



## Software de Business Intelligence en la analítica del Monitoreo de los contaminantes del aire en Lima Metropolitana – SENAMHI

### Autores:

**Cavero Gomero, Sandro Luis** (<https://orcid.org/0000-0002-7942-6543>)  
sacaverog@ucvvirtual.edu.pe, Ingeniería de Sistemas, Ciclo VII

**Castro Quicaña, Eduardo Franco** (<https://orcid.org/0000-0002-2133-0522>)  
ecastroqu01@ucvvirtual.edu.pe, Ingeniería de Sistemas, VIII

**Cruz Laos, Piero Fabrizio** (<https://orcid.org/0000-0003-0575-9836>)  
pcruzla26@ucvvirtual.edu.pe, Ingeniería de Sistemas, Ciclo VII

**Diaz Asto, Franz Jhamir** (<https://orcid.org/0000-0002-4489-0768>)  
fdiazas@ucvvirtual.edu.pe, Ingeniería de Sistemas, Ciclo VII

**Gonzales Lopez, Benjamin Elivelton** (<https://orcid.org/0000-0003-3271-158X>)  
bgonzaleslo@ucvvirtual.edu.pe, Ingeniería de Sistemas, VIII

**Soto Romero, Jack Steven Francesco** (<https://orcid.org/0000-0002-7092-7683>)  
jsotoro02@ucvvirtual.edu.pe, Ingeniería de Sistemas, VII

### Asesor:

Dr. Flores Chacón, Erick Giovanny ([orcid.org/0000-0002-4028-8059](https://orcid.org/0000-0002-4028-8059))

San Juan de Lurigancho - 2025

<b>Generalidades:</b>	
• Nivel:	III
• Objetivo de Desarrollo Sostenible y Meta:	Industria, Innovación e Infraestructura
• Línea de Investigación:	Tecnologías de la información y comunicación
• Línea de Responsabilidad Social Universitaria:	Desarrollo económico, empleo y emprendimiento

## **Resumen**

El presente informe describe el diseño e implementación de un software de Business Intelligence para el monitoreo y análisis de los contaminantes del aire en Lima Metropolitana, utilizando datos abiertos del Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) y del sistema de vigilancia de infecciones respiratorias agudas (IRA). El proyecto considera la construcción de un Data Mart con modelo dimensional que integra dimensiones de estación, ubicación, tiempo y contaminante, así como una tabla de hechos de contaminación alimentada mediante procesos ETL desarrollados en SQL Server y SSIS. Sobre esta arquitectura se implementan cinco dashboards en Power BI que permiten realizar análisis descriptivos, temporales, comparativos y de alertas respecto a los principales contaminantes atmosféricos. Adicionalmente, se propone un modelo analítico integrado Aire-Salud mediante la tabla de hechos Fact\_AireSalud, que relaciona los promedios anuales de PM2.5, PM10 y NO<sub>2</sub> con los episodios de IRA, neumonías, hospitalizaciones y defunciones en siete distritos de Lima Metropolitana. Los resultados muestran que la solución BI facilita la exploración interactiva de los datos, la identificación de periodos y zonas con mayor carga de contaminación y la generación de evidencia exploratoria sobre posibles asociaciones entre calidad del aire y salud respiratoria, constituyendo una herramienta de apoyo para la gestión ambiental y sanitaria.

## **Abstract**

This report presents the design and implementation of a Business Intelligence software solution for monitoring and analysing air pollutants in Metropolitan Lima, using open data from the Peruvian National Meteorology and Hydrology Service (SENAMHI) and the acute respiratory infection (ARI) surveillance system. The project includes the construction of a dimensional Data Mart that integrates station, location, time and pollutant dimensions, together with a fact table for air pollution fed through ETL processes implemented in SQL Server and SSIS. On top of this architecture, five dashboards were developed in Power BI to support descriptive, temporal, comparative and alert-based analyses of the main atmospheric pollutants. In addition, an integrated Air-Health analytical model is proposed through the Fact\_AireSalud fact table, which links annual averages of PM2.5, PM10 and NO<sub>2</sub> with ARI episodes, pneumonia cases, hospitalizations and deaths in seven districts of Metropolitan Lima. The results show that the BI solution enables interactive data exploration, identification of periods and areas with higher pollution burden, and exploratory evidence of potential associations between air quality and respiratory health, thus providing a decision-support tool for environmental and public health management.

## ÍNDICE

<b>Resumen</b>	<b>2</b>
<b>Abstract</b>	<b>3</b>
<b>1. Análisis</b>	<b>5</b>
1.1. Ficha técnica de la base de datos SENAMHI	5
1.2. Ficha técnica de la base de datos de salud (IRA)	6
1.3. Consultas analíticas	7
1.4. Diccionario de datos	8
1.5. Relación entre calidad del aire e infecciones respiratorias agudas	9
1.6. Arquitectura tecnológica	10
<b>2. Diseño</b>	<b>11</b>
2.1. Modelo Dimensional	11
2.2. Modelo Conceptual	12
2.3. Modelo Lógico	12
2.4. Modelo analítico integrado Aire-Salud	13
2.5. Diseño de migración de datos Senamhi	14
<b>3. Construcción</b>	<b>17</b>
3.1. Código OLTP	17
3.2. Script Data Mart	18
3.3. Código SSIS	20
3.4. Modelo Lógico final de Data Mart	31
3.5. Esquema del ETL	32
<b>4. Desarrollo de Dashboards</b>	<b>36</b>
<b>4.1. Interfaz Análisis de Contaminantes por Estación</b>	<b>36</b>
<b>4.1.1. Análisis de requerimientos</b>	<b>36</b>
4.1.2. Esquema de Data Mart necesario para el requerimiento	36
4.2. Diseño de estructura e integración	37
4.2.1. Estructura y componentes de Interfaz Gráfica de Usuario	37
4.2.2. Integración de Campos (datos) y Objetos gráficos de visualización (OGV)	37
4.3. Construcción	38
4.3.1. Tablero de resultados (Dashboard)	38
4.4. Despliegue	41
4.5. Dashboard “Correlación Aire-Salud”	44
4.5.1. Requerimientos funcionales	44
4.5.2. Esquema de datos utilizado	44
4.5.3. Diseño de la interfaz	44
4.5.4. Despliegue	45
<b>5. Conclusiones</b>	<b>45</b>

## 1. Análisis

### 1.1. Ficha técnica de la base de datos SENAMHI

**Proyecto:** Software de Business Intelligence en la analítica del Monitoreo de los contaminantes del aire en Lima Metropolitana – SENAMHI

OBJETIVO GENERAL	Diseñar un software de Business Intelligence que permita analizar, visualizar y optimizar la gestión ambiental mediante la integración de datos provenientes del monitoreo de contaminantes atmosféricos en Lima Metropolitana.
PROPÓSITO DEL REGISTRO	Ofrecer un acceso unificado, seguro y al día a la información sobre la calidad del aire obtenida de las estaciones automáticas del SENAMHI, facilitando decisiones rápidas y fundamentadas para salvaguardar la salud de la población y para la gestión ambiental.
AUTOR	Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI).
RECOLECCIÓN DE DATOS	<a href="https://www.datosabiertos.gob.pe/dataset/monitoreo-de-los-contaminantes-del-aire-en-lima-metropolitana-servicio-nacional-de">https://www.datosabiertos.gob.pe/dataset/monitoreo-de-los-contaminantes-del-aire-en-lima-metropolitana-servicio-nacional-de</a>
PERIODO DE RECOLECCIÓN	2018 - 2025

COBERTURA GEOGRÁFICA	Lima Metropolitana (estaciones de Ate, San Borja, Carabayllo, Villa María del Triunfo, San Juan de Lurigancho, y otras).
VARIABLES	14
FECHA DE ÚLTIMA ACTUALIZACIÓN	08/09/25
TIPO DE CONTROLES	No se realizó imputación de datos

## 1.2. Ficha técnica de la base de datos de salud (IRA)

Campo	Descripción
Objetivo general	Describir la situación epidemiológica de las infecciones respiratorias agudas (IRA) mediante el registro sistemático de casos, neumonías, hospitalizaciones y defunciones por distrito, año y semana epidemiológica.
Propósito del registro	Brindar información oportuna para la vigilancia y el control de las IRA, facilitando la detección temprana de brotes y la priorización de intervenciones en grupos y territorios de mayor riesgo.
Autor / responsable	Centro Nacional de Epidemiología, Prevención y Control de Enfermedades (CDC Perú) – Ministerio de Salud.
Plataforma de acceso	Conjunto de datos de vigilancia de IRA disponible en el portal de datos abiertos de salud pública.
Periodo de recolección	Serie histórica multianual con registro semanal de eventos de IRA a nivel nacional.
Cobertura geográfica	Todo el territorio peruano; para este proyecto se seleccionan siete distritos de Lima Metropolitana: Carabayllo, Jesús María, San Borja, San Juan de

	Lurigancho, San Martín de Porres, Santa Anita y Villa María del Triunfo.
Principales variables	Departamento, provincia, distrito, código de ubigeo, año ( <a href="#">año</a> ), semana epidemiológica ( <a href="#">semana</a> ), número de casos de IRA sin neumonía ( <a href="#">ira_no_neumonia</a> ), neumonías en menores de 5 años ( <a href="#">neumonias_men5</a> ), neumonías en 60 años a más ( <a href="#">neumonias_60mas</a> ), hospitalizaciones y defunciones asociadas.
Tipo de datos	Datos numéricos agregados por semana y distrito, obtenidos de la notificación de establecimientos de salud.

### 1.3. Consultas analíticas

¿Cuál es el valor medio mensual de PM2. 5 y PM10 en Lima Metropolitana para el año 2024?

¿Qué estación de monitoreo muestra los niveles más altos de dióxido de nitrógeno (NO<sub>2</sub>) entre 2020 y 2025?

¿Cómo ha cambiado anualmente la concentración de ozono troposférico (O<sub>3</sub>) en las zonas norte, central y sur de Lima?

¿En qué meses del año se superan los límites de PM2. 5 definidos por la OMS en la estación de San Juan de Lurigancho?

¿Qué relación existe entre las concentraciones de material particulado (PM10) y la temperatura promedio del aire?

¿Qué estación reporta la mayor cantidad de alertas por contaminación (PM2. 5 > 25 µg/m<sup>3</sup>) durante 2023?

¿Cómo fluctúan los niveles de NO<sub>2</sub> en función de la humedad relativa observada en las estaciones urbanas?

¿Qué patrón temporal muestra el dióxido de azufre (SO<sub>2</sub>) entre 2018 y 2025 en Lima Metropolitana?

¿Cuál es el promedio del índice de calidad del aire por distrito y contaminante en los últimos tres meses de 2024?

¿Qué área de Lima presenta una mayor frecuencia de sobrepasar el límite diario de PM10 establecido por el MINAM?

#### 1.4. Diccionario de datos

Variable	Descripción	Tipo de dato	Tamaño de dato
ID	Identificador único del registro	INT	4
Estación	Nombre de la estación de monitoreo	VARCHAR	200
Distrito	Distrito de ubicación de la estación	VARCHAR	50
Fecha	Fecha de registro de la medición	DATE	YYYY-MM-DD
Hora	Hora de la medición	TIME	HH:MM
PM10	Concentracion de partículas menores a 10 micras	DECIMAL	(10,2)
PM2.5	Concentracion de partículas menores a 2.5 micras	DECIMAL	(10,2)
NO <sub>2</sub>	Concentración de dióxido de nitrógeno	DECIMAL	(10,2)
SO <sub>2</sub>	Concentración de dióxido de azufre	DECIMAL	(10,2)
O <sub>3</sub>	Concentración de	DECIMAL	(10,2)

	ozono troposférico		
Temperatura	Temperatura ambiente registrada	DECIMAL	(5,2)
Humedad relativa	Porcentaje de humedad ambiental	DECIMAL	(5,2)
Velocidad del viento	Intensidad del viento promedio	DECIMAL	(5,2)
Dirección del viento	Dirección media del viento	INT	3
Índice de calidad del aire	Clasificación calculada de la calidad del aire (Bueno, Moderado, Malo, Muy malo)	VARCHAR	20

### 1.5. Relación entre calidad del aire e infecciones respiratorias agudas

Las infecciones respiratorias agudas (IRA) constituyen una de las principales causas de consulta y hospitalización en el sistema de salud, particularmente en menores de cinco años y adultos mayores, quienes presentan mayor susceptibilidad frente a agentes infecciosos y condiciones ambientales adversas. Su adecuada vigilancia es prioritaria porque un incremento inusual de casos puede indicar la presencia de brotes o cambios en los patrones de transmisión.

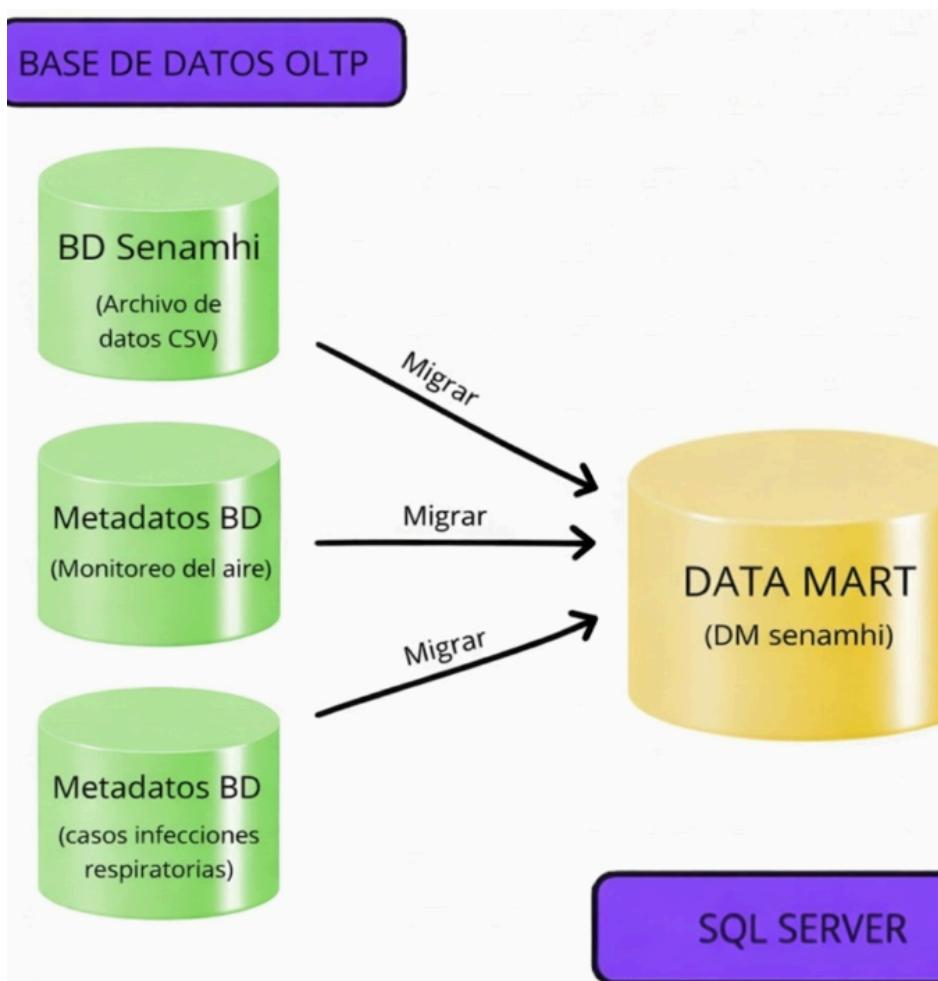
Al mismo tiempo, la literatura científica ha mostrado que la exposición a contaminantes atmosféricos como el material particulado fino (PM2.5), el material particulado respirable (PM10) y el dióxido de nitrógeno ( $\text{NO}_2$ ) se vincula con un aumento en la incidencia de síntomas respiratorios, exacerbaciones de enfermedades crónicas y episodios de infección respiratoria. En ciudades densamente pobladas, la combinación de altos niveles de emisión y condiciones meteorológicas desfavorables puede llevar a concentraciones de contaminantes que superan las recomendaciones internacionales y agravan los riesgos para la salud.

Integrar la información de calidad del aire y de vigilancia de IRA permite abordar el problema desde la perspectiva de la ciencia de datos: se combinan hechos medibles de

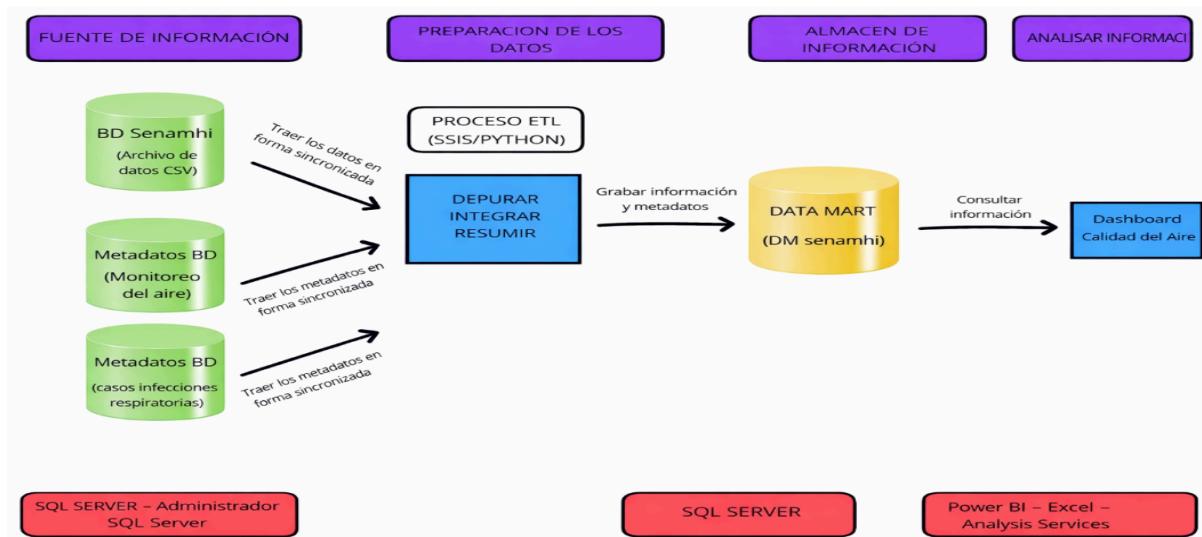
dos dominios distintos (ambiente y salud) para generar evidencia cuantitativa sobre posibles asociaciones. Este enfoque facilita la construcción de indicadores conjuntos Aire-Salud y de modelos exploratorios de correlación que aportan insumos a políticas de prevención, alerta temprana y gestión ambiental.

### 1.6. Arquitectura tecnológica

#### Diseño de la arquitectura tecnológica del Data Mart

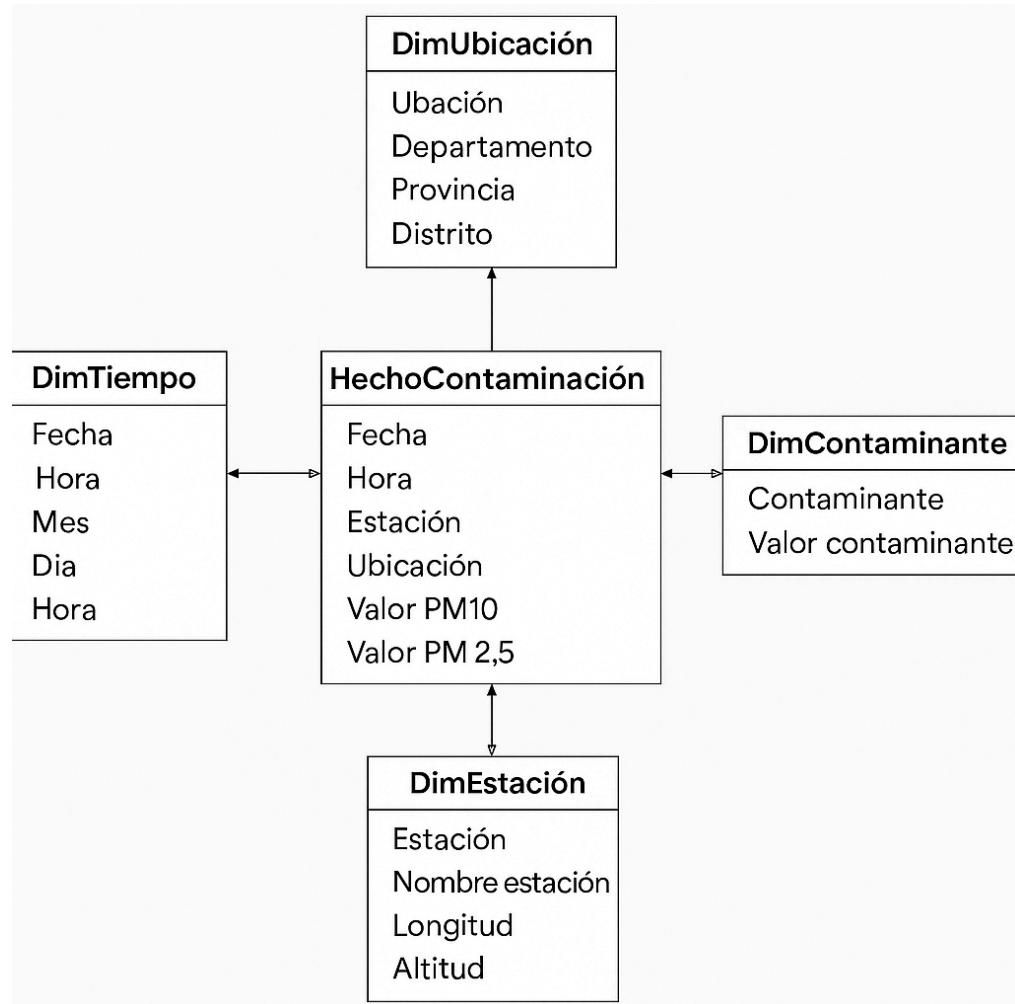


## Arquitectura tecnológica general del proyecto

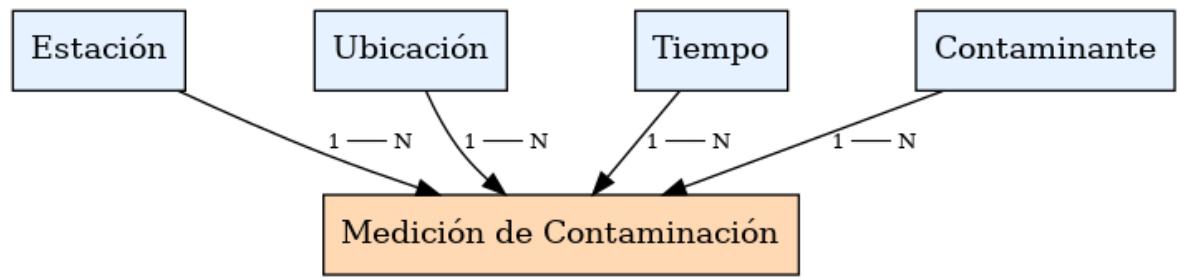


## 2. Diseño

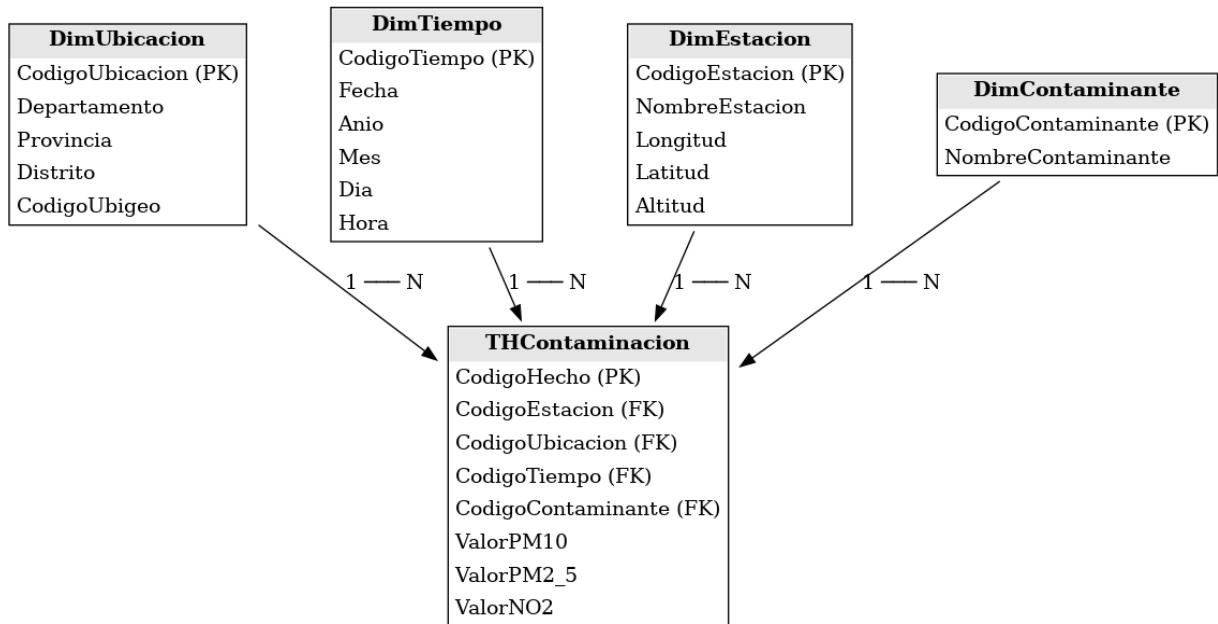
### 2.1. Modelo Dimensional



## 2.2. Modelo Conceptual



## 2.3. Modelo Lógico



## 2.4. Modelo analítico integrado Aire-Salud

El modelo integrado Aire-Salud se construye a partir de dos estructuras principales:

- Stage\_IRA, tabla de staging que almacena la open data de vigilancia epidemiológica con los campos

Stage_IRA
departamento
provincia
distrito
ano
semana
sub_reg_nt
ubigeo
ira_no_neumonia
neumonias_men5
neumonias_60mas
hospitalizados_men5
hospitalizados_60mas
defunciones_men5
defunciones_60mas

- Fact\_AireSalud, tabla de hechos diseñada específicamente para el análisis anual por distrito, donde se integran indicadores ambientales y de salud.

Fact_AireSalud	
	Distrito
	CodigoUrbigeo
	Ano
	Prom_PM10_Anual
	Prom_PM25_Anual
	Prom_NO2_Anual
	Total_IRA_NoNeum_Anual
	Total_Neum_Men5_Anual
	Total_Neum_60Mas_Anual
	Total_Hosp_Men5_Anual
	Total_Hosp_60Mas_Anual
	Total_Def_Men5_Anual
	Total_Def_60Mas_Anual

Fact\_AireSalud se obtiene combinando las agregaciones anuales calculadas sobre Stage\_IRA con los promedios anuales de contaminantes derivados de THContaminacion y DimTiempo. De esta manera, cada fila de la tabla representa el resumen Aire-Salud de un distrito en un año determinado y se convierte en la fuente principal de datos para el quinto dashboard.

## 2.5. Diseño de migración de datos Senamhi

a. Migrando hacia Dimensión Estación → DimEstacion

ORIGEN - OLTP - BDSenamhi			DESTINO - DM- BMSenamhi		
Tabla	Atributo	Metadata	Dimensión	Atributo	Metadata
Datos_Aire	ESTACION	varchar(50)	DimEstacion	NombreEstacion	nvarchar(200)
	ALTITUD	numeric(10, 0)		Altitud	float
	LONGITUD	numeric(20, 0)		Longitud	float
	LATITUD	numeric(20, 0)		Latitud	float

b. Migrando hacia Dimensión Ubicación → DimUbicacion

ORIGEN - OLTP - BDSenamhi			DESTINO - DM- BMSenamhi		
Tabla Origen (OLTP)	Atributo Origen	Metadata Origen	Dimensión (DM)	Atributo Dimensión	Metadata DM
Datos_Aire	DEPARTAMENTO	varchar(4)	DimUbicacion	Departamento	NVARCHAR (50)
Datos_Aire	PROVINCIA	varchar(4)	DimUbicacion	Provincia	NVARCHAR (50)
Datos_Aire	DISTRITO	varchar(100)	DimUbicacion	Distrito	NVARCHAR (100)
Datos_Aire	UBIGEO	varchar(10)	DimUbicacion	CodigoUbigeo	NVARCHAR (10) PK

c. Migrando hacia Dimensión Tiempo → DimTiempo

ORIGEN - OLTP - BDSenamhi			DESTINO - DM- BMSenamhi		
Tabla Origen (OLTP)	Atributo Origen	Metadata Origen	Dimensión (DM)	Atributo Dimensión	Metadata DM
Datos_Air e	FECHA	numeric(8 ,0)	DimTiempo	Año	int
	FECHA	numeric(8 ,0)		Mes	int
	FECHA	numeric(8 ,0)		Día	int
	HORA	varchar(6)		Hora	TIME(0)

d. Migrando hacia Dimensión Contaminante → DimContaminante

ORIGEN - OLTP - BDSenamhi			DESTINO - DM- BMSenamhi		
Tabla	Atributo	Metadata	Dimensión	Atributo	Metadata
Datos_Air e	PM10	float	DimContaminan te	NombreContamina nte	nvarchar(50 )
	PM2.5	float			nvarchar(50 )
	NO2	float			nvarchar(50 )

e. Migrando hacia Tabla de Hechos → THContaminacion

ORIGEN			DESTINO - DM		
Dimension/Tabla	Atributo	Metad ata	Dimensión	Atributo	Metad ata
DMSenamhi.DimEstacion	CodigoEstacion	int	THContaminacion	CodigoEstacion	int
DMSenamhi.DimUbicacion	CodigoUbicacion	int		CodigoUrbigeo	int
DMSenamhi.DimTiempo	CodigoTiempo	int		CodigoTiempo	int
DMSenamhi.DimContaminante	CodigoContaminante	int		CodigoContaminante	int
BDSenamhi.Datos_Aire	PM10	float		ValorPM10	float
BDSenamhi.Datos_Aire	PM2_5	float		ValorPM2_5	float
BDSenamhi.Datos_Aire	NO2	float		ValorNO2	float

### 3. Construcción

#### 3.1. Código OLTP

El entorno OLTP corresponde a la base de datos BDSenamhi, que almacena los datos originales obtenidos del Portal de Datos Abiertos del SENAMHI.

En esta base se cargaron los archivos CSV/Excel con registros históricos de monitoreo de calidad del aire en Lima Metropolitana.

**Estructura principal:**

## Tabla: Datos\_Aire

### Campos:

ID, Estacion, Fecha, Hora, Longitud, Latitud, Altitud, PM10, PM2\_5, NO2, Departamento, Provincia, Distrito, Ubigeo, Fecha\_Corte

ID	ESTACION	FECHA	HORA	LONGITUD	LATITUD	ALTITUD	PM10	PM2_5	NO2	DEPARTAMENTO	PROVINCIA	DISTRITO	UBIGEO	FECHA_CORTE
320987	508187	VILLA_MARI...	20160623	30000	-76.92	-12.1664	272.0	30.3	25	31.2	LIMA	VILLA_MARI...	150143.0	20240531
320988	508188	VILLA_MARI...	20160623	40000	-76.92	-12.1664	272.0	51.73	15.3	26.1	LIMA	VILLA_MARI...	150143.0	20240531
320989	508189	VILLA_MARI...	20160623	50000	-76.92	-12.1664	272.0	29.19	15.2	20	LIMA	VILLA_MARI...	150143.0	20240531
320990	508190	VILLA_MARI...	20160623	60000	-76.92	-12.1664	272.0	27.7	15	15.3	LIMA	VILLA_MARI...	150143.0	20240531
320991	508191	VILLA_MARI...	20160623	70000	-76.92	-12.1664	272.0	26.2	21.7	16.2	LIMA	VILLA_MARI...	150143.0	20240531
320992	508192	VILLA_MARI...	20160623	80000	-76.92	-12.1664	272.0	24.17	19.4	14.7	LIMA	VILLA_MARI...	150143.0	20240531
320993	508193	VILLA_MARI...	20160623	90000	-76.92	-12.1664	272.0	22.15	15.5	21.5	LIMA	VILLA_MARI...	150143.0	20240531
320994	508194	VILLA_MARI...	20160623	100000	-76.92	-12.1664	272.0	29.06	13.9	23.8	LIMA	VILLA_MARI...	150143.0	20240531
320995	508195	VILLA_MARI...	20160623	110000	-76.92	-12.1664	272.0	33.32	25.9	27.4	LIMA	VILLA_MARI...	150143.0	20240531
320996	508196	VILLA_MARI...	20160623	120000	-76.92	-12.1664	272.0	35.74	22.8	29.8	LIMA	VILLA_MARI...	150143.0	20240531
320997	508197	VILLA_MARI...	20160623	130000	-76.92	-12.1664	272.0	50.89	37.7	31.4	LIMA	VILLA_MARI...	150143.0	20240531
320998	508198	VILLA_MARI...	20160623	140000	-76.92	-12.1664	272.0	64.24	46.9	26.2	LIMA	VILLA_MARI...	150143.0	20240531
320999	508199	VILLA_MARI...	20160623	150000	-76.92	-12.1664	272.0	53.89	38.3	24.4	LIMA	VILLA_MARI...	150143.0	20240531
321000	508200	VILLA_MARI...	20160623	160000	-76.92	-12.1664	272.0	43.54	23.2	24.6	LIMA	VILLA_MARI...	150143.0	20240531
321001	508201	VILLA_MARI...	20160623	170000	-76.92	-12.1664	272.0	36.35	25.8	23.3	LIMA	VILLA_MARI...	150143.0	20240531
321002	508202	VILLA_MARI...	20160623	180000	-76.92	-12.1664	272.0	29.17	9.7	24	LIMA	VILLA_MARI...	150143.0	20240531
321003	508203	VILLA_MARI...	20160623	190000	-76.92	-12.1664	272.0	40.67	13	21.3	LIMA	VILLA_MARI...	150143.0	20240531
321004	508204	VILLA_MARI...	20160623	200000	-76.92	-12.1664	272.0	35.81	10.8	22.4	LIMA	VILLA_MARI...	150143.0	20240531
321005	508205	VILLA_MARI...	20160623	210000	-76.92	-12.1664	272.0	50	16.5	26.7	LIMA	VILLA_MARI...	150143.0	20240531

## 3.2. Script Data Mart

CREATE DATABASE DMSenamhi;

GO

USE DMSenamhi;

- Tabla Dimension Estación

```
CREATE TABLE DimEstacion (
    CodigoEstacion INT IDENTITY(1,1) PRIMARY KEY,
    NombreEstacion NVARCHAR(200),
    Altitud FLOAT,
    Longitud FLOAT,
    Latitud FLOAT
);
```

- Tabla Dimension Ubicación

```
CREATE TABLE DimUbicacion (
    CodigoUbigeo NVARCHAR(10) PRIMARY KEY,
    Departamento NVARCHAR(50),
    Provincia NVARCHAR(50),
    Distrito NVARCHAR(100)
);
```

- **Tabla Dimension Tiempo**

```
create TABLE dbo.DimTiempo (
   CodigoTiempo INT IDENTITY(1,1) PRIMARY KEY,
    Año INT NOT NULL,
    Mes INT NOT NULL,
    Día INT NOT NULL,
    Hora TIME(0) NOT NULL,
    FechaCorte DATE NOT NULL
);
```

- **Tabla Dimension Contaminante**

```
CREATE TABLE DimContaminante (
   CodigoContaminante INT IDENTITY(1,1) PRIMARY KEY,
    NombreContaminante NVARCHAR(50)
);
```

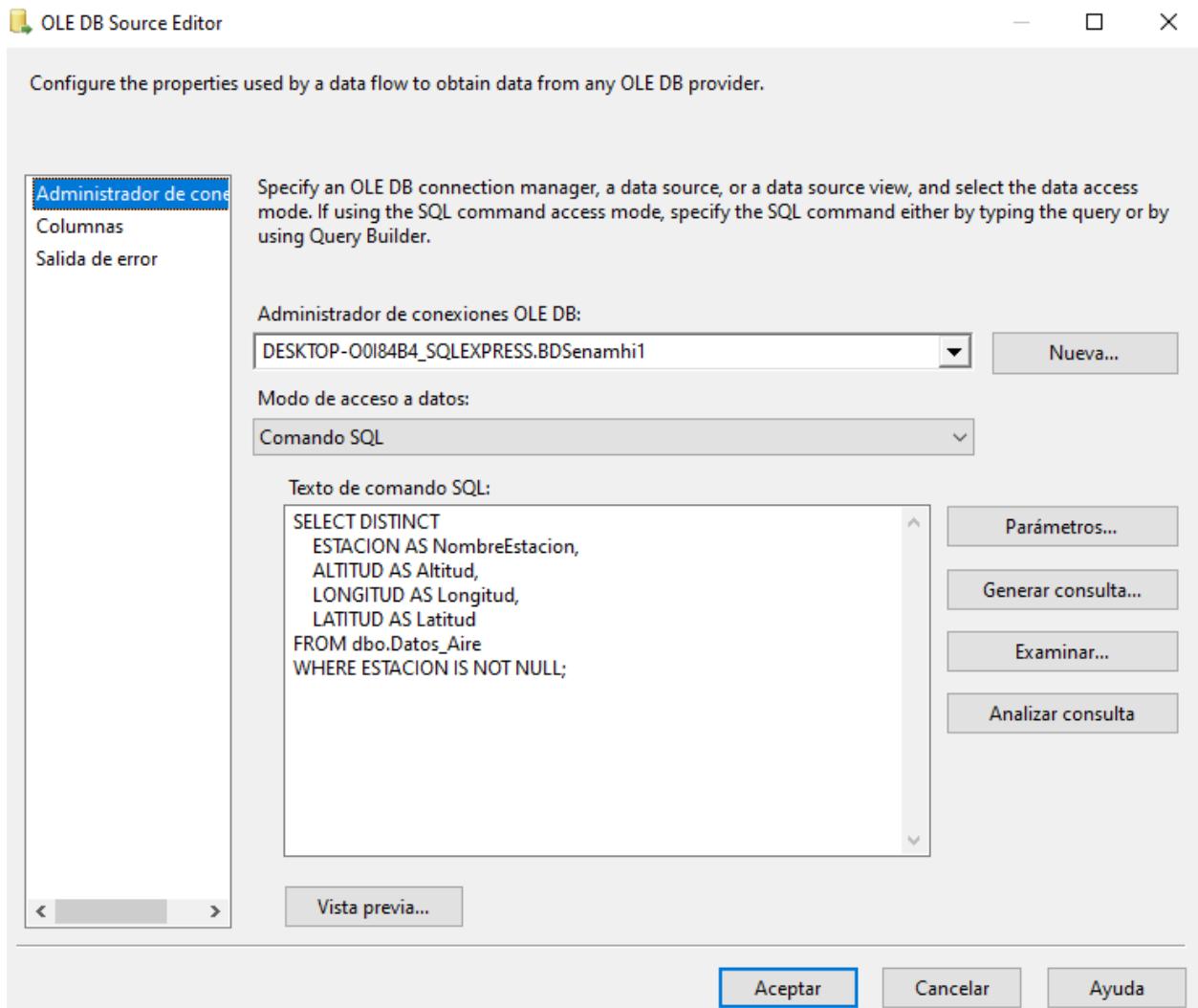
- **Tabla Hechos Contaminacion**

```
CREATE TABLE THContaminacion (
   CodigoHecho INT IDENTITY(1,1) PRIMARY KEY,
    CódigoEstacion INT FOREIGN KEY REFERENCES DimEstacion(CódigoEstacion),
    CódigoUbigeo NVARCHAR(10) FOREIGN KEY REFERENCES
DimUbicacion(CódigoUbigeo),
    CódigoTiempo INT FOREIGN KEY REFERENCES DimTiempo(CódigoTiempo),
    CódigoContaminante INT FOREIGN KEY REFERENCES
DimContaminante(CódigoContaminante),
    ValorPM10 FLOAT,
    ValorPM2_5 FLOAT,
    ValorNO2 FLOAT
);
```

### 3.3. Código SSIS

#### Dimensión Estación

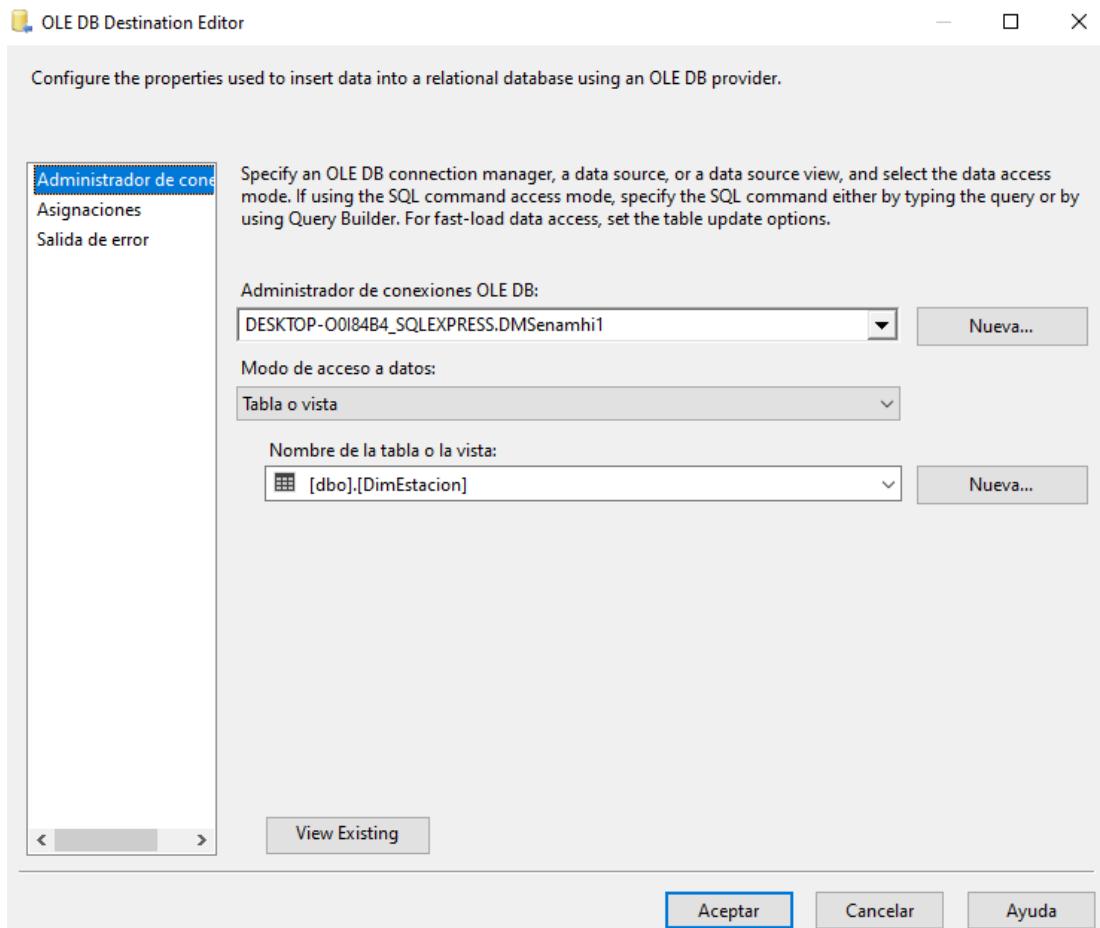
##### Origen:

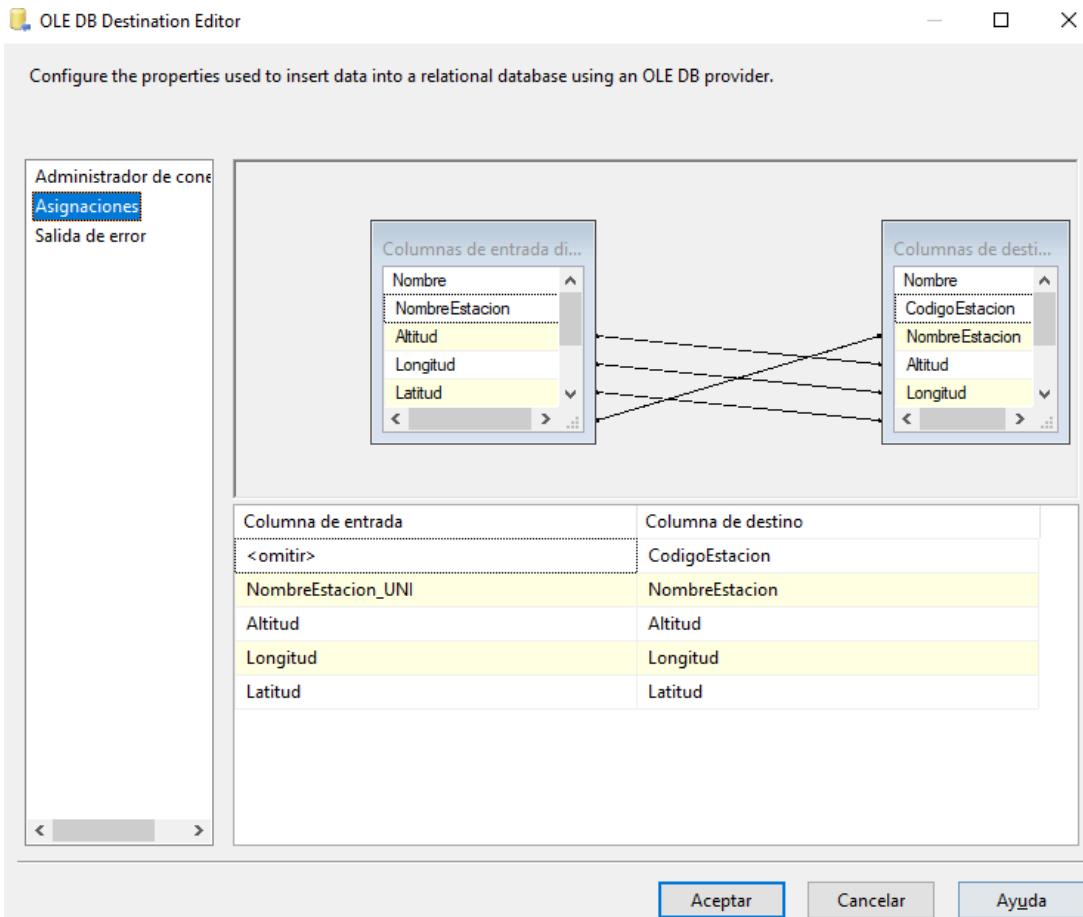


##### Script:

```
SELECT DISTINCT
    ESTACION AS NombreEstacion,
    ALTITUD AS Altitud,
    LONGITUD AS Longitud,
    LATITUD AS Latitud
FROM dbo.Datos_Aire
WHERE ESTACION IS NOT NULL;
```

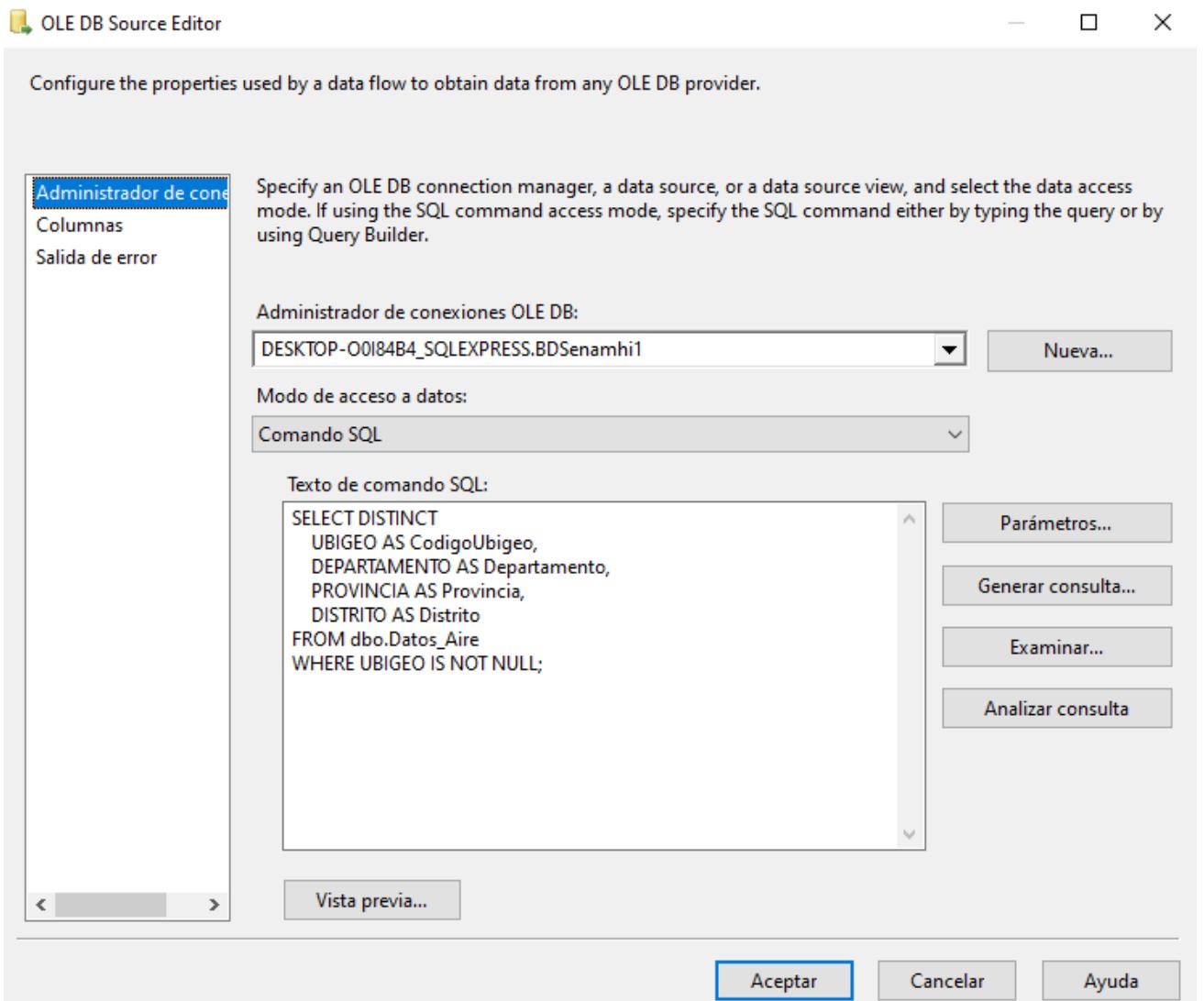
## Destino:





## Dimensión Ubicación

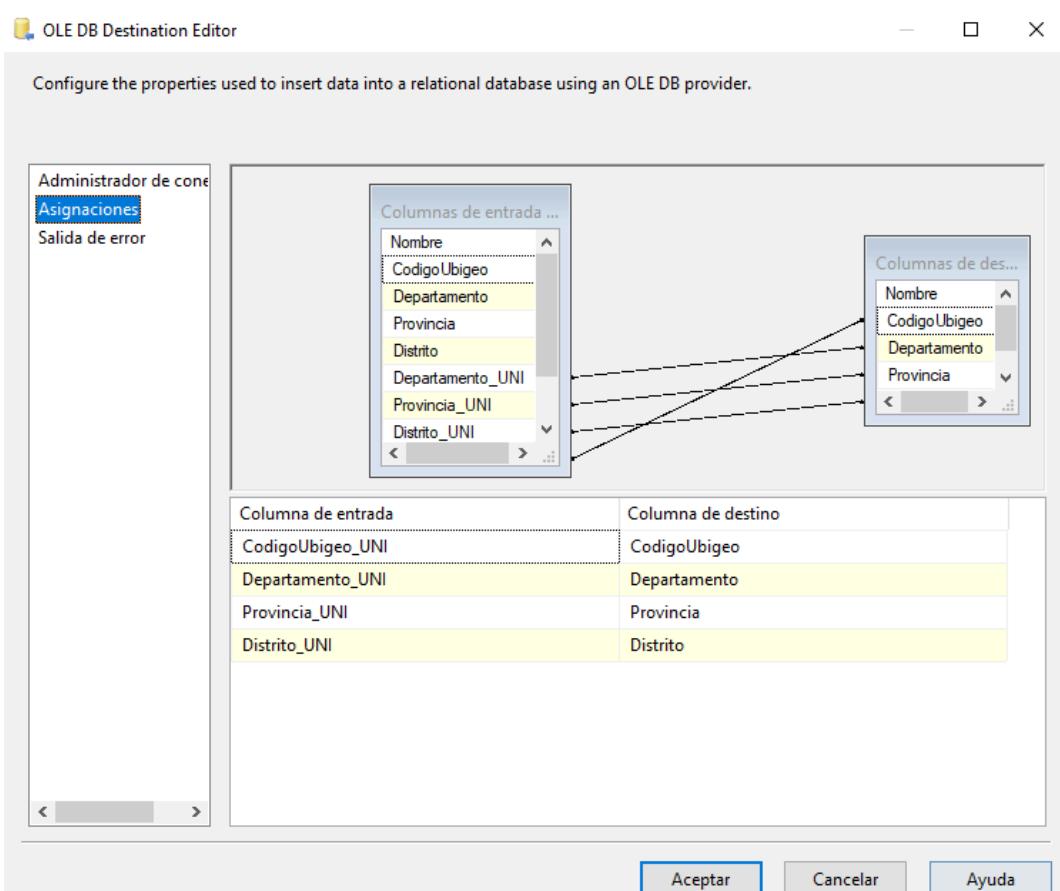
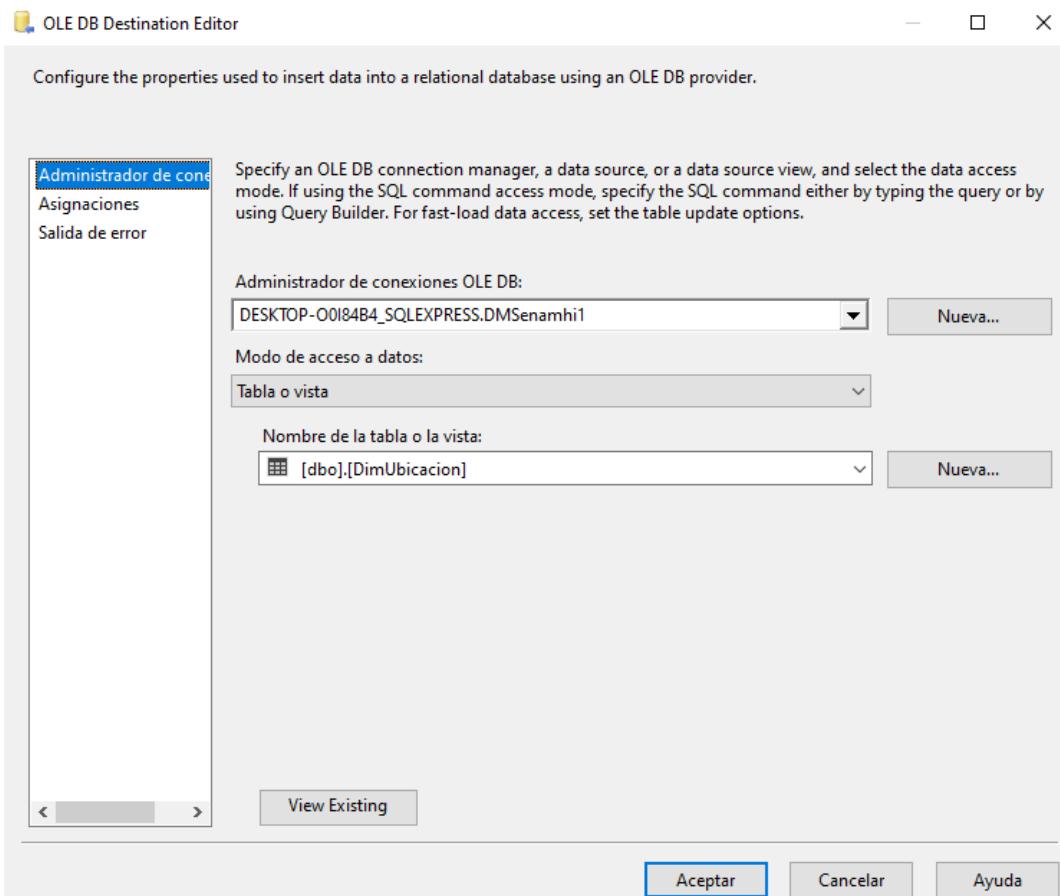
### Origen:



### Script:

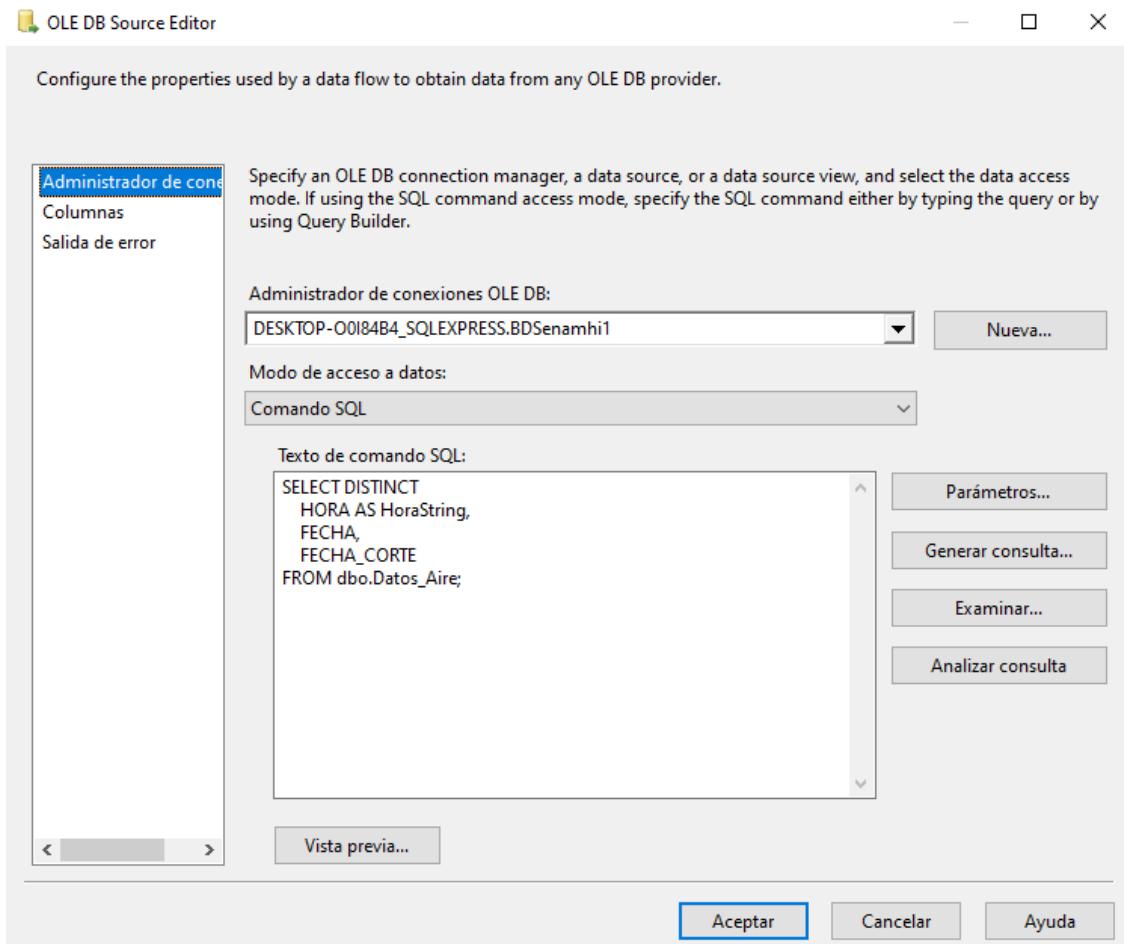
```
SELECT DISTINCT
    UBIGEO AS CódigoUrbano,
    DEPARTAMENTO AS Departamento,
    PROVINCIA AS Provincia,
    DISTRITO AS Distrito
FROM dbo.Datos_Aire
WHERE UBIGEO IS NOT NULL;
```

## Destino:



## Dimensión Tiempo

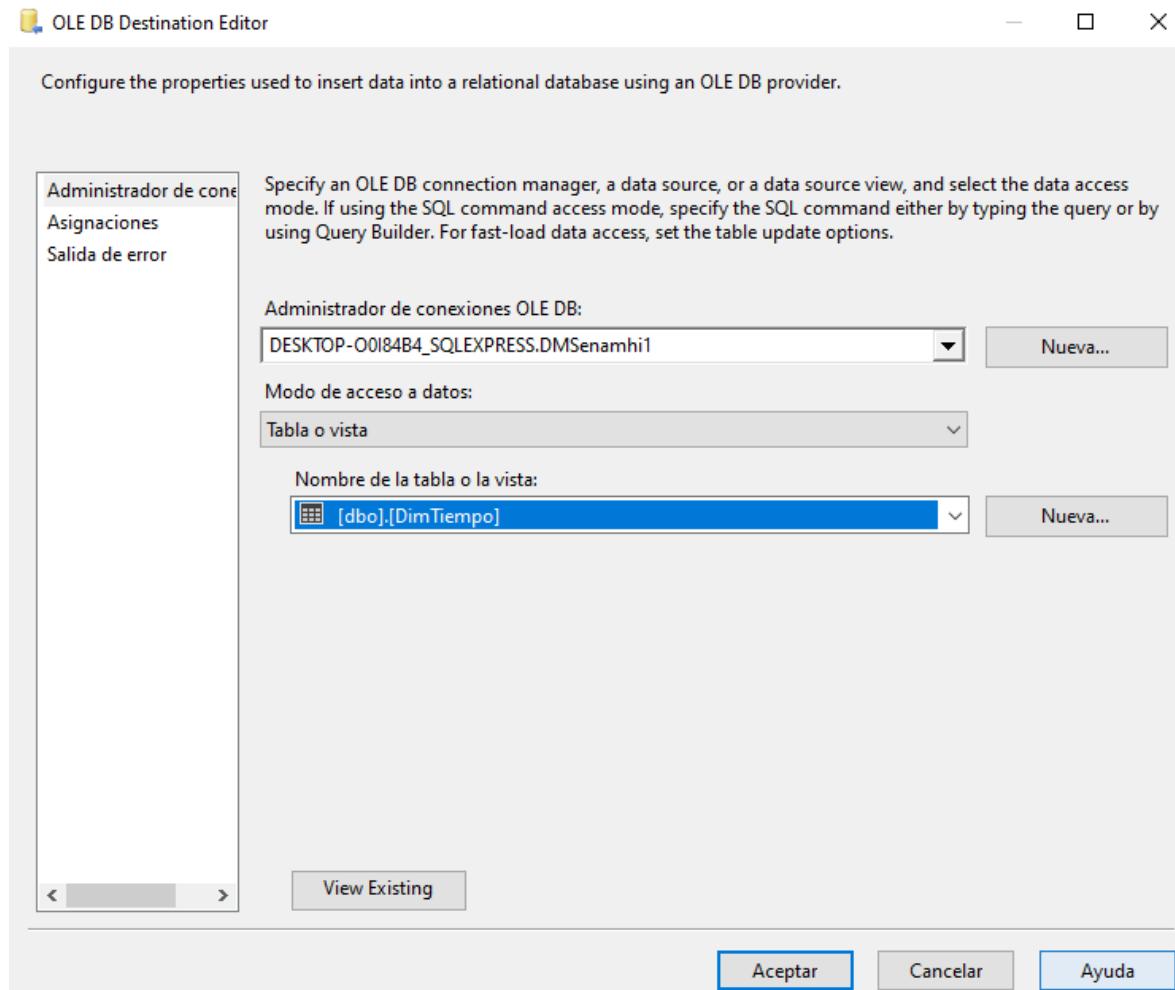
### Origen:

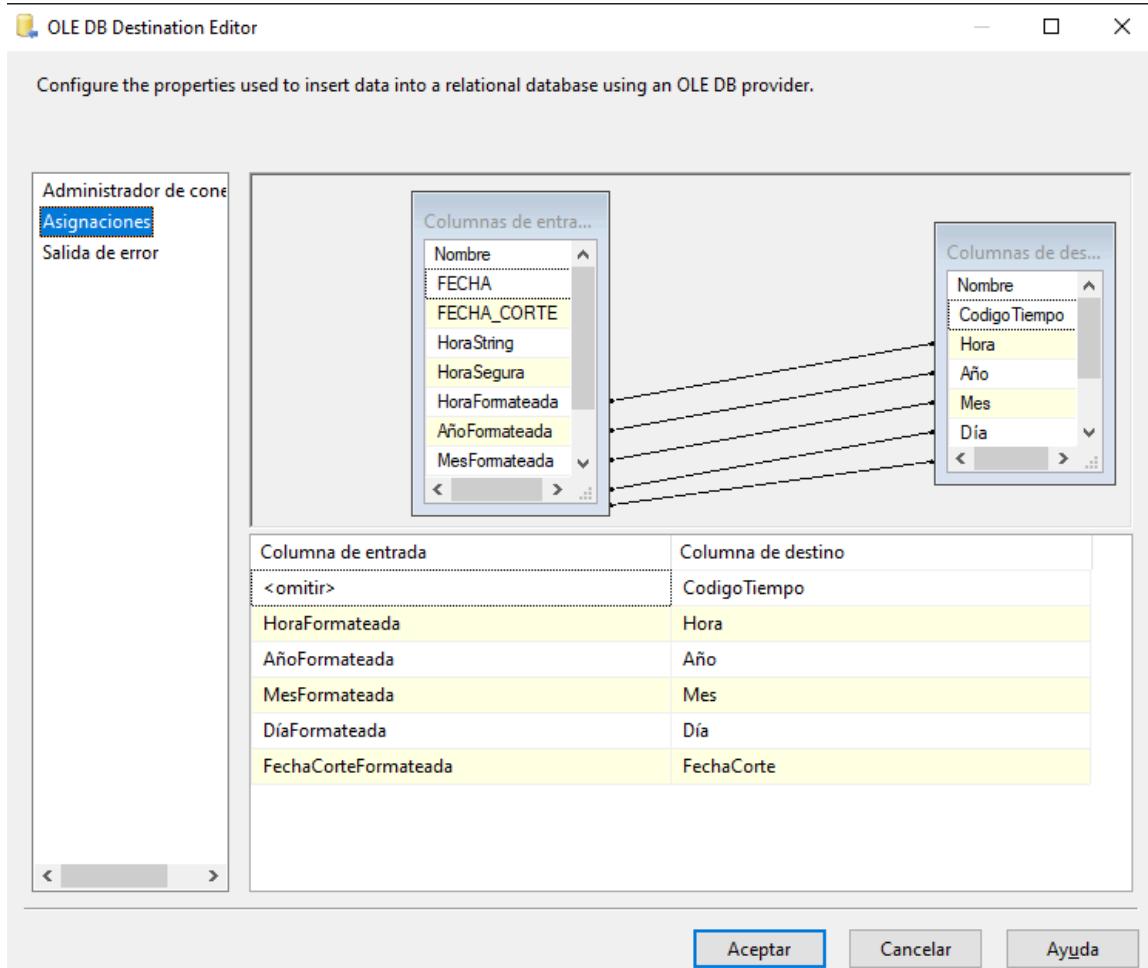


### Script:

```
SELECT DISTINCT
    HORA AS HoraString,
    FECHA,
    FECHA_CORTE
FROM dbo.Datos_Aire;
```

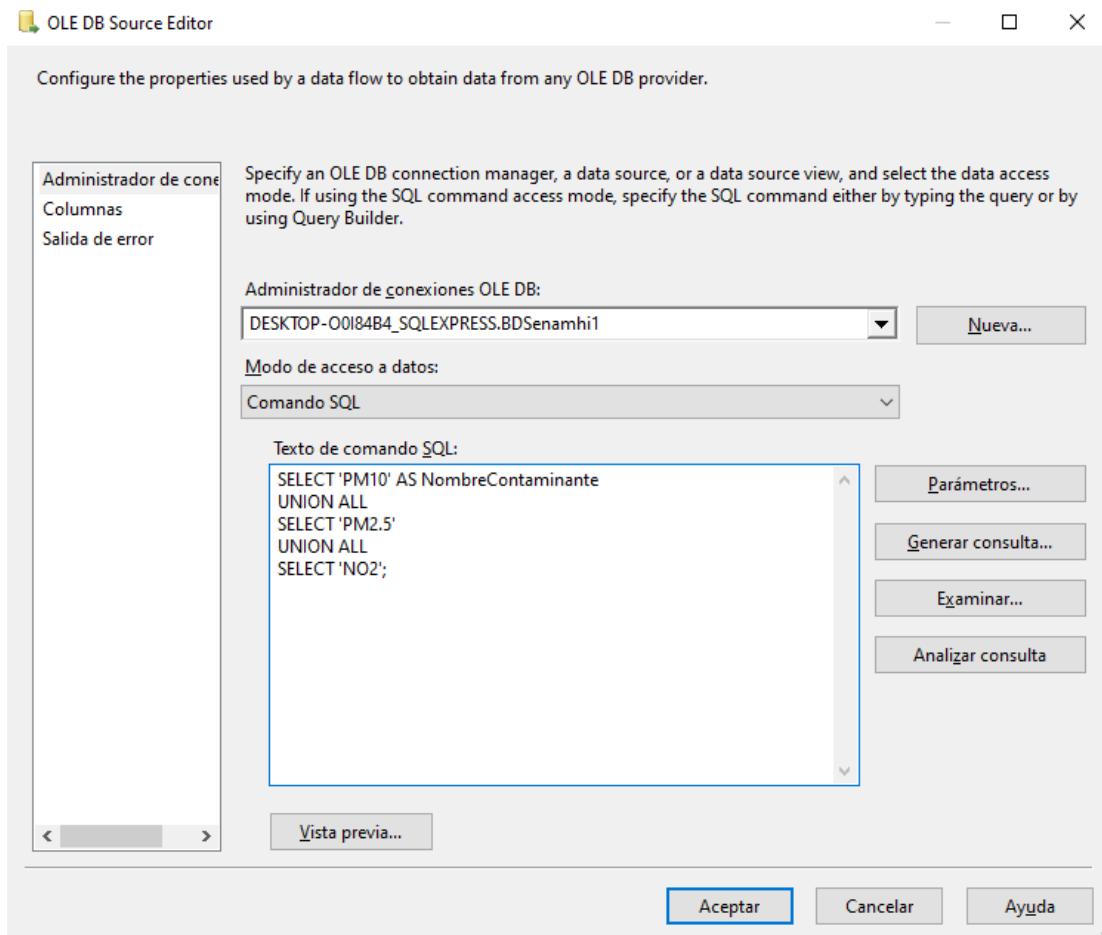
## Destino:





## Dimensión Contaminante

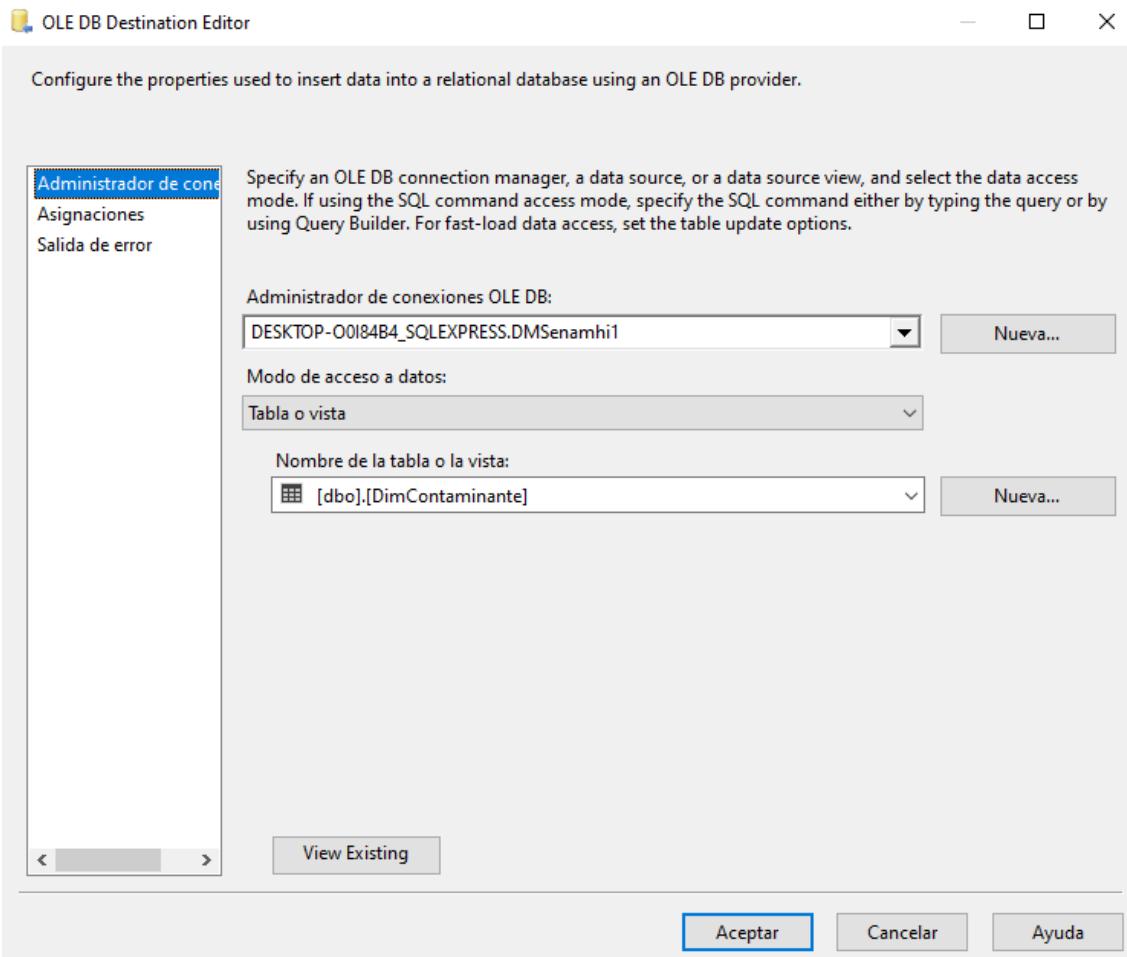
Origen:

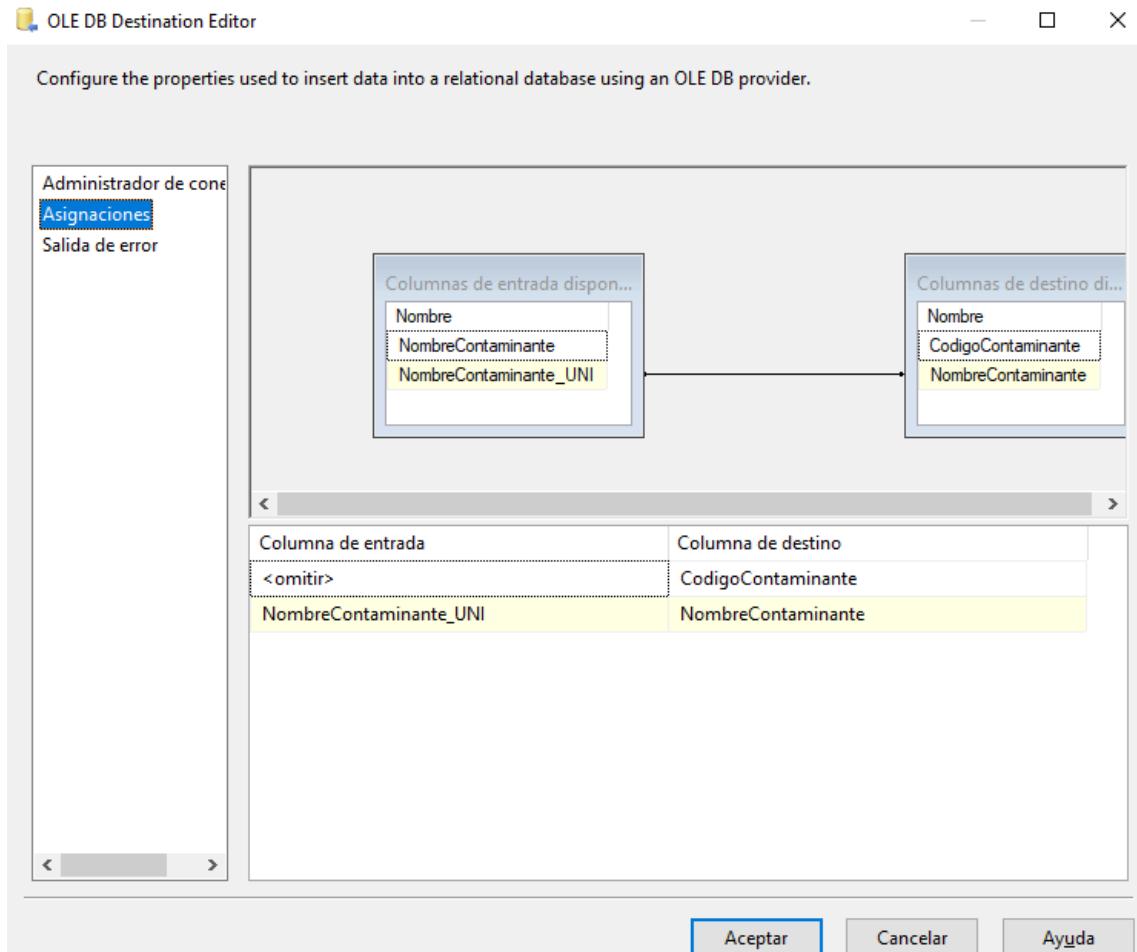


## Código:

```
SELECT 'PM10' AS NombreContaminante
UNION ALL
SELECT 'PM2.5'
UNION ALL
SELECT 'NO2';
```

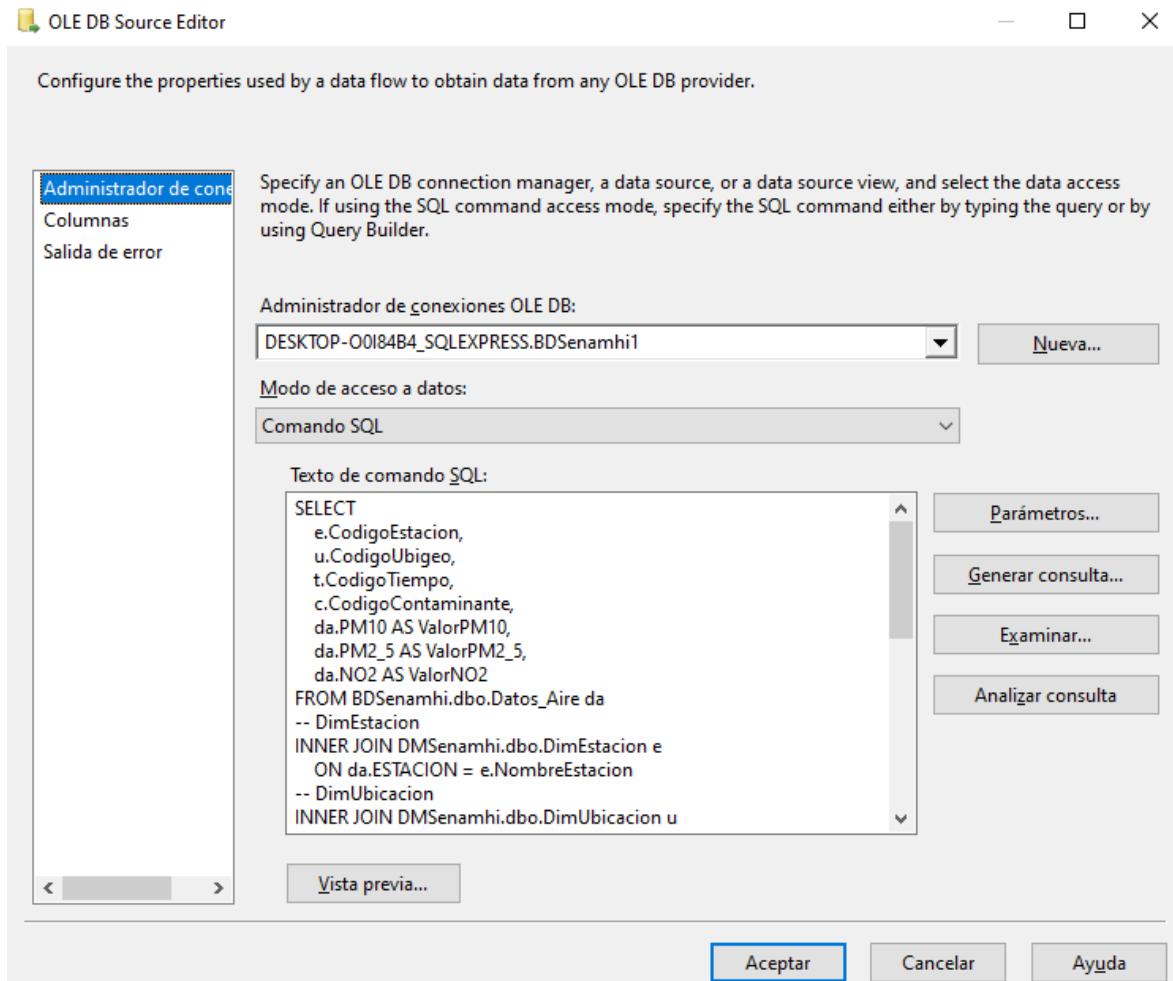
## Destino:





### Tabla Hechos Contaminación del Aire:

**Origen:**

**Script:**

SELECT

```
e.CodigoEstacion,  
u.CodigoUrbgeo,  
t.CodigoTiempo,  
c.CodigoContaminante,  
da.PM10 AS ValorPM10,  
da.PM2_5 AS ValorPM2_5,  
da.NO2 AS ValorNO2
```

FROM BDSenamhi.dbo.Datos\_Aire da

```
INNER JOIN DMSenamhi.dbo.DimEstacion e  
ON da.ESTACION = e.NombreEstacion
```

```
INNER JOIN DMSenamhi.dbo.DimUbicacion u  
ON da.UBIGEO = u.CodigoUrbgeo
```

```
INNER JOIN DMSenamhi.dbo.DimTiempo t  
ON t.Año = CAST(SUBSTRING(CAST(da.FECHA AS VARCHAR(8)),1,4) AS  
INT)
```

```

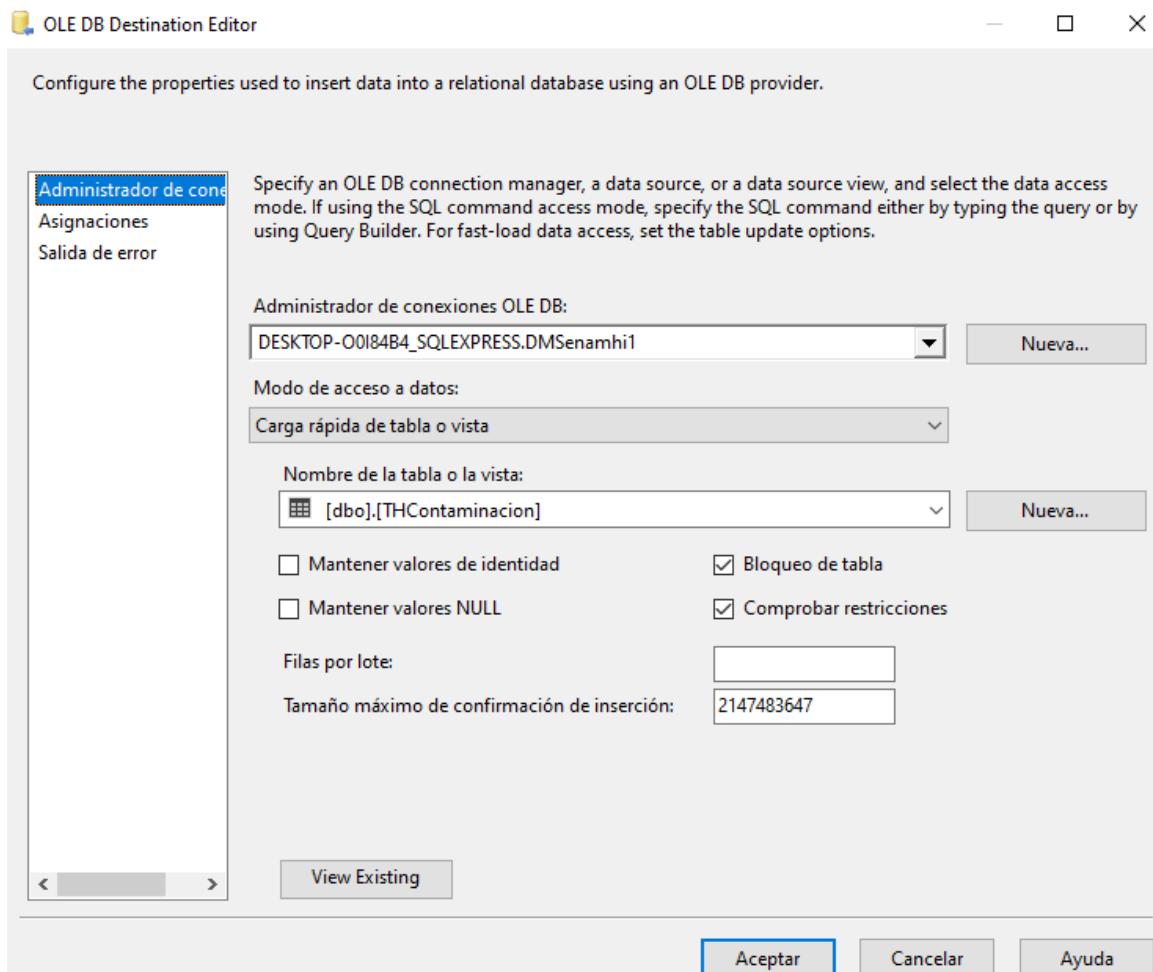
        AND t.Mes = CAST(SUBSTRING(CAST(da.FECHA AS VARCHAR(8)),5,2) AS
INT)
        AND t.Día = CAST(SUBSTRING(CAST(da.FECHA AS VARCHAR(8)),7,2) AS
INT)
        AND t.Hora = CAST(STUFF(STUFF(RIGHT('000000' + da.HORA,6),3,0,'.'),6,0,'.')
AS TIME(0))
        AND t.FechaCorte = CONVERT(DATE, STUFF(STUFF(CAST(da.FECHA_CORTE
AS VARCHAR(8)),5,0,'-'),8,0,'-'))

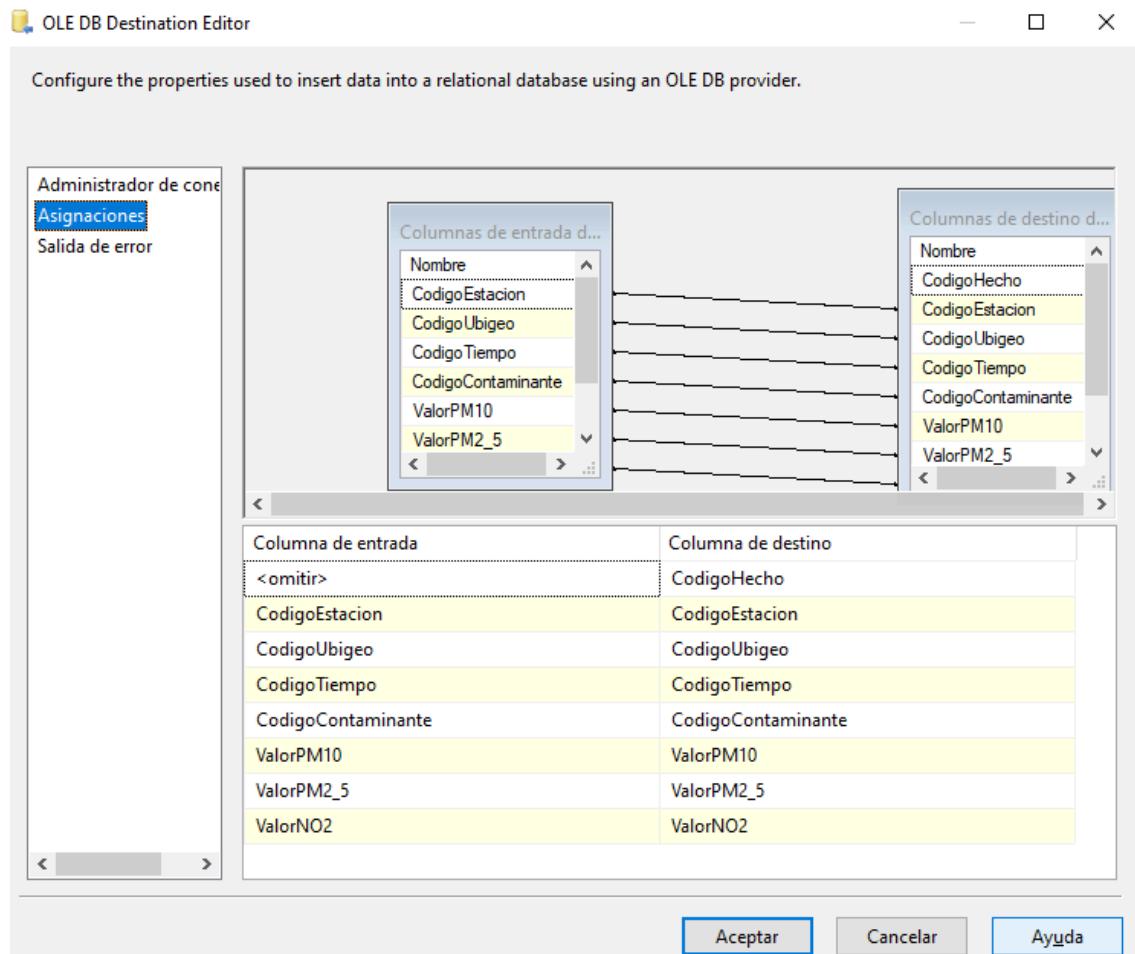
INNER JOIN DMSenamhi.dbo.DimContaminante c
ON c.NombreContaminante IN ('PM10', 'PM2.5', 'NO2')

WHERE da.PM10 IS NOT NULL
OR da.PM2_5 IS NOT NULL
OR da.NO2 IS NOT NULL;

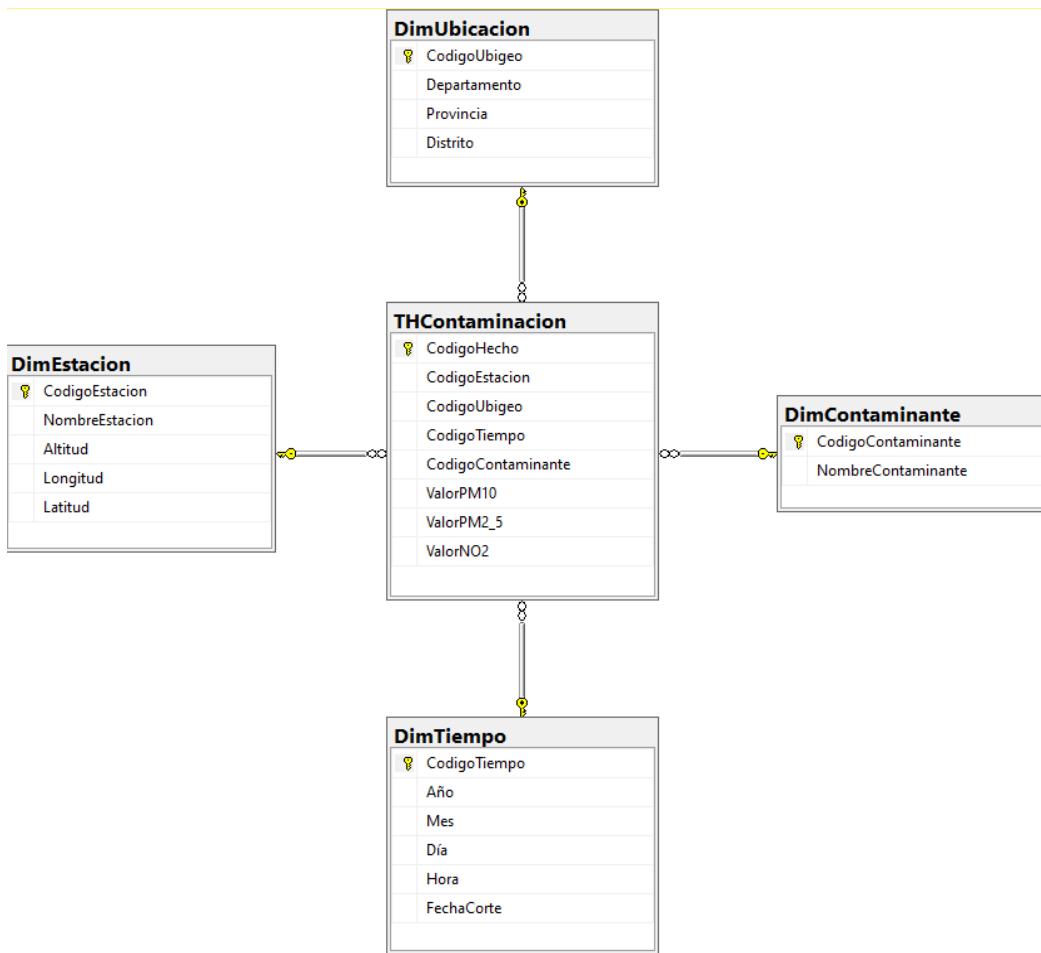
```

### **Destino:**

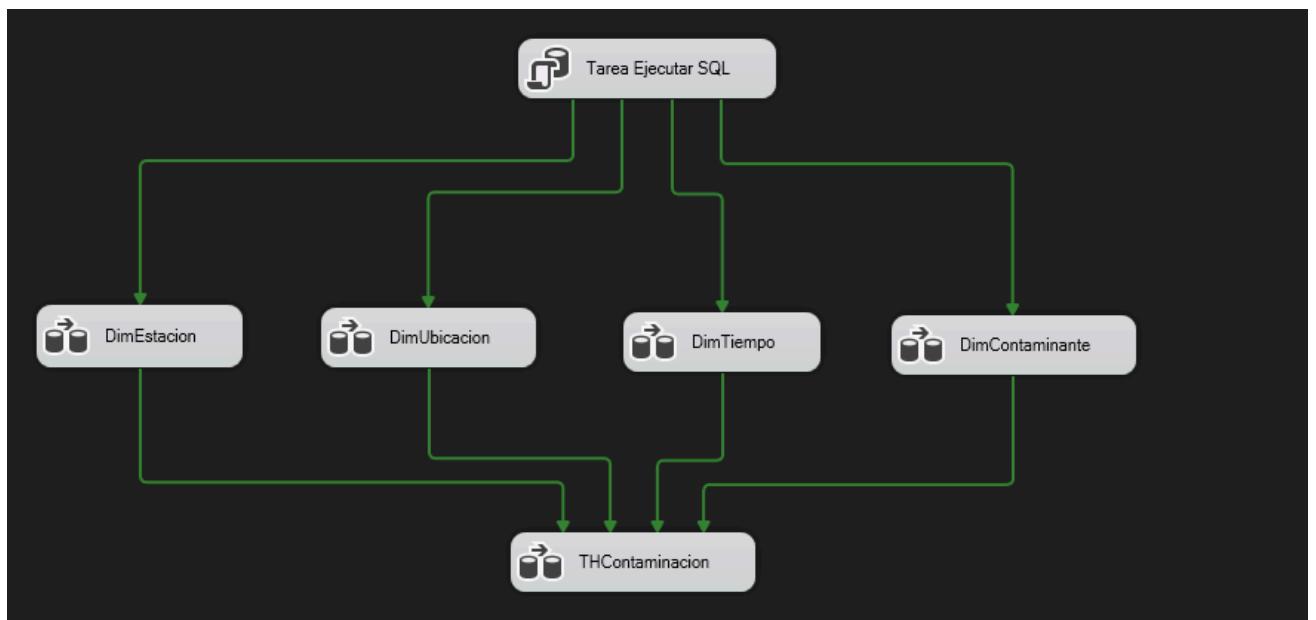




### 3.4. Modelo Lógico final de Data Mart

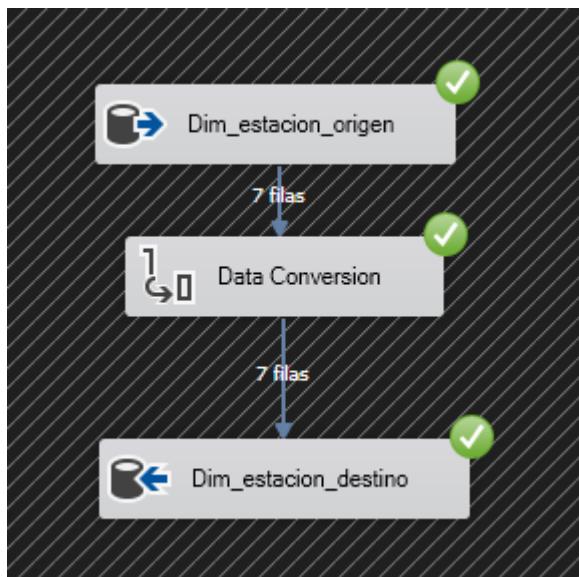


### 3.5. Esquema del ETL

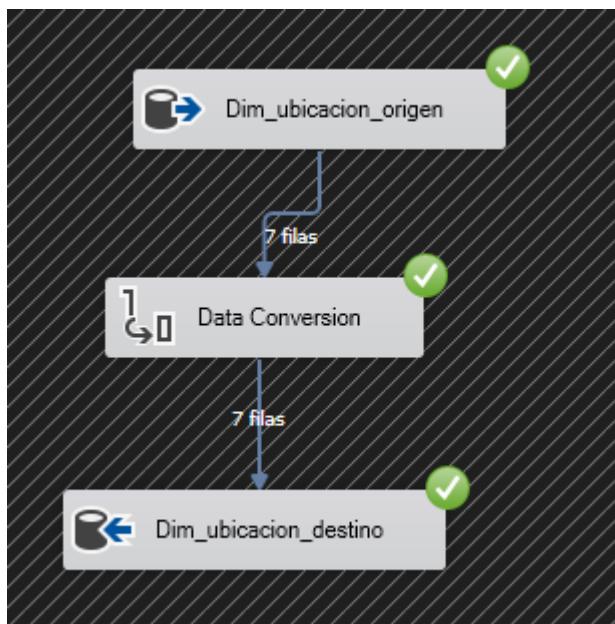


## EJECUCION:

Dimension Estacion:



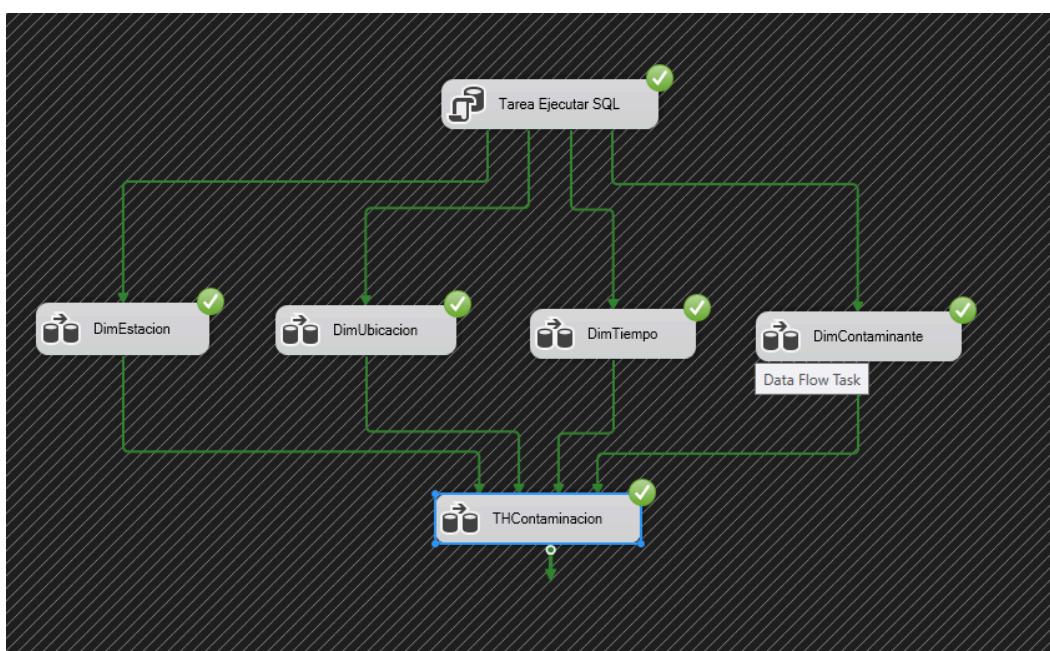
Dimension Ubicación:



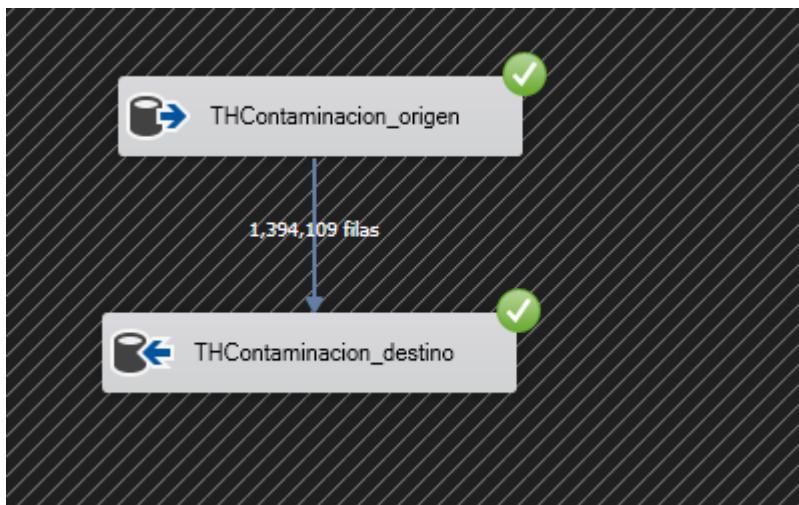
## Dimensión Tiempo:



## Dimension Ubicación:



## Dimension Hechos Contaminacion



## 4. Desarrollo de Dashboards

### 4.1. Interfaz Análisis de Contaminantes por Estación

#### 4.1.1. Análisis de requerimientos

El requerimiento de este informe tiene como objetivo proporcionar información detallada sobre la calidad del aire por estación, considerando los contaminantes medidos, períodos de tiempo y ubicación geográfica. Entre las consultas analíticas que se requieren (en lenguaje natural) se encuentran, posiblemente:

- ¿Cuál es el nivel promedio de PM2.5, PM10 y NO<sub>2</sub> por estación y por mes?
- ¿Qué estaciones presentan niveles de contaminación superiores a los límites recomendados por la OMS y MINAM?
- ¿Cuál es la evolución temporal de cada contaminante por distrito?
- ¿Cuál es el ranking de las estaciones según el nivel promedio de contaminantes?

#### 4.1.2. Esquema de Data Mart necesario para el requerimiento

A partir del Data Mart DMSenamhi, se seleccionaron las siguientes tablas y campos para satisfacer el requerimiento:

- DimEstacion: CódigoEstacion, NombreEstacion, Distrito, Ubigeo
- DimContaminante: CódigoContaminante, NombreContaminante, LimiteMaximo

- DimTiempo: Fecha, Año, Mes, DíaSemana
- THContaminacion: CódigoEstación, CódigoContaminante, CódigoTiempo, ValorMedido

Estas tablas permiten construir un modelo relacional que integra las dimensiones de tiempo, contaminante y ubicación con los hechos de medición.

## **4.2. Diseño de estructura e integración**

### **4.2.1. Estructura y componentes de Interfaz Gráfica de Usuario**

La interfaz de usuario se estructuró en zonas:

- Título: “BI – DMSenamhi”
- Filtros de consulta: Estación, Contaminante, Periodo (Año-Mes), Distrito
- Totales / globales: Promedios de contaminantes, número de alertas, valor máximo registrado
- Análisis en detalle: Gráficos de líneas, barras y mapas para evolución temporal, comparación entre estaciones y alertas

### **4.2.2. Integración de Campos (datos) y Objetos gráficos de visualización (OGV)**

Campo (dato) del Data Mart	Integración y propósito	Objeto gráfico visual (OGV)
Estación	En el Slicer, permite seleccionar estaciones específicas para filtrar la información	Slicer

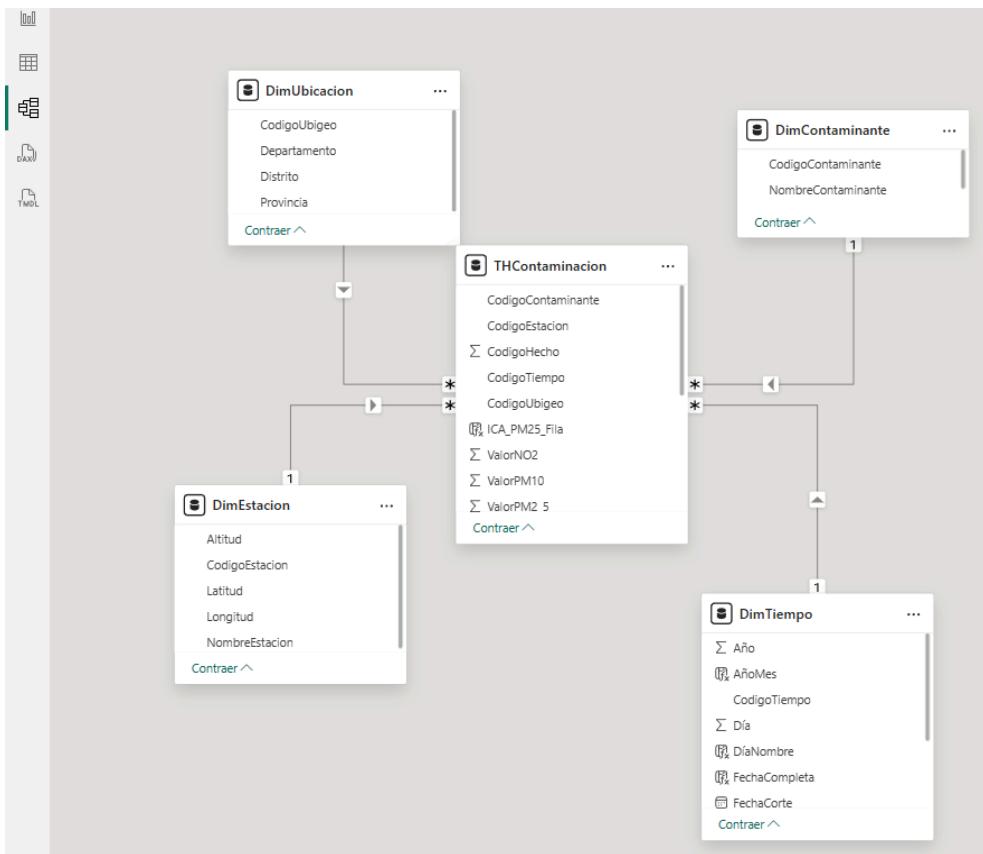
Contaminante	En el Slicer, permite seleccionar contaminantes específicos para filtrar datos	Slicer
Año-Mes (Tiempo)	En el Slicer, permite filtrar información por periodo	Slicer
Distrito	En el Slicer, filtra por distrito para análisis geográfico	Slicer
ValorMedido	En Card, se muestra el promedio del contaminante seleccionado	Card
ValorMedido	En Card, se muestra el valor máximo registrado por estación	Card
ValorMedido, Estación, Fecha	En gráfico de líneas, se visualiza la evolución temporal del contaminante por estación	Line Chart
ValorMedido, Estación	En gráfico de barras, se comparan los niveles promedio entre estaciones	Clustered Column Chart
ValorMedido, Estación	En Funnel, se muestra ranking de estaciones según nivel promedio de contaminantes	Funne

#### 4.3. Construcción

##### 4.3.1. Tablero de resultados (Dashboard)

Para la construcción de la interfaz gráfica de usuario Dashboard de Contaminantes por Estación, se utilizaron las siguientes acciones en Power BI Desktop:

- a) Conexión a SQL Server con la base DMSenamhi
- b) Selección de tablas: DimEstacion, DimContaminante, DimTiempo, THContaminacion
- c) Carga de datos y transformación con Power Query (tipos de datos correctos, eliminación de valores nulos y errores)
- d) Creación de relaciones entre tablas y configuración de cardinalidad



- e) Definición de medidas DAX: promedio por contaminante, número de alertas, valor máximo registrado

Nombre de la Medida	Código DAX	Descripción
<b>Promedio PM2.5</b>	<code>DAX\nPromedio PM25 =\nAVERAGE(THContaminacion[PM2_5])\n</code>	Calcula el promedio general de PM2.5 según filtros.

<b>Promedio PM10</b>	<b>DAX</b> Promedio PM10 = <code>AVERAGE(THContaminacion[PM10])</code>	Calcula el promedio de PM10 según filtros.
<b>Promedio NO2</b>	<b>DAX</b> Promedio NO2 = <code>AVERAGE(THContaminacion[NO2])</code>	Calcula el promedio de NO <sub>2</sub> .
<b>Promedio Mensual PM2.5</b>	<b>DAX</b> Promedio Mensual PM25 = <code>CALCULATE(AVERAGE(THContaminacion[PM2_5]), ALLEXCEPT(DimTiempo, DimTiempo[Año], DimTiempo[Mes]))</code>	Promedio por mes respetando año y mes.
<b>Promedio Mensual PM10</b>	<b>DAX</b> Promedio Mensual PM10 = <code>CALCULATE(AVERAGE(THContaminacion[PM10]), ALLEXCEPT(DimTiempo, DimTiempo[Año], DimTiempo[Mes]))</code>	Promedio PM10 mensual.
<b>Promedio Mensual NO2</b>	<b>DAX</b> Promedio Mensual NO2 = <code>CALCULATE(AVERAGE(THContaminacion[NO2]), ALLEXCEPT(DimTiempo, DimTiempo[Año], DimTiempo[Mes]))</code>	Promedio mensual NO <sub>2</sub> .
<b>Última Fecha Registrada</b>	<b>DAX</b> Ultima Fecha = <code>MAX(DimTiempo[Fecha])</code>	Muestra el último día con datos.
<b>Promedio por Estación (Dynamic)</b>	<b>DAX</b> Promedio Estacion = <code>AVERAGEX(values(DimEstacion[Estacion]), [Promedio PM25])</code>	Recalcula dinámicamente por estación.

- f) Integración de objetos visuales según la lógica de datos para su visualización
- g) Configuración de la interfaz de usuario: Título, filtros, totales/globales y análisis detallado

En la construcción se obtuvo la siguiente interfaz gráfica de usuario, Dashboard de Contaminantes por Estación:

- Los analistas y gestores ambientales pueden consultar los niveles de contaminación por estación, contaminante, periodo de tiempo y distrito.
- Las consultas analíticas pueden ser resueltas seleccionando filtros como Estación, Contaminante y Año-Mes (Conocimiento explícito), permitiendo el análisis de resultados y la toma de decisiones estratégicas (Conocimiento implícito).

#### **4.4. Despliegue**

- El dashboard fue publicado en Power BI Service, disponible para usuarios autorizados.
- Se configuró actualización automática mediante On-premises data gateway, asegurando que los datos reflejan mediciones recientes.
- La interfaz permite filtrar, comparar y analizar los datos de manera interactiva, apoyando la toma de decisiones respecto a políticas de control de la contaminación y alertas ambientales.

A continuación nuestro software:

## → Dashboard 1 : Resumen general

**Sistema BI**  
**Calidad del Aire - Lima**

Data Mart  
DMSenamhi  
SENAMHI 2018-2025

**MODULO DE ANALISIS**

- Resumen General
- Análisis Temporal
- Comparación Estaciones
- Alertas y Reportes
- Correlación Aire-Salud

### Resumen General

Año

Estación Monitoreo

- Seleccionar todo
- CARABAYLLO
- JESUS\_MARIA
- SAN\_BORJA
- SAN\_JUAN\_DE\_LURIGANCHO
- SAN\_MARTIN\_DE\_PORRES

Última Actualización

viernes, 31 de mayo de 2024

PM10 Promedio Total

**58,04**

PM2.5 Promedio Total

**30,30**

NO2 Promedio Total

**21,57**

Estaciones Totales

**7**

NombreEstacion	Promedio_PM25	Promedio_PM10	Promedio_NO2
SAN_JUAN_DE_LURIGANCHO	35,98	65,80	
CAMPO_DE_MARTE	17,11	24,33	14,77
SAN_BORJA	26,07	77,17	18,59
VILLA_MARIA_DEL_TRIUNFO	34,89	67,88	18,72
SAN_MARTIN_DE_PORRES	23,02	36,92	22,74
CARABAYLLO	35,59	64,32	27,39
SANTA_ANITA	38,68	70,47	27,91
Total	<b>30,30</b>	<b>58,04</b>	<b>21,57</b>

Distribución calidad de aire

Promedio_NO2	52,8%
Promedio_PM10	19,6%
Promedio_PM2.5	27,6%

## → Dashboard 2 : Análisis Temporal

**Sistema BI**  
**Calidad del Aire - Lima**

Data Mart  
DMSenamhi  
SENAMHI 2018-2025

**MODULO DE ANALISIS**

- Resumen General
- Análisis Temporal
- Comparación Estaciones
- Alertas y Reportes

**Analisis Temporal**

NO2

PM10

PM2.5

Año

 2015
  2016
  2017
  2018

Estación Monitoreo

- Seleccionar todo
- CAMPO\_DE\_MARTE
- CARABAYLLO
- SAN\_BORJA
- SAN\_JUAN\_DE\_LURIGANCHO

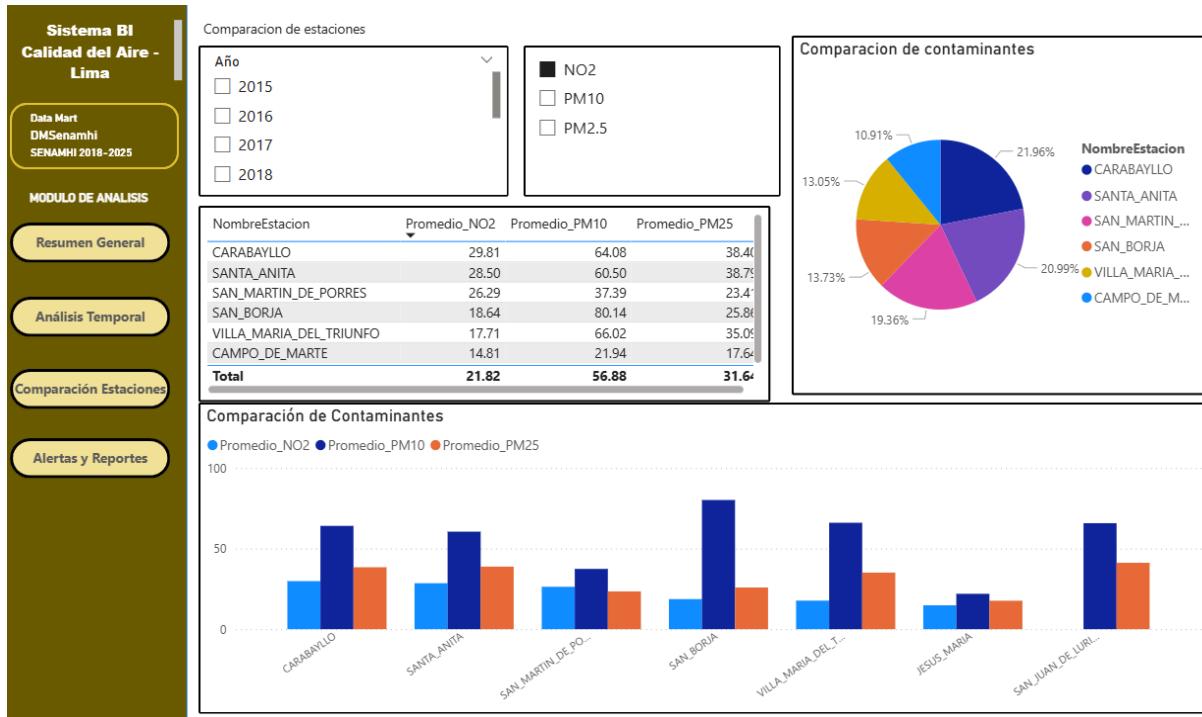
Evolución Mensual de Contaminantes - Últimos 12 meses

Mes	Concentración (ug/m³)
jun	27
may	26
jul	25
sep	24
ago	24
abr	23
oct	22
nov	21
mar	21
dic	20
feb	20
ene	19

Patrón Horario Promedio

Hora	MedidaContaminanteSeleccionado
12:00 a.m.	25
03:00 a.m.	24
06:00 a.m.	20
09:00 a.m.	18
12:00 p.m.	22
03:00 p.m.	28
06:00 p.m.	22
09:00 p.m.	23

## → Dashboard 3 : Comparación de estaciones



## → Dashboard 4 : Alertas y Reportes

**Sistema BI  
Calidad del Aire - Lima**

Data Mart  
DMSenamhi  
SENAMHI 2018-2025

**MÓDULO DE ANÁLISIS**

- Resumen General
- Análisis Temporal
- Comparación Estaciones
- Alertas y Reportes
- Correlación Aire-Salud

NombreEstacion	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Total
SAN_JUAN_DE_LURIGANCHO	12654	11583	15654	9126	8163	6606	13158	19830	6693	2265	<b>105732</b>
SANTA_ANITA	11049	13029	14574	7749	3201	561	17130	9222	2580	79101	
CARABAYLLO	7794	9930	6	6396	1269	3429	3138	15960	12819	2292	<b>63033</b>
VILLA_MARIA_DEL_TRIUNFO	4515	7095	9156	7110	4815	1281	276	16869	7998	900	<b>60015</b>
SAN_BORJA	3624	3630	2265	3015	597	690	2601	9021	2544	1341	<b>29328</b>
SAN_MARTIN_DE_PORRES	2556	2826	4467	2394	1119	201	8526	2823	1644	26556	
CAMPO_DE_MARTE	2169	2046	1350	4926	5151	2358	1569	3687	510	1497	<b>25263</b>
<b>Total</b>	<b>44361</b>	<b>50139</b>	<b>47472</b>	<b>40716</b>	<b>24315</b>	<b>14364</b>	<b>21504</b>	<b>91023</b>	<b>42609</b>	<b>12525</b>	<b>389028</b>

## **4.5. Dashboard “Correlación Aire-Salud”**

### **4.5.1. Requerimientos funcionales**

Este dashboard tiene como finalidad mostrar, de forma integrada, la situación de calidad del aire y la carga de infecciones respiratorias agudas en los siete distritos seleccionados de Lima Metropolitana. Los requerimientos principales son:

- Visualizar el promedio anual de los contaminantes PM2.5, PM10 y NO<sub>2</sub> por distrito.
- Mostrar el número anual de episodios de IRA, neumonías, hospitalizaciones y defunciones por grupo etario.
- Permitir la comparación entre distritos y años mediante filtros interactivos.
- Proporcionar visualizaciones que faciliten la interpretación de posibles asociaciones entre los niveles de contaminación y los eventos de salud.

### **4.5.2. Esquema de datos utilizado**

El dashboard se alimenta de la tabla de hechos Fact\_AireSalud, conectada en Power BI con las dimensiones de tiempo y ubicación. Fact\_AireSalud provee, para cada combinación de distrito y año, los valores de:

- Prom\_PM10\_Anual, Prom\_PM25\_Anual, Prom\_NO2\_Anual.
- Total\_IRA\_NoNeum\_Anual, Total\_Neum\_Men5\_Anual, Total\_Neum\_60Mas\_Anual.
- Total\_Hosp\_Men5\_Anual, Total\_Hosp\_60Mas\_Anual.
- Total\_Def\_Men5\_Anual, Total\_Def\_60Mas\_Anual.

Esta estructura permite que todos los objetos visuales del dashboard respondan a los mismos filtros y entreguen una vista coherente del vínculo Aire-Salud.

### **4.5.3. Diseño de la interfaz**

La interfaz se organiza en tres áreas:

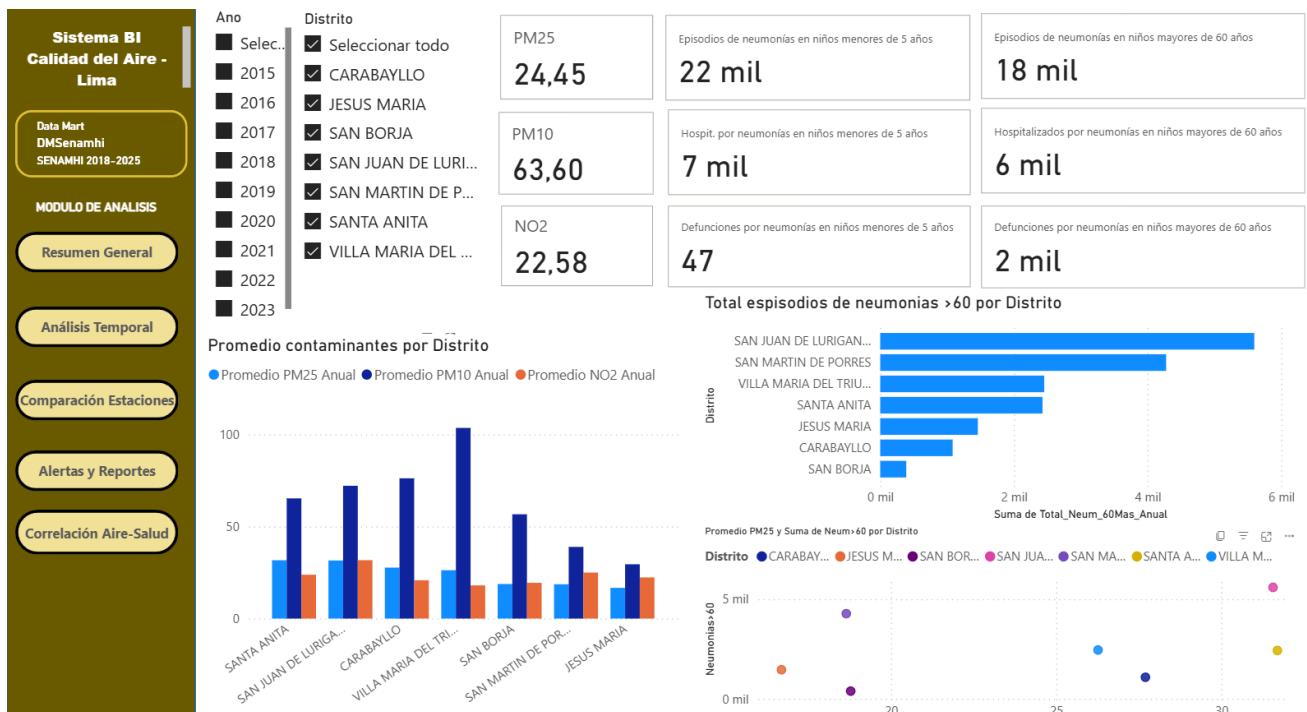
- Filtros laterales: segmentadores para Año y Distrito, que permiten seleccionar uno o varios distritos y períodos de análisis.
- Indicadores clave (tarjetas): muestran, para el contexto filtrado, los promedios anuales de PM2.5, PM10 y NO<sub>2</sub>, así como los totales anuales de neumonías y hospitalizaciones por grupo etario.
- Visualizaciones analíticas:

- Gráfico de columnas agrupadas con los promedios de contaminantes por distrito.
- Gráfico de barras con el total anual de neumonías en mayores de 60 años por distrito.
- Gráfico combinado o de dispersión que relaciona un contaminante (por ejemplo Prom\_PM25\_Anual) con un indicador de salud (como Total\_Neum\_Men5\_Anual), facilitando la identificación de patrones de asociación.

#### 4.5.4. Despliegue

El dashboard “Correlación Aire-Salud” se integra al sistema BI mediante un botón adicional en el menú lateral de la aplicación, manteniendo la misma línea gráfica y esquema de navegación que los otros cuatro tableros. El informe se publica en el servicio de visualización utilizado por el proyecto y se configura su actualización periódica en función de la disponibilidad de nuevos datos en el Data Mart DMSenamhi y en Stage\_IRA.

#### → Dashboard 5 : Correlación Aire-Salud



## **5. Conclusiones**

El desarrollo del software de Business Intelligence permitió consolidar en un solo entorno la información de monitoreo de contaminantes atmosféricos de las estaciones automáticas del SENAMHI para Lima Metropolitana, pasando de archivos planos dispersos a un Data Mart estructurado con dimensiones de estación, ubicación, tiempo y contaminante. Esta arquitectura facilitó la construcción de dashboards que responden de manera ágil a consultas analíticas sobre niveles de PM2.5, PM10 y NO<sub>2</sub>, evolución temporal por distrito y comparación entre estaciones, apoyando la vigilancia de la calidad del aire y la identificación de escenarios críticos.

La integración de la open data de vigilancia de infecciones respiratorias agudas en la tabla de hechos Fact\_AireSalud amplió el alcance del proyecto, al vincular indicadores ambientales con la carga de enfermedad respiratoria en siete distritos de Lima Metropolitana. A través del dashboard “Correlación Aire-Salud” se logró visualizar, por distrito y año, los promedios de contaminantes junto con los totales anuales de IRA, neumonías, hospitalizaciones y defunciones por grupo etario, lo que aporta una perspectiva conjunta Ambiente–Salud que no estaba disponible en los reportes originales de cada institución.

Desde el punto de vista de ingeniería de datos, el proyecto demuestra la viabilidad de integrar fuentes heterogéneas mediante un modelo dimensional coherente y tablas de hechos agregadas, manteniendo la trazabilidad desde los datos originales hasta los indicadores mostrados en los dashboards. El uso de medidas analíticas y visualizaciones comparativas permite realizar análisis exploratorios sobre la posible asociación entre niveles de contaminación y eventos de salud, constituyendo un primer paso hacia la aplicación sistemática de técnicas de ciencia de datos en la gestión ambiental y sanitaria.