monitoreo para su posterior gestión, actualización y soporte por parte del Servicio de Calidad de Aire.
5. Prover documentación funcional del sistema, un manual de usuario y tutoriales web para su posterior utilización por parte del Servicio de Calidad de Aire.

# Capítulo 5

## Diseño

En este capítulo se presenta la base de herramientas y componentes que formarán la arquitectura del sistema de BI y el diseño de cada uno de los elementos analizados.

### 5.1. Arquitectura

Como se puede identificar en la figura 5.1, donde se pueden visualizar los distintos componentes que forman parte de cada capa, se presenta una arquitectura en varios niveles, logrando desacoplar las responsabilidades de cada nivel. Fuentes de datos, integración y limpieza, repositorio de datos, herramientas de análisis de datos y metadatos, son algunos de los niveles identificados.

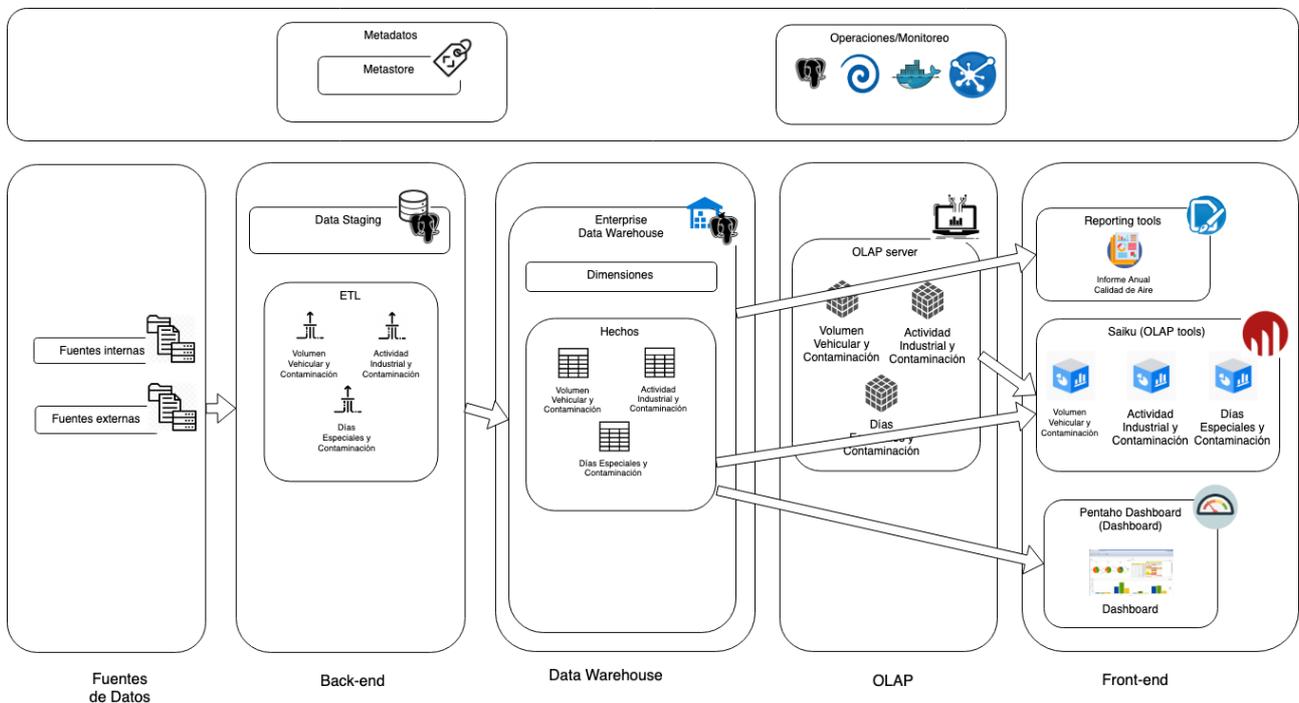


Figura 5.1: Arquitectura sistema de BI

### 5.1.1. Fuentes de datos

La capa de Fuentes de datos representa las diferentes fuentes de datos que alimentan los datos del DW. La fuente de datos puede estar en cualquier formato: archivo de texto plano, base de datos relacional, otros tipos de base de datos, planillas, etc.

Como se explicitó en la etapa de Análisis, y se muestra en la figura 5.2, se logran identificar y describir fuentes de datos internas del Servicio de Calidad de Aire y externas de Datos Abiertos, ANCAP y ADME.

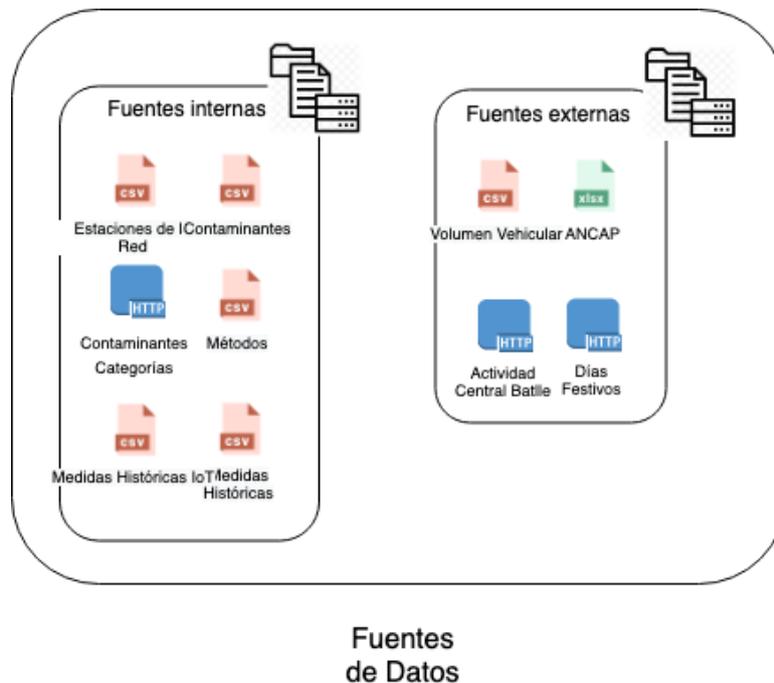


Figura 5.2: Fuentes de Datos

### 5.1.2. Back-end

La capa de Back-end, ver figura 5.3, contiene tres secciones importantes: Capa de extracción de datos, Área de pruebas y procesos ETL.

Dentro del Área de pruebas (Data Staging) podemos visualizar las distintas tablas basadas en las fuentes de datos que se identificaron, tanto internas como externas.

En los procesos de ETL se define uno por cada requerimiento identificado.

### 5.1.3. DW

En la capa de almacenamiento de datos se tiene una base de datos relacional que oficia de repositorio de datos.

En el mismo se pueden diferenciar dos grandes grupos: las tablas que sirven como Dimensiones y las que sirven como Hechos, como se muestra en la figura 5.4.

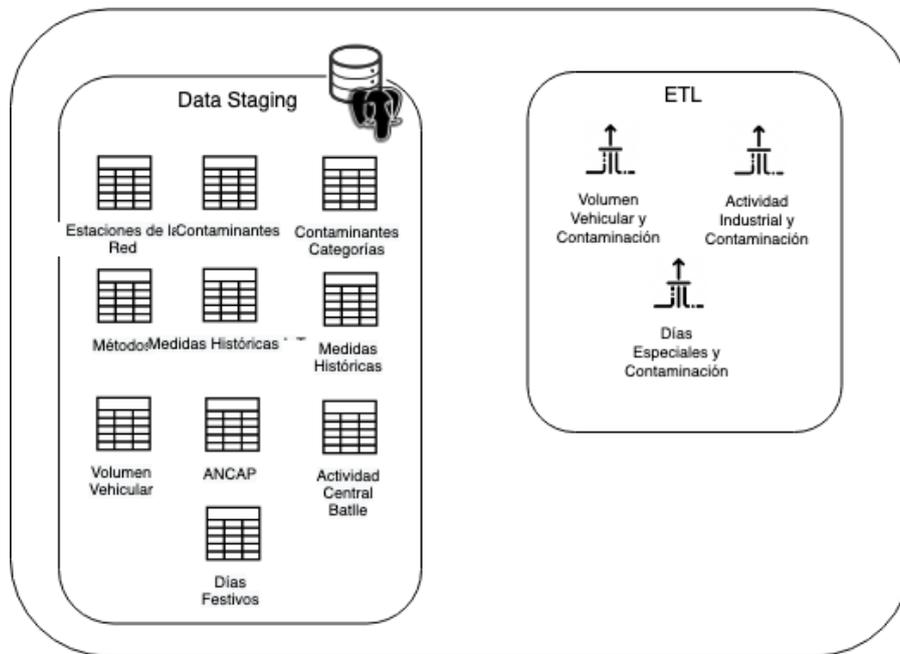
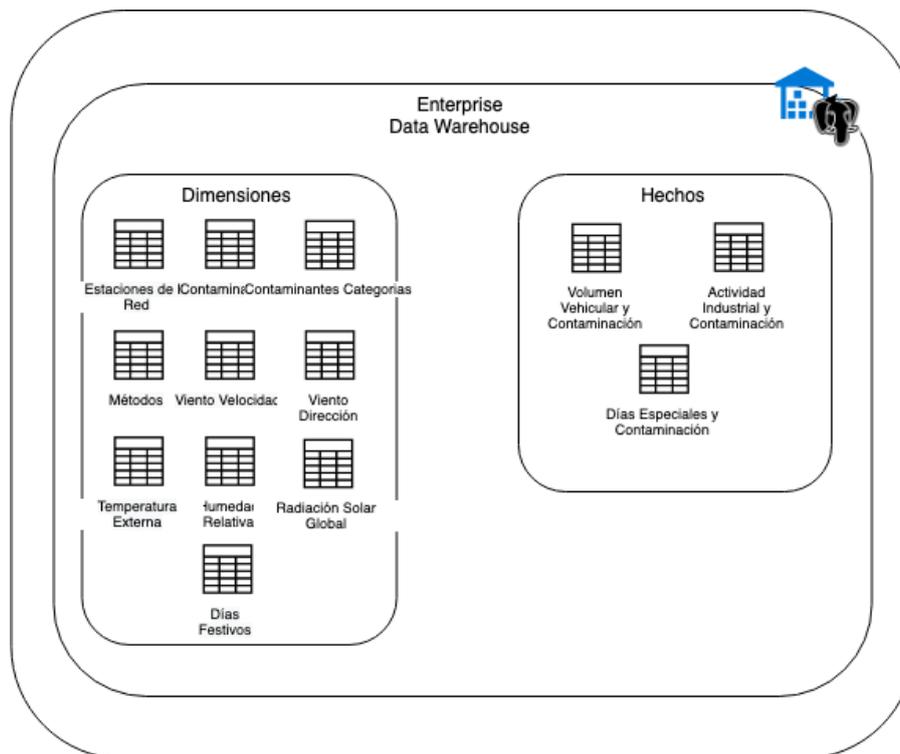


Figura 5.3: Back-end



Data Warehouse

Figura 5.4: DW

#### 5.1.4. OLAP

En la capa OLAP se mantienen las reglas de negocio que se aplican a la información almacenada en el Enterprise DW para ser presentada de forma intuitiva por las herramientas de análisis de

datos.

En el mismo, ver figura 5.5, se definen tres cubos, uno por cada requerimiento identificado.

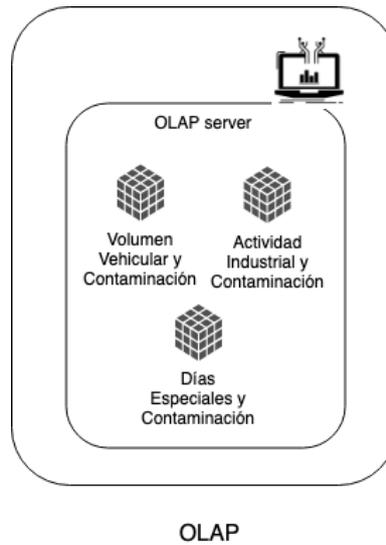


Figura 5.5: OLAP

### 5.1.5. Front-end

Finalmente, en la capa de análisis de datos, se ofrecen al usuario herramientas con los siguientes objetivos:

1. Mecanismos de acceso a la información eficaces.
2. Tener conexión eficaz al DW.

Primero se utiliza Saiku, una herramienta de la suite Pentaho, para que los especialistas puedan realizar las consultas de forma dinámica logrando la mayor flexibilidad posible para analizar los datos.

A su vez, se disponibiliza la capacidad de generar un informe, similar al publicado anualmente, para automatizar la generación del mismo.

Finalmente, se genera un dashboard a medida para poder visualizar de forma rápida y directa los indicadores preestablecidos.

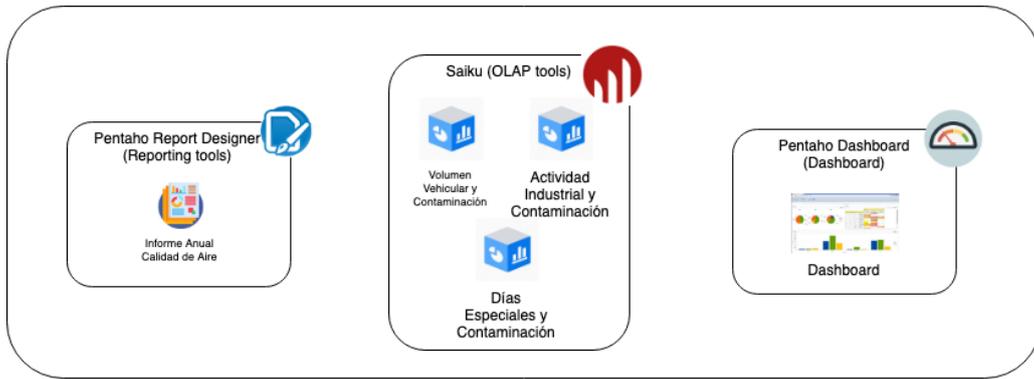
En la figura 5.6 se pueden diferenciar las distintas herramientas de análisis.

### 5.1.6. Operaciones del sistema

Esta capa incluye información sobre cómo está funcionando el sistema de BI, cuál es el estado de trabajo ETL, cuál es el rendimiento del sistema y el historial de acceso de los usuarios.

Como se puede visualizar en la figura 5.7, en esta capa se pueden visualizar múltiples elementos de operaciones que sirven para el monitoreo:

1. Monitoreo de solución en Docker [28] mediante Grafana [24] y Prometheus [34].



Front-end

Figura 5.6: Front-end

2. Status de PBI.
3. Status de PDI (Carte).
4. Aplicación web desarrollada a medida.



Figura 5.7: Operaciones

## 5.2. Diseño

Dado que el área de Servicio de Calidad de Aire se enfoca especialmente en el monitoreo y evaluación de la calidad de aire, un área de negocio con objetivos específicos, se utiliza la metodología bottom-up.

A continuación se especifican cada una de las etapas de diseño realizadas.

### 5.2.1. Diseño Conceptual

La etapa de Diseño Conceptual se realiza utilizando un enfoque combinado, haciendo uso paralelamente de los requerimientos identificados y el modelo conceptual de las fuentes de datos generado, ambas salidas generadas en la etapa de Análisis. Finalmente se integran los resultados de ambos enfoques.

Se incluye la representación gráfica de los distintos componentes del diseño conceptual utilizando el modelo CMDM [9].

#### Dimensiones y jerarquías

Para modelar la realidad del problema es necesario construir las dimensiones que sirven como criterios de análisis de los datos. Se pueden distinguir distintos grupos de dimensiones como las pertenecientes a la Red de Monitoreo en la figura 5.8, de tipo temporal en la figura 5.9, de tipo meteorológico en la figura 5.10, y externos a la Red de Monitoreo en la figura 5.11.

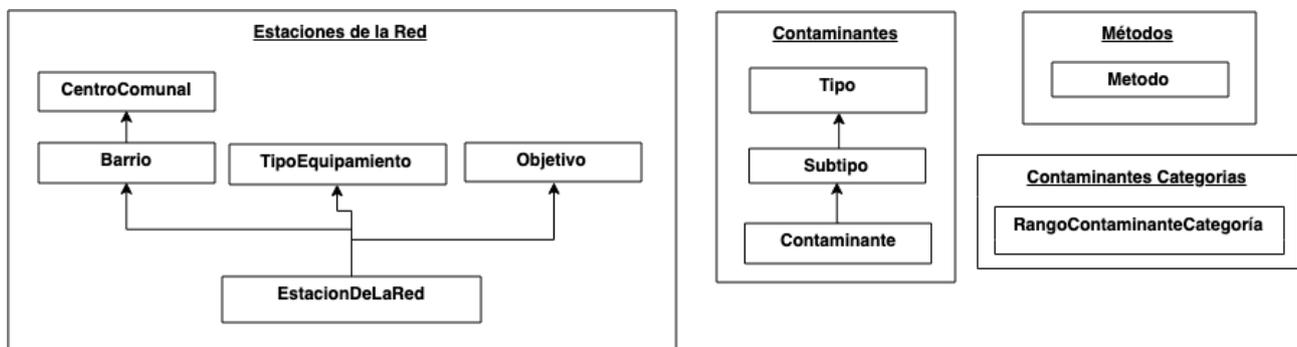


Figura 5.8: Dimensiones de Red de Monitoreo de Calidad de Aire de Montevideo

A continuación se describen en detalle las dimensiones:

#### **Estaciones de la Red**

La dimensión Estaciones de la Red representa las estaciones que permiten evaluar la concentración de material particulado en diversas fracciones y de gases que habitualmente se encuentran en ambientes urbanos (dióxido de azufre y dióxido de nitrógeno).

A continuación se describen los niveles y las jerarquías definidas para la dimensión que se encuentran representadas en la figura 5.13.

#### Niveles

*EstacionDeLaRed*: Compuesto por las estaciones de la red de monitoreo.

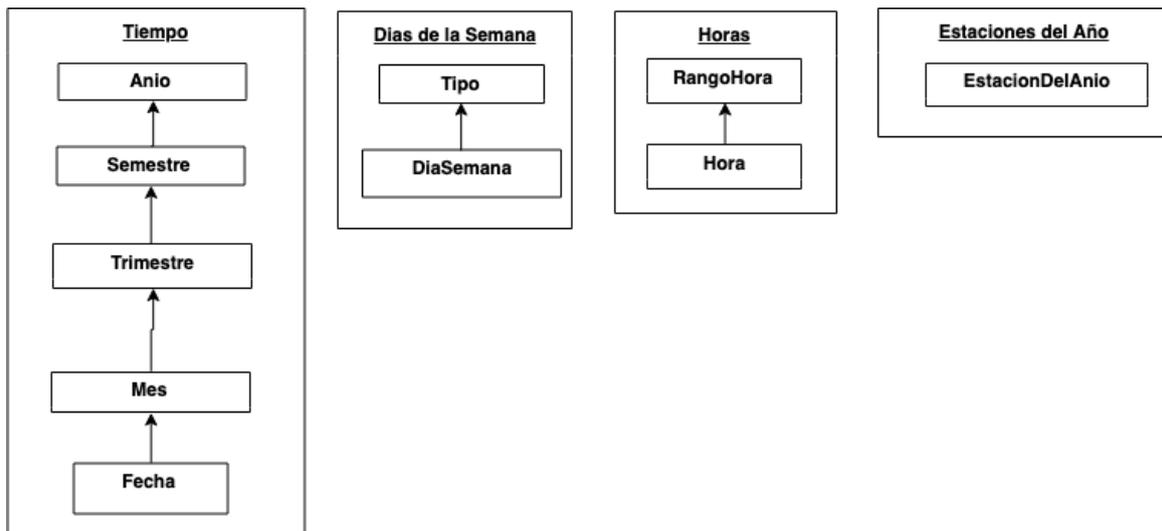


Figura 5.9: Dimensiones Temporales

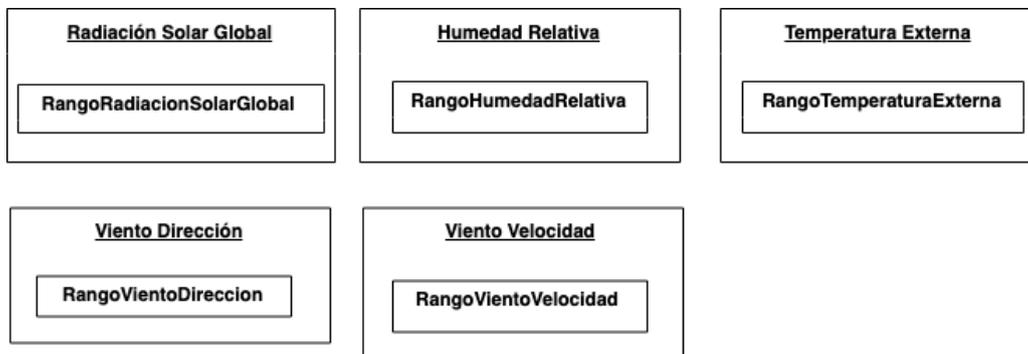


Figura 5.10: Dimensiones Meteorológicas

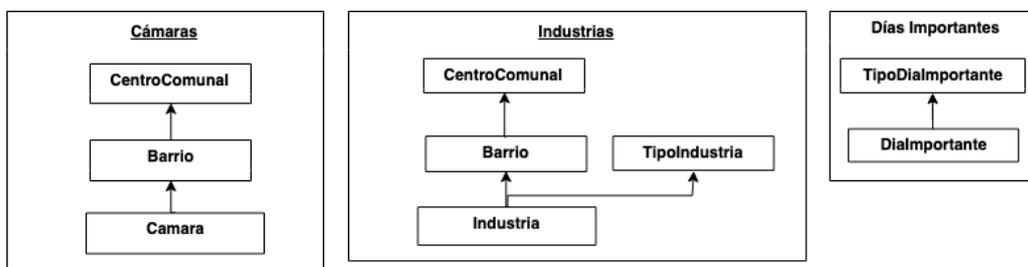


Figura 5.11: Dimensiones Fuentes Externas

*Barrio:* Compuesto por los barrios de Montevideo.

*CentroComunal:* Compuesto por los centros comunales zonales de Montevideo.

*TipoEquipamiento:* Compuesto por los métodos de medición que emplean las estaciones de la red.

*Objetivo:* Compuesto por los propósitos que cumplen las estaciones de la red.

### Jerarquías

*EstacionDeLaRed ->Barrio ->CentroComunal:* Se diseñó de forma tal que fuera posible agregar las medidas por su ubicación física, específicamente, barrios y centros comunales zonales.

Como se puede visualizar en la figura 5.12 la jerarquía no es de tipo Estricta y Balanceada debido a que existen barrios que se encuentran en varios CCZ.

20	Punta de Rieles, Bella Italia	9	F
21	Jardines del Hipódromo	9	F
22	Ituzaingó	9	F
23	Unión	6 y 11	D, E y F
24	Villa Española	9 y 11	D y F
25	Mercado Modelo, Bolívar	3 y 11	C y D
26	Castro, Pérez Castellanos	11	D
27	Cerrito de la Victoria	11	D
28	Las Acacias	11	D
29	Aires Puros	10 y 15	C y D
30	Casavalle	11	D
31	Piedras Blancas	9 y 10	D y F
32	Manga, Toledo Chico	10	D
33	Paso de las Duranas	13	G
34	Peñarol, Lavalleja	13	G
35	Villa del Cerro	17	A
36	Casabó, Pajas Blancas	17	A
37	La Paloma, Tomkinson	17	A
38	La Teja	14	A
39	Prado, Nueva Savona	15 y 16	A y C
40	Capurro, Bella Vista, Arroyo Seco	16	C
41	Aguada	1, 2 y 3	B y C
42	Reducto	3	C
43	Atahualpa	15	C
44	Jacinto Vera	3	C
45	La Figurita	3	C
46	Larrañaga	4	C y CH
47	La Blanqueada	4 y 6	CH y E
48	Villa Muñoz, Retiro, Goes	3	B y C
49	La Comercial	2 y 3	B y C
50	Tres Cruces	4	B y CH
51	Brazo Oriental	3	C
52	Sayago	13	G
53	Conciliación	13	G
54	Belvedere, Paso Molino	13 y 14	A y G
55	Nuevo París	13 y 14	A y G
56	Tres Ombúes, Pueblo Victoria	14	A
57	Paso de la Arena, Santiago Vázquez	18	A
58	Colón Sureste, Abayubá	12	G
59	Colón Centro y Noroeste	12	G
60	Lezica, Melilla	12	A y G
61	Villa García, Manga Rural	9	F
62	Manga	9	F

Figura 5.12: Barrios y CCZ

*EstacionDeLaRed -> TipoEquipamiento:* Las estaciones de la red se pueden agrupar según el método de medición que emplean en automáticas o manuales. La jerarquía es de tipo Estricta y Balanceada.

*EstacionDeLaRed -> Objetivo:* Las estaciones de la red se pueden diferenciar según el propósito u objetivo que cumplen en evaluar la calidad base de la Ciudad o evaluar el impacto generado por algunas de las fuentes más significativas del departamento. La jerarquía es de tipo Estricta y Balanceada.

En la figura 5.14, en la figura 5.15 y en la figura 5.16 se presentan ejemplos de los datos que cada nivel de cada jerarquía contiene.

## Contaminantes

La dimensión Contaminantes representan tanto los contaminantes gaseosos (dióxido de carbono, el monóxido de carbono, los hidrocarburos, los óxidos de nitrógeno, los óxidos de azufre y el ozono) como el material particulado (mezcla heterogénea de partículas sólidas o líquidas suspendidas en un gas).

### Niveles

*Contaminante:* Compuesto por los contaminantes medidos por las estaciones de la red.

*Subtipo:* Compuesto por los subtipos de los contaminantes.

*Tipo:* Compuesto por los tipos de los contaminantes.

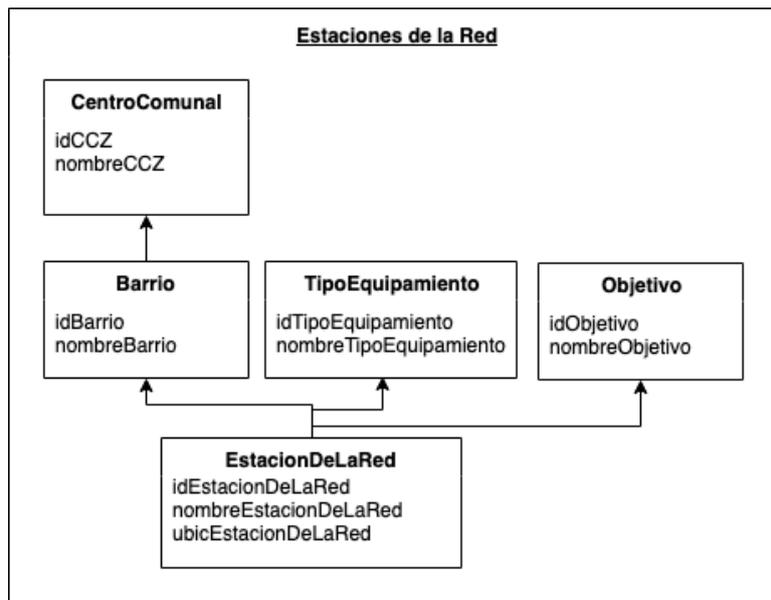


Figura 5.13: Estaciones de la Red

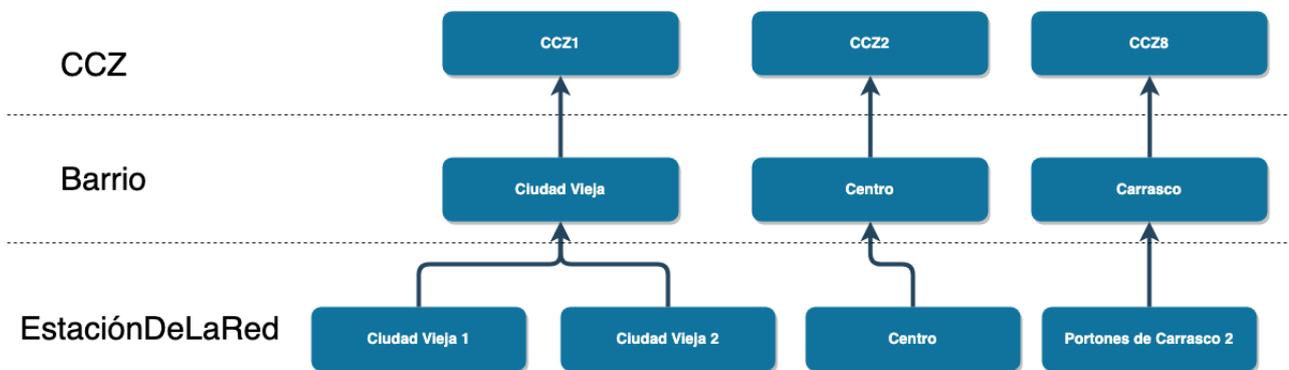


Figura 5.14: Estaciones de la Red - Ubicación física

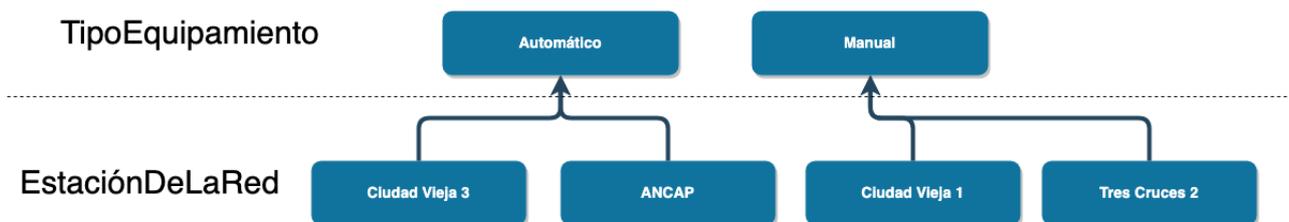


Figura 5.15: Estaciones de la Red - Tipo Equipamiento



Figura 5.16: Estaciones de la Red - Objetivo

## Jerarquías

*Contaminante -> Subtipo -> Tipo*: Los contaminantes se pueden agrupar según el subtipo y tipo en contaminantes gaseosos y material particulado. La jerarquía es de tipo Estricta y Balanceada.

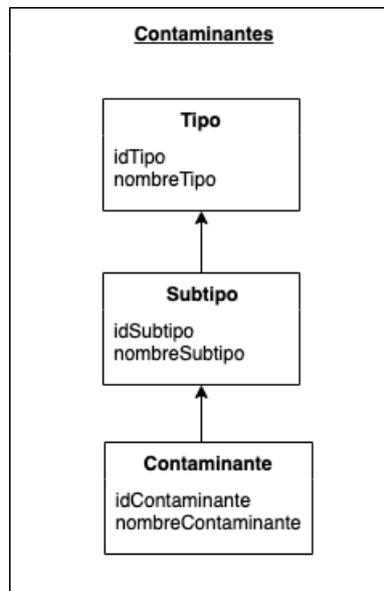


Figura 5.17: Contaminantes

En la figura 5.18 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

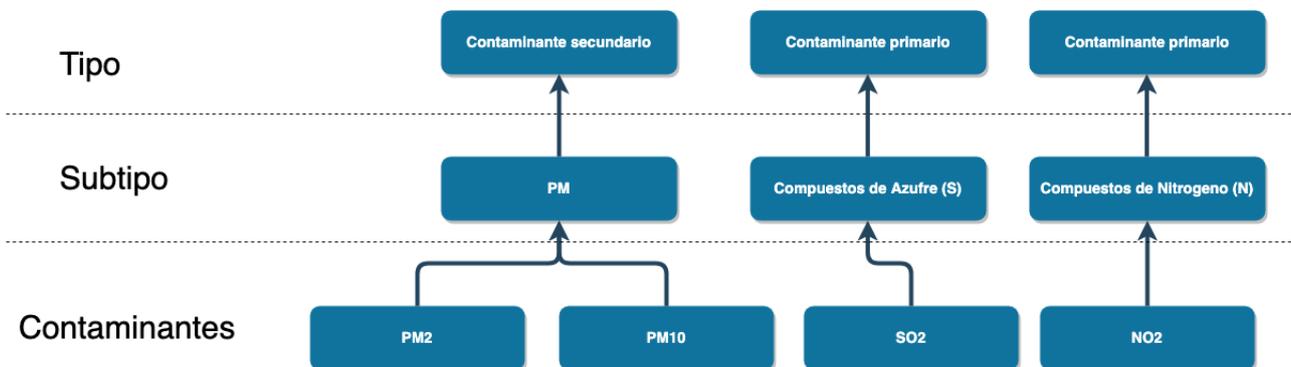


Figura 5.18: Contaminantes Ejemplo

## **Contaminantes Categorías**

La dimensión Contaminantes Categorías representa los rangos que clasifican la concentración de cada contaminante en el aire. Las categorías son Muy Buena, Buena, Aceptable, Inadecuada, Mala y Muy Mala.

## Niveles

*ContaminanteCategoría*: Compuesto por rangos de concentración definidos por el Servicio de Calidad de Aire para cada contaminante.

## Jerarquías

Para este criterio de análisis se definió una jerarquía compuesta por un único nivel.



Figura 5.19: Contaminantes Categorías

En la figura 5.20 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.



Figura 5.20: Contaminantes Categorías Ejemplo

## Métodos

La dimensión Métodos representan los distintos métodos de medida de los contaminantes en el aire, donde se abarcan métodos químicos simples ó métodos de algunas técnicas electrónicas más sofisticadas.

### Niveles

*Método:* Compuesto por los métodos utilizados para la realización de las mediciones por las estaciones de la red.

### Jerarquías

Para este criterio de análisis se definió una jerarquía compuesta por un único nivel.

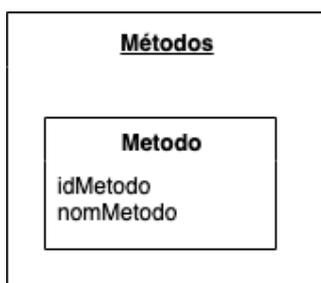


Figura 5.21: Métodos

En la figura 5.22 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.



Figura 5.22: Métodos Ejemplo

## Días de la Semana

La dimensión Días de la Semana representan los 7 días de la semana.

### Niveles

*DiaDeLaSemana*: Compuesto por los 7 días de la semana.

*Tipo*: Compuesto por los distintos tipos de día de la semana.

### Jerarquías

*DiaDeLaSemana* -> *Tipo*: Los días de la semana se diferencian entre días laborables y de descanso. La jerarquía es de tipo Estricta y Balanceada.

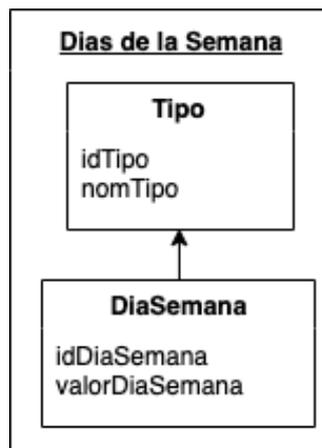


Figura 5.23: Días de la Semana

En la figura 5.24 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

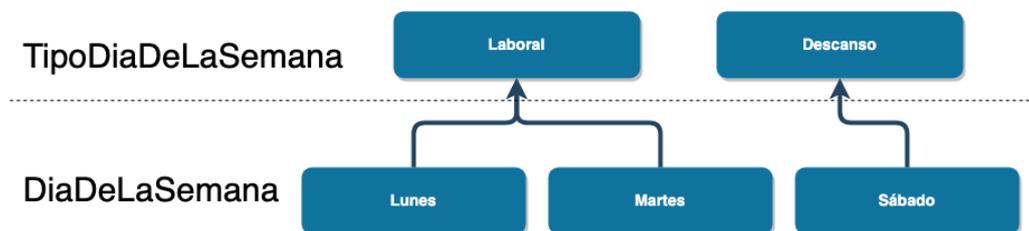


Figura 5.24: Días de la Semana Ejemplo

## **Estaciones del Año**

La dimensión Estaciones del Año representa los períodos en los que las condiciones climáticas imperantes se mantienen, en una determinada región, dentro de un cierto rango. En este caso, nos basamos en las estaciones que aplican a Uruguay.

### Niveles

*EstacionDelAnio*: Compuesto por las estaciones del año según el contexto uruguayo.

### Jerarquías

Para este criterio de análisis se definió una jerarquía compuesta por un único nivel.

En la figura 5.26 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

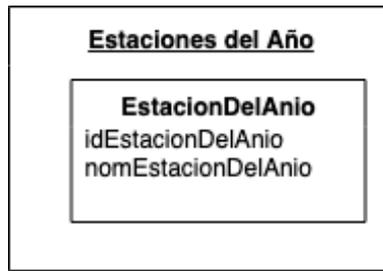


Figura 5.25: Estaciones del Año



Figura 5.26: Estaciones del Año Ejemplo

## Tiempo

La dimensión Tiempo representa el criterio de análisis temporal.

### Niveles

*Fecha*: Compuesto por todos los días de cada año.

*Mes*: Compuesto por todos los meses de cada año.

*Trimestre*: Compuesto por todos trimestres de cada año.

*Semestre*: Compuesto por todos semestres de cada año.

*Año*: Compuesto por todos años.

### Jerarquías

*Fecha ->Mes ->Trimestre ->Semestre ->Año*: Se diseñó de forma tal que fuera posible agregar las medidas por los rangos de tiempo Mes, Trimestre, Semestre y Año. La jerarquía es de tipo Estricta y Balanceada.

En la figura 5.28 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

## Horas

La dimensión Horas representa la una unidad de tiempo.

### Niveles

*Hora*: Compuesto por todas las horas del día.

*RangoHora*: Compuesto por los rangos de horas según la parte del día.

### Jerarquías

*Hora ->RangoHora*: Se se diseñó de forma tal que fuera posible agregar las horas por los rangos según las partes del día en, entre otros, Mañana, Tarde y Noche. La jerarquía es de tipo Estricta y Balanceada.

En la figura 5.30 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

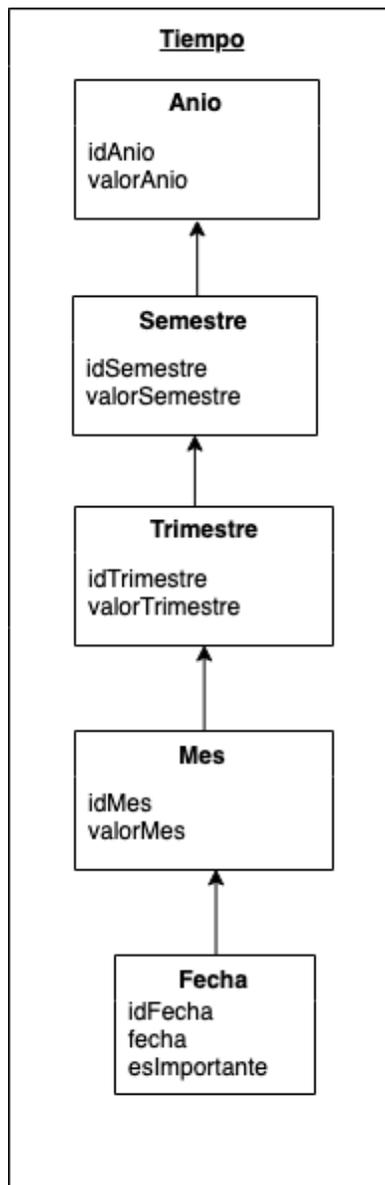


Figura 5.27: Tiempo

## Radiación Solar Global

La dimensión Radiación Solar Global representa la energía proveniente del sol que incide en una superficie plana (suelo) en forma de ondas electromagnéticas.

### Niveles

*RangoRadiacionSolarGlobal*: Compuesto por rangos de radiación solar definidos con los especialistas. Los valores de la radiación solar [22] se pueden agrupar según la intensidad.

### Jerarquías

Para este criterio de análisis se definió una jerarquía compuesta por un único nivel.

En la figura 5.32 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

## Humedad Relativa

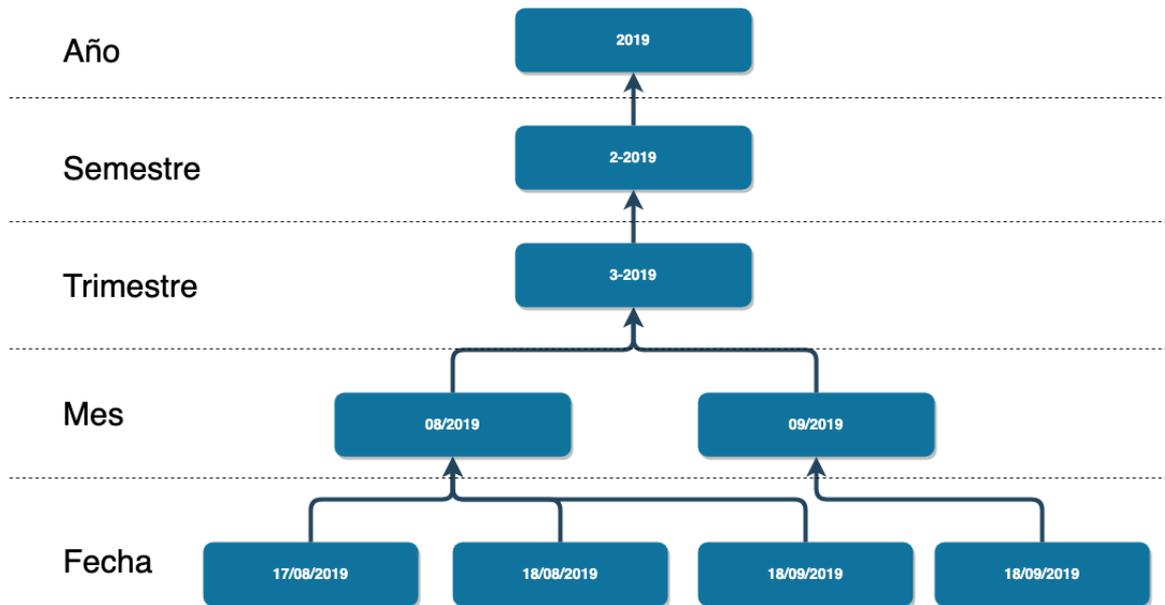


Figura 5.28: Tiempo Ejemplo

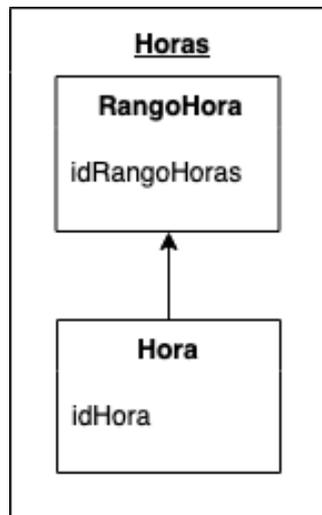


Figura 5.29: Horas

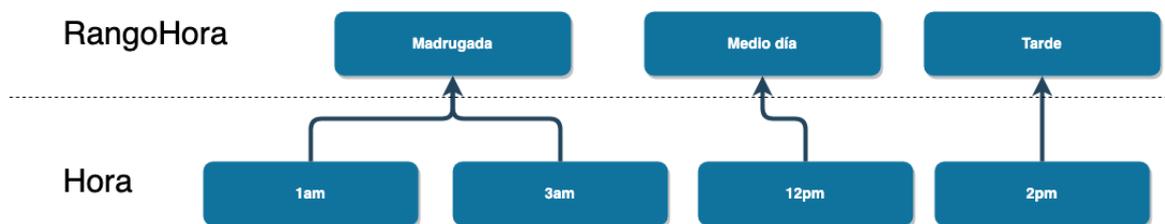


Figura 5.30: Horas Ejemplo

La dimensión Humedad Relativa representa la relación entre la presión parcial del vapor de agua y la presión de vapor de equilibrio del agua a una temperatura dada.

Niveles

*RangoHumedadRelativa*: Compuesto por rangos de humedad relativa estándar. Los valores de

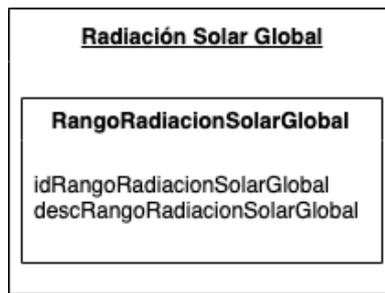


Figura 5.31: Radiación Solar Global

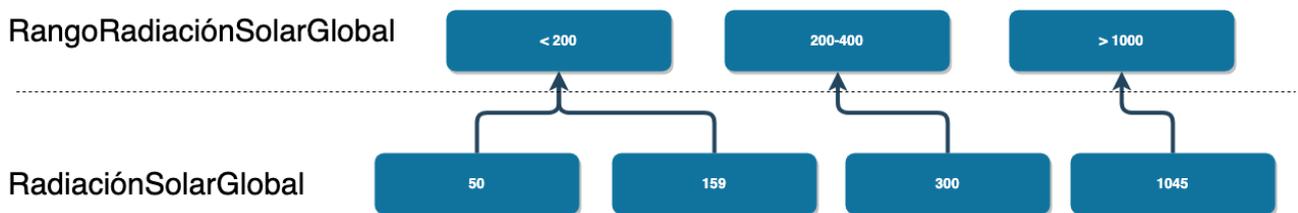


Figura 5.32: Radiación Solar Global Ejemplo

la humedad relativa se pueden agrupar según el orden de saturación basado en el porcentaje [35].

### Jerarquías

Para este criterio de análisis se definió una jerarquía compuesta por un único nivel.

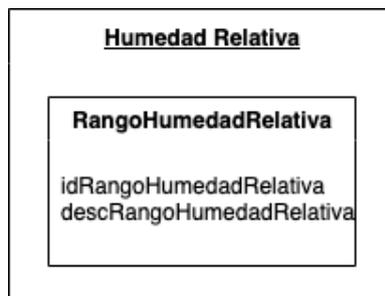


Figura 5.33: Humedad Relativa

En la figura 5.34 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

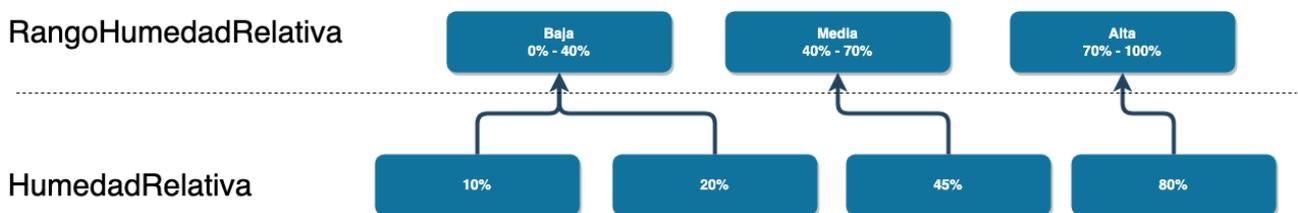


Figura 5.34: Humedad Relativa Ejemplo

### **Temperatura Externa**

La dimensión Temperatura Externa representa la magnitud referida a la noción de calor medible mediante un termómetro.

## Niveles

*Rango Temperatura Externa*: Compuesto por rangos de temperatura externa estándar. Los valores de la temperatura externa se pueden agrupar según el grado de nivel térmico de la atmósfera.

## Jerarquías

Para este criterio de análisis se definió una jerarquía compuesta por un único nivel.

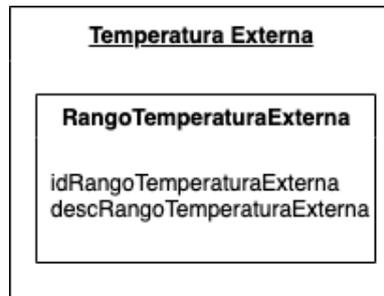


Figura 5.35: Temperatura Externa

En la figura 5.36 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

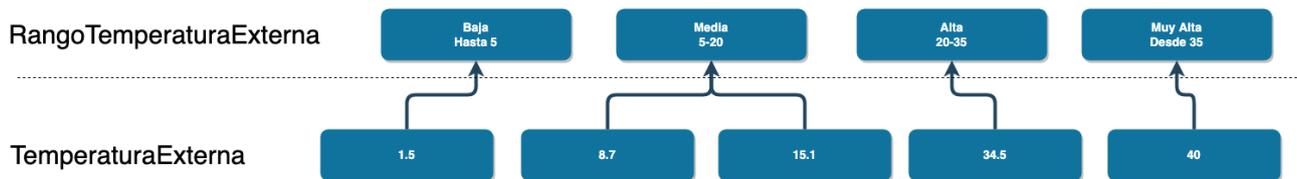


Figura 5.36: Temperatura Externa Ejemplo

## **Viento Dirección**

La dimensión Viento Dirección representa la dirección desde la cual sopla el viento, y se mide en grados en la dirección de las agujas del reloj a partir del norte verdadero.

## Niveles

*Rango Viento Dirección*: Compuesto por rangos según la rosa de los vientos. Los valores de la dirección del viento se pueden agrupar en los rumbos en que se divide la circunferencia del horizonte [38].

## Jerarquías

Para este criterio de análisis se definió una jerarquía compuesta por un único nivel.

En la figura 5.38 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

## **Viento Velocidad**

La dimensión Viento Velocidad representa la medición de la componente horizontal del desplazamiento del aire en un punto y en un instante determinados.

## Niveles



Figura 5.37: Viento Dirección

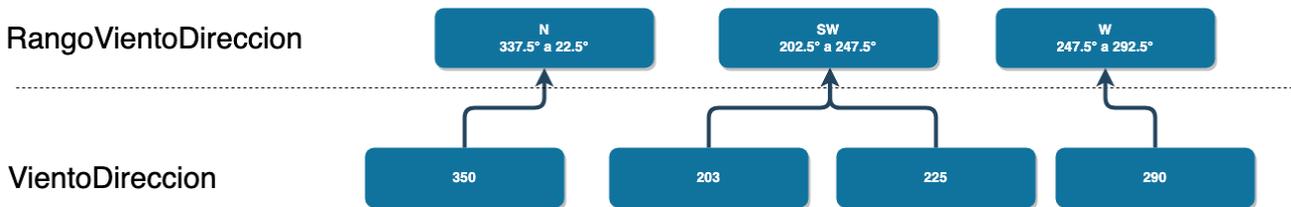


Figura 5.38: Viento Dirección Ejemplo

*RangoVientoVelocidad*: Compuesto por rangos según la Escala de Beaufort [29] en relación a las distintas velocidad [30]. Los valores de la velocidad del viento se pueden agrupar según la intensidad [36] del viento basada principalmente en el estado del mar, de sus olas y la fuerza del viento.

### Jerarquías

Para este criterio de análisis se definió una jerarquía compuesta por un único nivel.

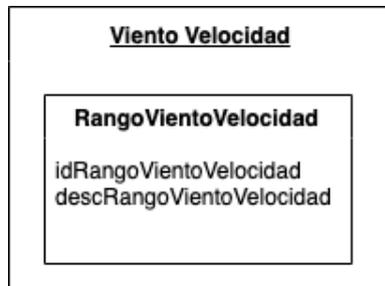


Figura 5.39: Viento Velocidad

En la figura 5.40 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

### **Cámaras**

La dimensión Cámaras representa las cámaras de la IM que sirven a monitorear el tránsito y tráfico en distintos puntos de la ciudad.

### Niveles

*Cámaras*: Compuesto por todas las cámaras de tránsito.

*Barrio*: Compuesto por los barrios de Montevideo.

*CentroComunal*: Compuesto por los centros comunales zonales de Montevideo.

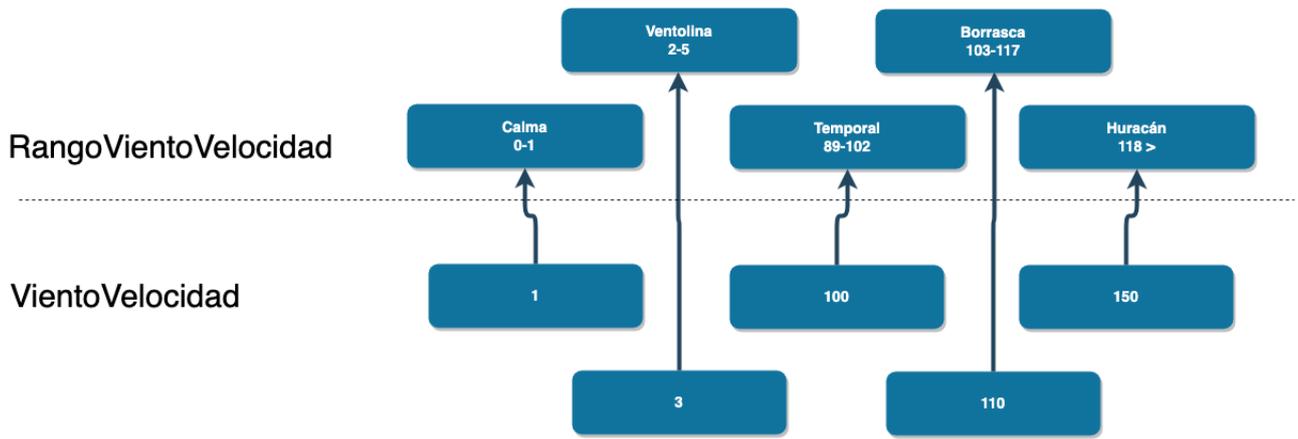


Figura 5.40: Viento Velocidad Ejemplo

### Jerarquías

*Cámaras ->Barrio ->CentroComunal*: Se diseñó de forma tal que fuera posible agregar las medidas por su ubicación física, específicamente, barrios y centros comunales zonales. La jerarquía no es de tipo Estricta y Balanceada debido a que existen barrios que se encuentran en varios CCZ.

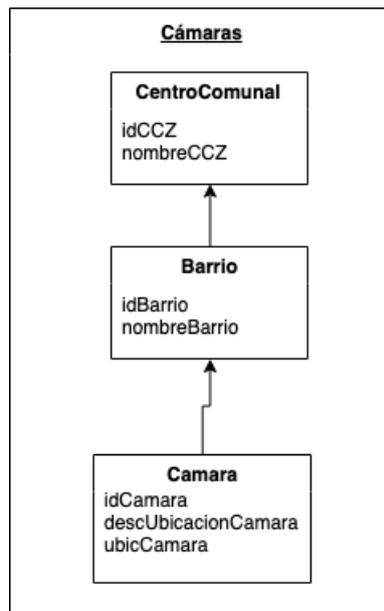


Figura 5.41: Cámaras

En la figura 5.42 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

### **Industrias**

La dimensión Industrias representan las industrias que se encuentran en el área de Montevideo.

### Niveles

*Industria*: Compuesto por los todas las industrias.

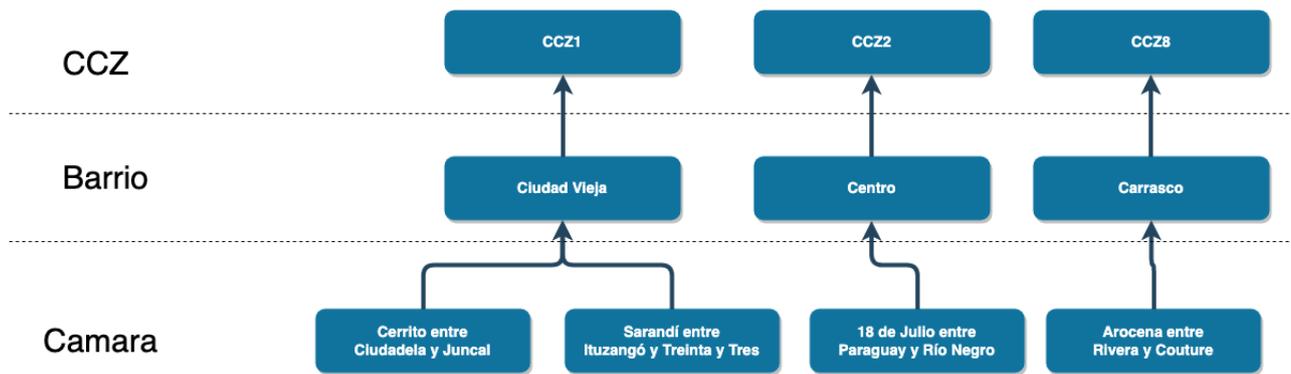


Figura 5.42: Cámaras Ejemplo

*Barrio*: Compuesto por los barrios de Montevideo.

*CentroComunal*: Compuesto por los centros comunales zonales de Montevideo.

*Tipo*: Compuesto por los tipos de industrias existentes según su rubro.

### Jerarquías

*Industria ->Barrio ->CentroComunal*: Se diseñó de forma tal que fuera posible agregar las medidas por su ubicación física, específicamente, barrios y centros comunales zonales. La jerarquía no es de tipo Estricta y Balanceada debido a que existen barrios que se encuentran en varios CCZ.

*Industria ->Tipo*: Las industrias se pueden agrupar según el rubro en el cual se encuentren. La jerarquía es de tipo Estricta y Balanceada.

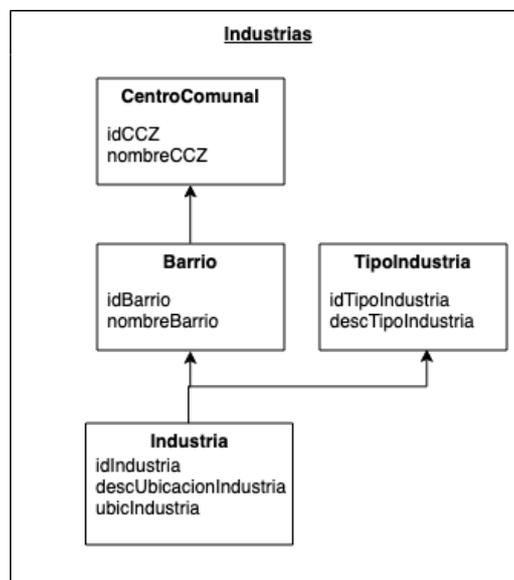


Figura 5.43: Industrias

En la figura 5.44 y figura 5.45 se presentan ejemplos de los datos que cada nivel de cada jerarquía contiene.

### **Días Importantes**

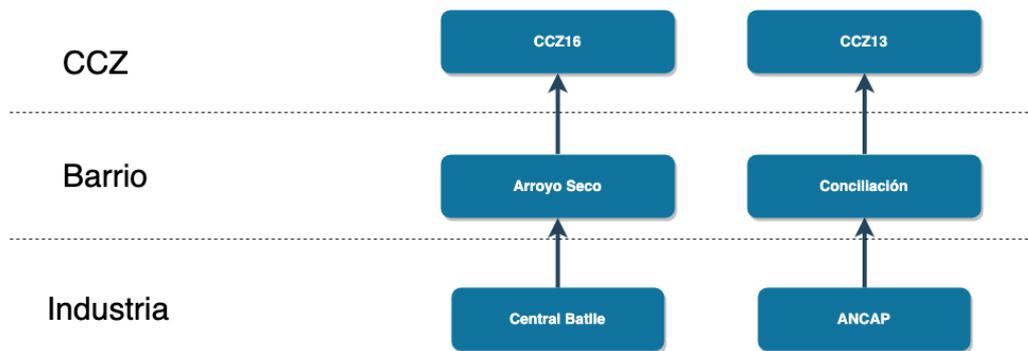


Figura 5.44: Industrias Ejemplo

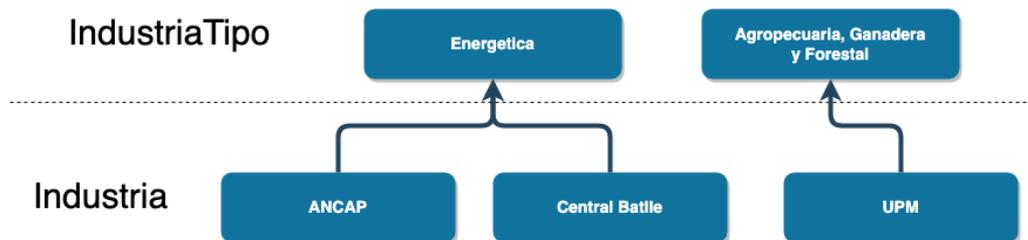


Figura 5.45: Industrias Ejemplo

La dimensión Días Importantes representan los días que interesan analizar debido a su carácter festivo o de descanso que aumenta considerablemente la posibilidad de combustión, específicamente la quema de leña generada por los asados.

### Niveles

*Día Importante:* Compuesto por los todos los días importantes definidos.

*Tipo:* Compuesto por los tipos de días importantes según sean festivos, patrios o eventos relevantes.

### Jerarquías

*Día Importante -> Tipo:* Los días importantes se pueden generalizar según el motivo por el cual es importante. La jerarquía es de tipo Estricta y Balanceada.

En la figura 5.47 se presenta un ejemplo de los datos que cada nivel de cada jerarquía contiene.

### Medidas

Se presentan las medidas definidas para el análisis según la especificación de requerimientos realizada.

### **Vehículos Contaminación**

Basado en el RFA1, relativo a analizar la evolución en el tiempo de la contaminación en el aire relacionándolo con la densidad vehicular, se definen los indicadores de contaminación y densidad vehicular, como se puede visualizar en la figura 5.49.

### **Industrias Contaminación**

Basado en el RFA2, relativo a analizar la evolución en el tiempo de la contaminación en el

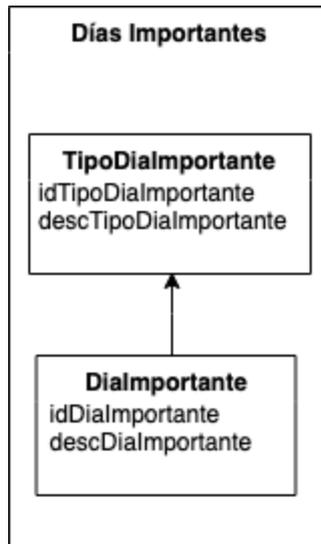


Figura 5.46: Días Importantes

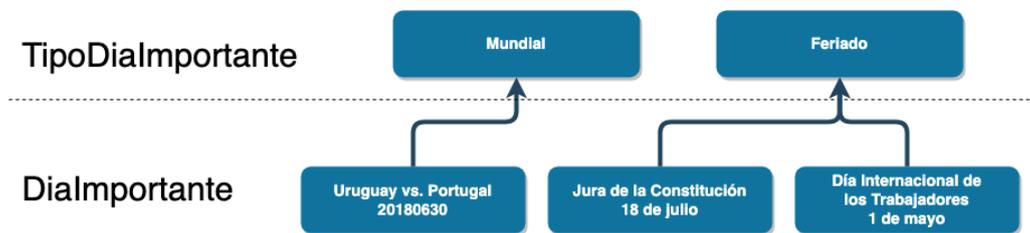


Figura 5.47: Días Importantes Ejemplo

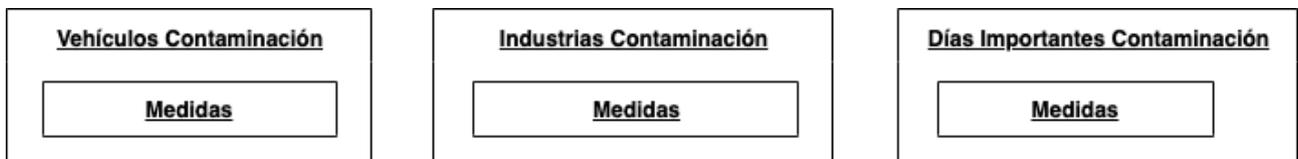


Figura 5.48: Medidas

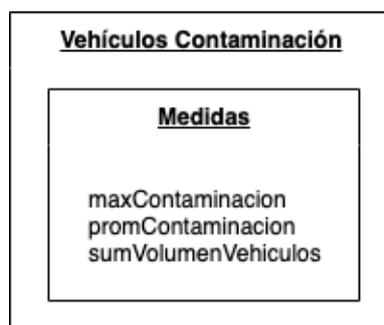


Figura 5.49: Vehículos Contaminación

aire relacionándolo con la actividad industrial, se definen los indicadores de contaminación y actividad industrial, como se puede visualizar en la figura 5.50.

### Días Importantes Contaminación

Basado en el RFA3, relativo a analizar la evolución en el tiempo de la contaminación en el aire



Figura 5.50: Industrias Contaminación

relacionándolo con la los días especiales, se define el indicador de contaminación, como se puede visualizar en la figura 5.51.

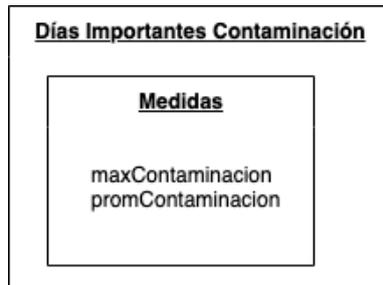


Figura 5.51: Días Importantes Contaminación

## Relaciones Dimensionales

A continuación se presentan las relaciones dimensionales a partir de las cuales se confeccionaron los cubos para el análisis.

### **Vehículos Contaminación**

La relación dimensional Vehículos Contaminación permite examinar la relación entre la contaminación y la densidad vehicular según los distintos criterios dimensionales indicados en la figura 5.52.

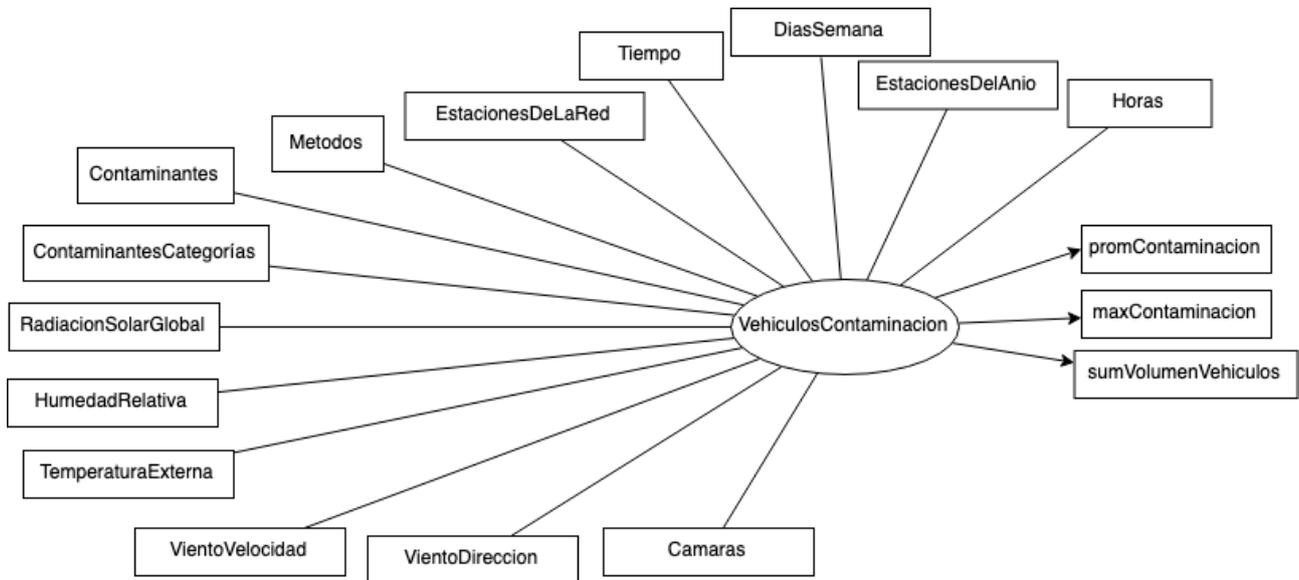


Figura 5.52: Vehículos Contaminación

### *Cubos*

En la figura 5.53 se visualiza un cubo de la relación dimensional que permite examinar la relación entre la contaminación y la densidad vehicular según distintos criterios como: los barrios de Montevideo, los tipos de contaminante, los métodos de medición, los días de la semana, las estaciones del año, los rangos definidos de concentración de cada contaminante según los parámetros de calidad de aire, los meses del año, los períodos del día, los rangos de radiación solar global, humedad relativa, temperatura externa, velocidad y dirección del viento y las cámaras de conteo vehicular.

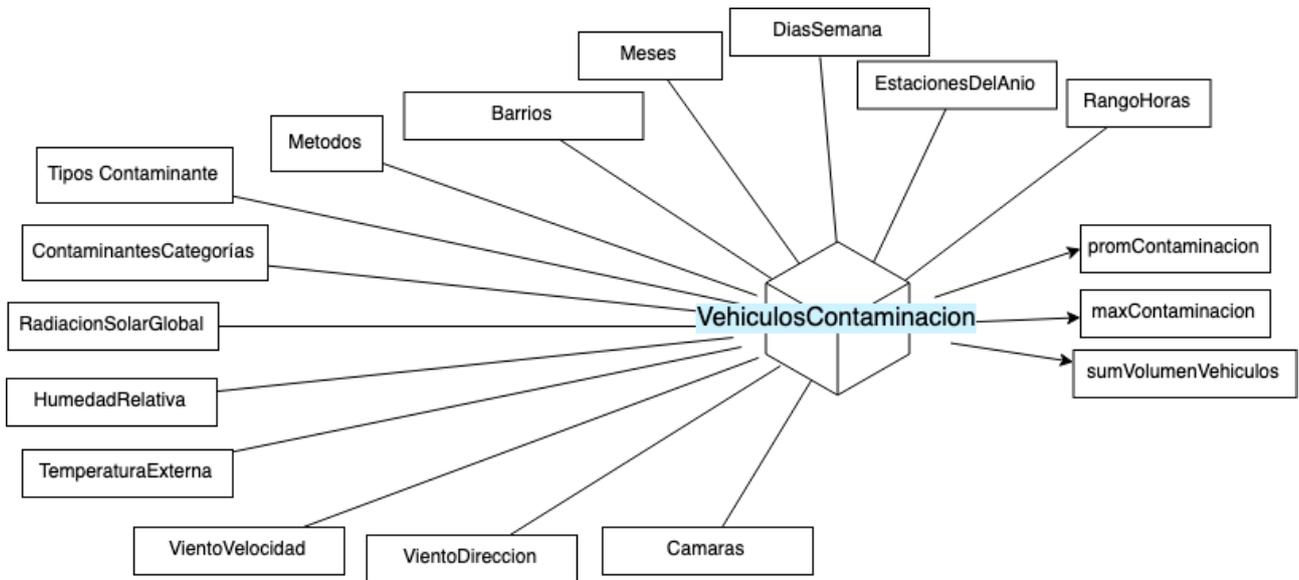


Figura 5.53: Cubo - Vehículos Contaminación

### Industrias Contaminación

La relación dimensional Industrias Contaminación permite examinar la relación entre la contaminación y la actividad industrial según los distintos criterios dimensionales indicados en la figura 5.54.

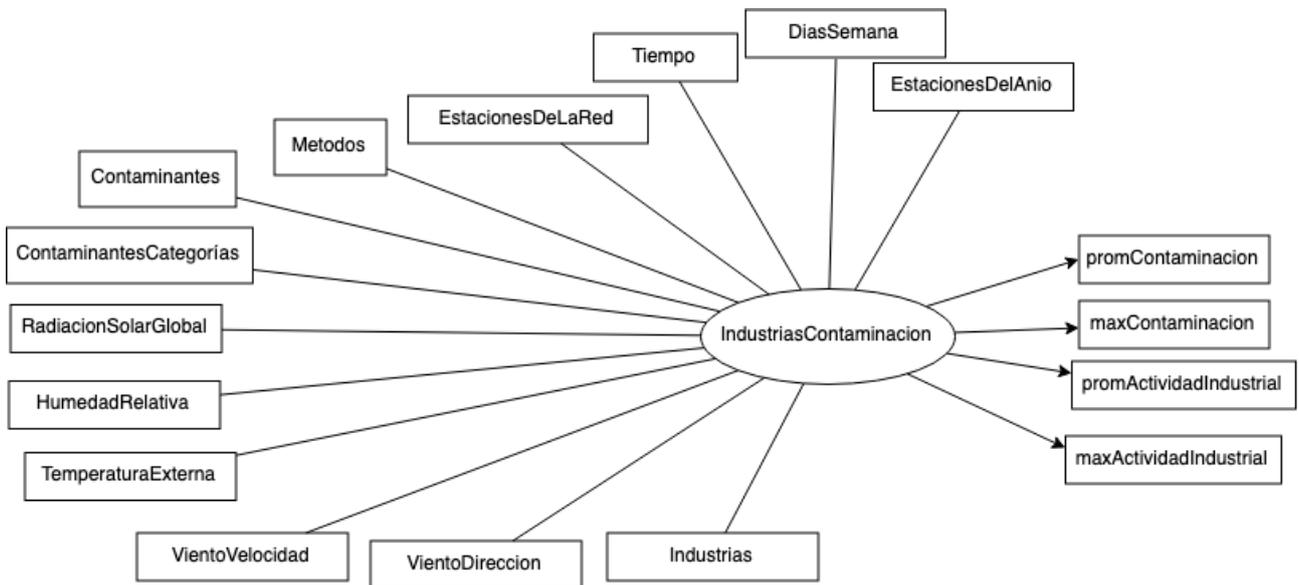


Figura 5.54: Industrias Contaminación

### Cubos

En la figura 5.55 se visualiza un cubo de la relación dimensional que permite examinar la relación entre la contaminación y la actividad industrial según distintos criterios como: las estaciones de la red, los subtipos de contaminante, los métodos de medición, las estaciones del año, los rangos definidos de concentración de cada contaminante según los parámetros de calidad de aire, las fechas de medición, los rangos de radiación solar global, humedad relativa, temperatura externa, velocidad y dirección del viento y las industrias.

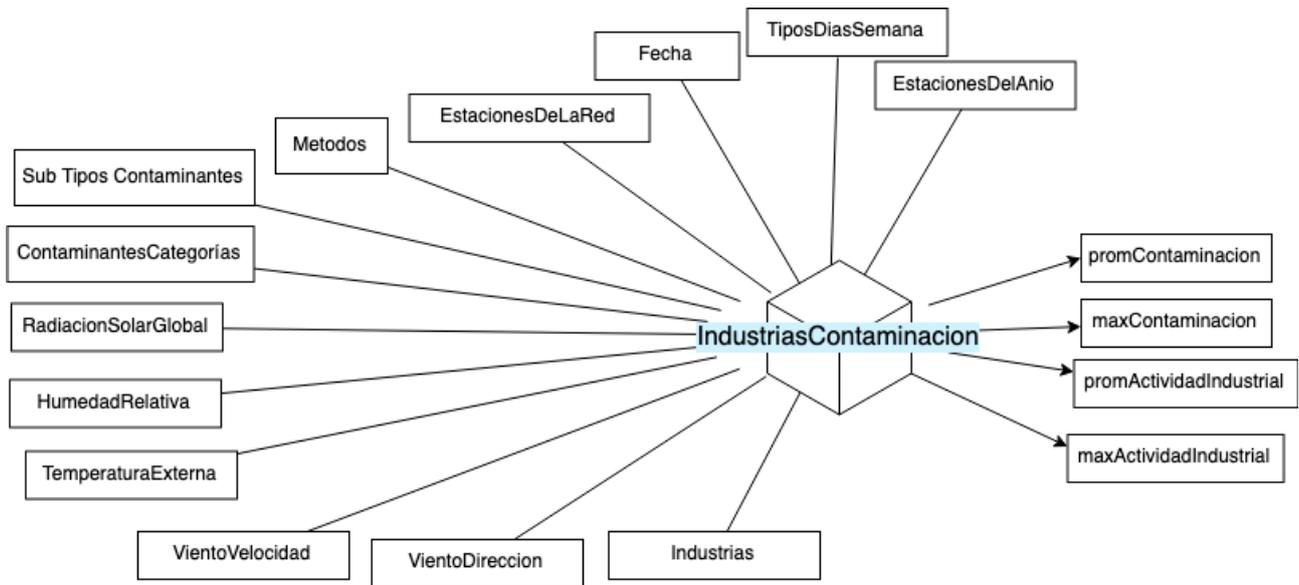


Figura 5.55: Cubo - Industrias Contaminación

### Días Importantes Contaminación

La relación dimensional Días Importantes Contaminación permite examinar la contaminación según los distintos criterios dimensionales indicados en la figura 5.56.

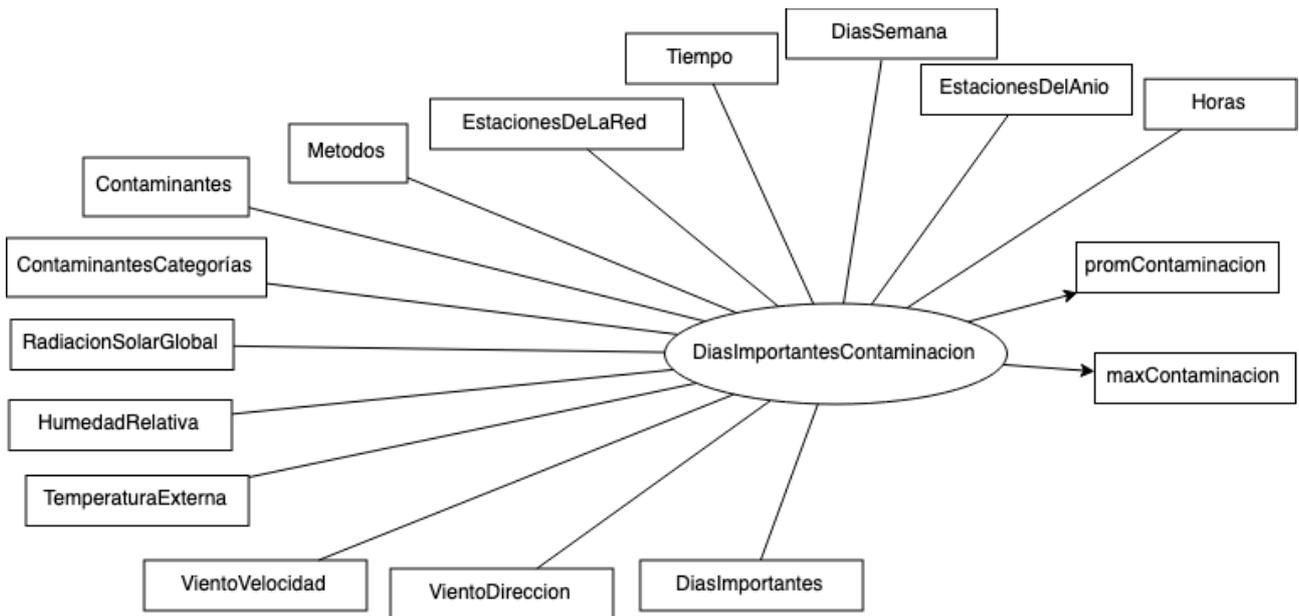


Figura 5.56: Días Importantes Contaminación

### Cubos

En la figura 5.57 se visualiza un cubo de la relación dimensional que permite examinar la contaminación según distintos criterios como: los centros comunales zonales, los contaminantes, los métodos de medición, los días de la semana, las estaciones del año, los rangos definidos de concentración de cada contaminante según los parámetros de calidad de aire, los semestres, los rangos de radiación solar global, humedad relativa, temperatura externa, velocidad y dirección del viento y los días importantes.

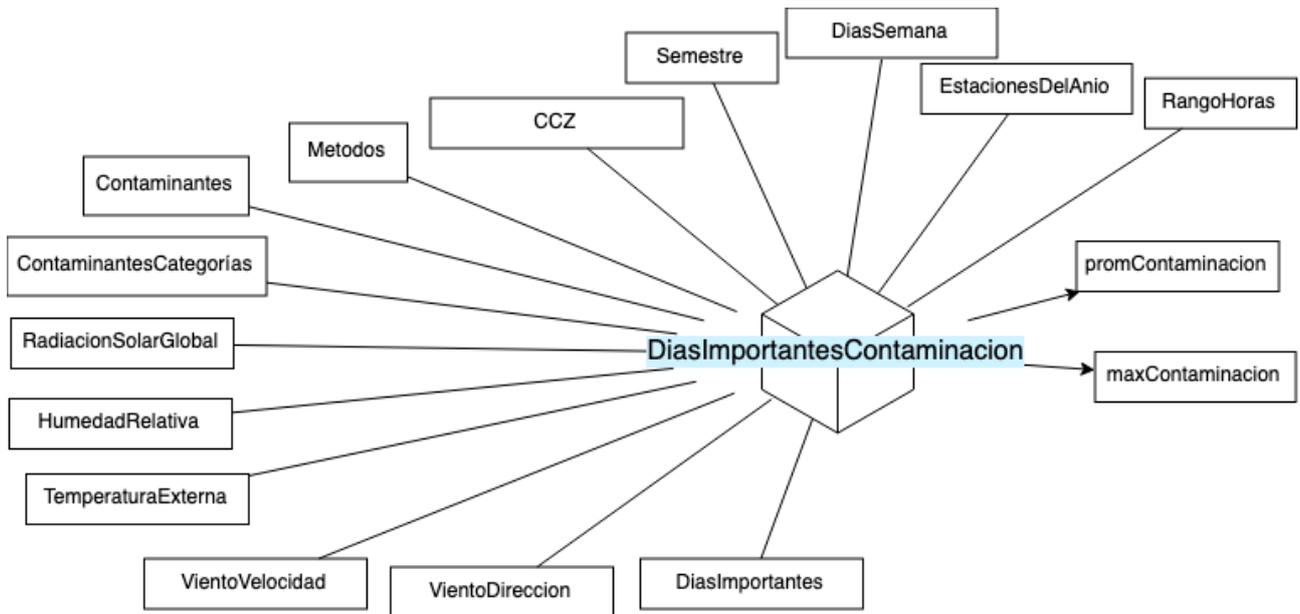


Figura 5.57: Cubo - Días Importantes Contaminación

### Dimensiones Compartidas entre Relaciones Dimensionales

A continuación, en cuadro 5.1, se presenta una tabla que indica qué dimensiones pertenecen a qué relaciones dimensionales y cuáles son compartidas entre ellas.

De la misma se puede visualizar que durante el análisis multidimensional es posible utilizar la operación Drill-Across en las dimensiones compartidas.

Dimensión	RFA1	RFA2	RFA3
Estaciones de la Red	X	X	X
Contaminantes	X	X	X
Métodos	X	X	X
Días de la Semana	X	X	X
Estaciones del Año	X	X	X
Contaminantes Categorías	X	X	X
Tiempo	X	X	X
Horas	X	X	X
Radiación Solar Global	X	X	X
Humedad Relativa	X	X	X
Temperatura Externa	X	X	X
Viento Dirección	X	X	X
Viento Velocidad	X	X	X
Cámaras	X	-	-
Industrias	-	X	-
Días Importantes	-	-	X

Cuadro 5.1: Dimensiones en Requerimientos Funcionales de Análisis

### Estudio de aditividad

A continuación se presenta para cada relación dimensional la tabla de roll-up de las medidas analizadas de cada una de ellas. Este cuadro muestra para cada operación de roll-up qué función de agregación se debe utilizar. En la primera columna se presentan todas las dimensiones. En la segunda columna se indican todas las operaciones de drill-up posibles dentro de las jerarquías de cada dimensión. Finalmente hay una columna por cada medida a 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 de roll-up, mientras que el símbolo  $+$  representa la suma, el símbolo *avg* el promedio y el símbolo *max* el máximo.

En cuanto a la aditividad de las medidas, se puede visualizar que *sumVolumenVehículos* es aditiva ya que la función de agregación en todos los casos es la suma, mientras que las medidas *promContaminación*, *maxContaminación*, *promActividadIndustrial* y *maxActividadIndustrial* no son aditivas ya que en ningún caso es posible aplicar la función de agregación suma y conservar la semántica de las mismas. A su vez, es importante aclarar que para las medidas de contaminación no aplica la agregación por subtipo y tipo de contaminantes dado que aplicar el promedio de dos contaminantes distintos no es correcto.

## Vehículos Contaminación

En figura 5.58 y figura 5.59 se presenta el estudio de aditividad para la relación Vehículos Contaminación.

		VehiculosContaminación		
		promContami nacion	maxContami nacion	sumVolumenV ehiculos
<b>Horas</b>	Horas -> RangoHora	avg	max	+
	RangoHora -> All	avg	max	+
<b>Tiempo</b>	Fecha -> Mes	avg	max	+
	Mes -> Trimestre	avg	max	+
	Trimestre -> Semestre	avg	max	+
	Semestre -> Año	avg	max	+
	Año -> All	avg	max	+
	<b>EstacionesDelAnio</b>	EstacionesDelAnio -> All	avg	max
<b>DiasSemana</b>	DiasSemana -> TipoDiasSemana	avg	max	+
	TipoDiasSemana -> All	avg	max	+
<b>Métodos</b>	Método -> All	avg	max	+
<b>Contaminantes</b>	Contaminante -> Subtipo	N/A	N/A	+
	Subtipo -> Tipo	N/A	N/A	+
	Tipo -> All	N/A	N/A	+
<b>Contaminantes Categorías</b>	Contaminante Categoría -> All	avg	max	+

Figura 5.58: Estudio Aditividad - Vehículos Contaminación

		VehiculosContaminación		
		promContami nacion	maxContami nacion	sumVolumenV ehiculos
<b>EstacionesDeLaRed</b>	EstacionDeLaRed -> Barrio	avg	max	+
	Barrio -> CentroComunal	avg	max	+
	CentroComunal -> All	avg	max	+
	EstacionDeLaRed -> TipoEquipamiento	avg	max	+
	TipoEquipamiento -> All	avg	max	+
	EstacionDeLaRed -> Objetivo	avg	max	+
	Objetivo -> All	avg	max	+
<b>RadiacionSolarGlobal</b>	RadiacionSolarGlobal -> All	avg	max	+
<b>HumedadRelativa</b>	HumedadRelativa -> All	avg	max	+
<b>TemperaturaExterna</b>	TemperaturaExterna -> All	avg	max	+
<b>VientoVelocidad</b>	VientoVelocidad -> All	avg	max	+
<b>VientoDireccion</b>	VientoDireccion -> All	avg	max	+
<b>Camaras</b>	Camaras -> Barrios	avg	max	+
	Barrios -> CCZ	avg	max	+
	CCZ -> All	avg	max	+
<b>Industrias</b>	Industrias -> Tipo	-	-	-
	Tipo -> All	-	-	-
	Industrias -> Barrios	-	-	-
	Barrios -> CCZ	-	-	-
	CCZ -> All	-	-	-
<b>Dialimportante</b>	DiasImportantes -> Tipo	-	-	-
	Tipo -> All	-	-	-

Figura 5.59: Estudio Aditividad - Vehículos Contaminación

## Industrias Contaminación

En figura 5.60 y figura 5.61 se presenta el estudio de aditividad para la relación Industrias Contaminación.

		IndustriasContaminación			
		promContami nacion	maxContami nacion	promActividad	maxActividad
<b>Horas</b>	Horas -> RangoHora	avg	max	avg	max
	RangoHora -> All	avg	max	avg	max
<b>Tiempo</b>	Fecha -> Mes	avg	max	avg	max
	Mes -> Trimestre	avg	max	avg	max
	Trimestre -> Semestre	avg	max	avg	max
	Semestre -> Año	avg	max	avg	max
	Año -> All	avg	max	avg	max
<b>EstacionesDelAño</b>	EstacionesDelAño -> All	avg	max	avg	max
<b>DiasSemana</b>	DiasSemana -> TipoDiasSemana	avg	max	avg	max
	TipoDiasSemana -> All	avg	max	avg	max
<b>Métodos</b>	Método -> All	avg	max	avg	max
<b>Contaminantes</b>	Contaminante -> Subtipo	N/A	N/A	avg	max
	Subtipo -> Tipo	N/A	N/A	avg	max
	Tipo -> All	N/A	N/A	avg	max
<b>Contaminantes Categorías</b>	Contaminante Categoría -> All	avg	max	avg	max

Figura 5.60: Estudio Aditividad - Industrias Contaminación

		IndustriasContaminación			
		promContami nacion	maxContami nacion	promActividad	maxActividad
<b>EstacionesDeLaRed</b>	EstacionDeLaRed -> Barrio	avg	max	avg	max
	Barrio -> CentroComunal	avg	max	avg	max
	CentroComunal -> All	avg	max	avg	max
	EstacionDeLaRed -> TipoEquipamiento	avg	max	avg	max
	TipoEquipamiento -> All	avg	max	avg	max
	EstacionDeLaRed -> Objetivo	avg	max	avg	max
	Objetivo -> All	avg	max	avg	max
<b>RadiacionSolarGlobal</b>	RadiacionSolarGlobal -> All	avg	max	avg	max
<b>HumedadRelativa</b>	HumedadRelativa -> All	avg	max	avg	max
<b>TemperaturaExterna</b>	TemperaturaExterna -> All	avg	max	avg	max
<b>VientoVelocidad</b>	VientoVelocidad -> All	avg	max	avg	max
<b>VientoDireccion</b>	VientoDireccion -> All	avg	max	avg	max
<b>Camaras</b>	Camaras -> Barrios	-	-	-	-
	Barrios -> CCZ	-	-	-	-
	CCZ -> All	-	-	-	-
<b>Industrias</b>	Industrias -> Tipo	avg	max	avg	max
	Tipo -> All	avg	max	avg	max
	Industrias -> Barrios	avg	max	avg	max
	Barrios -> CCZ	avg	max	avg	max
	CCZ -> All	avg	max	avg	max
<b>DíaImportante</b>	DiasImportantes -> Tipo	-	-	-	-
	Tipo -> All	-	-	-	-

Figura 5.61: Estudio Aditividad - Industrias Contaminación

## Días Importantes Contaminación

En figura 5.62 y figura 5.63 se presenta el estudio de aditividad para la relación Industrias Contaminación.

		DiasImportantesContaminación	
		promContaminacion	maxContaminacion
<b>Horas</b>	Horas -> RangoHora	avg	max
	RangoHora -> All	avg	max
<b>Tiempo</b>	Fecha -> Mes	avg	max
	Mes -> Trimestre	avg	max
	Trimestre -> Semestre	avg	max
	Semestre -> Año	avg	max
	Año -> All	avg	max
<b>EstacionesDelAño</b>	EstacionesDelAño -> All	avg	max
<b>DiasSemana</b>	DiasSemana -> TipoDiasSemana	avg	max
	TipoDiasSemana -> All	avg	max
<b>Métodos</b>	Método -> All	avg	max
<b>Contaminantes</b>	Contaminante -> Subtipo	N/A	N/A
	Subtipo -> Tipo	N/A	N/A
	Tipo -> All	N/A	N/A
<b>Contaminantes Categorías</b>	Contaminante Categoría -> All	avg	max

Figura 5.62: Estudio Aditividad - Días Importantes Contaminación

		DiasImportantesContaminación	
		promContaminacion	maxContaminacion
<b>EstacionesDeLaRed</b>	EstacionDeLaRed -> Barrio	avg	max
	Barrio -> CentroComunal	avg	max
	CentroComunal -> All	avg	max
	EstacionDeLaRed -> TipoEquipamiento	avg	max
	TipoEquipamiento -> All	avg	max
	EstacionDeLaRed -> Objetivo	avg	max
	Objetivo -> All	avg	max
<b>RadiacionSolarGlobal</b>	RadiacionSolarGlobal -> All	avg	max
<b>HumedadRelativa</b>	HumedadRelativa -> All	avg	max
<b>TemperaturaExterna</b>	TemperaturaExterna -> All	avg	max
<b>VientoVelocidad</b>	VientoVelocidad -> All	avg	max
<b>VientoDireccion</b>	VientoDireccion -> All	avg	max
<b>Camaras</b>	Camaras -> Barrios	-	-
	Barrios -> CCZ	-	-
	CCZ -> All	-	-
<b>Industrias</b>	Industrias -> Tipo	-	-
	Tipo -> All	-	-
	Industrias -> Barrios	-	-
	Barrios -> CCZ	-	-
	CCZ -> All	-	-
<b>Dialimportante</b>	DiasImportantes -> Tipo	avg	max
	Tipo -> All	avg	max

Figura 5.63: Estudio Aditividad - Días Importantes Contaminación

### 5.2.2. Diseño Lógico

En la etapa de Diseño Lógico se diseña la base de datos relacional que almacenará el DW.

Como puntos de entrada se toman:

1. Modelo Conceptual generado en el Diseño Conceptual.
2. Características de las Fuentes de Datos presentadas en la etapa de Análisis.
3. Estrategias para resolver los requerimientos de performance y almacenamiento indicados en la Especificación de Requerimientos No Funcionales.

El enfoque tomado para la realización del Diseño Lógico es el pasaje/traducción del Modelo Conceptual Multi-Dimensional al Modelo Lógico. Los pasos a realizar para construir las estructuras del DW son los siguientes:

1. Traducir al modelo lógico, las dimensiones a tablas de dimensión y las relaciones dimensionales a tablas de hecho.
2. Incorporar al modelo conceptual los requerimientos no funcionales:
  - a) Operaciones críticas.
  - b) Frecuencias de operaciones.
  - c) Volúmenes de datos.

Basados en el Modelo Dimensional presentado por Kimball, en el modo de almacenamiento ROLAP, ya que así lo requería la herramienta Pentaho utilizada para la implementación del proyecto, y en los aspectos y características del sistema de BI en cuanto a frecuencia y criticidad de operaciones y volúmenes de datos, se presenta el modelo lógico para cada una de las Relaciones Dimensionales.

El esquema utilizado para definir la estructura de la base de datos relacional del DW para cada relación fue el Star Cluster, combinación de los esquemas Star y Snowflake. A su vez, debido a que se identifican dimensiones compartidas entre los Star Cluster, se agrupan las mismas para generar una Constelación.

La figura 5.64 muestra el modelo lógico donde se comparte la tabla de Barrios y CCZ entre las tablas de dimensión de Cámaras, Industrias y Estaciones de la Red, que forman parte de los esquemas de cada requerimiento.

Teniendo en cuenta las Relaciones Dimensionales presentadas en el Diseño Conceptual, se deben seguir los siguientes pasos para cada una:

1. Identificar las tablas de dimensión y la tabla de hecho. Las tablas de dimensión se encuentran en su mayoría desnormalizadas, al contener sus jerarquías embebidas, menos las que son compartidas entre distintas relaciones dimensionales.
2. Unir la tabla de hecho con cada tabla de dimensión por una relación 1:N y asignar como clave de la tabla de hecho la unión de las claves de las dimensiones relacionadas.

Durante la etapa de diseño conceptual se definieron tres relaciones dimensionales: Vehículos Contaminación, Industrias Contaminación y Días Importantes Contaminación. A continuación se detalla el resultado de traducir cada relación dimensional a su correspondiente modelo lógico utilizando el procedimiento descrito.

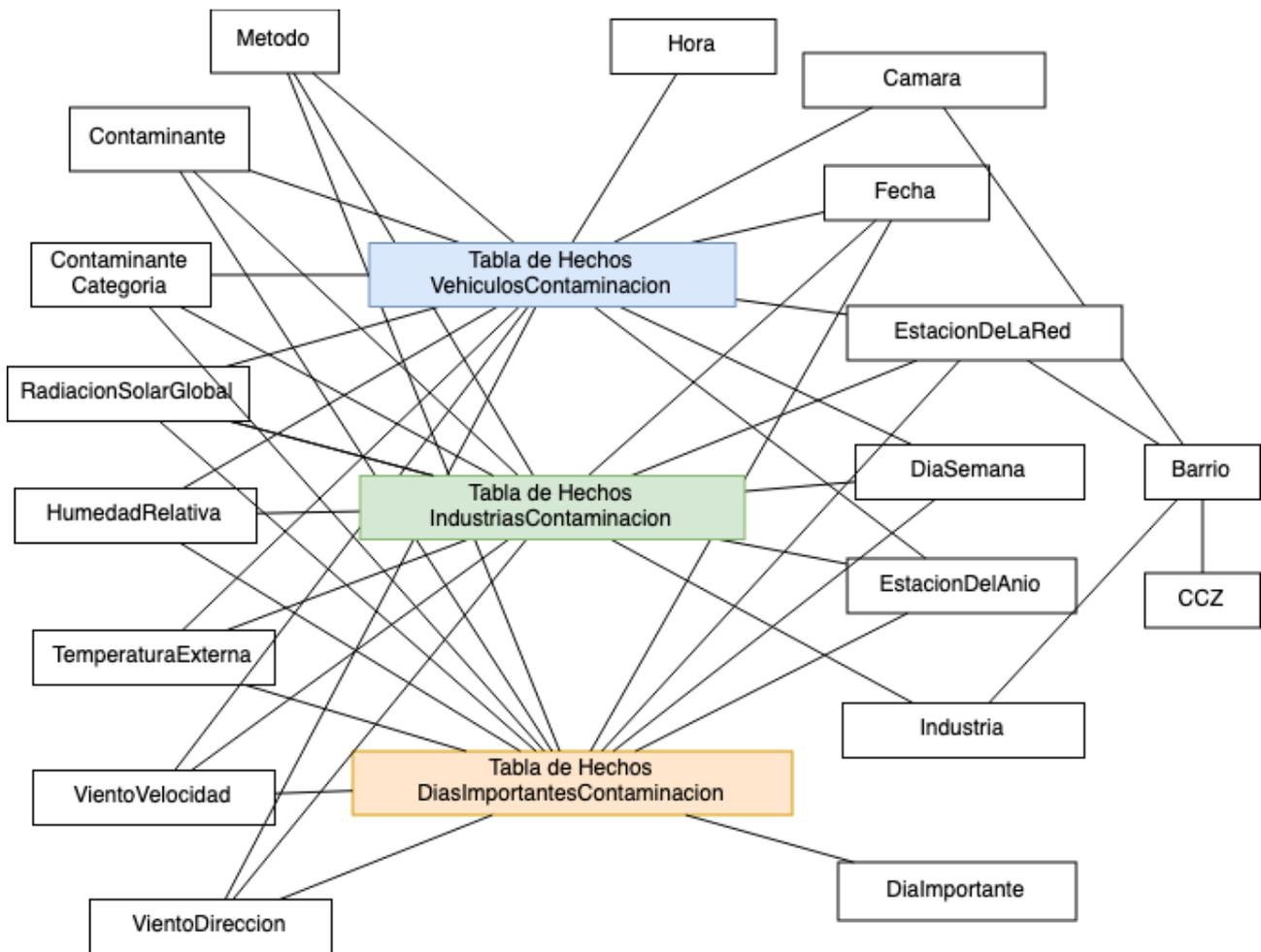


Figura 5.64: Diseño Lógico - Star Cluster

## Vehículos Contaminación

Como se puede apreciar en la figura 5.65, el modelo lógico que se obtiene a partir de la relación dimensional Vehículos Contaminación se estructura en torno a la tabla de hecho *Vehiculos-Contaminacion*. En cuanto a las medidas, existen atributos que sirven a efectos del análisis definido para la relación dimensional. El atributo *contaminacion* representa el valor de una medida histórica de calidad de aire para un momento en específico, mientras que el atributo *volumenVehiculos* corresponden a la suma de todos los vehículos de la hora anterior a la fecha y hora del registro. En cuanto al resto de los atributos son las claves foráneas a las tablas dimensionales relacionadas, estas representan los objetos específicos relacionados a cada valor como el contaminante que se mide, las unidades, el método de medición y la cámara vehicular.

Asociado a la tabla de hechos, se pueden visualizar las distintas tablas de dimensión, de las cuales es fácil notar que en su mayoría contienen los atributos de todas sus jerarquías embebidos en ellas. Como ejemplo se puede observar la tabla de la dimensión *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*. La única tabla de dimensión que no se encuentra desnormalizada es la *Barrios* y *CCZ* debido a que es compartida por múltiples jerarquías de otras dimensiones como *Estaciones de la Red* y *Cámaras*.

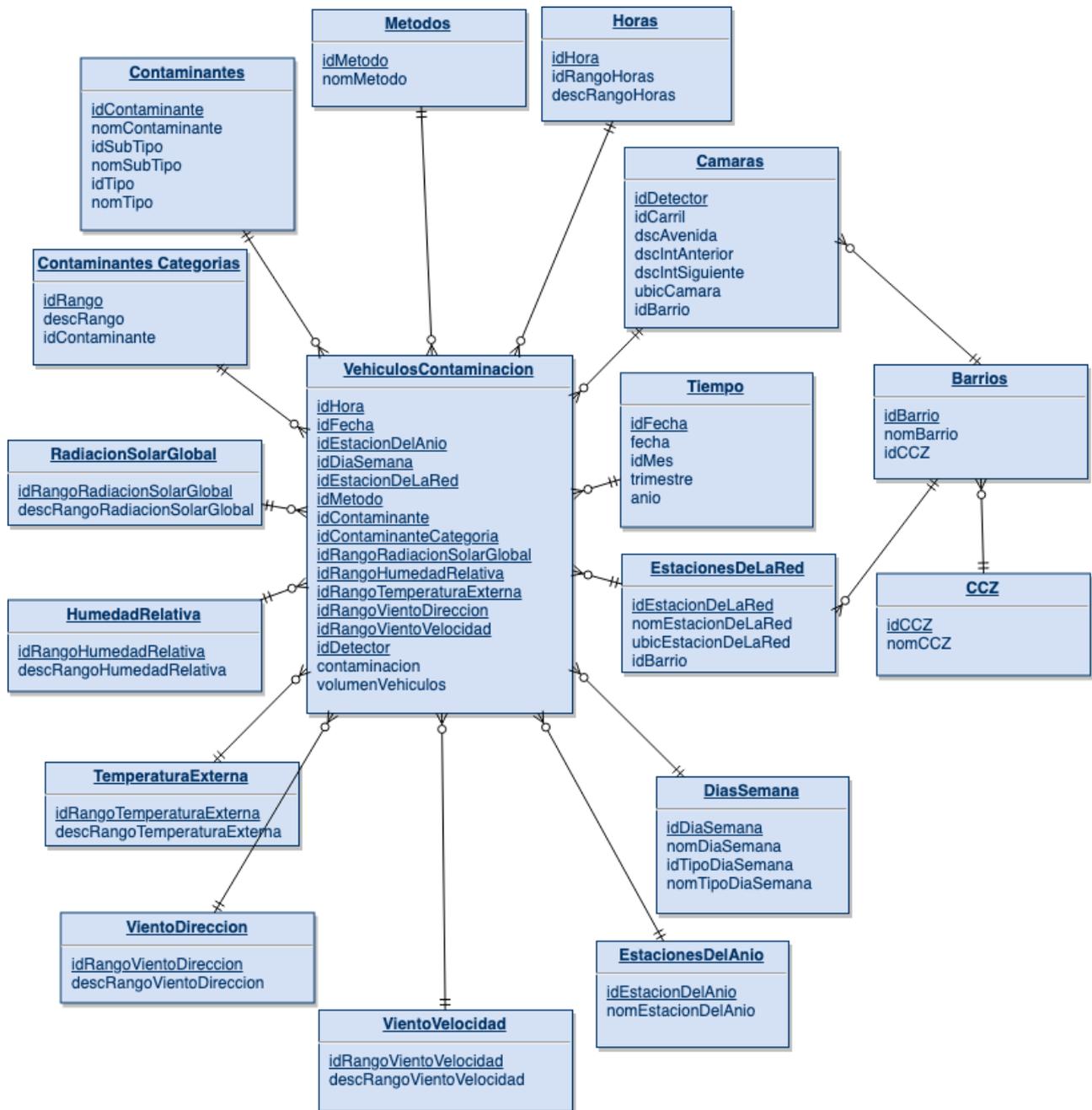


Figura 5.65: Diseño Lógico - Vehículos Contaminación

## Industrias Contaminación

Como se puede apreciar en la figura 5.66, el modelo lógico que se obtiene a partir de la relación dimensional Industrias Contaminación se estructura en torno a la tabla de hecho *Industrias-Contaminacion*. En cuanto a las medidas, existen atributos que sirven a efectos del análisis definido para la relación dimensional. El atributo *contaminacion* representa el valor de una medida histórica de calidad de aire promedio para el día, mientras que el atributo *actividadIndustrial* corresponde al nivel de actividad industrial promedio del día. En cuanto al resto de los atributos son las claves foráneas a las tablas dimensionales relacionadas, estas representan los objetos específicos relacionados a cada valor como el contaminante que se mide, las unidades, el método de medición y la industria.

Asociado a la tabla de hechos, se pueden visualizar las distintas tablas de dimensión, de las cuales es fácil notar que en su mayoría contienen los atributos de todas sus jerarquías embebidos en ellas. Como ejemplo se puede observar la tabla de la dimensión *Contaminantes* que contiene los atributos *idContaminante* y *nomContaminante* pertenecientes al nivel Contaminante, *idSubTipo* y *nomSubTipo* pertenecientes al nivel SubTipo e *idTipo* y *nomTipo* perteneciente al nivel Tipo. La única tabla de dimensión que no se encuentra desnormalizada es la *Barrios* y *CCZ* debido a que es compartida por múltiples jerarquías de otras dimensiones como *Estaciones de la Red* y en dimensiones de otras relaciones dimensionales como *Cámaras*.

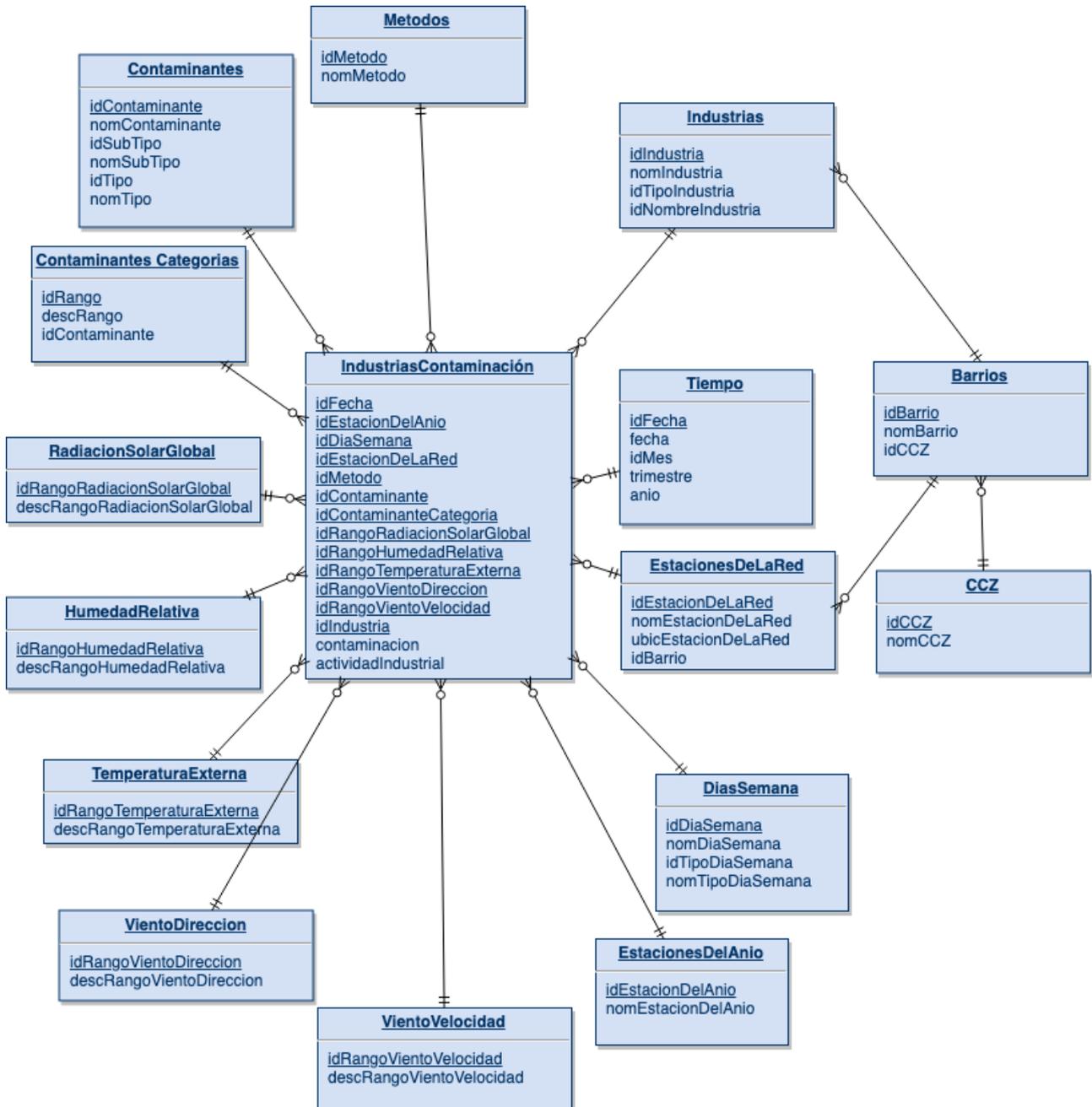


Figura 5.66: Diseño Lógico - Industrias Contaminación

## Días Importantes Contaminación

Como se puede apreciar en la figura 5.67, el modelo lógico que se obtiene a partir de la relación dimensional Días Importantes Contaminación se estructura en torno a la tabla de hecho *DíasImportantesContaminacion*. En cuanto a las medidas, existen atributos que sirven a efectos del análisis definido para la relación dimensional. El atributo *contaminacion* representa el valor de una medida histórica de calidad de aire promedio para el día y el único atributo en la medida. En cuanto al resto de los atributos son las claves foráneas a las tablas dimensionales relacionadas, estas representan los objetos específicos relacionados a cada valor como el contaminante que se mide, las unidades, el método de medición y el día importante específico.

De forma análoga a los casos anteriores, asociado a la tabla de hechos, se pueden visualizar las distintas tablas de dimensión, de las cuales es fácil notar que en su mayoría contienen los atributos de todas sus jerarquías embebidos en ellas, menos la jerarquía *Estaciones de la Red, Barrios y CCZ*.

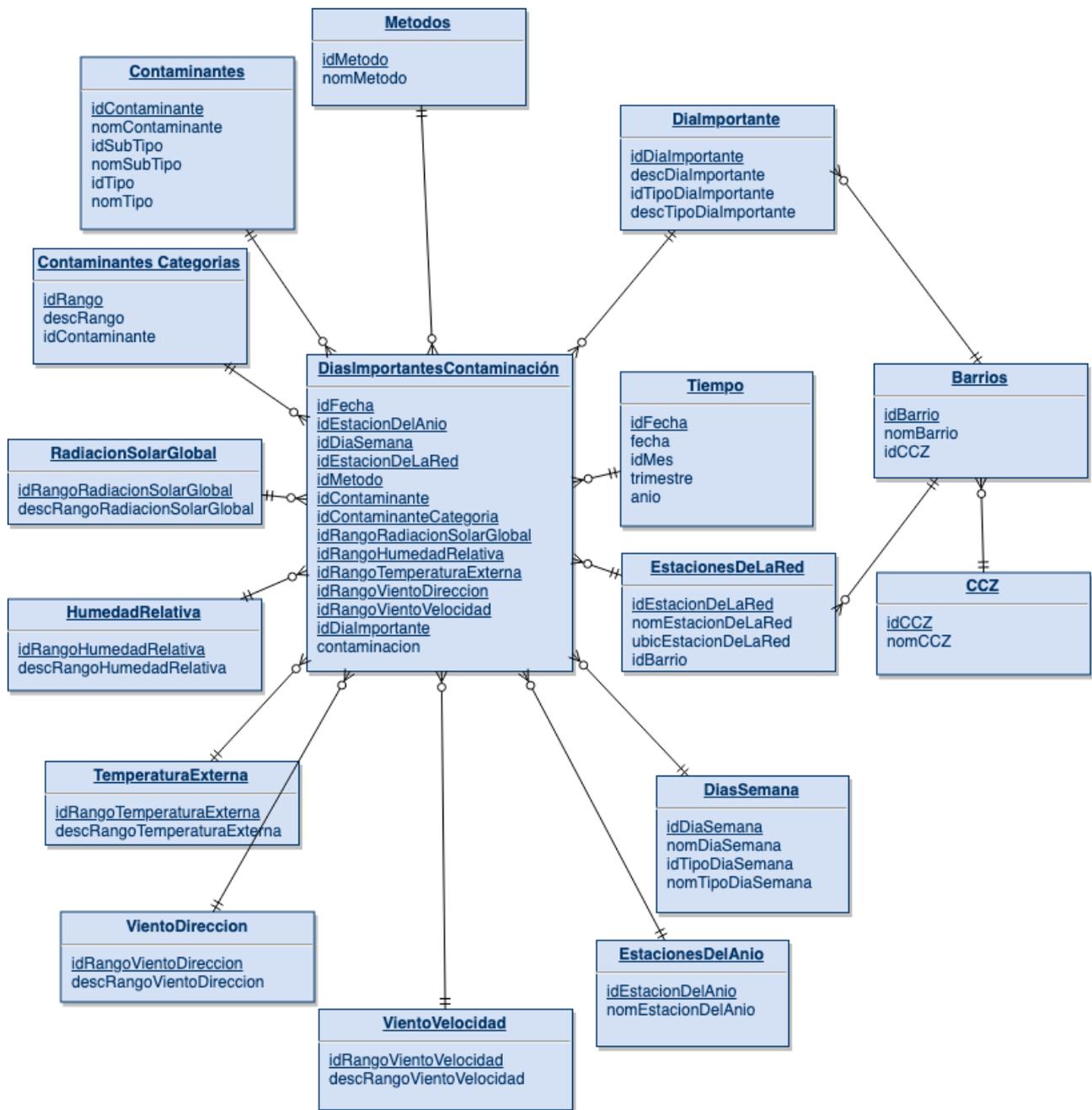


Figura 5.67: Diseño Lógico - Días Importantes Contaminación

### 5.2.3. Diseño Físico

En la etapa de Diseño Físico se presentan las decisiones relacionadas al manejador de base de datos que se utilizará para el DW tomando en cuenta las características deseables y necesarias para los requerimientos.

Como se especificó en los requerimientos no funcionales, el manejador de bases de datos elegido es PostgreSQL. A su vez, se realiza la instalación de la extensión PostGIS para convertirla en una base de datos geográfica necesaria para almacenar los datos espaciales y utilizar funciones geográficas requeridas para los procesos ETL.

Con los objetivos de lograr un diseño de base de datos eficiente y eficaz, teniendo en cuenta

los tipos de datos que se almacenarán y los requerimientos de carga de trabajo y tiempos de respuesta requeridos, se realiza una investigación a fondo de los aspectos de la base de datos seleccionada para cumplir los objetivos.

PostgreSQL provee varios tipos de datos con los que se puede trabajar, facilitando de esta forma el diseño eficiente de la base de datos [33].

Name	Aliases	Description
<b>bigint</b>	<b>int8</b>	signed eight-byte integer
<b>bigserial</b>	<b>serial8</b>	autoincrementing eight-byte integer
<b>boolean</b>	<b>bool</b>	logical Boolean (true/false)
<b>bytea</b>		binary data ("byte array")
<b>character [ (n) ]</b>	<b>char [ (n) ]</b>	fixed-length character string
<b>character varying [ (n) ]</b>	<b>varchar [ (n) ]</b>	variable-length character string
<b>date</b>		calendar date (year, month, day)
<b>double precision</b>	<b>float8</b>	double precision floating-point number (8 bytes)
<b>integer</b>	<b>int, int4</b>	signed four-byte integer
<b>numeric [ (p, s) ]</b>	<b>decimal [ (p, s) ]</b>	exact numeric of selectable precision
<b>real</b>	<b>float4</b>	single precision floating-point number (4 bytes)
<b>smallint</b>	<b>int2</b>	signed two-byte integer
<b>smallserial</b>	<b>serial2</b>	autoincrementing two-byte integer
<b>serial</b>	<b>serial4</b>	autoincrementing four-byte integer
<b>text</b>		variable-length character string
<b>time [ (p) ] [ without time zone ]</b>		time of day (no time zone)
<b>time [ (p) ] with time zone</b>	<b>timetz</b>	time of day, including time zone
<b>timestamp [ (p) ] [ without time zone ]</b>		date and time (no time zone)
<b>timestamp [ (p) ] with time zone</b>	<b>timestamptz</b>	date and time, including time zone

Figura 5.68: PostgreSQL - Data Types

Adicionalmente, se definieron múltiples claves subrogadas. 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 DW, de tipo numérico sin signo y no tienen ningún significado de negocio en la base de datos fuente. Se presentan las claves subrogadas más relevantes definidas en las tablas de dimensión:

1. *idRango* de la tabla *ContaminantesCategorias*.
2. *idDiaImportante* de la tabla *DiasImportantes*.
3. *idRangoRadiacionSolarGlobal*, *idRangoHumedadRelativa*, *idRangoTemperaturaExterna*, *idRangoVientoDireccion*, *idRangoVientoVelocidad* de las tablas *RangoRadiacionSolarGlobal*, *RangoHumedadRelativa*, *RangoTemperaturaExterna*, *RangoVientoDireccion*, *RangoVientoVelocidad* respectivamente.

# Capítulo 6

## Implementación

Este capítulo engloba aspectos de implementación que incluyen los procesos de Extracción, Transformación y Carga (ETL), los cubos multidimensionales definidos, las herramientas de software utilizados, la generación automática del reporte anual y la interfaz de usuario utilizada por los especialistas.

### 6.1. Extracción, Transformación y Carga (ETL)

Los procesos ETL son cruciales en la integración de datos, en definitiva, es la fase de la etapa de Implementación que requiere una mayor dedicación, esfuerzos y recursos para su realización.

La presentación de los procesos ETL se divide en dos secciones; la primera para la carga de las tablas de dimensión y la segunda para la carga de las tablas de hecho del DW.

La misma se realiza con un esquema que abstrae las actividades específicas de cada proceso, encontrándose la especificación completa en el anexo del capítulo C.

Para lograr un mejor entendimiento de cada etapa se presentan utilizando colores con el siguiente significado:

1. **Azul:** Paso inicial. Obtener datos desde un archivo de entrada.
2. **Amarillo:** Paso intermedio. Consolidación, cálculo o generación de campos.
3. **Verde:** Paso final. Almacenamiento en la base de datos del DW.
4. **Violeta:** Paso inicial. Obtener datos desde la base de datos del DW.

#### Tablas de Dimensión

##### Estaciones de la Red

En la figura 6.1 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Estaciones de la Red.

##### *Obtener Estaciones de la Red*

En primera instancia se extraen los datos del archivo CSV fuente de Estaciones de la Red.

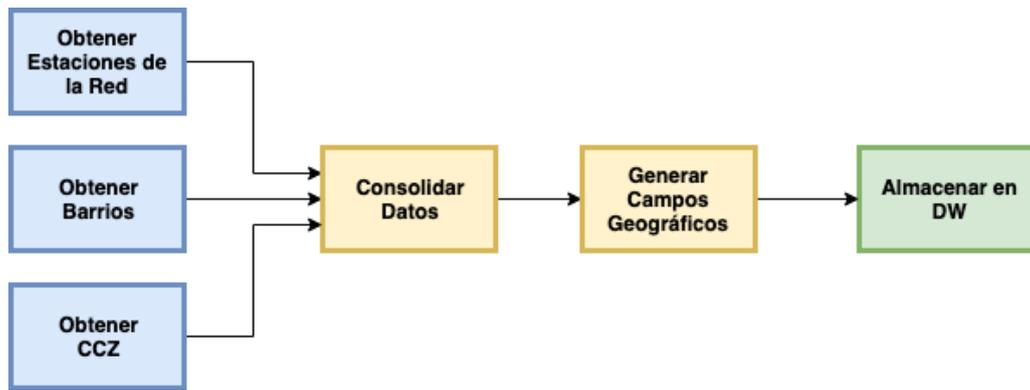


Figura 6.1: Carga Estaciones De La Red

#### *Obtener Barrios*

Se obtienen los Barrios de Montevideo desde la fuente de datos obtenida.

#### *Obtener CCZ*

Se obtienen los Centros Comunales Zonales de Montevideo desde la fuente de datos obtenida.

#### *Consolidar Datos*

Se asocian los datos de Estaciones de la Red con los Barrios y los CCZ de Montevideo.

#### *Generar Campos Geográficos*

Basado en los datos de ubicación de la estación de la red se genera el campo de la base de datos geográfica para su posterior uso en los ETLs.

#### *Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.estacionesdelared*.

### **Contaminantes**

En la figura 6.2 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Contaminantes.

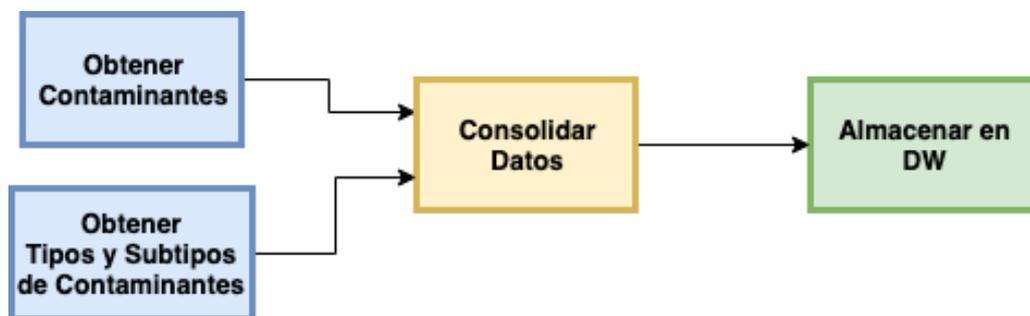


Figura 6.2: Carga Contaminantes

#### *Obtener Contaminantes*

En primera instancia se extraen los datos del archivo CSV fuente de Contaminantes.

### Obtener Tipos y Subtipos de Contaminantes

Se obtienen los datos de los distintos tipos y subtipos de contaminantes.

### Consolidar Datos

Se asocian los datos de Contaminantes con los tipos y subtipos de los contaminantes.

### Almacenar en DW

En última instancia, se guardan los datos generados en la tabla *dw.contaminantes*.

## Contaminantes Categorías

En la figura 6.3 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Contaminantes Categorías.

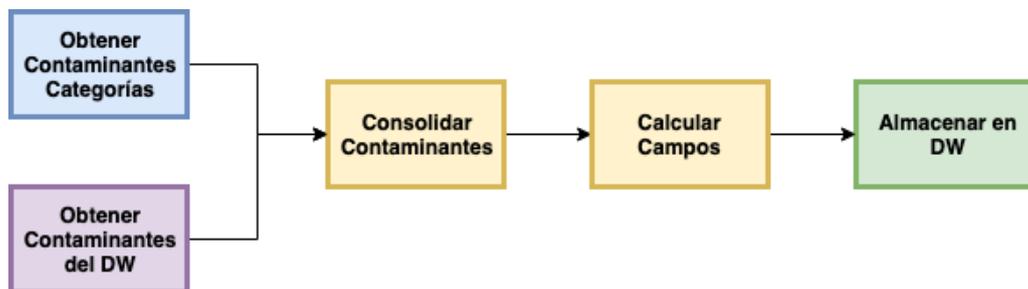


Figura 6.3: Carga Contaminantes Categorías

### Obtener Categorías Contaminantes

En primera instancia se obtienen los datos de las distintas franjas para categorías de calidad de aire según el valor de contaminación.

Se puede visualizar en la figura 6.4 las categorías para cada contaminante.

Lapso muestreo en horas	PTS	PM10	Humo Negro	SO2	NO2	CO	O3	Indice	Calidad
	ug/m <sup>3</sup>								
	24	24	24	24	1	8	8		
Nivel 1	60	50	50	20	40	4,5	80	25	Muy Buena
Nivel 2	100	75	75	50	75	7	100	50	Buena
Nivel 3	150	100	100	125	200	10	160	100	Aceptable
Nivel 4	375	150	150	365	500	15	240	200	Inadecuada
Nivel 5	563	225	225	550	1130	22	540	300	Mala
Nivel 6	845	340	340	825	2260	33	1080	400	Muy mala

Figura 6.4: Carga Contaminantes Obtener Categorías Contaminantes

### Obtener Contaminantes del DW

Se obtienen los datos de la tabla de dimensión Contaminantes con el objetivo de ser utilizado posteriormente.

### Consolidar Datos

Se asocian los datos de Categorías Contaminantes con los datos de Contaminantes para almacenar el identificador que los relaciona.

*Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.contaminantescategorias*.

### Métodos

En la figura 6.5 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Métodos.

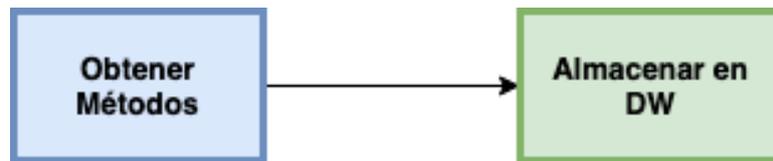


Figura 6.5: Carga Métodos

*Obtener Métodos*

En primera instancia se extraen los datos del archivo CSV fuente de Métodos.

*Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.metodos*.

### Días de la Semana

En la figura 6.6 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Días de la Semana.

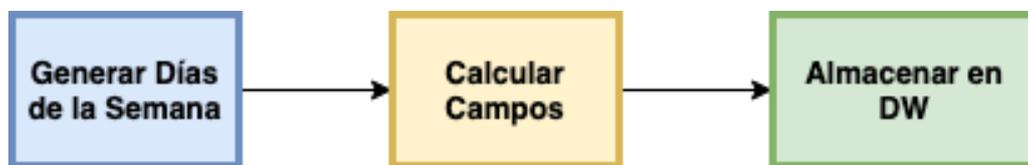


Figura 6.6: Carga Días de la Semana

*Generar Días de la Semana*

En primera instancia, se generan 7 filas (cantidad de días que hay en 1 semana). Se define para cada fila un único identificador secuencial que comienza en 1 y finaliza en 7 para diferenciar los días.

*Calcular Campos*

Una vez generados los días de la semana, se calculan los demás campos que forman parte de la tabla de dimensión. Para identificar el día de la semana se utiliza una función de Date para poder definir el día de la semana específico.

*Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.diassemana*.

### Estaciones del Año

En la figura 6.7 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Estaciones del Año.

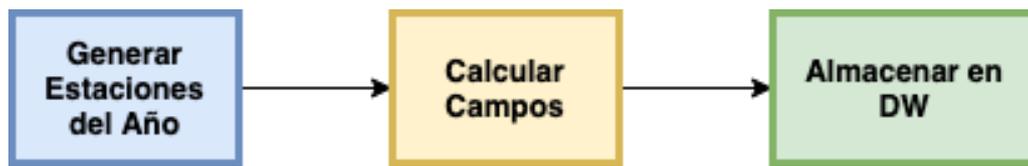


Figura 6.7: Carga Estaciones del Año

#### *Generar Estaciones del Año*

En primera instancia, se generan 4 filas (cantidad de estaciones que hay en 1 año). Se define para cada fila un único identificador secuencial que comienza en 1 y finaliza en 4 para diferenciar las estaciones del año.

#### *Calcular Campos*

Una vez generadas las estaciones del año, se calculan los demás campos que forman parte de la tabla de dimensión. Para identificar la estación del año se utiliza una función de Date para poder definir la estación del año específica.

#### *Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.estacionesdelanio*.

### **Tiempo**

En la figura 6.8 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Tiempo.

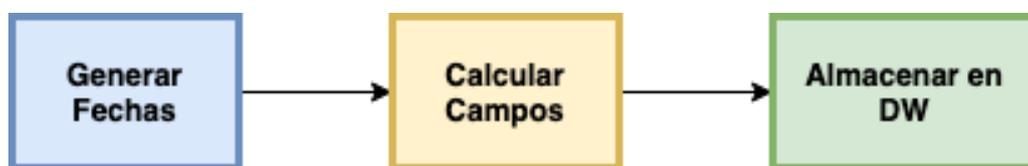


Figura 6.8: Carga Tiempo

#### *Generar Fechas*

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 que la Red de Monitoreo comenzó a operar en 2005 pero existen datos desde 2003, es 01/01/2000.

#### *Calcular Campos*

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 formato AAAAMMDD, donde AAAA representa el año, MM el mes y DD el día. En cuanto al nivel Mes se calculan los dos campos: el identificador, que es un número de 6 dígitos con el 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.

#### *Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.tiempo*.

## Horas

En la figura 6.9 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Horas.

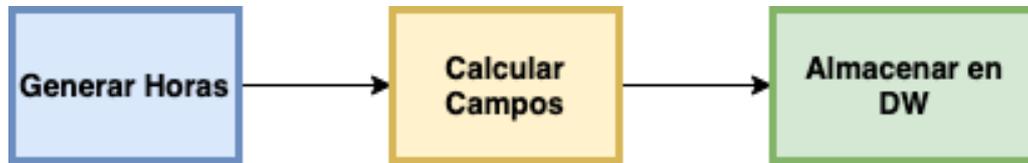


Figura 6.9: Carga Horas

### *Generar Horas*

En primera instancia, se generan 24 filas (cantidad de horas que hay en 1 día). Se define para cada fila un único identificador secuencial que comienza en 0 y finaliza en 23 para diferenciar las horas.

### *Calcular Campos*

Una vez generadas las horas, se calculan los demás campos que forman parte de la jerarquía. Para identificar los rangos de horas según la parte del día se realiza el cálculo basado en el cuadro 6.1.

Hora	Descripción
00 - 06	Madrugada
07 - 12	Mañana
13 - 19	Tarde
20 - 23	Noche

Cuadro 6.1: Rango de Horas

### *Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.horas*.

## Radiación Solar Global

En la figura 6.10 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Radiación Global Solar.

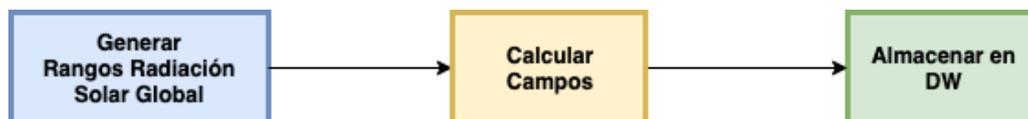


Figura 6.10: Carga Radiación Global Solar

### *Generar Rangos Radiación Global Solar*

En primera instancia, se generan 3 filas (número de rangos) para identificar cada rango. Se define para cada fila un único identificador secuencial que comienza en 1 y finaliza en 3 para diferenciar los rangos.

### Calcular Campos

Una vez generados los rangos, se calculan los demás campos que forman parte de la tabla de dimensión. Para identificar los rangos de radiación según sus valores se realiza el cálculo basado en el cuadro 6.2.

Valor	Descripción
0 - 250	Bajo
750 - 1000	Medio
1000 - 1500	Alto

Cuadro 6.2: Rango de Radiación Solar Global

### Almacenar en DW

En última instancia, se guardan los datos generados en la tabla *dw.radiacionsolarglobal*.

### Humedad Relativa

En la figura 6.11 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Humedad Relativa.

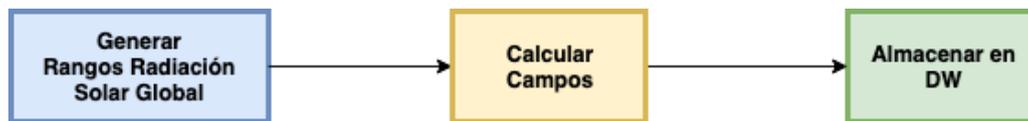


Figura 6.11: Carga Humedad Relativa

### Generar Rangos Humedad Relativa

En primera instancia, se generan 3 filas (número de rangos) para identificar cada rango. Se define para cada fila un único identificador secuencial que comienza en 1 y finaliza en 3 para diferenciar los rangos.

### Calcular Campos

Una vez generados los rangos, se calculan los demás campos que forman parte de la tabla de dimensión. Para identificar los rangos de humedad relativa según sus valores se realiza el cálculo basado en el cuadro 6.3.

Valor	Descripción
0 % - 40 %	Baja
40 % - 70 %	Media
70 % - 100 %	Alta

Cuadro 6.3: Rango de Humedad Relativa

### Almacenar en DW

En última instancia, se guardan los datos generados en la tabla *dw.humedadrelativa*.

### Temperatura Externa

En la figura 6.12 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Temperatura Externa.

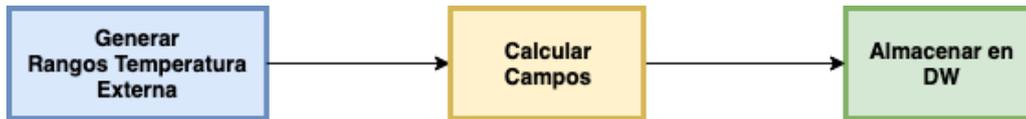


Figura 6.12: Carga Temperatura Externa

#### *Generar Rangos Temperatura Externa*

En primera instancia, se generan 4 filas (número de rangos) para identificar cada rango. Se define para cada fila un único identificador secuencial que comienza en 1 y finaliza en 4 para diferenciar los rangos.

#### *Calcular Campos*

Una vez generados los rangos, se calculan los demás campos que forman parte de la tabla de dimensión. Para identificar los rangos de temperatura externa según sus valores se realiza el cálculo basado en el cuadro 6.4.

Valor	Descripción
Hasta 5	Baja
5 - 20	Media
20 - 35	Alta
Desde 35	Muy Alta

Cuadro 6.4: Rango de Temperatura Externa

#### *Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.temperaturaexterna*.

### **Viento Dirección**

En la figura 6.13 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Viento Dirección.

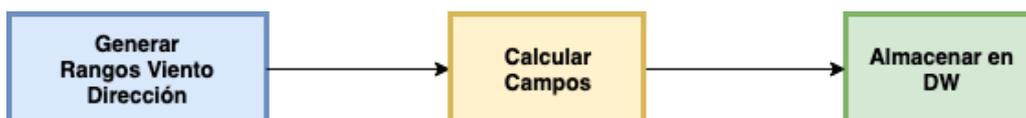


Figura 6.13: Carga Viento Dirección

#### *Generar Rangos Viento Dirección*

En primera instancia, se generan 8 filas (número de rangos) para identificar cada rango. Se define para cada fila un único identificador secuencial que comienza en 1 y finaliza en 8 para diferenciar los rangos.

#### *Calcular Campos*

Una vez generados los rangos, se calculan los demás campos que forman parte de la tabla de dimensión. Para identificar los rangos de la dirección del viento según su rumbo se realiza el cálculo basado en el cuadro 6.5.

#### *Almacenar en DW*

Valor	Descripción
337.5° - 22.5°	N
22.5° - 67.5°	NE
67.5° - 112.5°	E
112.5° - 157.5°	SE
157.5° - 202.5°	S
202.5° - 247.5°	SW
247.5° - 292.5°	W
292.5° - 337.5°	NW

Cuadro 6.5: Rango de Viento Dirección

En última instancia, se guardan los datos generados en la tabla *dw.vientodireccion*.

### Viento Velocidad

En la figura 6.14 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Viento Velocidad.

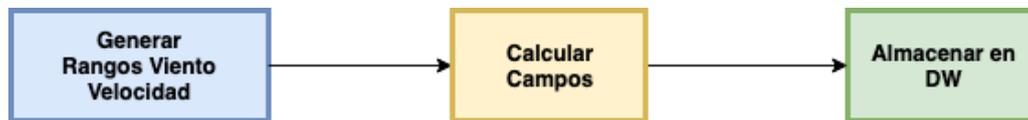


Figura 6.14: Carga Viento Velocidad

#### *Generar Rangos Viento Velocidad*

En primera instancia, se generan 13 filas (número de rangos) para identificar cada rango. Se define para cada fila un único identificador secuencial que comienza en 1 y finaliza en 13 para diferenciar los rangos.

#### *Calcular Campos*

Una vez generados los rangos, se calculan los demás campos que forman parte de la tabla de dimensión. Para identificar los rangos de la velocidad del viento según su intensidad se realiza el cálculo basado en el cuadro 6.6.

#### *Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.vientovelocidad*.

Valor	Descripción
0-1	Calma
2-5	Ventolina
6-11	Brisa muy débil
12-19	Brisa ligera
20-28	Brisa moderada
29-38	Brisa fresca
39-49	Brisa fuerte
50-61	Viento fuerte
62-74	Viento duro
75-88	Viento muy duro
89-102	Temporal
103-117	Borrasca
118 >	Huracán

Cuadro 6.6: Rango de Viento Velocidad

## Cámaras

En la figura 6.15 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Cámaras.

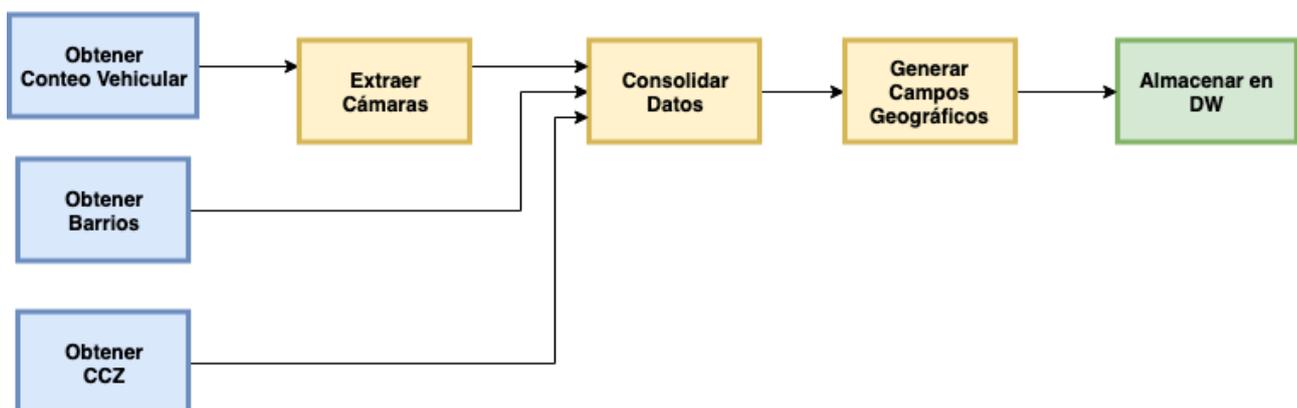


Figura 6.15: Carga Cámaras

### *Obtener Conteo Vehicular*

En primera instancia se extraen los datos del archivo CSV fuente de Conteo Vehicular.

### *Extraer Cámaras*

Se agrupan los datos para obtener las cámaras que se encuentran asociados a los valores de conteo vehicular. Para ello, se utiliza como clave los datos de identificación de las cámaras.

### *Obtener Barrios*

Se obtienen los Barrios de Montevideo desde la fuente de datos obtenida.

### *Obtener CCZ*

Se obtienen los Centros Comunales Zonales de Montevideo desde la fuente de datos obtenida.

### *Consolidar Datos*

Se asocian los datos de Cámaras con los Barrios y los CCZ de Montevideo.

### *Generar Campos Geográficos*

Basado en los datos de ubicación de la cámara se genera el campo de la base de datos geográfica para su posterior uso en los ETLs.

### *Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.camaras*.

## **Industrias**

En la figura 6.16 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Industrias.

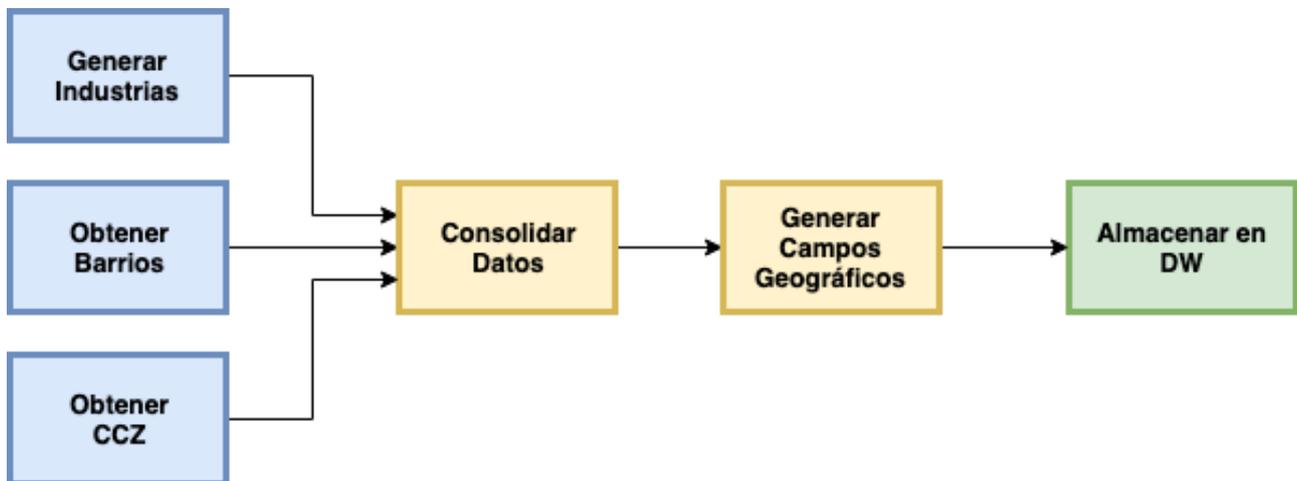


Figura 6.16: Carga Industrias

### *Generar Industrias*

Con el objetivo de cargar industrias a medida que se obtienen datos de actividad, y en una primer instancia se obtienen datos de ADME, se genera 1 fila para ingresarla. Se adicionan datos de descripción y ubicación geográfica.

### *Obtener Barrios*

Se obtienen los Barrios de Montevideo desde la fuente de datos obtenida.

### *Obtener CCZ*

Se obtienen los Centros Comunales Zonales de Montevideo desde la fuente de datos obtenida.

### *Consolidar Datos*

Se asocian los datos de Industrias con los Barrios y los CCZ de Montevideo.

### *Generar Campos Geográficos*

Basado en los datos de ubicación de la industria se genera el campo de la base de datos geográfica para su posterior uso en los ETLs.

*Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.industrias*.

### **Días Importantes**

En la figura 6.17 se ilustra, en términos generales, el proceso de carga de la tabla de dimensión Días Importantes.

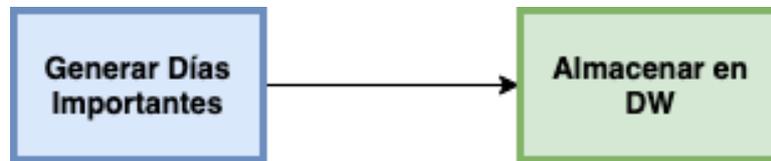


Figura 6.17: Carga Días Importantes

*Generar Días Importantes*

Con el objetivo de cargar días importantes a medida que se identifican, en una primer instancia se obtienen datos días festivos en Uruguay y se generan tantas filas para ingresarlos. Se adicionan datos de descripción y su tipo. Posteriormente los especialistas de la IM podrán cargar los días importantes que deseen.

*Almacenar en DW*

En última instancia, se guardan los datos generados en la tabla *dw.diasimportantes*.

## Tablas de Hecho

### **Vehículos Contaminación**

En la figura 6.18 se ilustra, en términos generales, el proceso de carga de la tabla de hechos Vehículos Contaminación.

#### *Obtener Conteo Vehicular*

Se extraen los datos del archivo CSV fuente de Conteo Vehicular.

#### *Obtener Cámaras del DW*

Se obtienen los datos de la tabla de dimensión Cámaras con el objetivo de ser utilizado posteriormente.

#### *Obtener Mediciones Históricas*

Se extraen los datos del archivo CSV de Mediciones Históricas.

#### *Obtener Contaminantes, Contaminantes Categorías, Métodos y Estaciones de la Red del DW*

Se obtienen los datos de las distintas dimensiones relacionadas a las Mediciones Históricas con el objetivo de ser utilizado posteriormente.

#### *Obtener Datos Meteorológicos*

Se extraen los datos del archivo XLS provisto por ANCAP de Datos Meteorológicos obtenidos en la refinería de ANCAP.

#### *Obtener Radiación Solar Global, Humedad Relativa, Temperatura Externa, Viento Dirección y Viento Velocidad del DW*

Se obtienen los datos de las distintas dimensiones relacionadas a los Datos Meteorológicos con el objetivo de ser utilizado posteriormente.

#### *Consolidar Datos Conteo Vehicular*

Se asocian los datos obtenidos de Conteo Vehicular con los datos de la tabla de dimensión Cámaras.

#### *Consolidar Datos Mediciones Históricas*

Se asocian los datos obtenidos de Mediciones Históricas con los datos de las distintas dimensiones relacionadas.

#### *Consolidar Datos Meteorológicos*

Se asocian los datos Meteorológicos obtenidos con los datos de las distintas dimensiones relacionadas. Cabe destacar que no se realiza un filtrado de los datos meteorológicos teniendo en cuenta la distancia de la fuente de medición de los mismos con las cámaras o estaciones de la red debido a que se indica por los especialistas que los datos medidos son representativos para el departamento de Montevideo en general.

#### *Join por distancia <150m*

Se realiza un join entre los datos obtenidos de Conteo Vehicular y Mediciones Históricas para filtrar y relacionar los datos obtenidos de estaciones de monitoreo y cámaras de conteo vehicular que se encuentren a un radio de más de 150m, como se definió en la etapa de Análisis.

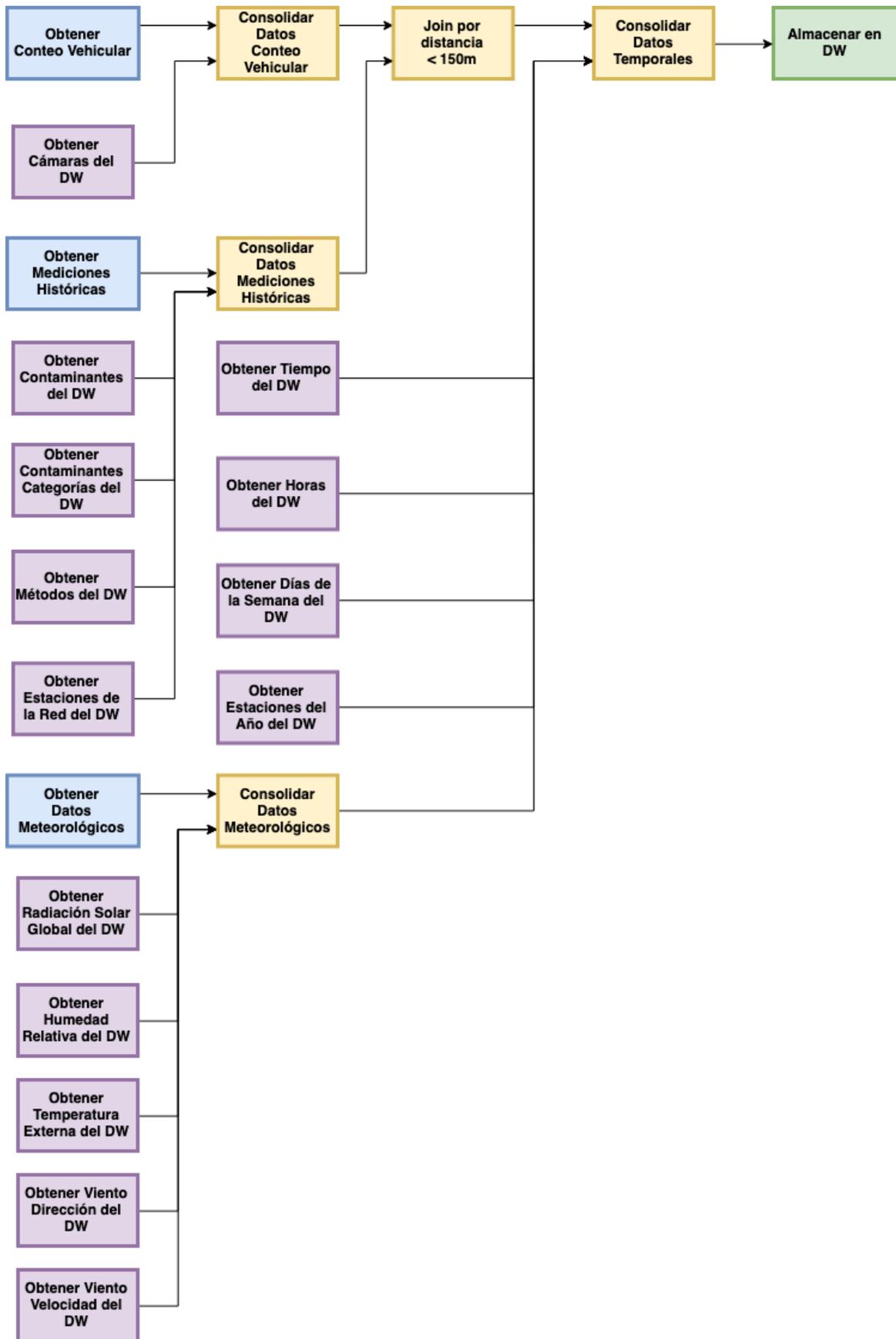


Figura 6.18: Carga Vehículos Contaminación

### *Consolidar Datos Temporales*

Se asocian, los datos relacionados y filtrados de contaminación y volumen vehicular, con los datos temporales para su posterior uso.

### *Almacenar en DW*

En última instancia, se guardan los datos relacionados en la tabla *dw.vehiculoscontaminacion*.

## **Industrias Contaminación**

En la figura 6.19 se ilustra, en términos generales, el proceso de carga de la tabla de hechos Industrias Contaminación.

### *Obtener Actividad Industrial*

Se extraen los datos del archivo CSV fuente de Actividad Industrial.

### *Obtener Industrias del DW*

Se obtienen los datos de la tabla de dimensión Industrias con el objetivo de ser utilizado posteriormente.

### *Obtener Mediciones Históricas*

Se extraen los datos del archivo CSV de Mediciones Históricas.

### *Obtener Contaminantes, Contaminantes Categorías, Métodos y Estaciones de la Red del DW*

Se obtienen los datos de las distintas dimensiones relacionadas a las Mediciones Históricas con el objetivo de ser utilizado posteriormente.

### *Obtener Datos Meteorológicos*

Se extraen los datos del archivo XLS provisto por ANCAP de Datos Meteorológicos obtenidos en la refinería de ANCAP.

### *Obtener Radiación Solar Global, Humedad Relativa, Temperatura Externa, Viento Dirección y Viento Velocidad del DW*

Se obtienen los datos de las distintas dimensiones relacionadas a los Datos Meteorológicos con el objetivo de ser utilizado posteriormente.

### *Consolidar Datos Actividad Industrial*

Se asocian los datos obtenidos de Actividad Industrial con los datos de la tabla de dimensión Industrias.

### *Consolidar Datos Mediciones Históricas*

Se asocian los datos obtenidos de Mediciones Históricas con los datos de las distintas dimensiones relacionadas.

### *Consolidar Datos Meteorológicos*

Se asocian los datos Meteorológicos obtenidos con los datos de las distintas dimensiones relacionadas. Cabe destacar que no se realiza un filtrado de los datos meteorológicos teniendo en cuenta la distancia de la fuente de medición de los mismos con las industrias o estaciones de la red debido a que se indica por los especialistas que los datos medidos son representativos para el departamento de Montevideo en general.

### *Join por distancia < 3km*

Se realiza un join entre los datos obtenidos de Actividad Industrial y Mediciones Históricas para filtrar y relacionar los datos obtenidos de estaciones de monitoreo e industrias de actividad industrial que se encuentren a un radio de más de 3km, como se definió en la etapa de Análisis.

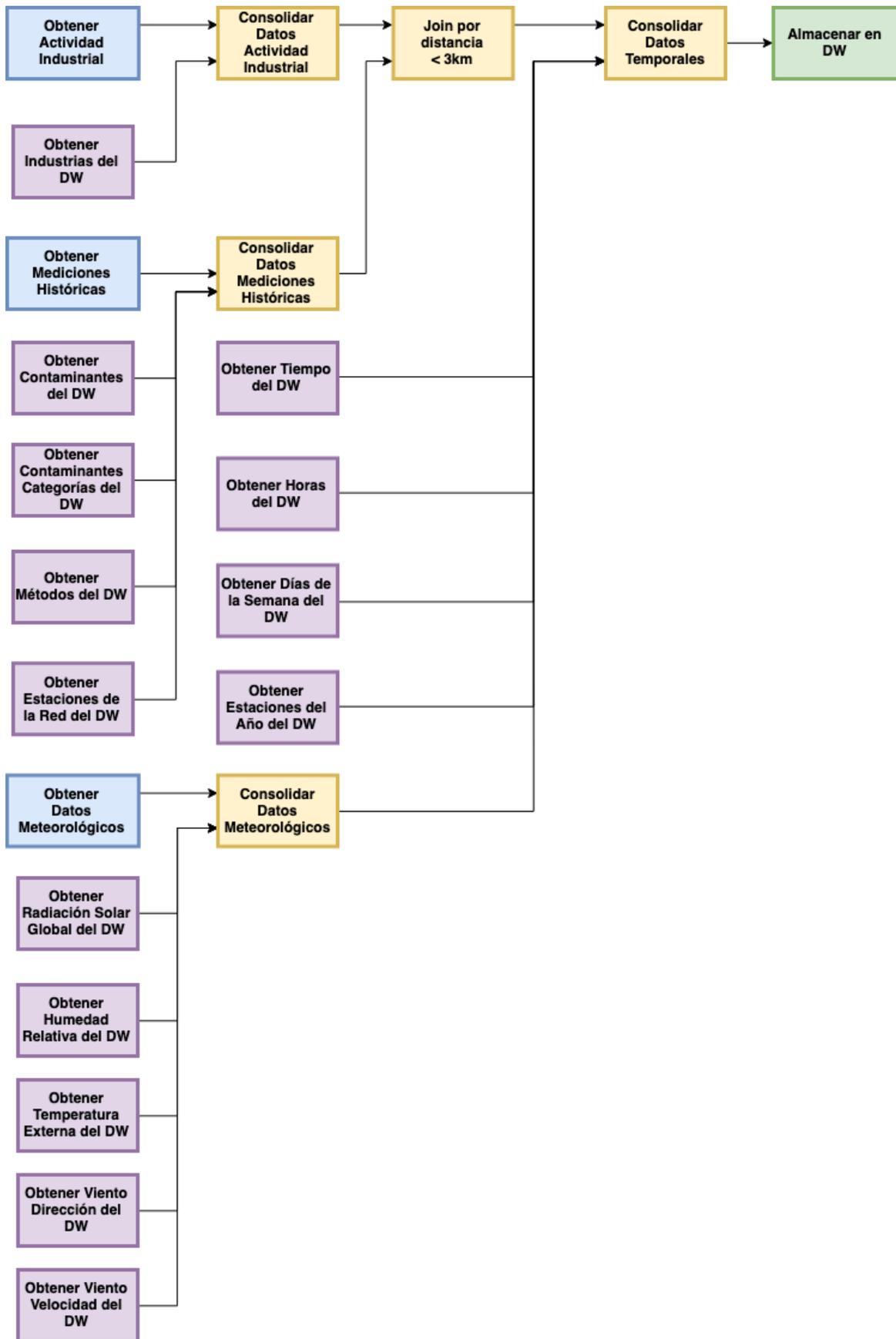


Figura 6.19: Carga Industrias Contaminación

### *Consolidar Datos Temporales*

Se asocian, los datos relacionados y filtrados de contaminación y actividad industrial, con los datos temporales para su posterior uso.

### *Almacenar en DW*

En última instancia, se guardan los datos relacionados en la tabla *dw.industriascontaminacion*.

## **Días Importantes Contaminación**

En la figura 6.20 se ilustra, en términos generales, el proceso de carga de la tabla de hechos Días Importantes Contaminación.

### *Obtener Días Importantes del DW*

Se obtienen los datos de la tabla de dimensión Días Importantes con el objetivo de ser utilizado posteriormente.

### *Obtener Mediciones Históricas*

Se extraen los datos del archivo CSV de Mediciones Históricas.

### *Obtener Contaminantes, Contaminantes Categorías, Métodos y Estaciones de la Red del DW*

Se obtienen los datos de las distintas dimensiones relacionadas a las Mediciones Históricas con el objetivo de ser utilizado posteriormente.

### *Obtener Datos Meteorológicos*

Se extraen los datos del archivo XLS provisto por ANCAP de Datos Meteorológicos obtenidos en la refinería de ANCAP.

### *Obtener Radiación Solar Global, Humedad Relativa, Temperatura Externa, Viento Dirección y Viento Velocidad del DW*

Se obtienen los datos de las distintas dimensiones relacionadas a los Datos Meteorológicos con el objetivo de ser utilizado posteriormente.

### *Consolidar Datos Mediciones Históricas*

Se asocian los datos obtenidos de Mediciones Históricas con los datos de las distintas dimensiones relacionadas.

### *Consolidar Datos Meteorológicos*

Se asocian los datos Meteorológicos obtenidos con los datos de las distintas dimensiones relacionadas. Cabe destacar que no se realiza un filtrado de los datos meteorológicos teniendo en cuenta la distancia de la fuente de medición de los mismos estaciones de la red debido a que se indica por los especialistas que los datos medidos son representativos para el departamento de Montevideo en general.

### *Consolidar Datos Temporales*

Se asocian los datos relacionados y filtrados de contaminación con los datos temporales para su posterior uso.

### *Almacenar en DW*

En última instancia, se guardan los datos relacionados en la tabla *dw.diasimportantescontaminacion*.

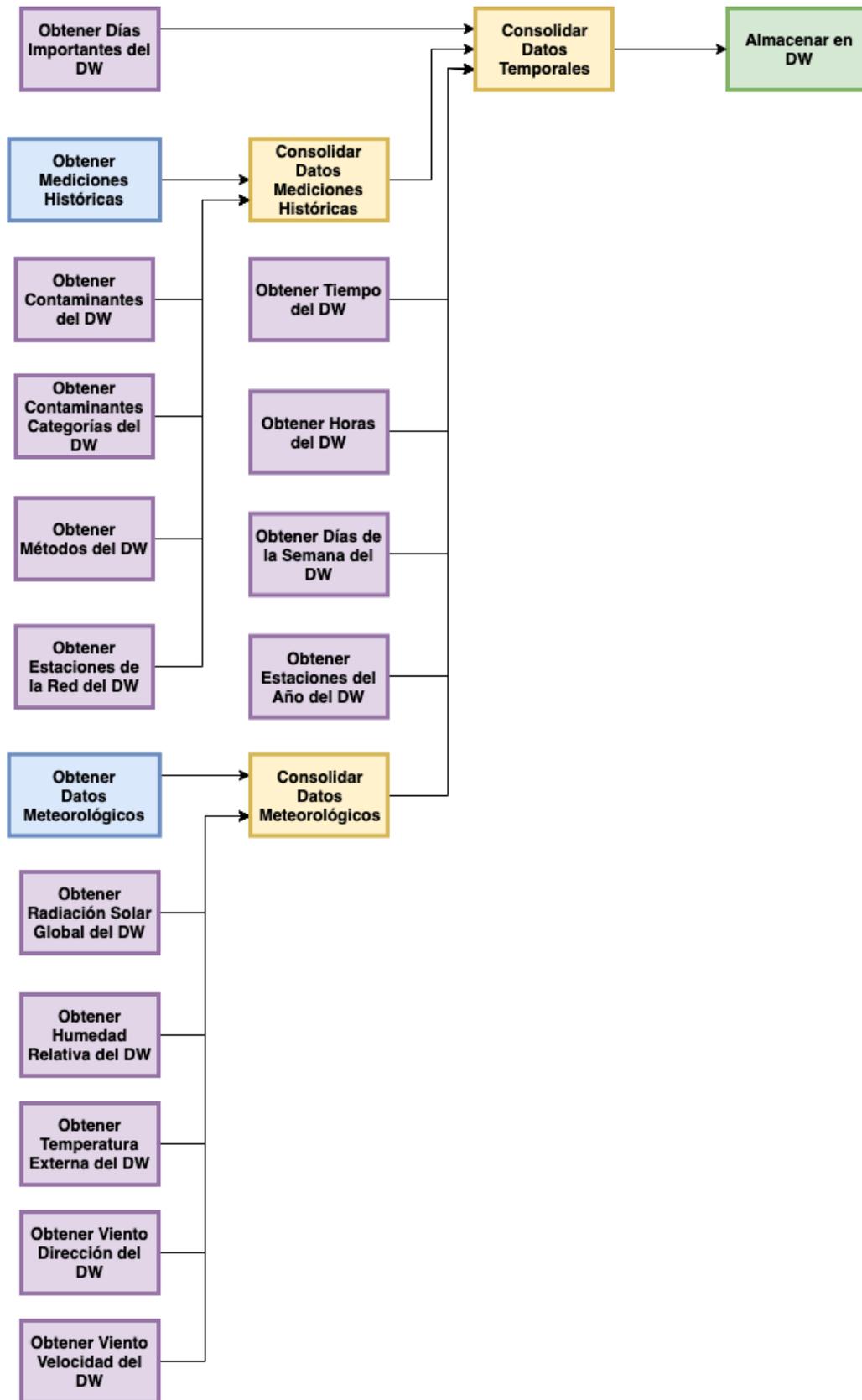


Figura 6.20: Carga Días Importantes Contaminación

## 6.2. Implementación de cubos

A continuación se presenta la definición de los cubos multidimensionales utilizando la herramienta Pentaho Workbench Schema.

Para cada una de las relaciones dimensionales, especificadas en la etapa de Diseño Conceptual, se define un cubo con el mayor nivel de granularidad para las dimensiones participantes.

Los tres cubos presentados comparten la mayoría de las dimensiones como se especificaba en el cuadro 5.1. Estas son Estaciones de la Red, Contaminantes, Métodos, Días de la Semana, Estaciones del Año, Contaminantes Categorías, Tiempo, Horas, Radiación Solar Global, Humedad Relativa, Temperatura Externa, Viento Dirección, Viento Velocidad.

### 6.2.1. Dimensiones compartidas

En la figura 6.21 y la figura 6.22 se especifica en detalle cómo están definidas las dimensiones compartidas.

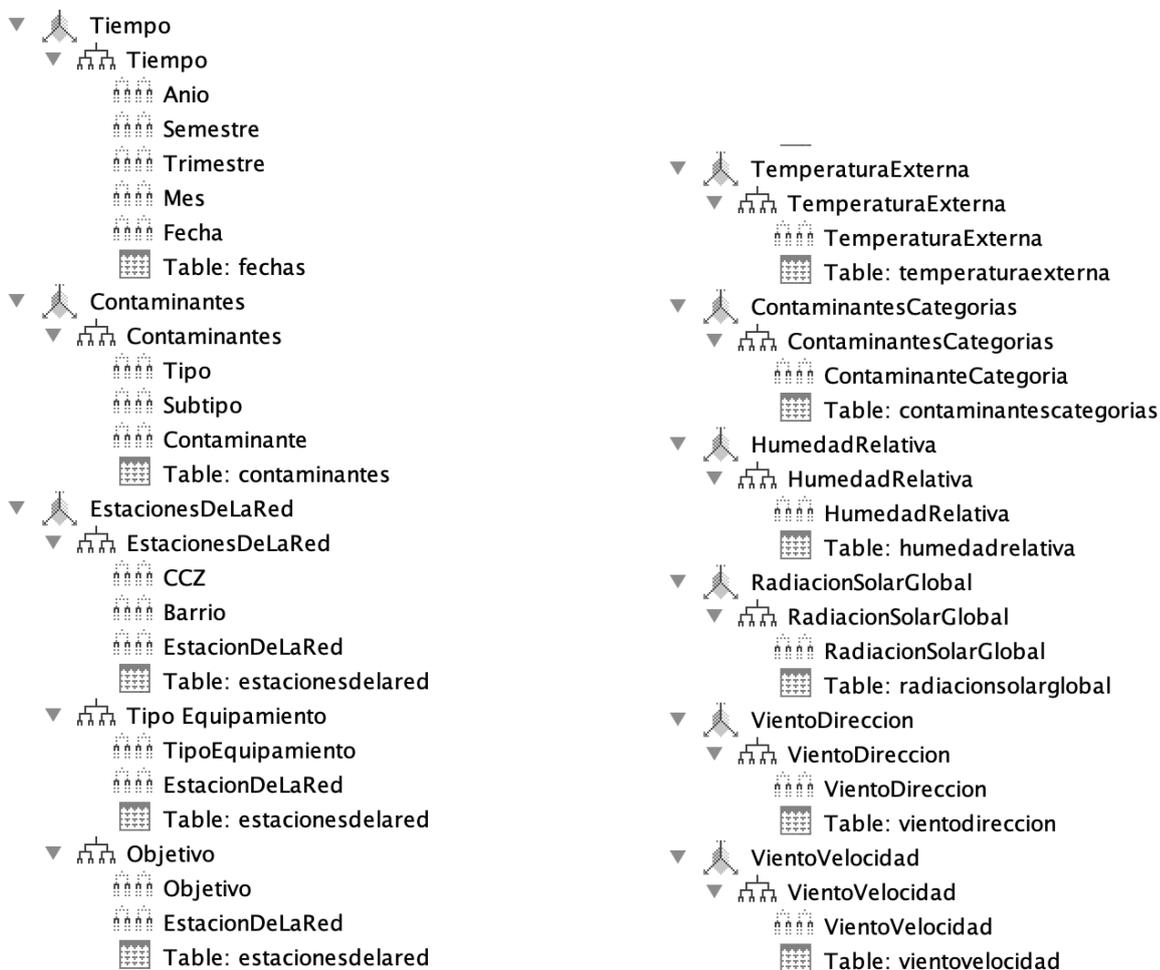


Figura 6.21: Dimensiones Compartidas

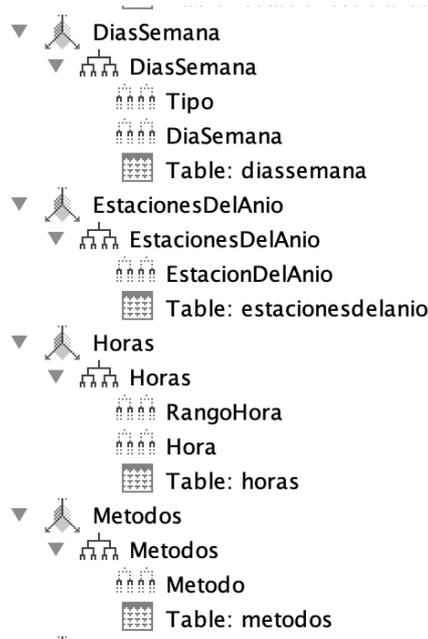


Figura 6.22: Dimensiones Compartidas

### 6.2.2. Cubo Vehículos Contaminación

A continuación se presenta el cubo generado en base a la relación dimensional Vehículos Contaminación. En figura 6.23 se puede visualizar la tabla de hechos *dw.vehiculoscontaminacion*, las tablas de dimensión y las medidas definidas.

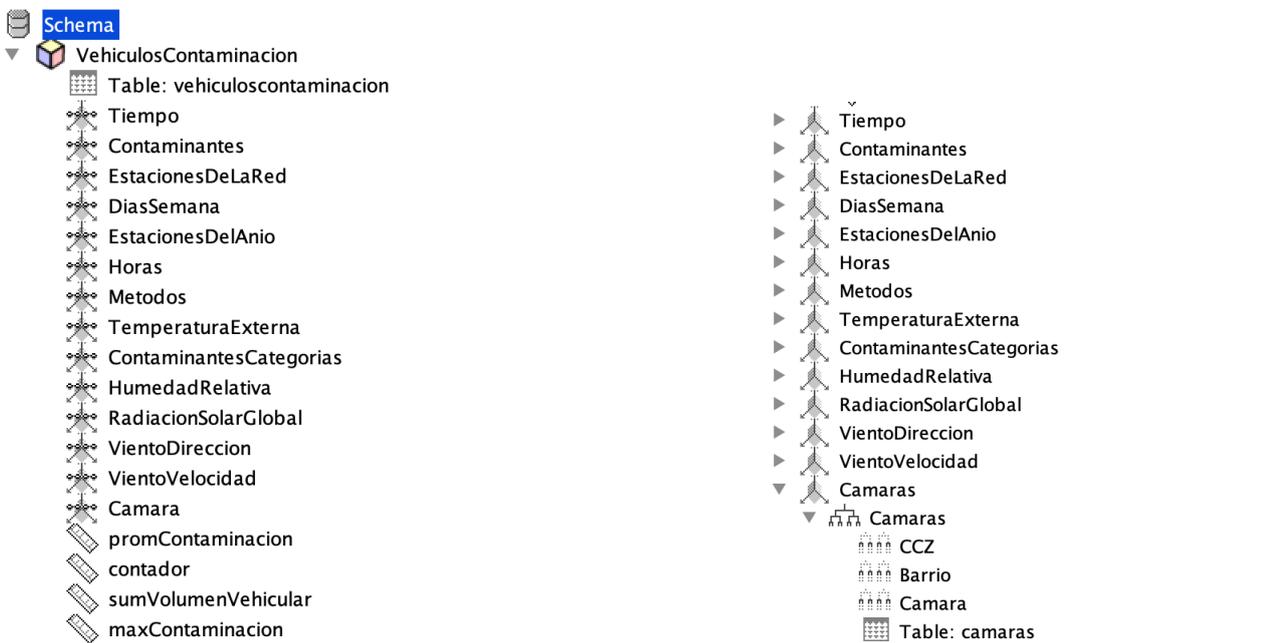


Figura 6.23: Cubo Vehículos Contaminación

### 6.2.3. Cubo Industrias Contaminación

A continuación se presenta el cubo generado en base a la relación dimensional Industrias Contaminación. En figura 6.24 se puede visualizar la tabla de hechos *dw.industriascontaminacion*, las tablas de dimensión y las medidas definidas.



Figura 6.24: Cubo Industrias Contaminación

## 6.2.4. Cubo Días Importantes Contaminación

A continuación se presenta el cubo generado en base a la relación dimensional Días Importantes Contaminación. En figura 6.25 se puede visualizar la tabla de hechos *dw.diasimportantescontaminacion*, las tablas de dimensión y las medidas definidas.



Figura 6.25: Cubo Días Importantes Contaminación

### 6.3. Generación de reporte de BI de sección del Informe Anual

Con el objetivo de automatizar ciertas tareas ejecutadas manualmente por los especialistas de la IM, como se especifica en los requerimientos funcionales de uso, se destaca la automatización en la realización de la sección Resultados de Red de Monitoreo del Informe Anual.

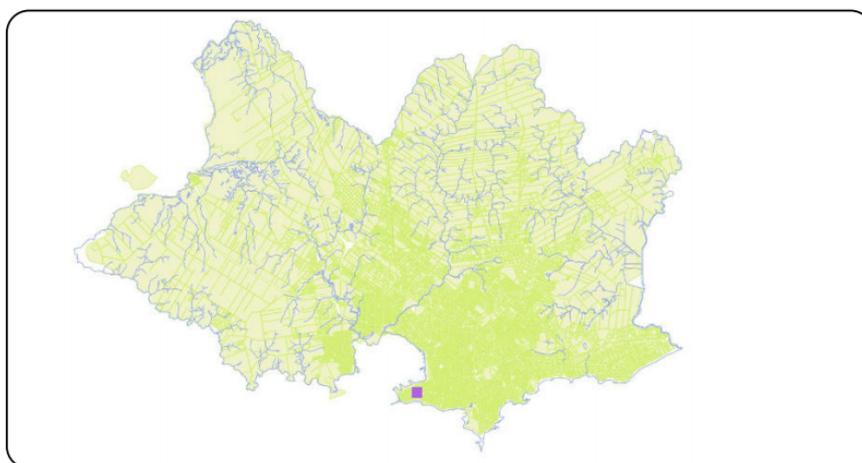
En la misma se presentan, para cada estación, los siguientes datos:

1. Información básica descriptiva de cada estación como nombre, ubicación, contaminantes y métodos utilizados, como se puede visualizar en la figura 6.26.
2. Una tabla con información de resultados anuales por cada contaminante medido por la estación.
3. Gráficas de los resultados expresados utilizando las categorías definidas por cada contaminante como se presenta en la figura 6.27.
4. Un extracto de texto libre que resume los resultados anuales de la estación.

## 2.2 Estación 1: Ciudad Vieja

Estación de Base

Academia Uruguay			
SIRGAS 2000	X= 572795	Altura sobre Nivel del mar 20 m	
UTM ZONE 21S	Y= 6137122	Elevación desde la calzada 9 m	
Sede de Asociación de Empleados Bancarios del Uruguay			
SIRGAS 2000	X= 572831	Altura sobre Nivel del mar 4 m	
UTM ZONE 21S	Y= 6136774	Elevación desde la calzada 3 m	
Parámetro	Unidades	Método de Medida	Frecuencia
Material Particulado PM2.5	PM2.5 ug/m <sup>3</sup>	Lightscattering	Horario
Humo Negro	HN ug/m <sup>3</sup>	Reflectometría	24 horas cada 12 días



Mapa 2.2 Ciudad Vieja

Parámetro	Promedio	IC_Inf	IC_sup	Mínimo	Máximo	Percentil 95	Mediana	Datos
	ug/m3	ug/m3	ug/m3	ug/m3	ug/m3	ug/m3	ug/m3	Número
Humo Negro	13	11	16	0	32	27	11	29
PM2.5	10	9	11	2	80	23	7	365

Figura 6.26: Informe - Información básica

A continuación se describe el diseño, desarrollo y disponibilización de la generación automática del reporte de BI sobre la sección Resultados de Red de Monitoreo del Informe Anual.

### Diseño

En primera instancia, para definir la estructura visual, contenidos y cálculos del reporte, se utiliza la herramienta Pentaho Report Designer.

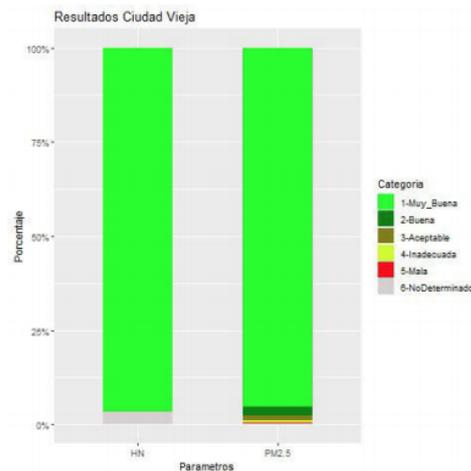
Debido a la necesidad de presentar la misma información para cada estación se especifica una subsección esqueleto de los componentes visuales utilizados, con variables que en tiempo de ejecución serán cargadas con los datos de cada estación, y se repite la generación de cada subsección por cada estación de la red.

En la figura 6.28 se visualiza la definición en Pentaho Report Designer de los siguientes com-

CIUDAD_VIEJA		
Categoría	HUMO NEGRO	MATERIAL PARTICULADO PM2.5
Muy Buena	96.7%	95.3%
Buena	0.0%	2.2%
Aceptable	0.0%	1.4%
Inadecuada	0.0%	0.5%
Mala	0.0%	0.3%
No_Determinada	3.3%	0.3%

**Tabla 2.3 Resumen de resultados de la Estación**

**Ilustración 2.1 Distribución de resultados de la Estación**



En el año 2018, se tomaron 29 muestras del tren de monitoreo, y en la estación automática se registraron valores durante 365 días, lo que representa el 99% del año.

Los porcentajes de las diversas categorías de calidad de aire reportadas se muestran en la Ilustración 2. El 97% de muestras de Humo Negro, y el 95% de PM 2.5 se encuentran dentro de la categoría MUY BUENA.

Tres días del año 2018 ( 0.8%) se registraron valores de PM2.5 promedios diarios mayores de 50  $\mu\text{g}/\text{m}^3$ , por lo que la categoría fue INADCUADA uno de estos días, la categoría fue MALA ya que el valor observado fue superior a 75  $\mu\text{g}/\text{m}^3$ .

Estos eventos se registraron en junio, agosto y setiembre.

Figura 6.27: Informe - Resumen resultados utilizando el ICAire

ponentes y variables:

1. Año del informe presentado.
2. Nombre e imagen representativa de la ubicación de la estación de la red en Montevideo.
3. Cuadro de resultados con los siguientes valores:
  - a) Nombre del contaminante.



## Desarrollo

Tomando en cuenta las variables especificadas, es necesario obtener los datos que conforman el contenido de la sección del informe. Los mismos se toman desde tres fuentes:

1. Consultas específicas a la base de datos del DW para obtener las estaciones de la red, los contaminantes y los datos de mediciones históricas de contaminación para realizar ciertos cálculos.
2. Variables obtenidas desde el aplicativo web como el año de generación del informe y el resumen en texto por cada estación.
3. Ejecución de funciones estadísticas en R.

Como aspecto relevante a destacar se utiliza la operación estadística denominada *bootstrapping* del paquete *openair* [10] de R [37] para calcular el intervalo de confianza para cada contaminante utilizando como entrada los datos de mediciones históricas de cada contaminante.

El método estadístico utilizado es común para caracterizar la variabilidad de un indicador en asociación a un nivel de confianza.

A continuación se presenta el código fuente de la ejecución que se disponibiliza mediante un servicio web:

```
library(openair)

#' @post_/bootstrapping
#' @html
function(a) {
  v <- as.numeric(unlist(strsplit(a, split=",")))
  bootstrapping <- bootMeanDF(v)
  percentile <- quantile(v, c(.95))
  result <- cbind(bootstrapping,percentile)
  toString(result)
}
```

## Disponibilización

Para su disponibilización se utilizan múltiples herramientas y componentes de software, a saber:

1. Se define en Pentaho Report Designer un archivo de configuración con la estructura visual, contenidos y cálculos del reporte.
2. Se desarrolla un servicio en R para el cálculo estadístico.
3. Se genera un proceso ETL encargado de realizar la ejecución del reporte utilizando la definición generada en el Pentaho Report Designer, de la conexión al DW para obtener los datos de las mediciones históricas y de la llamada al servicio web disponible en R para realizar el cálculo estadístico. En la figura 6.29 se visualizan la definición de los procesos con los pasos explicitados.
4. Se disponibiliza el ETL como servicios REST, mediante el servidor Carte de la plataforma PBI.

5. En la aplicación web se crea una página para la ejecución de la generación del reporte en PDF donde el usuario inicia la ejecución indicando el año y un resumen de resultados para cada una de las estaciones de la red.
6. Finalmente el archivo PDF se disponibiliza en el sistema de archivos provisto.

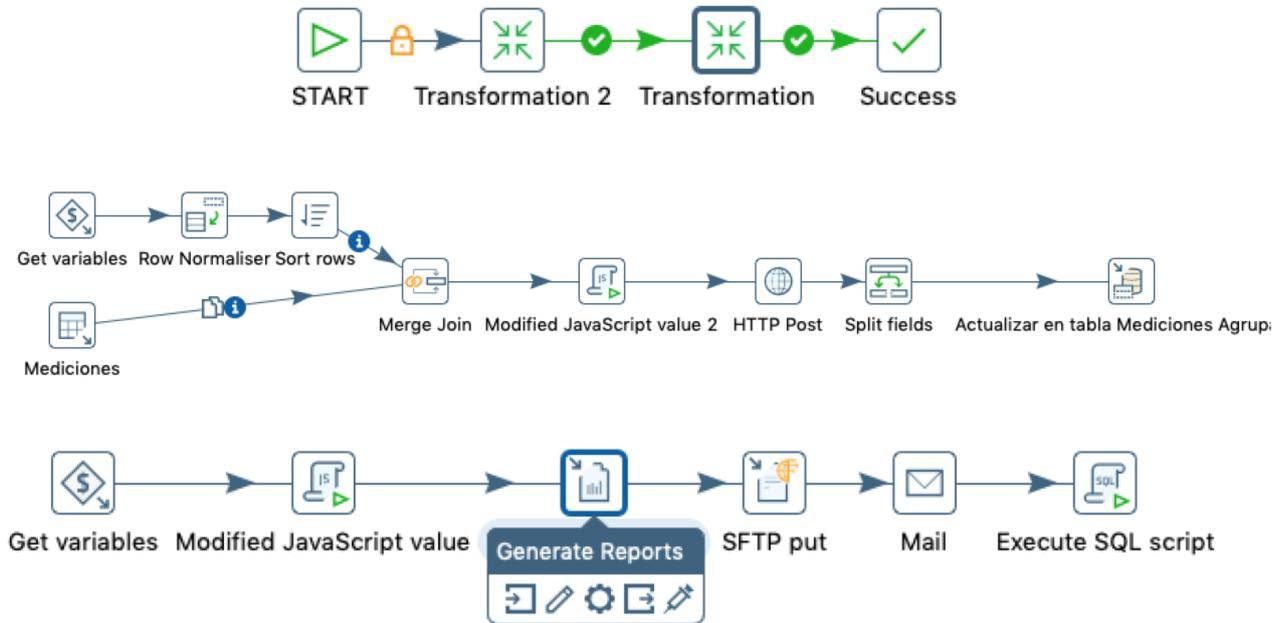


Figura 6.29: Informe - ETL

## Verificación

Para realizar la verificación de la correcta ejecución de la generación de la sección del informe anual se disponen de los informes oficiales disponibilizados por el servicio de ECCA públicamente en su página web.

Durante el desarrollo del proyecto se realizan múltiples ejecuciones de generación de la sección del Informe Anual para los años 2017 y 2018, de los cuales se dispone del informe oficial y de anexos con los datos y cálculos utilizados, y se los compara entre ellos.

En las pruebas se pudieron constatar diferencias entre los cálculos teóricos realizados manualmente, basados en los datos de mediciones históricas para cada año, y los resultados que se presentan en los informes oficiales.

Luego de sucesivas instancias con los especialistas se logran identificar y realizar ajustes sobre los cálculos presentados en los informes oficiales y se los publican nuevamente a fin de corregir las diferencias.

## 6.4. Herramientas utilizadas

Se describen las herramientas de software que componen el sistema mediante un Diagrama de Despliegue, en la figura 6.30, y se especifican las interacciones entre los mismos.

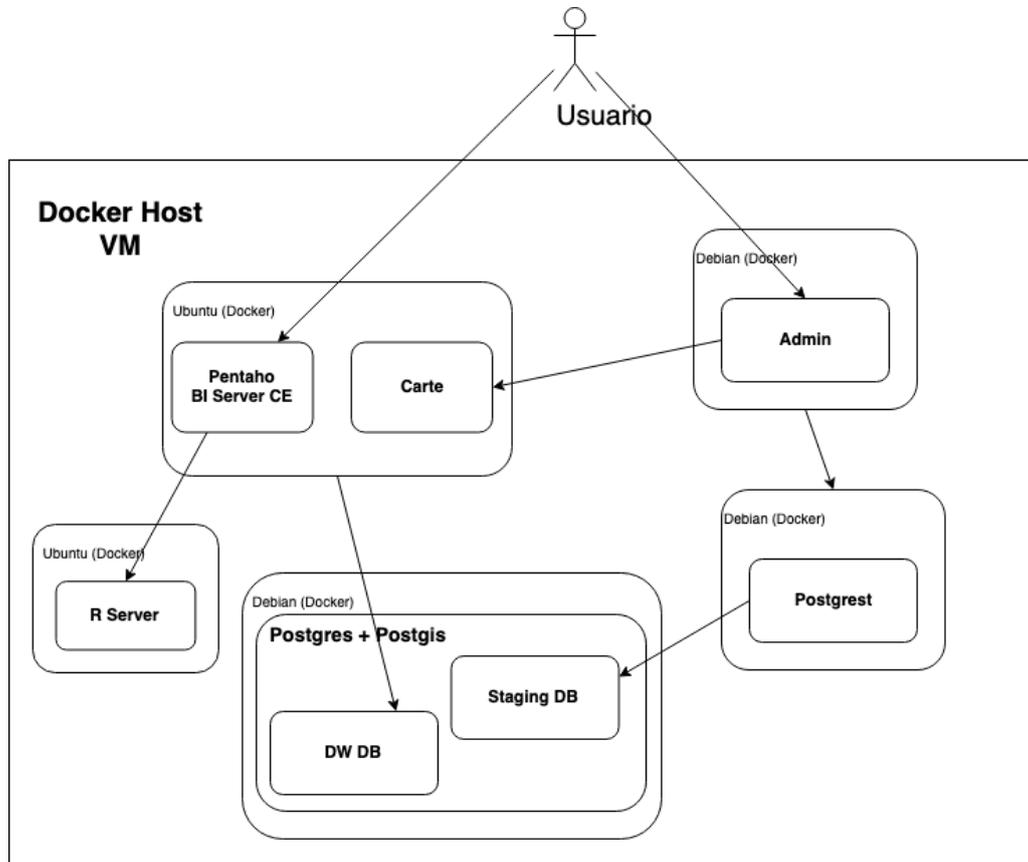


Figura 6.30: Despliegue

### Aplicación web

El aplicativo web de carga, actualización y reportes se instala dentro de un contenedor de Debian Linux en un servidor web NGINX. Dicho aplicativo web es desarrollado utilizando tecnologías estándar como HTML/JS/CSS y la librería React para facilitar el desarrollo y mantenimiento.

### Base de Datos

Existen distintos contenedores, relacionados al almacenamiento de datos, que forman parte de la solución:

1. PostgreSQL + PostGIS: Este contenedor Debian hace parte como la Base de Datos principal del sistema desplegado. Es utilizado para almacenar las fuentes de datos, para las bases de staging y, como objetivo primordial, para almacenar los datos del DW.
2. PostgREST: Este contenedor Debian hace parte como la capa de servicios utilizada por el aplicativo web para la gestión de las fuentes de datos.
3. pgAdmin: Este contenedor Debian sirve una aplicación web para la gestión de la base de datos PostgreSQL de forma más eficiente.

## PBI CE

La herramienta core de la suite de Pentaho utilizada es la solución de BI, específicamente, la versión CE. La misma se despliega en un contenedor Ubuntu que compone un servidor y un contenedor de bases de datos MySQL Server, corriendo en un Oracle Linux, para el almacenamiento de los datos y configuración del servidor.

Tan indispensable como el servidor de BI es el componente denominado Kettle, encargado de la planificación y ejecución de los ETL. El mismo brinda la capacidad de ejecución de los procesos de forma remota utilizando una interfaz web vía servicios REST denominada Carte.

## Documentación en línea

Con el objetivo de brindar documentación en línea se despliega un contenedor, basado en Debian, que provee una herramienta web con videos y tutoriales sobre el uso de la plataforma. El mismo es accesible desde la aplicación web de carga, actualización y reportes y puede visualizarse en la figura 6.31.

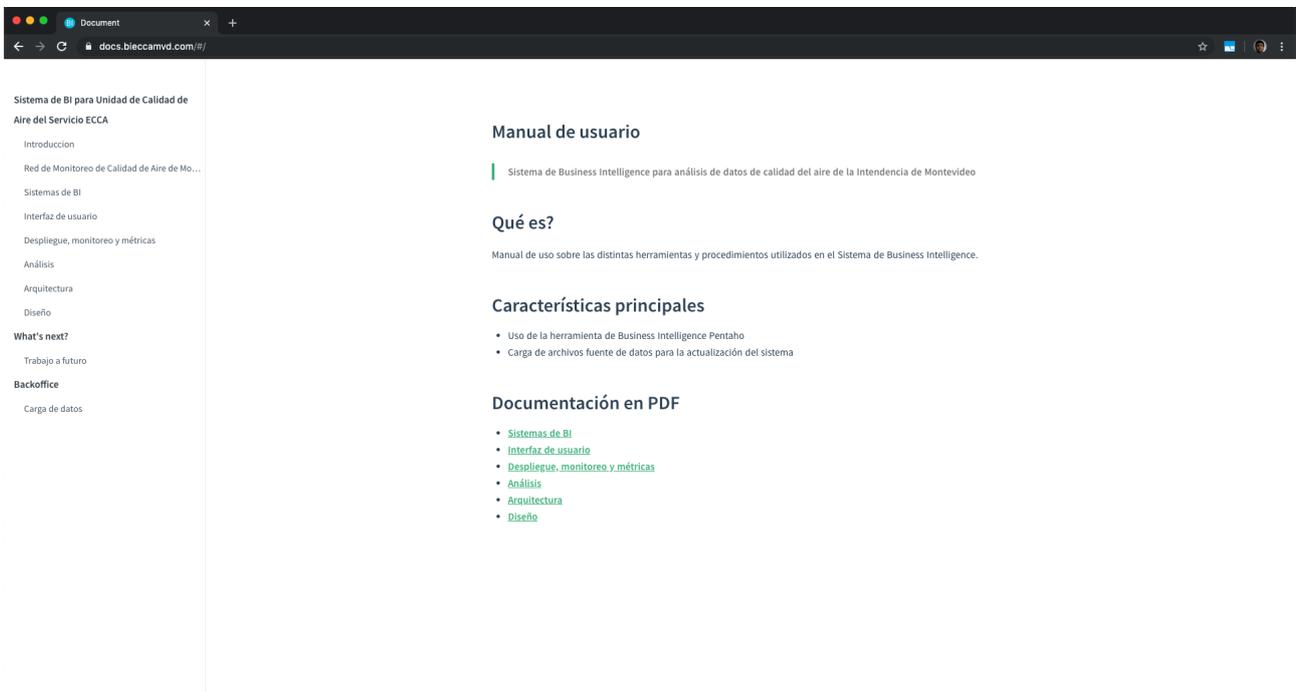


Figura 6.31: Documentación en línea

## Sistema de archivos

Para lograr desacoplar las distintas responsabilidades que tienen los componentes, se dispone de un contenedor, basado en Debian, que sirve una aplicación web para la gestión de archivos a un SFTP, que se encuentra desplegado en otro contenedor basado en Debian.

Con esta decisión, los archivos que se manejan en la plataforma, a saber, los archivos fuentes de datos y los reportes generados por la aplicación web, se almacenan en un repositorio de archivos que no comparte el espacio lógico de los otros componentes, por ejemplo, el espacio lógico de PDI que es quien utiliza dichos archivos como entrada y salida de la ejecución de los procesos.

En la figura 6.32 se puede visualizar la herramienta web.

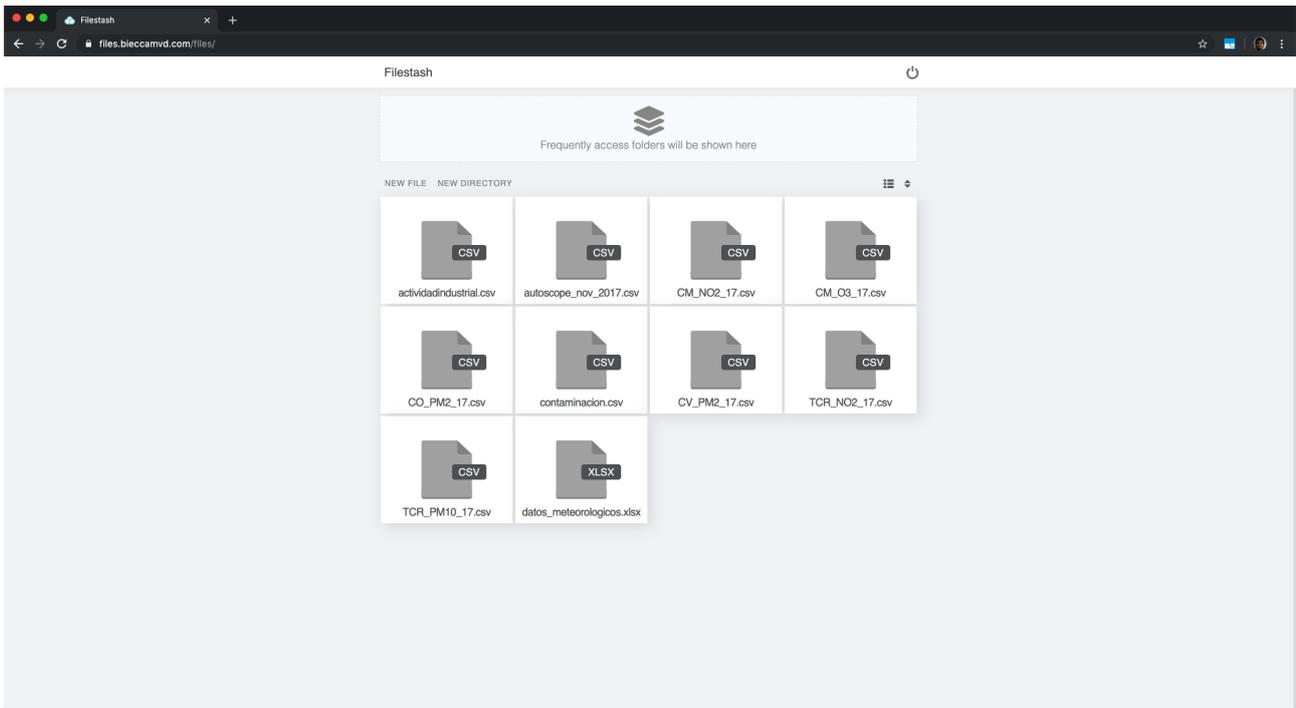


Figura 6.32: Sistema de archivos web

## R

Para la generación de la sección Resultados de Red de Monitoreo del Informe Anual es necesario realizar cálculos estadísticos que los especialistas de la IM realizan manualmente en la aplicación de escritorio R.

Dada la necesidad de ejecución de los métodos estadísticos en los procesos ETL de PDI, es necesario brindarle la capacidad de utilizar dichos métodos en tiempo de ejecución de los procesos. Esto es posible mediante la disponibilización de un contenedor servidor de R con una capa de servicios REST, provista por Plumber, para la integración.

## **6.5. Interfaz de usuario**

Debido a que la plataforma de BI presentada será de uso exclusivo de los especialistas de la IM, delegándose por completo su carga, actualización y utilización, es necesario brindar distintas herramientas que sirvan a cada caso.

### **6.5.1. Aplicación web de carga, actualización y reportes**

Con el fin de brindar una herramienta que facilite la carga y actualización del DW y mediante la cual se pueda generar los reportes de BI requeridos, se implementó una aplicación web que permite realizar dichas tareas de forma centralizada. Basado en lineamientos estandarizados de diseño y experiencia de usuario, se provee una interfaz limpia e intuitiva para la gestión de los datos del DW y la generación de reportes de BI.

En dicha aplicación web es posible:

1. Realizar la gestión de datos del DW.
2. Ejecutar la generación de un reporte de BI de la sección del Informe Anual.

## Carga y actualización

Para realizar la gestión de los datos del DW se visualizan y disponen dos opciones bien diferenciadas:

1. Realizar una carga inicial del DW para inicializarlo.
2. Ejecutar una actualización del DW con nuevos datos que pueden ser cargados directamente en la web.

En la figura 6.33 se muestra la pantalla de inicio de la aplicación y los distintos items del menú.

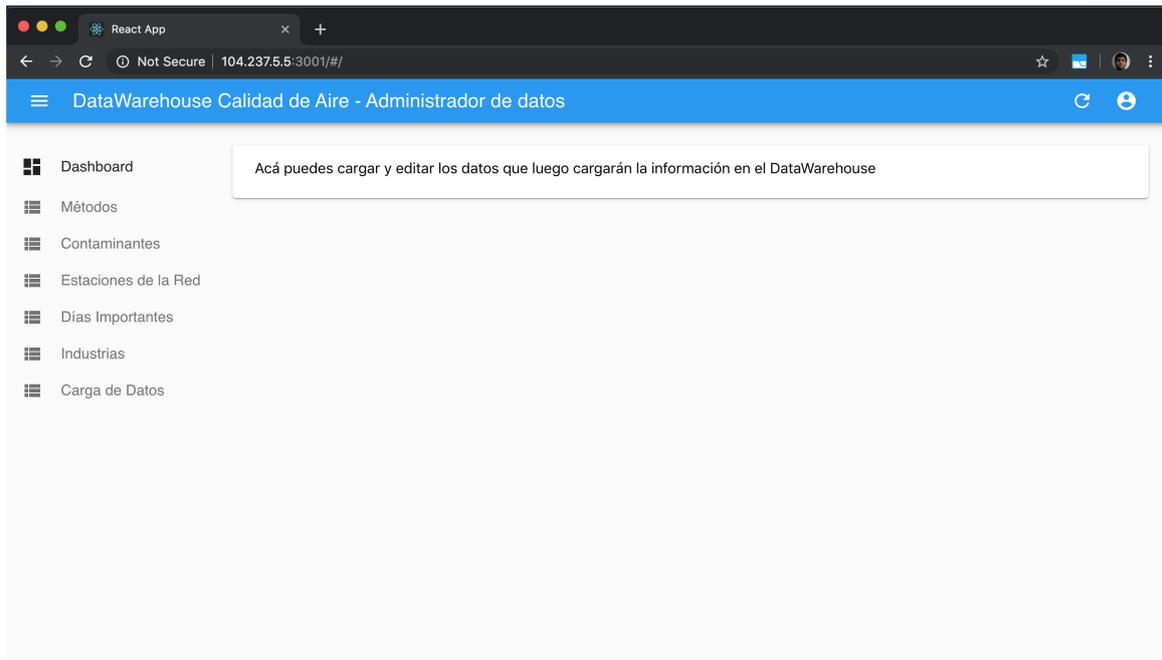


Figura 6.33: Aplicación web - Carga y actualización

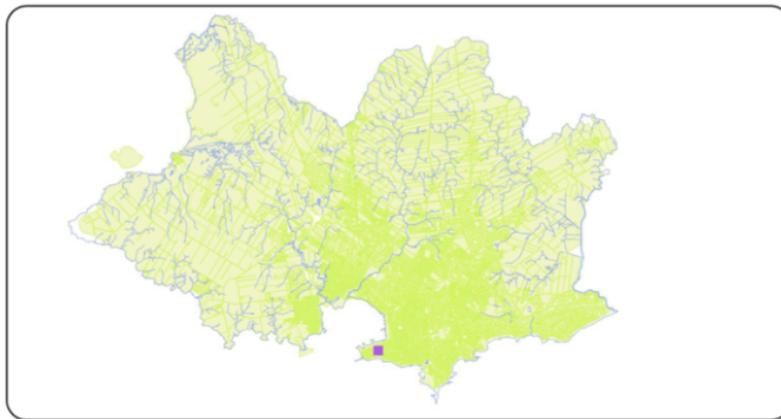
## Generación de reporte de BI de sección del Informe Anual

Como requerimiento funcional de uso, específicamente RFU2, se dispone una acción dentro del aplicativo web para ejecutar la generación de un reporte de BI de la sección del Informe Anual para el año que se indique con los datos que se encuentren almacenados en el DW.

En la figura 6.34 se presenta un ejemplo de la sección del informe generado.

### RESULTADOS DE RED MONITOREO 2017

Estacin de la red **Ciudad Vieja 3**



Mapa 2.2 Ciudad Vieja

Contaminante	Promedio (ug/m3)	IC (ug/m3)		Nmero de cuentas	Mximo	Mnimo	Percentil 95 (ug/m3)
PM2	10	10	11	347	67.0	3.0	26
SO2	7	5	11	14	21.0	0.0	20

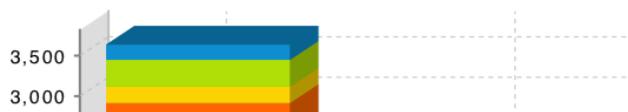


Figura 6.34: Aplicación web - Reporte de BI de sección del Informe Anual

## 6.5.2. Análisis

Con el objetivo de lograr flexibilidad y completitud para la realización del análisis de datos sobre los requerimientos definidos y basados en el antipatrón de no reimplementar componentes que se pueden conseguir prefabricando de antemano, se utilizan distintas herramientas provistas por la suite de Pentaho:

### Saiku

Para poder realizar consultas de forma dinámica, logrando la mayor flexibilidad posible para analizar los datos, se definen y cargan múltiples cubos, uno por cada requerimiento definido, y se disponibilizan para su uso.

### Vehículos Contaminación

En la figura 6.35 se muestra cómo se visualiza la relación entre el volumen vehicular y la contaminación promedio según las distintas horas del día en distintas fechas.

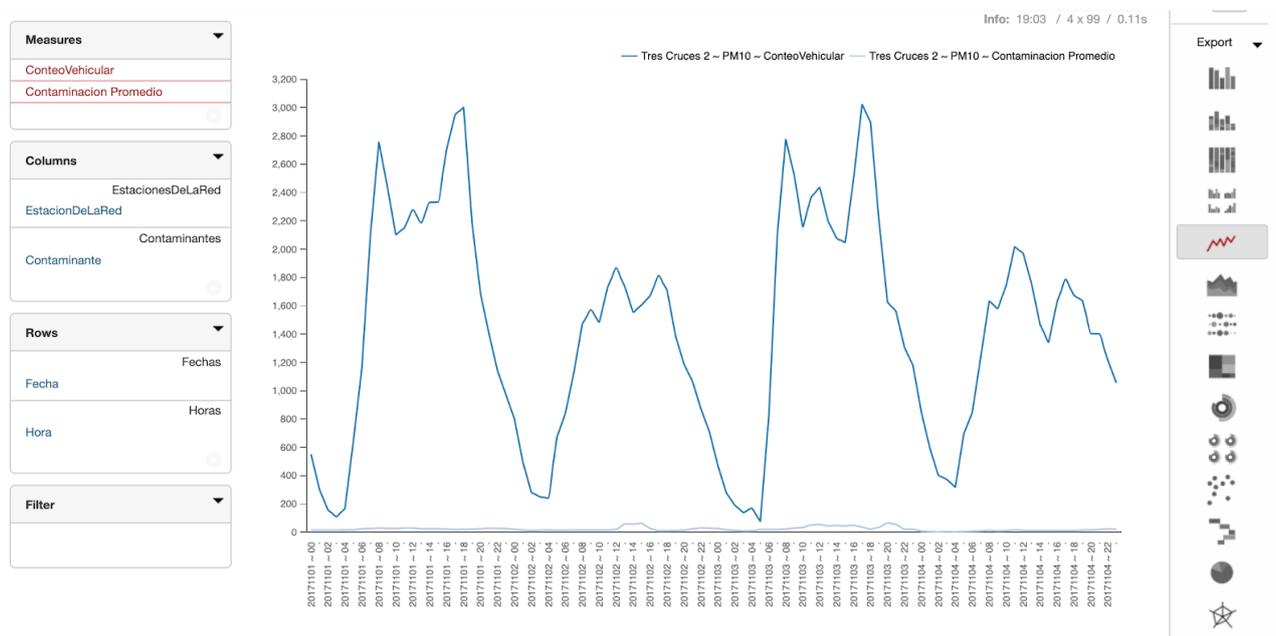


Figura 6.35: Análisis de Datos - Volumen Vehicular

## Industrias Contaminación

En la figura 6.36 se muestra cómo se visualiza, en los distintos días de la semana, la relación entre la actividad industrial de la Central Batlle y la contaminación de PM10 en cada franja de la velocidad del viento.

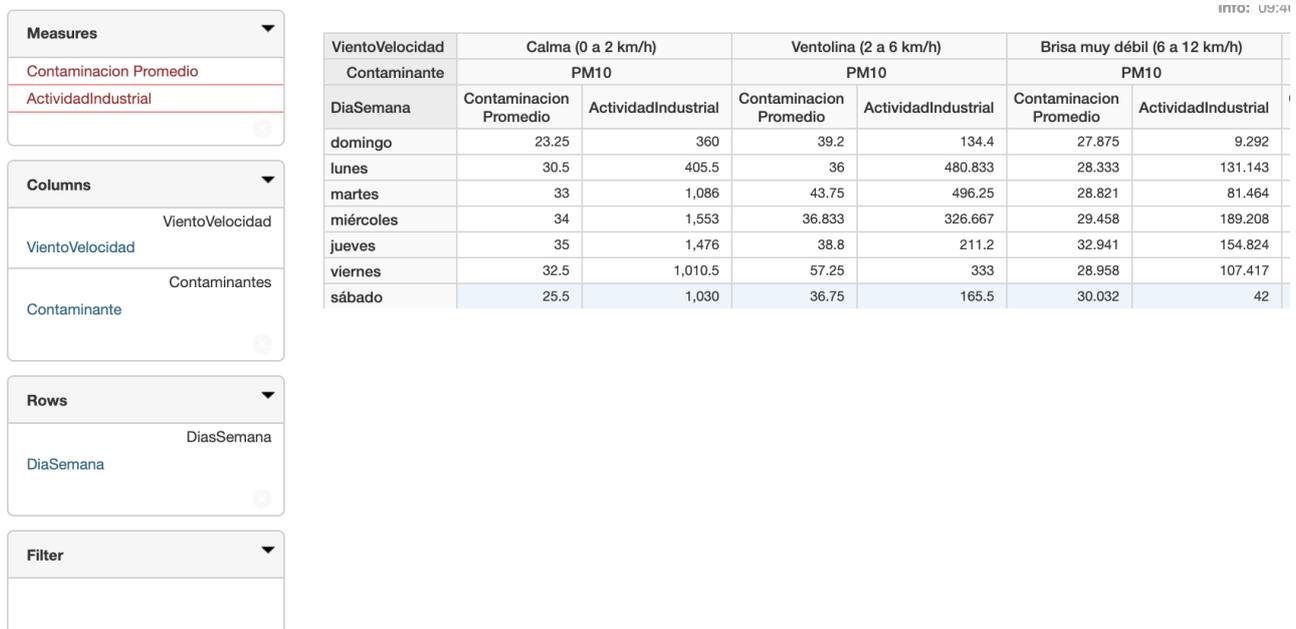


Figura 6.36: Análisis de Datos - Actividad Industrial

## Días Importantes Contaminación

En la figura 6.37 se puede analizar gráficamente el porcentaje del número de mediciones dentro de cada categoría por hora independientemente del día.

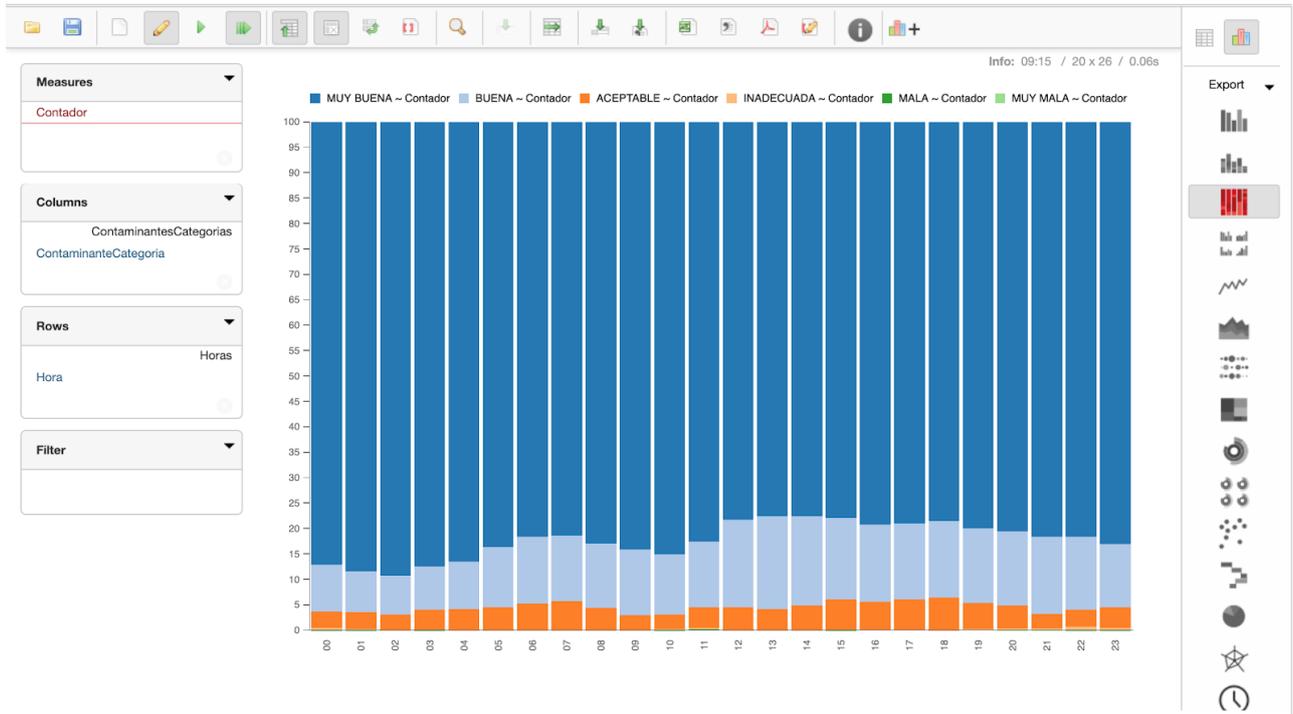


Figura 6.37: Análisis de Datos - Días Importantes



# Capítulo 7

## Conclusiones y trabajo futuro

En el proyecto realizado se ha podido cumplir con los objetivos planteados al comienzo del mismo y se sientan las bases para el descubrimiento de nuevas oportunidades de mejora sobre el sistema de BI.

### 7.1. Conclusiones

En el marco del proyecto se logran cumplir los distintos objetivos planteados, a saber:

1. Prototipo de sistema de BI en ambiente productivo.
2. Aplicación web a medida.
3. Características de software deseables.
4. Documentación detallada del trabajo realizado.
5. Guía de instalación y manual de usuario de las herramientas a utilizar.

Como objetivo primordial del proyecto se implanta y disponibiliza, en un ambiente productivo, un prototipo de la herramienta de BI con datos de mediciones actuales e históricas de las estaciones de la red, de la actividad de distintas industrias y del volumen vehicular en las principales arterias de Montevideo, permitiendo al Servicio de Calidad de Aire el acceso inmediato a los mismos en forma adecuada e intuitiva para el posterior análisis y generación de reportes para la toma de decisiones.

Para promover el uso del sistema de BI es necesario brindar herramientas simples e intuitivas para las tareas que son indispensables de realizar periódicamente para su gestión, como lo son la carga y actualización de los datos desde las fuentes. Para cumplir dicho objetivo se disponibiliza una aplicación web fácilmente accesible y simple de utilizar basada en lineamientos estandarizados de diseño y experiencia de usuario. A su vez, es posible lanzar la generación del reporte de BI de la sección de resultados del Informe Anual.

Las herramientas brindadas para la carga, actualización, generación de reportes y análisis fueron utilizadas en múltiples instancias del proyecto por los especialistas de la IM sugiriendo ajustes y mejoras que fueron implementadas según su feedback, logrando validar el uso del sistema.

El sistema de BI y DW disponibilizado cumple múltiples aspectos deseados en las soluciones de software que cabe destacar:

1. Modularización con un alto grado de desacople de sus componentes principales.
2. Independencia de la infraestructura local o Cloud en la cual se despliega.
3. Solución integral, logrando unificar y abarcar los distintos componentes de un sistema de software completo en una solución autocontenida.

El sistema de software desplegado presenta:

1. Distintos mecanismos de análisis de datos como indicadores, gráficas, consultas dinámicas y tablas.
2. Reportes de estado situación de la calidad de aire.
3. Herramientas de análisis estadístico y predictivo de calidad de aire.
4. Sistema de gestión de archivos.
5. Notificaciones mediante emails.
6. Seguimiento mediante métricas de uso.
7. Distintos métodos de auditoría.
8. Monitoreo proactivo mediante alertas.
9. Instrumentos de medición de performance de los distintos componentes.

Dentro de las limitaciones que presenta el sistema se pueden identificar:

1. Debido a que los archivos de fuentes de datos de mediciones históricas no tienen el mismo formato, la solución presentada está fuertemente acoplada a ese formato, requiriendo que ante cambios del mismo haya que realizar modificaciones en la implementación del sistema.
2. Existen problemas de calidad de datos que afectan directamente a la definición de los procesos ETL, los cuales resuelven dichos problemas, pero que no son una solución genérica, sirviendo para un conjunto conocido de casos particulares identificados en el momento de desarrollo del proyecto.

Finalmente se brinda documentación extensiva del proyecto para posteriores capacitaciones o transferencias de conocimiento y un manual de usuario web con videos y tutoriales sobre el uso de la plataforma al que se puede acceder desde la misma aplicación web desarrollada.

A modo personal, y como principal punto requerido, se entiende que el sistema desplegado simplifica las tareas realizadas por el personal de la Unidad de Calidad de Aire, permitiéndoles cumplir con las solicitudes, tanto desde la interna en la IM como de terceros, de forma sencilla y rápida con el apoyo de una interfaz intuitiva y flexible que permite la realización de consultas eficazmente.

Al brindar una solución basada en contenedores, las distintas tareas técnicas indispensables para el correcto funcionamiento de la herramienta en producción, en la etapa de soporte y mantenimiento post-despliegue, son claras y se encuentran especificadas y probadas. Esto es clave para un proyecto que se enmarque en el contexto de un Proyecto de Grado dado que no existe una etapa de soporte y mantenimiento como en la realidad.

Dada la naturaleza de las soluciones en contenedores es posible, mediante un conjunto reducido de comandos, desplegar la misma solución en otros ambientes para distintos fines como: instalación en la intranet de la IM o en otra cloud, generar distintos entornos para el desarrollo del proyecto (testing, validación de usuario), utilizar la base de la plataforma para otras unidades en ECCA, entre otros.

## 7.2. Trabajo futuro

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

### **Nuevos requerimientos de análisis**

Los sistemas de BI sufren cambios constantemente por la necesidad de apoyar la evolución de la aplicación debido a nuevos requerimientos de análisis o cambios en los requerimientos existentes.

La habilidad del sistema para evolucionar requiere un esfuerzo especial para anticipar cuándo y cómo pueden ocurrir estos cambios. Basado en las métricas de uso de las consultas dinámicas más utilizados por los especialistas de la IM en Saiku es posible obtener retroalimentación del usuario para entender cómo debe evolucionar el sistema con nuevos requerimientos o cambios a los requerimientos ya existentes.

Con ello es posible identificar las consultas más frecuentes y generar indicadores o gráficas pre-armadas para hacer la tareas de análisis más efectiva.

### **Aumentar capacidades para monitorear la calidad del agua, suelo y efluentes industriales.**

El Servicio cumple la función de evaluación y vigilancia ambiental, y tiene como principales áreas de trabajo el monitoreo de la calidad de agua, aire, suelo y efluentes industriales, la fiscalización de plantas de tratamiento y la gestión para una adecuada disposición final de residuos sólidos industriales.

Con el desarrollo del proyecto actual se disponibiliza un sistema de BI que cubre las necesidad de monitoreo del aire pero que es extensible a las otras áreas que son monitoreadas por el servicio ECCA debido a que las herramientas y componentes de software utilizados son los mismos.

Se dejan sentadas las bases para la realización de un pre-proyecto para evaluar los requerimientos y necesidades de las otras unidades dentro del servicio ECCA y definir el impacto y los ajustes requeridos para abarcarlos.

### **Minería de datos (*Data mining*)**

La minería de datos es una técnica de análisis y procesamiento de grandes volúmenes de datos con el objeto de extraer información útil y patrones de fácil comprensión, que sería imposible conseguir por los medios y herramientas estadísticas tradicionales.

En la actualidad existen múltiples trabajos, basados en dichas técnicas, con foco en la calidad de aire de distintas ciudades del mundo.

Utilizando los datos disponibles en el momento del desarrollo del proyecto es posible utilizar técnicas de minería de datos para encontrar relaciones entre las fuentes de datos utilizadas, como variables climatológicas, volumen vehicular, actividad industrial y emisiones contaminantes, y otras fuentes de datos que se puedan descubrir y utilizar. Además, sobre esa información nueva o sobre la ya disponible, aplicar técnicas de análisis predictivo para realizar proyecciones de futuro.

Cabe destacar que para su realización es posible utilizar el paquete *openair*, que presenta herramientas de minería sencillas y muy atractivas, que ya se encuentra disponibilizada en la plataforma y se utiliza para el cálculo estadístico en la generación de la sección del Informe Anual.

### **Instalación en la intranet e integración con bases de datos de la Red de Monitoreo de la IM**

Actualmente la carga de fuentes de datos generados por la Red de Monitoreo se realiza de forma manual debiendo los especialistas de la IM ingresar al aplicativo web para realizar la carga y actualización de los datos del sistema de BI. Este procedimiento debe su naturaleza a que los datos agregados disponibilizados son públicos y se encuentran disponibles en el transcurso del proyecto.

Este procedimiento tiene múltiples desventajas como ser propenso a errores humanos, ocupar horas/persona en tareas que pueden ser automatizables, entre otros.

Se propone el desarrollo de una integración directa desde los procesos ETL hacia las bases de datos fuentes utilizadas por los aplicativos de producción que se dispone desde la IM para la Red de Monitoreo. Para ello sería necesario la instalación del sistema completo de BI en la intranet de la IM logrando otros aspectos importantes como seguridad y gobernanza de los datos.

Dicha integración facilitaría la carga y actualización del sistema de BI, aspecto clave en los proyectos de BI, y podría abrir las puertas a la incorporación de nuevos tipos de datos que se dispongan en la interna y no sean públicos.

### **Disponibilizar y popularizar el sistema de BI**

Periódicamente los especialistas generan reportes de resultados que son compartidos a las distintas áreas interesadas dentro de la IM y a otras organizaciones con las que se mantiene una relación de colaboración.

Actualmente la interfaz de usuario esta enfocada al uso interno del Servicio de Evaluación de la Calidad y Control Ambiental por parte del personal de la Unidad Calidad de Aire.

Resulta de interés permitir el ingreso al sistema de BI posibilitando a usuarios externos a la Unidad Calidad de Aire poder realizar el análisis de los datos cuándo y cómo lo deseen.

El objetivo primordial es el de popularizar la herramienta y enmarcarla como punto de referencia para el análisis de la información y como complemento de ayuda a la toma de decisiones.

### **Automatizar generación completa del Informe Anual e Informe Semanal**

Simplificar el proceso de redacción de informes es crucial para mejorar una organización; las horas de trabajo pueden utilizarse indebidamente durante el proceso de redacción de informes si los datos no son claros y los gráficos individuales deben generarse manualmente.

Utilizar una solución que proporcione informes automatizados reducirá fácilmente las horas de redacción de informes mediante la recopilación, interpretación y presentación automática de datos de una manera comprensible y automática para que todos los involucrados puedan entender lo que se intenta transmitir en el informe.

Dicho esto, como resultado inicial del desarrollo del proyecto se logró automatizar la generación de una sección específica del Informe Anual y se brinda la posibilidad, utilizando las mismas herramientas, de extender las automatizaciones para cubrir la generación total del Informe Anual y además el Informe Semanal.

### **Generar un Dashboard utilizando Pentaho Dashboards**

Con el objetivo de brindar un acceso eficaz y eficiente a la información se sugiere el desarrollo de un dashboard con gráficas e indicadores preestablecidos para poder ser consultado de manera rápida e intuitiva.

Dentro de las posibilidades brindadas por los dashboards es posible estructurar los datos geográficos, visualizando los datos desplegados en mapas.



# Bibliografía

- [1] Icaire. <http://montevideo.gub.uy/areas-tematicas/ambiente/calidad-del-aire/icaire>, May 2014.
- [2] Informes de calidad de aire anuales. <http://montevideo.gub.uy/areas-tematicas/ambiente/calidad-del-aire/informes-historicos-de-calidad/informes-anuales-de-calidad-de-aire>, May 2014.
- [3] Informes de calidad de aire semanales. <http://montevideo.gub.uy/areas-tematicas/ambiente/calidad-del-aire/informes-de-calidad-de-aire>, Sep 2014.
- [4] Principales contaminantes del aire. <http://montevideo.gub.uy/areas-tematicas/ambiente/calidad-del-aire/principales-contaminantes-del-aire>, May 2014.
- [5] Calidad del aire. <http://montevideo.gub.uy/areas-tematicas/ambiente/calidad-del-aire>, Oct 2017.
- [6] Magic quadrant for data integration tools. <https://www.gartner.com/en/documents/3955823/magic-quadrant-for-data-integration-tools>, 2019.
- [7] ADME. Adme. <https://adme.com.uy>. Accedido: 2019-08-29.
- [8] AGESIC. Catálogo de datos abiertos. <https://catalogodatos.gub.uy/>. Accedido: 2019-08-24.
- [9] Fernando Carpani. Cmdm: Un modelo conceptual para la especificación de bases multi-dimensionales. *Instituto de Computación-Facultad de Ingeniería. Universidad de la República, Uruguay*, 2000.
- [10] David C Carslaw and Karl Ropkins. Openair—an r package for air quality data analysis. *Environmental Modelling & Software*, 27:52–61, 2012.
- [11] Centro de Gestión de Movilidad. Conteo vehicular en las principales avenidas de monteideo. <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-conteo-de-vehiculos-del-centro-de-gestion-de-la-movilidad>. Accedido: 2019-08-24.
- [12] Centro de Gestión de Movilidad. Metadatos de conteo vehicular en las principales avenidas de monteideo. <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-conteo-de-vehiculos-del-centro-de-gestion-de-la-movilidad/resource/e548c71e-1d4e-4a33-bcaf-d94ed8bfac7b>. Accedido: 2019-08-24.
- [13] Servicio Evaluación de la Calidad y Control Ambiental. Descripción de campos. <https://catalogodatos.gub.uy/dataset/>

- intendencia-montevideo-red-de-monitoreo-de-la-calidad-del-aire-de-montevideo/resource/81e717b0-9cf2-4ab0-947f-3d8fc5391100. Accedido: 2019-08-24.
- [14] Servicio Evaluación de la Calidad y Control Ambiental. Encuesta continua de hogares. <https://catalogodatos.gub.uy/dataset/ine-encuesta-continua-de-hogares-ano-2017>. Accedido: 2019-08-24.
- [15] Servicio Evaluación de la Calidad y Control Ambiental. Estaciones de medición. <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-red-de-monitoreo-de-la-calidad-del-aire-de-montevideo/resource/688ddb7c-7e9c-47cb-9cc6-c57256141f36>. Accedido: 2019-08-24.
- [16] Servicio Evaluación de la Calidad y Control Ambiental. Historico de sensores de la plataforma iot - pm10. <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-red-de-monitoreo-de-la-calidad-del-aire-de-montevideo/resource/140abf93-374f-4f67-9261-e613bfc457bb>. Accedido: 2019-08-24.
- [17] Servicio Evaluación de la Calidad y Control Ambiental. Lista de identificación de contaminantes. <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-red-de-monitoreo-de-la-calidad-del-aire-de-montevideo/resource/8c8b8110-9ee7-4382-9fdb-8fb74fd4562c>. Accedido: 2019-08-24.
- [18] Servicio Evaluación de la Calidad y Control Ambiental. Medidas históricas de la calidad del aire. <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-red-de-monitoreo-de-la-calidad-del-aire-de-montevideo/resource/8f8fc2a3-8c16-4f6d-a15f-3d8afd5c1da9>. Accedido: 2019-08-24.
- [19] Servicio Evaluación de la Calidad y Control Ambiental. Métodos de medición. <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-red-de-monitoreo-de-la-calidad-del-aire-de-montevideo/resource/e7433f67-ce42-4e56-82ba-0cd52ce9db86>. Accedido: 2019-08-24.
- [20] Servicio Evaluación de la Calidad y Control Ambiental. Vías de tránsito. <https://catalogodatos.gub.uy/dataset/intendencia-montevideo-vias-de-transito-montevideo>. Accedido: 2019-08-24.
- [21] Ministerio de Turismo. Horarios y días festivos. <https://turismo.gub.uy/index.php/horarios-y-dias-festivos>. Accedido: 2019-08-29.
- [22] Ministerio de Turismo. Rangos para radiación solar global. <http://www.climatol.eu/reclim/reclim15e.pdf>. Accedido: 2019-08-29.
- [23] Josep Curto Díaz. *Introducción al business intelligence*. Editorial UOC, 2012.
- [24] Grafana. Grafana. <https://grafana.com/>. Accedido: 2019-11-03.
- [25] Ralph Kimball and Margy Ross. *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons, 2011.
- [26] Elzbieta Malinowski and Esteban Zimanyi. *Advanced data warehouse design, advanced data warehouse design: From conventional to spatial and temporal applications, data-centric systems and applications*, 2008.
- [27] Adriana Marotta. *Diapositivas del Curso de Diseño y Construcción de Data Warehouse*. Universidad de la República Uruguay, 2017.

- [28] Dirk Merkel. Docker: lightweight linux containers for consistent development and deployment. *Linux Journal*, 2014(239):2, 2014.
- [29] Oficina Meteorológica - Gobierno de UK. Escala de beaufort, 2019. <https://web.archive.org/web/20060409071345/http://www.met-office.gov.uk/education/secondary/students/beaufort.html>, Last accessed on 2019-10-06.
- [30] PCE. Tabla de las velocidades del viento, 2019. <https://www.pce-iberica.es/medidor-detalles-tecnicos/tablas-de-velocidades-del-viento.htm>, Last accessed on 2019-10-06.
- [31] Pentaho CE Audit. Pentaho ce audit, 2019. <https://github.com/it4biz/pentaho-ce-audit>, Last accessed on 2019-11-01.
- [32] Pentaho Performance Monitoring. Pentaho performance monitoring, 2019. <https://github.com/it4biz/pentaho-performance-monitoring>, Last accessed on 2019-11-01.
- [33] PostgreSQL. Postgresql - data types, 2019. <https://www.postgresql.org/docs/9.2/datatype.html>, Last accessed on 2019-10-06.
- [34] Prometheus. Prometheus. <https://prometheus.io/>. Accedido: 2019-10-21.
- [35] RAIG. Control de la humedad relativa, 2019. <https://www.raig.com/noticias/control-de-la-humedad-relativa>, Last accessed on 2019-10-06.
- [36] Servicio Meteorológico Nacional. Gobierno Argentino. Intensidad del viento, 2019. <https://www.smn.gob.ar/noticias/\T1\textquestiondowncómo-clasificamos-la-intensidad-del-viento>, Last accessed on 2019-10-06.
- [37] R Core Team et al. R: A language and environment for statistical computing. 2013.
- [38] UNAM. Rosa de los vientos, 2019. <https://web.archive.org/web/20031002005236/http://paidoteca.dgsca.unam.mx/estadistica/prac3.html>, Last accessed on 2019-10-06.



# Anexos



# Anexo A

## Monitoreo

Dentro de las operaciones del sistema requeridas para mantener el sistema de BI en correcto funcionamiento es posible monitorizar los distintos componentes de la solución.

### A.1. Herramientas

Se presentan distintas herramientas que sirven a monitorizar los distintos componentes según su naturaleza.

#### A.1.1. Portainer

Esta aplicación permite gestionar de forma muy fácil e intuitiva los contenedores Docker a través de una interfaz gráfica web. Mediante la misma un administrador puede tener una visión global más clara de los contenedores que está ejecutando y facilitar su gestión.

En la figura A.1 se presenta la pantalla de inicio de la herramienta donde se muestra el estado de cada componente.

A su vez, y como se presenta en la figura A.2, es posible visualizar los logs de cada uno de los componentes, desde el servidor PBI, PDI hasta la base de datos PostgreSQL.

#### A.1.2. Grafana + Prometheus

Prometheus es una plataforma de monitoreo de código abierto que es capaz de recopilar métricas de objetivos monitoreados al consultar las métricas.

Después de guardar los datos recopilados, Grafana provee una aplicación web donde se puede consultarlos utilizando su lenguaje de consulta y representar todos los resultados en gráficos.

En la figura A.3 se presenta la pantalla de inicio de la herramienta en la cual se muestran indicadores del Docker Host, esto es, la máquina virtual que corre los contenedores.

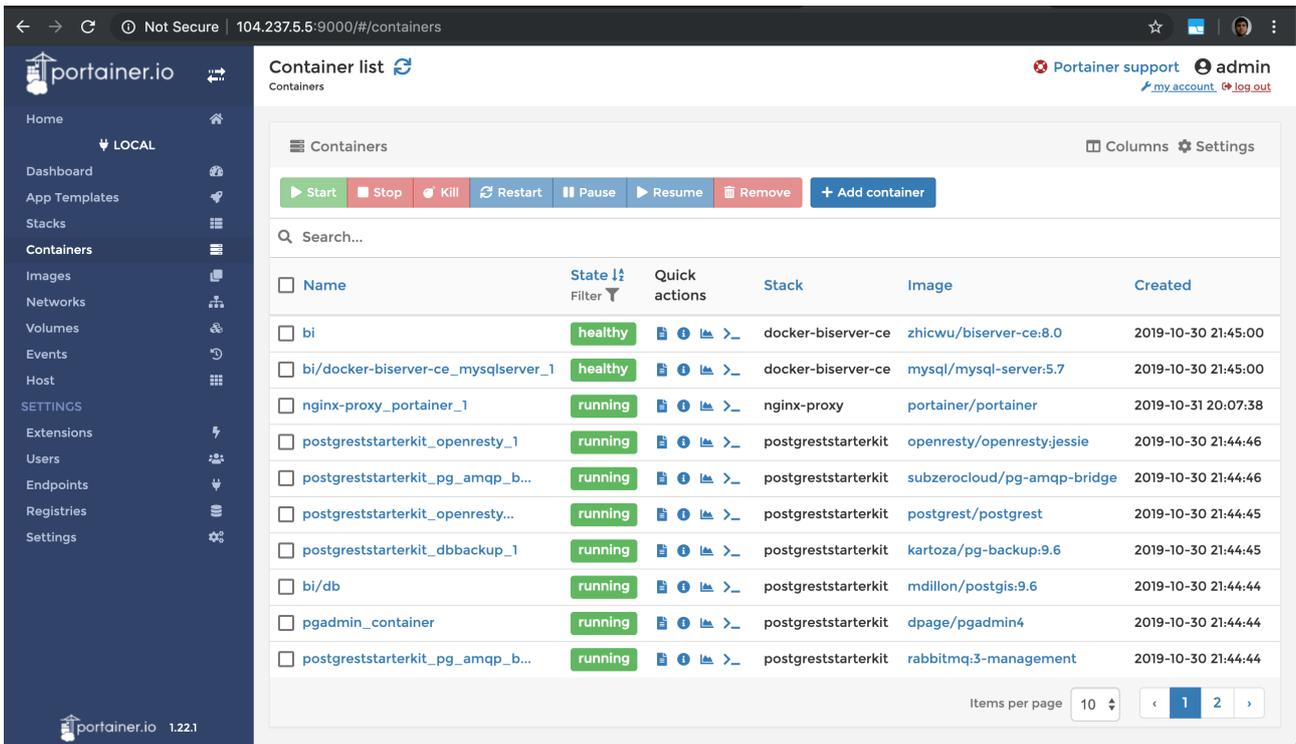


Figura A.1: Monitoreo - Portainer

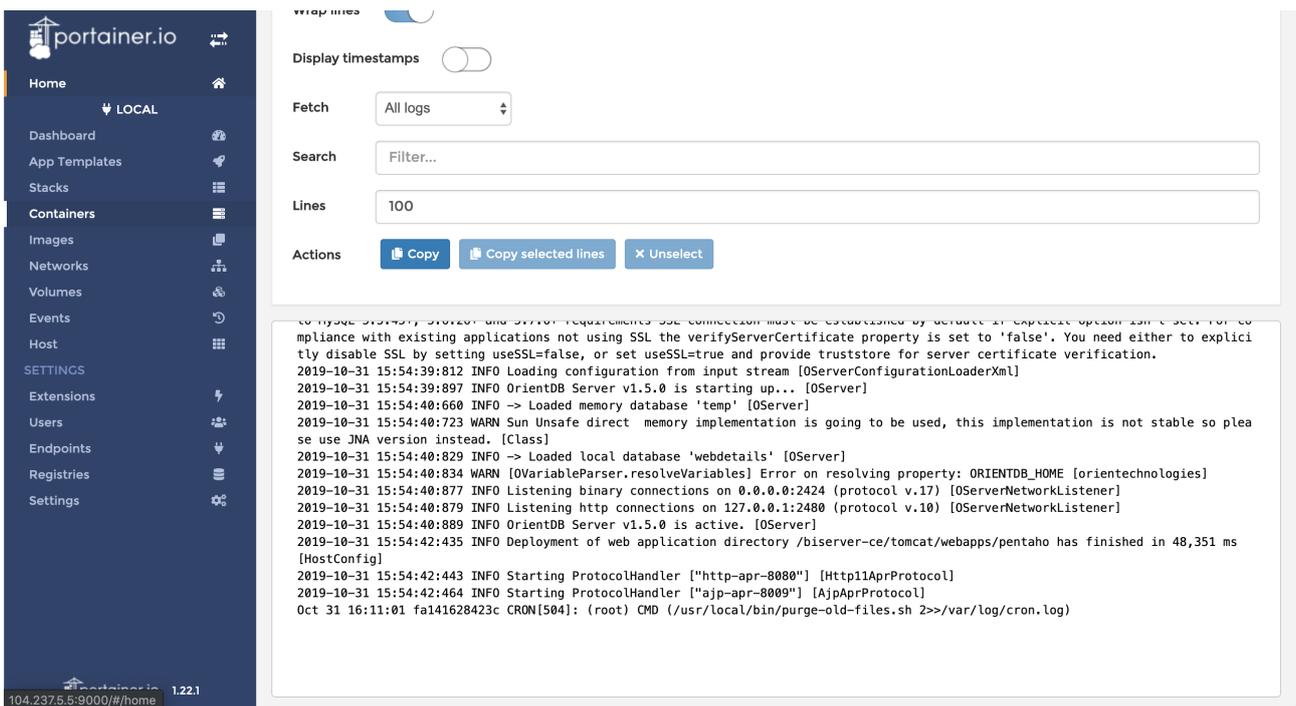


Figura A.2: Monitoreo - Portainer - Logs

En la figura A.4 se muestran indicadores de cada uno de los contenedores, aquí se puede ver que el contenedor que utiliza más memoria RAM es el servidor BI.

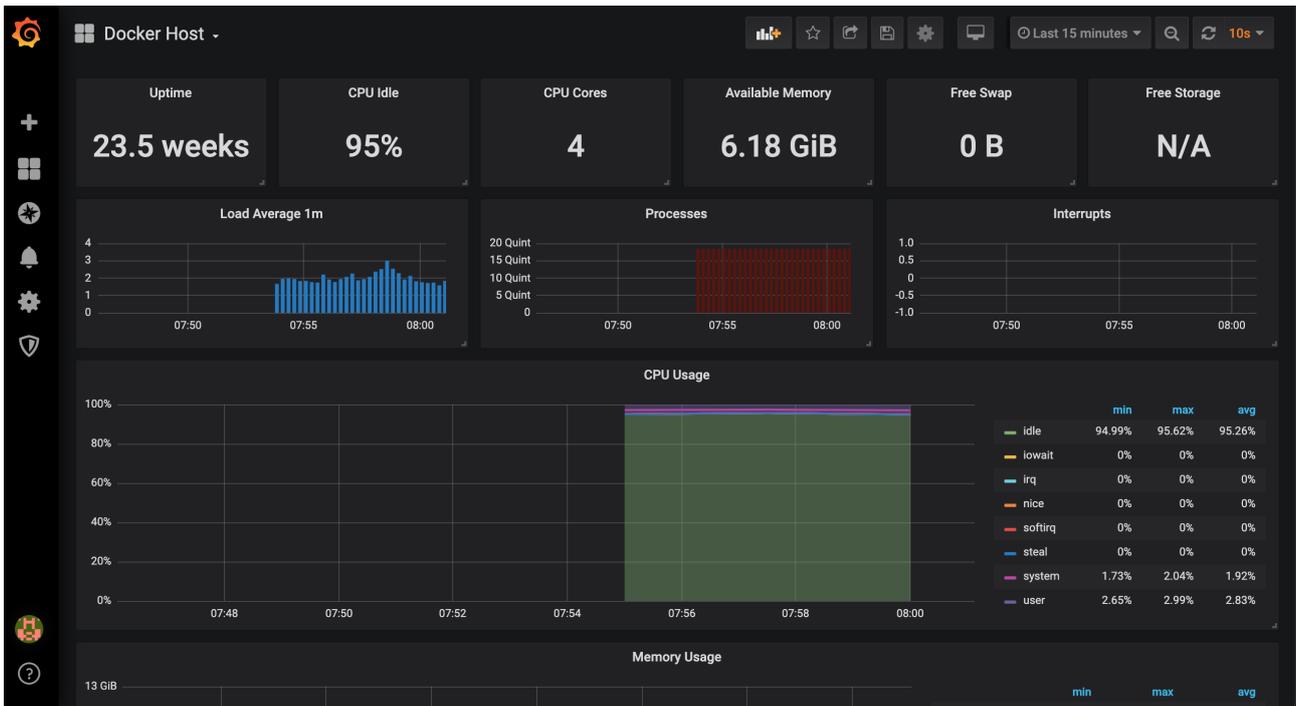


Figura A.3: Monitoreo - Grafana + Prometheus

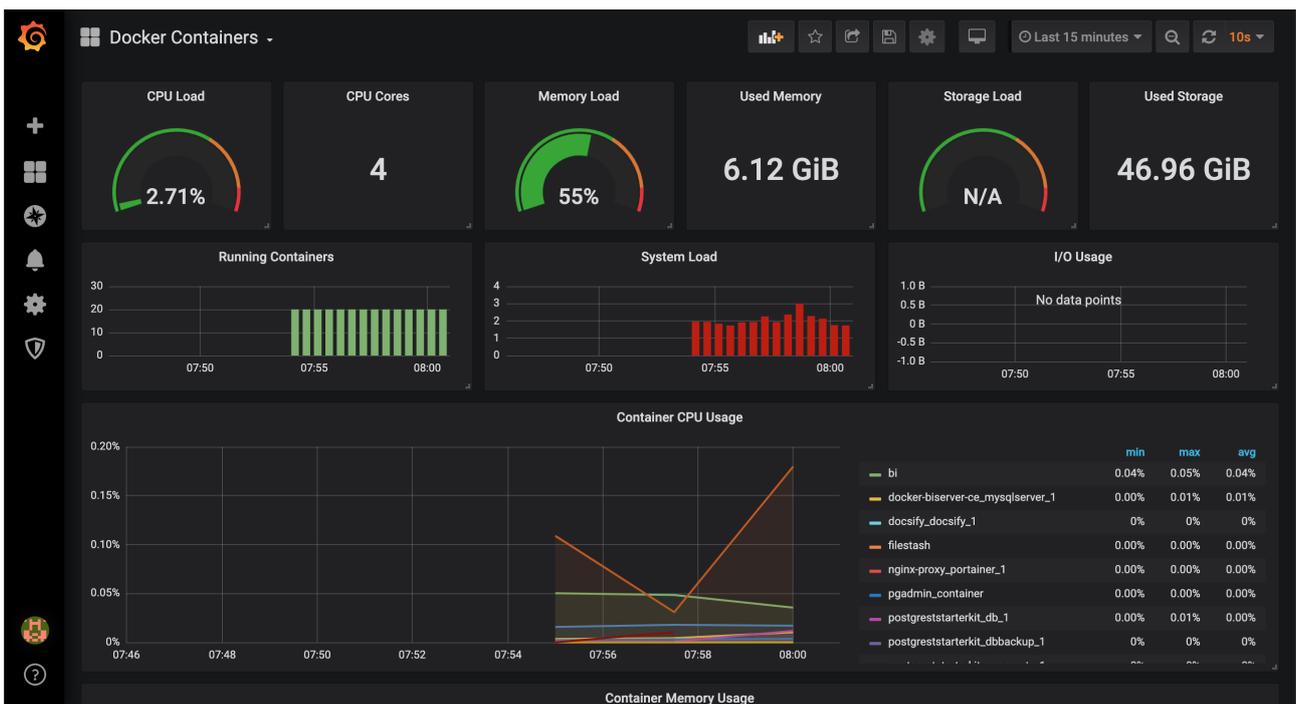


Figura A.4: Monitoreo - Grafana + Prometheus - Containers

### A.1.3. Status de Carte en PBI

Para visualizar el estado del componente encargado de la ejecución de los ETLs, específicamente, los jobs y transformaciones pre-definidas, es necesario ingresar a una página web donde se presenta dicha información, como se puede visualizar en la figura A.5.

The screenshot shows a web browser window with the title 'Kettle slave server status'. The address bar shows 'Not Secure 104.237.5.5:9090/kettle/status/'. The page content is as follows:

### PDI Status

Transformation name	Carte Object ID	Status	Last log date	Remove from list

Job name	Carte Object ID	Status	Last log date	Remove from list

### Configuration details:

Parameter	Value
The maximum size of the central log buffer	10000 lines
The maximum age of a log line	1440 minutes
The maximum age of a stale object	240 minutes
Repository name	Local

*These parameters can be set in the slave server configuration XML file: /data-integration/pwd/slave.xml*

[View site information](#)

Figura A.5: Monitorio - Status PDI

# Anexo B

## Métricas

Para poder visualizar y entender cómo los especialistas de la IM están utilizando la herramienta y las funcionalidades brindadas, se utilizan plugins Open Source de Pentaho de analítica web para luego, basado en dicho entendimiento, investigar, decidir y aplicar las oportunidades de mejora correctiva y evolutiva definidos en el sistema de BI.

### B.1. Herramientas

Se presentan distintas herramientas que sirven a auditar y presentar métricas de uso de la ejecución de análisis realizados en el servidor de BI.

#### B.1.1. Pentaho CE Audit [31]

Total control sobre quién y ingresó a la herramienta y cuándo lo hizo. Además se puede auditar a qué recursos se accedió y cuándo se realizó. En la figura B.1 se puede visualizar un ejemplo.

#### B.1.2. Pentaho Performance Monitoring [32]

Total control sobre qué consultas son las más realizadas, cuáles son las más rápidas y cuáles las más lentas. En la figura B.2 se puede visualizar un ejemplo.

### B.2. Interacción

El sistema de BI presentado consta de múltiples procesos en los cuales los distintos componentes de la solución interactúan para lograr distintos objetivos. A continuación se presentan algunos procesos y se describe la interacción de los componentes en cada uno.

Como se puede desprender del diagrama presentado el usuario accede a múltiples aplicaciones web:



Figura B.1: Métricas - Pentaho CE Audit

1. Aplicación web: Carga, actualización y reportes.
2. Documentación en línea sobre la plataforma.
3. Saiku (PBI CE): Realización de análisis dinámicos.

Para la gestión de las fuentes de datos que se utilizarán para cargar el DW la aplicación web se comunica con la base de datos mediante una capa de servicios realizada con PostgREST.

Para la carga y actualización de datos del DW, iniciada por el usuario, en primer paso carga al sistema de gestión de archivos las fuentes de datos a utilizar por los ETL. Luego, mediante la aplicación web, realiza la ejecución de los ETL cargados en PDI utilizando los servicios provistos con el servidor Carte. Los ETL se comunican con el sistema de archivos para tomar las fuentes de datos cargadas por el usuario, dado que no es posible cargarlos a los ETL mediante Carte, y con la base de datos fuentes previamente cargadas por el usuario. Finalmente el ETL impacta los datos en la base de datos del DW.

Para la generación del reporte de BI sobre la sección del Informe Anual, iniciada por el usuario desde la aplicación web, se ejecuta un ETL que se comunica con la base de datos del DW y con el servidor de R para obtener los datos necesarios y realizar los cálculos estadísticos respectivos.



Please help us do more, with just a "Like" on our Facebook page

[Curtis](#) [Compartilhar](#)  
or a "follow us" on twitter

[Tweet #PentahoCEAudit](#) [Follow @t4bizbrasil](#)

Powered by t4biz

Figura B.2: Métricas - Pentaho Performance Monitoring

Finalmente se genera el archivo PDF con el reporte y se lo envía al usuario por correo.



# Anexo C

## Carga y actualización

A continuación se detallan los procesos de extracción, transformación y carga más relevantes del DW que fueron descritos en el capítulo de Implementación en la sección Carga y actualización.

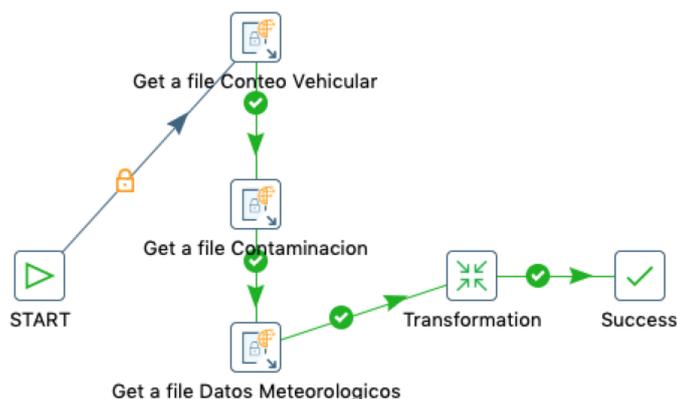


Figura C.1: Carga Vehículos Contaminación - Job

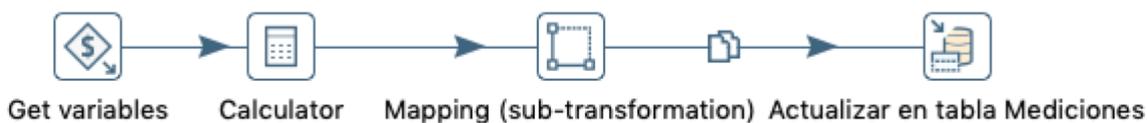


Figura C.2: Carga Vehículos Contaminación - Transformación Padre

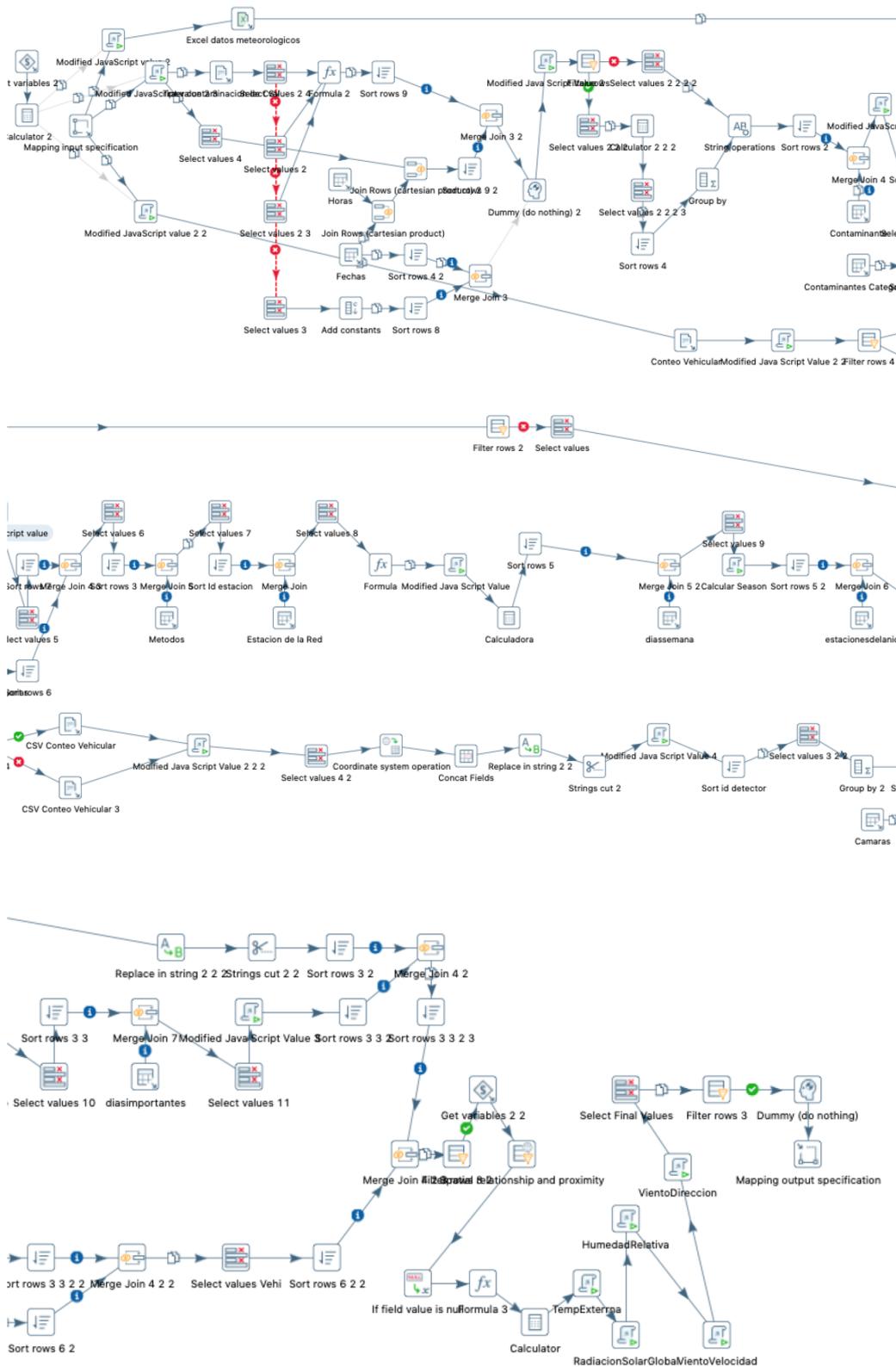


Figura C.3: Carga Vehículos Contaminación - Transformación Hija

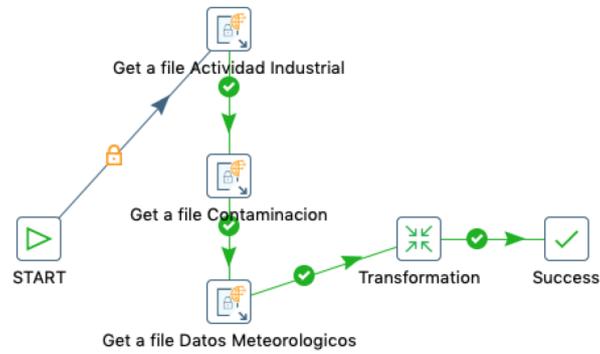


Figura C.4: Carga Industrias Contaminación - Job

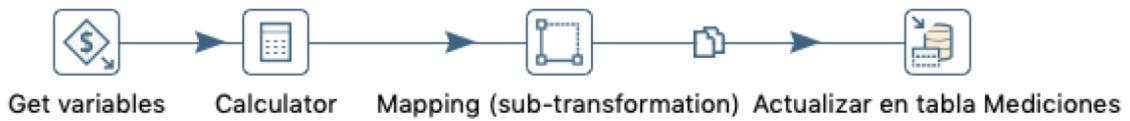


Figura C.5: Carga Industrias Contaminación - Transformación Padre

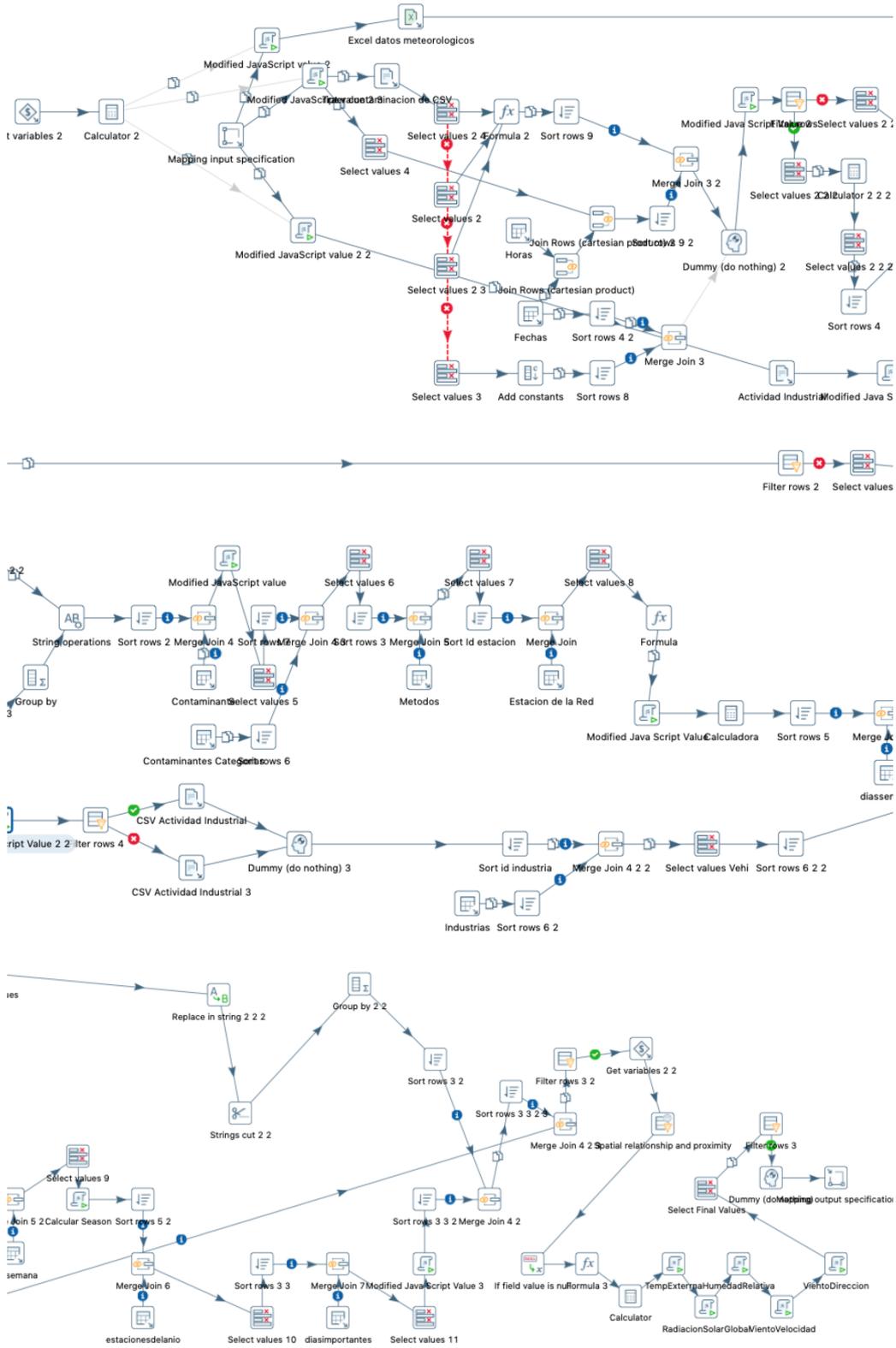


Figura C.6: Carga Industrias Contaminación - Transformación Hija

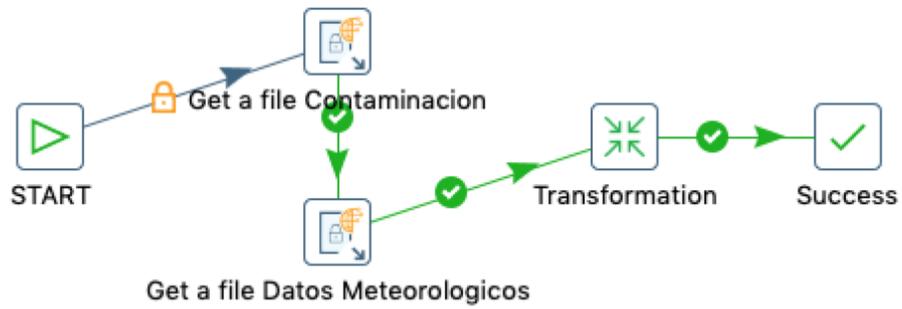


Figura C.7: Carga Días Importantes Contaminación - Job

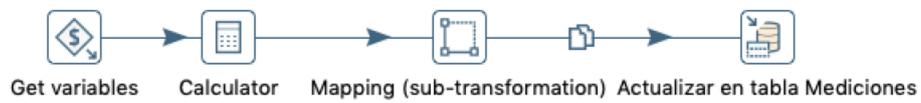


Figura C.8: Carga Días Importantes Contaminación - Transformación Padre

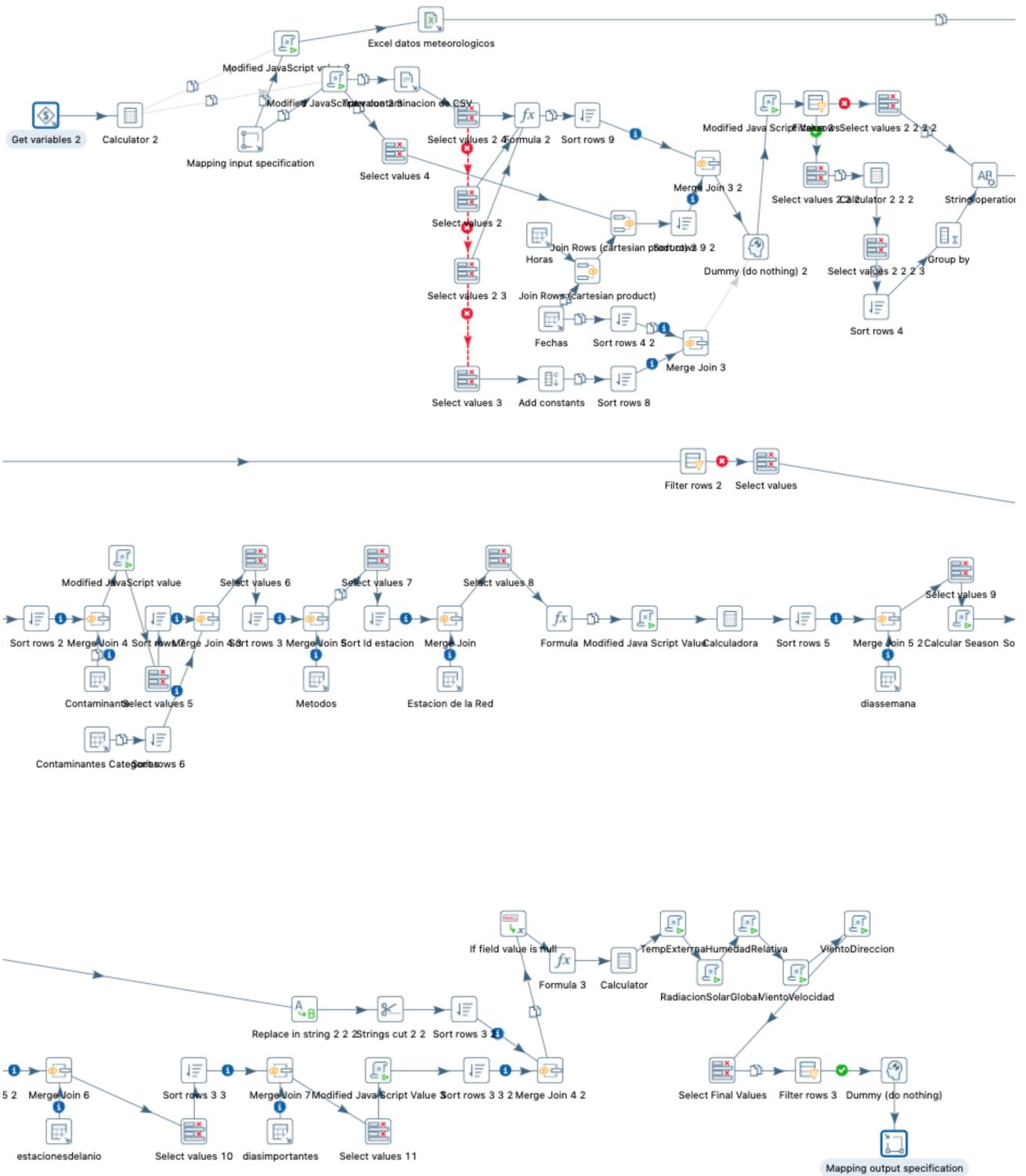


Figura C.9: Carga Días Importantes Contaminación - Transformación Hija

# Anexo D

## Implementación de cubos

A continuación se explicitan en detalle los cubos definidos en la etapa de Implementación.

```
<Schema name="CalidadDeAire">
  <Dimension type="StandardDimension" visible="true"
    highCardinality="false" name="Fechas">
    <Hierarchy name="Fechas" visible="true" hasAll="true">
      <Table name="fechas" schema="dw">
      </Table>
      <Level name="Anio" visible="true" table="fechas" column="
        anio" type="Integer" uniqueMembers="false" levelType="
        Regular" hideMemberIf="Never">
      </Level>
      <Level name="Mes" visible="true" table="fechas" column="
        idmes" type="String" uniqueMembers="false" levelType="
        Regular" hideMemberIf="Never" captionColumn="mes">
      </Level>
      <Level name="Fecha" visible="true" table="fechas" column="
        idfecha" type="Integer" uniqueMembers="false" levelType=
        "Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Dimension type="StandardDimension" visible="true"
    highCardinality="false" name="Contaminantes">
    <Hierarchy name="Contaminantes" visible="true" hasAll="true">
      <Table name="contaminantes" schema="dw">
      </Table>
      <Level name="Tipo" visible="true" table="contaminantes"
        column="idtipo" type="Integer" uniqueMembers="false"
        levelType="Regular" hideMemberIf="Never" captionColumn="
        nomtipo">
      </Level>
      <Level name="Subtipo" visible="true" table="contaminantes"
        column="idsubtipo" type="String" uniqueMembers="false">
      </Level>
    </Hierarchy>
  </Dimension>
</Schema>
```

```

        levelType="Regular" hideMemberIf="Never" captionColumn="
        nomsubtipo">
    </Level>
    <Level name="Contaminante" visible="true" table="
        contaminantes" column="idcontaminante" type="String"
        uniqueMembers="false" levelType="Regular" hideMemberIf="
        Never" captionColumn="nomcontaminante">
    </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
    highCardinality="false" name="EstacionesDeLaRed">
    <Hierarchy name="EstacionesDeLaRed" visible="true" hasAll="
        true">
        <Table name="estacionesdelared" schema="dw">
        </Table>
        <Level name="CCZ" visible="true" table="estacionesdelared"
            column="idccz" type="String" uniqueMembers="false"
            levelType="Regular" hideMemberIf="Never" captionColumn="
            nomccz">
        </Level>
        <Level name="Barrio" visible="true" table="
            estacionesdelared" column="idbarrio" type="String"
            uniqueMembers="false" levelType="Regular" hideMemberIf="
            Never" captionColumn="nombarrio">
        </Level>
        <Level name="EstacionDeLaRed" visible="true" table="
            estacionesdelared" column="nomestaciondelared" type="
            String" uniqueMembers="false" levelType="Regular"
            hideMemberIf="Never" captionColumn="nomestaciondelared">
        </Level>
    </Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
    highCardinality="false" name="DiasSemana">
    <Hierarchy name="DiasSemana" visible="true" hasAll="true">
        <Table name="diassemana" schema="dw">
        </Table>
        <Level name="DiaSemana" visible="true" table="diassemana"
            column="iddiasemana" type="String" uniqueMembers="false"
            levelType="Regular" hideMemberIf="Never" captionColumn=
            "nomdiasemana">
        </Level>
    </Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
    highCardinality="false" name="EstacionesDelAnio">

```

```

<Hierarchy name="EstacionesDelAnio" visible="true" hasAll="
  true">
  <Table name="estacionesdelanio" schema="dw">
  </Table>
  <Level name="EstacionDelAnio" visible="true" table="
    estacionesdelanio" column="idestaciondelanio" type="
    String" uniqueMembers="false" levelType="Regular"
    hideMemberIf="Never" captionColumn="nomestaciondelanio">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="Horas">
<Hierarchy name="Horas" visible="true" hasAll="true">
  <Table name="horas" schema="dw">
  </Table>
  <Level name="Hora" visible="true" table="horas" column="
    idhora" type="String" uniqueMembers="false" levelType="
    Regular" hideMemberIf="Never" captionColumn="hora">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="Metodos">
<Hierarchy name="Metodos" visible="true" hasAll="true">
  <Table name="metodos" schema="dw">
  </Table>
  <Level name="Metodo" visible="true" table="metodos" column=
    "idmetodo" type="String" uniqueMembers="false" levelType
    ="Regular" hideMemberIf="Never" captionColumn="nommetodo
    ">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="TemperaturaExterna">
<Hierarchy name="TemperaturaExterna" visible="true" hasAll="
  true">
  <Table name="temperaturaexterna" schema="dw">
  </Table>
  <Level name="TemperaturaExterna" visible="true" table="
    temperaturaexterna" column="idrangotemperaturaexterna"
    type="Numeric" uniqueMembers="false" levelType="Regular"
    hideMemberIf="Never" captionColumn="
    descrangotemperaturaexterna">
  </Level>
</Hierarchy>
</Dimension>

```

```

<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="ContaminantesCategorias">
  <Hierarchy name="ContaminantesCategorias" visible="true"
    hasAll="true">
    <Table name="contaminantescategorias" schema="dw">
    </Table>
    <Level name="ContaminanteCategoria" visible="true" table="
      contaminantescategorias" column="idcontaminante" type="
      String" uniqueMembers="false" levelType="Regular"
      hideMemberIf="Never" captionColumn="
      descrangocontaminantescategorias">
    </Level>
  </Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="HumedadRelativa">
  <Hierarchy name="HumedadRelativa" visible="true" hasAll="true
  ">
  <Table name="humedadrelativa" schema="dw">
  </Table>
  <Level name="HumedadRelativa" visible="true" table="
    humedadrelativa" column="idrangohumedadrelativa" type="
    Numeric" uniqueMembers="false" levelType="Regular"
    hideMemberIf="Never" captionColumn="
    descrangohumedadrelativa">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="RadiacionSolarGlobal">
  <Hierarchy name="RadiacionSolarGlobal" visible="true" hasAll=
  "true">
  <Table name="radiacionsolarglobal" schema="dw">
  </Table>
  <Level name="RadiacionSolarGlobal" visible="true" table="
    radiacionsolarglobal" column="
    idrangoradiacionsolarglobal" type="Numeric"
    uniqueMembers="false" levelType="Regular" hideMemberIf="
    Never" captionColumn="descrangoradiacionsolarglobal">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="VientoDireccion">
  <Hierarchy name="VientoDireccion" visible="true" hasAll="true
  ">
  <Table name="vientodireccion" schema="dw">
  </Table>

```

```

    <Level name="VientoDireccion" visible="true" table="
      vientodireccion" column="idrangovientodireccion" type="
      Numeric" uniqueMembers="false" levelType="Regular"
      hideMemberIf="Never" captionColumn="
      descrangovientodireccion">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="VientoVelocidad">
  <Hierarchy name="VientoVelocidad" visible="true" hasAll="true
    ">
    <Table name="vientovelocidad" schema="dw">
  </Table>
    <Level name="VientoVelocidad" visible="true" table="
      vientovelocidad" column="idrangovientovelocidad" type="
      Numeric" uniqueMembers="false" levelType="Regular"
      hideMemberIf="Never" captionColumn="
      descrangovientovelocidad">
  </Level>
  </Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="Industrias">
  <Hierarchy name="Industrias" visible="true" hasAll="true">
    <Table name="industrias" schema="dw">
  </Table>
    <Level name="CCZ" visible="true" table="industrias" column=
      "idindustria" type="Numeric" uniqueMembers="false"
      levelType="Regular" hideMemberIf="Never">
  </Level>
    <Level name="Barrio" visible="true" table="industrias"
      column="idindustria" type="Numeric" uniqueMembers="false
      " levelType="Regular" hideMemberIf="Never">
  </Level>
    <Level name="Industria" visible="true" table="industrias"
      column="idindustria" type="Numeric" uniqueMembers="false
      " levelType="Regular" hideMemberIf="Never" captionColumn
      ="nomindustria">
  </Level>
  </Hierarchy>
  <Hierarchy name="Tipo" visible="true" hasAll="true">
    <Table name="industrias" schema="dw">
  </Table>
    <Level name="Tipo" visible="true" table="industrias" column
      ="idindustria" type="Numeric" uniqueMembers="false"
      levelType="Regular" hideMemberIf="Never" captionColumn="
      nomindustria">

```

```

</Level>
<Level name="Industria" visible="true" table="industrias"
  column="idindustria" type="Numeric" uniqueMembers="false"
  levelType="Regular" hideMemberIf="Never" captionColumn
  ="nomindustria">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="DiasImportantes">
<Hierarchy name="DiasImportantes" visible="true" hasAll="true"
  ">
  <Table name="diasimportantes" schema="dw">
  </Table>
  <Level name="DiaImportante" visible="true" table="
  diasimportantes" column="iddiaimportante" type="String"
  uniqueMembers="false" levelType="Regular" hideMemberIf="
  Never" captionColumn="nomdiaimportante">
  </Level>
  <Level name="Tipo" visible="true" table="diasimportantes"
  column="idtipodiaimportante" uniqueMembers="false"
  levelType="Regular" hideMemberIf="Never" captionColumn="
  nomtipodiaimportante">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true"
  highCardinality="false" name="Camaras">
<Hierarchy name="Camaras" visible="true" hasAll="true">
  <Table name="camaras" schema="dw">
  </Table>
  <Level name="CCZ" visible="true" table="camaras" column="
  iddetector" type="Numeric" uniqueMembers="false"
  levelType="Regular" hideMemberIf="Never">
  </Level>
  <Level name="Barrio" visible="true" table="camaras" column=
  "iddetector" type="Numeric" uniqueMembers="false"
  levelType="Regular" hideMemberIf="Never">
  </Level>
  <Level name="Camara" visible="true" table="camaras" column=
  "iddetector" type="Numeric" uniqueMembers="false"
  levelType="Regular" hideMemberIf="Never" captionColumn="
  dscubicacion">
  </Level>
</Hierarchy>
</Dimension>
<Cube name="VehiculosContaminacion" visible="true" cache="true"
  enabled="true">

```

```

<Table name="vehiculoscontaminacion" schema="dw">
</Table>
<DimensionUsage source="Tiempo" level="Fecha" name="Tiempo"
  visible="true" foreignKey="idfecha" highCardinality="false
">
</DimensionUsage>
<DimensionUsage source="Contaminantes" level="Contaminante"
  name="Contaminantes" visible="true" foreignKey="
  idcontaminante" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="EstacionesDeLaRed" level="
  EstacionDeLaRed" name="EstacionesDeLaRed" visible="true"
  foreignKey="nomestaciondelared" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="DiasSemana" level="DiaSemana" name="
  DiasSemana" visible="true" foreignKey="iddiasemana"
  highCardinality="false">
</DimensionUsage>
<DimensionUsage source="EstacionesDelAnio" level="
  EstacionDelAnio" name="EstacionesDelAnio" visible="true"
  foreignKey="idestaciondelanio" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Horas" level="Hora" name="Horas"
  visible="true" foreignKey="idhora" highCardinality="false"
  >
</DimensionUsage>
<DimensionUsage source="Metodos" level="Metodo" name="Metodos
  " visible="true" foreignKey="idmetodo" highCardinality="
  false">
</DimensionUsage>
<DimensionUsage source="TemperaturaExterna" level="
  TemperaturaExterna" name="TemperaturaExterna" visible="
  true" foreignKey="idrangotemperaturaexterna"
  highCardinality="false">
</DimensionUsage>
<DimensionUsage source="ContaminantesCategorias" level="
  ContaminanteCategoria" name="ContaminantesCategorias"
  visible="true" foreignKey="surrogatekey" highCardinality="
  false">
</DimensionUsage>
<DimensionUsage source="HumedadRelativa" level="
  HumedadRelativa" name="HumedadRelativa" visible="true"
  foreignKey="idrangohumedadrelativa" highCardinality="false
  ">
</DimensionUsage>
<DimensionUsage source="RadiacionSolarGlobal" level="
  RadiacionSolarGlobal" name="RadiacionSolarGlobal" visible=
  "true" foreignKey="idrangoradiacionsolarglobal"

```

```

    highCardinality="false">
</DimensionUsage>
<DimensionUsage source="VientoDireccion" level="
    VientoDireccion" name="VientoDireccion" visible="true"
    foreignKey="idrangovientodireccion" highCardinality="false
    ">
</DimensionUsage>
<DimensionUsage source="VientoVelocidad" level="
    VientoVelocidad" name="VientoVelocidad" visible="true"
    foreignKey="idrangovientovelocidad" highCardinality="false
    ">
</DimensionUsage>
<DimensionUsage source="Camaras" level="Camara" name="Camara"
    visible="true" foreignKey="iddetector" highCardinality="
    false">
</DimensionUsage>
<Measure name="promContaminacion" column="contaminacion"
    datatype="Numeric" aggregator="avg" visible="true">
</Measure>
<Measure name="contador" column="surrogatekey" aggregator="
    count" visible="true">
</Measure>
<Measure name="sumVolumenVehicular" column="volumenvehiculos"
    aggregator="sum" visible="true">
</Measure>
<Measure name="maxContaminacion" column="contaminacion"
    aggregator="max" visible="true">
</Measure>
</Cube>
<Cube name="IndustriasContaminacion" visible="true" cache="true
    " enabled="true">
<Table name="industriascontaminacion" schema="dw">
</Table>
<DimensionUsage source="Fechas" level="Fecha" name="Fechas"
    visible="true" foreignKey="idfecha" highCardinality="false
    ">
</DimensionUsage>
<DimensionUsage source="Contaminantes" level="Contaminante"
    name="Contaminantes" visible="true" foreignKey="
    idcontaminante" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="EstacionesDeLaRed" level="
    EstacionDeLaRed" name="EstacionesDeLaRed" visible="true"
    foreignKey="nomestaciondelared" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="DiasSemana" level="DiaSemana" name="
    DiasSemana" visible="true" foreignKey="iddiasemana"
    highCardinality="false">

```

```

</DimensionUsage>
<DimensionUsage source="EstacionesDelAnio" level="
  EstacionDelAnio" name="EstacionesDelAnio" visible="true"
  foreignKey="idestaciondelanio" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Horas" level="Hora" name="Horas"
  visible="true" foreignKey="idhora" highCardinality="false"
  >
</DimensionUsage>
<DimensionUsage source="Metodos" level="Metodo" name="Metodos
  " visible="true" foreignKey="idmetodo" highCardinality="
  false">
</DimensionUsage>
<DimensionUsage source="TemperaturaExterna" level="
  TemperaturaExterna" name="TemperaturaExterna" visible="
  true" foreignKey="idrangotemperaturaexterna"
  highCardinality="false">
</DimensionUsage>
<DimensionUsage source="ContaminantesCategorias" level="
  ContaminanteCategoria" name="ContaminantesCategorias"
  visible="true" foreignKey="surrogatekey" highCardinality="
  false">
</DimensionUsage>
<DimensionUsage source="HumedadRelativa" level="
  HumedadRelativa" name="HumedadRelativa" visible="true"
  foreignKey="idrangohumedadrelativa" highCardinality="false
  ">
</DimensionUsage>
<DimensionUsage source="RadiacionSolarGlobal" level="
  RadiacionSolarGlobal" name="RadiacionSolarGlobal" visible=
  "true" foreignKey="idrangoradiacionsolarglobal"
  highCardinality="false">
</DimensionUsage>
<DimensionUsage source="VientoDireccion" level="
  VientoDireccion" name="VientoDireccion" visible="true"
  foreignKey="idrangovientodireccion" highCardinality="false
  ">
</DimensionUsage>
<DimensionUsage source="VientoVelocidad" level="
  VientoVelocidad" name="VientoVelocidad" visible="true"
  foreignKey="idrangovientovelocidad" highCardinality="false
  ">
</DimensionUsage>
<DimensionUsage source="Industrias" level="Industria" name="
  Industria" visible="true" foreignKey="idindustria"
  highCardinality="false">
</DimensionUsage>

```

```

<Measure name="promContaminacion" column="contaminacion"
  datatype="Numeric" aggregator="avg" visible="true">
</Measure>
<Measure name="contador" column="surrogatekey" aggregator="
  count" visible="true">
</Measure>
<Measure name="promActividadIndustrial" column="
  actividadindustrial" aggregator="avg" visible="true">
</Measure>
<Measure name="maxContaminacion" column="contaminacion"
  aggregator="max" visible="true">
</Measure>
<Measure name="maxActividadIndustrial" column="
  actividadindustrial" aggregator="max" visible="true">
</Measure>
</Cube>
<Cube name="DiasImportantesContaminacion" visible="true" cache=
  "true" enabled="true">
<Table name="diasimportantescontaminacion" schema="dw">
</Table>
<DimensionUsage source="Fechas" level="Fecha" name="Fechas"
  visible="true" foreignKey="idfecha" highCardinality="false
  ">
</DimensionUsage>
<DimensionUsage source="Contaminantes" level="Contaminante"
  name="Contaminantes" visible="true" foreignKey="
  idcontaminante" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="EstacionesDeLaRed" level="
  EstacionDeLaRed" name="EstacionesDeLaRed" visible="true"
  foreignKey="nomestaciondelared" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="DiasSemana" level="DiaSemana" name="
  DiasSemana" visible="true" foreignKey="iddiasemana"
  highCardinality="false">
</DimensionUsage>
<DimensionUsage source="EstacionesDelAnio" level="
  EstacionDelAnio" name="EstacionesDelAnio" visible="true"
  foreignKey="idestaciondelanio" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Horas" level="Hora" name="Horas"
  visible="true" foreignKey="idhora" highCardinality="false"
  >
</DimensionUsage>
<DimensionUsage source="Metodos" level="Metodo" name="Metodos
  " visible="true" foreignKey="idmetodo" highCardinality="
  false">
</DimensionUsage>

```

```

<DimensionUsage source="TemperaturaExterna" level="
  TemperaturaExterna" name="TemperaturaExterna" visible="
  true" foreignKey="idrangotemperaturaexterna"
  highCardinality="false">
</DimensionUsage>
<DimensionUsage source="DiasImportantes" level="DiaImportante
  " name="DiasImportantes" visible="true" foreignKey="
  iddiaimportante" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="ContaminantesCategorias" level="
  ContaminanteCategoria" name="ContaminantesCategorias"
  visible="true" foreignKey="surrogatekey" highCardinality="
  false">
</DimensionUsage>
<DimensionUsage source="HumedadRelativa" level="
  HumedadRelativa" name="HumedadRelativa" visible="true"
  foreignKey="idrangohumedadrelativa" highCardinality="false
  ">
</DimensionUsage>
<DimensionUsage source="RadiacionSolarGlobal" level="
  RadiacionSolarGlobal" name="RadiacionSolarGlobal" visible=
  "true" foreignKey="idrangoradiacionsolarglobal"
  highCardinality="false">
</DimensionUsage>
<DimensionUsage source="VientoDireccion" level="
  VientoDireccion" name="VientoDireccion" visible="true"
  foreignKey="idrangovientodireccion" highCardinality="false
  ">
</DimensionUsage>
<DimensionUsage source="VientoVelocidad" level="
  VientoVelocidad" name="VientoVelocidad" visible="true"
  foreignKey="idrangovientovelocidad" highCardinality="false
  ">
</DimensionUsage>
<Measure name="promContaminacion" column="contaminacion"
  datatype="Numeric" aggregator="avg" visible="true">
</Measure>
<Measure name="contador" column="surrogatekey" aggregator="
  count" visible="true">
</Measure>
<Measure name="maxContaminacion" column="contaminacion"
  aggregator="max" visible="true">
</Measure>
</Cube>
</Schema>

```