

# Documentación clara y detallada del CuboMovilidadSV y del proceso ETL con su análisis.

---

## ETL y Análisis OLAP en SQL Server

### Introducción

Esta guía te ayudará a entender y ejecutar **el proceso ETL (Extract, Transform, Load) y la creación de un cubo OLAP en SQL Server Analysis Services (SSAS)**. Abordaremos la estructura de datos, consultas en SQL y MDX, y visualización de resultados en herramientas como **Power BI y Excel**.

**Objetivo:** Analizar la movilidad en distintos departamentos de **El Salvador** durante la pandemia de **COVID-19** usando datos de un archivo CSV.

---

## 1. Proceso ETL

El **ETL (Extract, Transform, Load)** es el proceso de preparación de datos antes de analizarlos en el cubo OLAP. **Tiene tres fases fundamentales:**

**Extracción:** Se importan los datos desde un archivo CSV a SQL Server.

**Transformación:** Se limpian, organizan y estructuran los datos para análisis.

**Carga:** Se insertan en tablas dimensionales y de hechos.

---

### 1.1 Extracción de Datos

**Fuente de Datos:** El archivo movilidad.csv contiene las siguientes columnas:

- fecha
- pais
- departamento
- parks\_percent\_change
- grocery\_percent\_change
- transit\_percent\_change

- workplaces\_percent\_change
- residential\_percent\_change

### **Creación de la Tabla Temporal para Importar Datos**

```
CREATE TABLE MovilidadTemporal (  
    fecha DATE,  
    pais VARCHAR(50),  
    departamento VARCHAR(100),  
    parks FLOAT,  
    grocery FLOAT,  
    transit FLOAT,  
    workplaces FLOAT,  
    residential FLOAT  
);
```

### **Importación de Datos desde el Archivo CSV**

```
BULK INSERT MovilidadTemporal  
FROM 'C:\ruta_del_archivo\movilidad.csv'  
WITH (  
    FIELDTERMINATOR = ',',  
    ROWTERMINATOR = '\n',  
    FIRSTROW = 2  
);
```

**Este paso almacena los datos crudos antes de procesarlos.**

---

## **1.2 Transformación de Datos**

**Antes de insertar los datos en las tablas finales, validamos la calidad de los datos.**

### **Verificar valores nulos**

```
SELECT * FROM MovilidadTemporal WHERE fecha IS NULL OR departamento IS NULL;
```

### **Identificar duplicados**

```
SELECT fecha, departamento, COUNT(*)  
FROM MovilidadTemporal  
GROUP BY fecha, departamento  
HAVING COUNT(*) > 1;
```

## Revisar rangos de movilidad

```
SELECT MIN(parks) AS Min_Movilidad, MAX(parks) AS Max_Movilidad FROM  
MovilidadTemporal;
```

**Cualquier dato incorrecto se corrige antes de la carga.**

---

## 1.3 Creación del Modelo Dimensional

**Para optimizar el análisis OLAP, organizamos los datos en un esquema dimensional.**

### Tablas de Dimensiones

```
CREATE TABLE DimTiempo (  
    idTiempo INT PRIMARY KEY IDENTITY(1,1),  
    fecha DATE NOT NULL,  
    anio AS (YEAR(fecha)) PERSISTED,  
    mes AS (MONTH(fecha)) PERSISTED,  
    dia AS (DAY(fecha)) PERSISTED  
);
```

```
CREATE TABLE DimUbicacion (  
    idUbicacion INT PRIMARY KEY IDENTITY(1,1),  
    pais VARCHAR(50) NOT NULL,  
    departamento VARCHAR(100) NOT NULL  
);
```

```
CREATE TABLE DimMovilidad (  
    idCategoria INT PRIMARY KEY IDENTITY(1,1),  
    tipoMovilidad VARCHAR(50) NOT NULL  
);
```

### Tabla de Hechos

```
CREATE TABLE HechosMovilidad (  
    idHechosMovilidad INT PRIMARY KEY IDENTITY(1,1),  
    idTiempo INT FOREIGN KEY REFERENCES DimTiempo(idTiempo),  
    idUbicacion INT FOREIGN KEY REFERENCES DimUbicacion(idUbicacion),  
    idCategoria INT FOREIGN KEY REFERENCES DimMovilidad(idCategoria),  
    porcentajeCambio FLOAT NOT NULL  
);
```

**Este modelo facilita el análisis multidimensional.**

---

## 1.4 Carga de Datos en Tablas Finales

### Insertar registros en las dimensiones

```
INSERT INTO DimTiempo (fecha)
SELECT DISTINCT fecha FROM MovilidadTemporal;
```

```
INSERT INTO DimUbicacion (pais, departamento)
SELECT DISTINCT pais, departamento FROM MovilidadTemporal;
```

```
INSERT INTO DimMovilidad (tipoMovilidad) VALUES
('Parques'), ('Supermercados'), ('Transporte'), ('Trabajo'), ('Residencial');
```

### Insertar registros en la tabla de hechos

```
INSERT INTO HechosMovilidad (idTiempo, idUbicacion, idCategoria, porcentajeCambio)
SELECT dt.idTiempo, du.idUbicacion, dm.idCategoria, mt.parks
FROM MovilidadTemporal mt
JOIN DimTiempo dt ON mt.fecha = dt.fecha
JOIN DimUbicacion du ON mt.pais = du.pais AND mt.departamento = du.departamento
JOIN DimMovilidad dm ON dm.tipoMovilidad = 'Parques';
```

**Este paso consolida los datos en el esquema dimensional para análisis.**

---

## 2. Creación del Cubo OLAP

**En SQL Server Analysis Services (SSAS), configuramos un cubo multidimensional.**

### Pasos en SSAS

**Crear un proyecto en SQL Server Data Tools (SSDT).**

**Definir el origen de datos (MovilidadElSalvador).**

**Agregar las tablas DimTiempo, DimUbicacion, DimMovilidad, HechosMovilidad.**

**Configurar el cubo y seleccionar porcentajeCambio como medida principal.**

**Procesar el cubo OLAP.**

---

## 3. Consultas en MDX para OLAP

**consultas\_OLAP\_MDX.mdx**

## Comparación de movilidad por departamento

```
SELECT
  [DimUbicacion].[Departamento].Members ON ROWS,
  [Measures].[PorcentajeCambio] ON COLUMNS
FROM [CuboMovilidadSV];
```

## Análisis de movilidad en parques

```
SELECT
  [DimTiempo].[Mes].Members ON ROWS,
  [Measures].[PorcentajeCambio] ON COLUMNS
FROM [CuboMovilidadSV]
WHERE ([DimMovilidad].[TipoMovilidad].[Parques]);
```

---

## 4. Visualización de Datos

**Las consultas pueden visualizarse en:**

**Power BI** → Dashboards dinámicos con gráficos de movilidad.

**Excel (PivotTables)** → Conexión con SSAS para reportes.

**Python (Matplotlib)** → Gráficos personalizados en código.

Ejemplo en Python:

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
df = pd.read_sql("SELECT departamento, AVG(porcentajeCambio) AS movilidad FROM
HechosMovilidad JOIN DimUbicacion ON HechosMovilidad.idUbicacion =
DimUbicacion.idUbicacion GROUP BY departamento", conn)
```

```
plt.bar(df['departamento'], df['movilidad'])
plt.xlabel("Departamento")
plt.ylabel("Cambio de Movilidad (%)")
plt.title("Promedio de Movilidad por Departamento")
plt.xticks(rotation=45)
plt.show()
```

---