Logotipo, nombre de la empresa

Descripción generada automáticamente

**Bases de Datos Aplicada**

**Parcial Práctico**

**Alumnos:**

* Bravo Santiago
* Tittonell Ian
* Mardyks Lautaro
* Romero Agustin

**Docente:** Martin Miguel Rivas Sarquis

**Localización:** Lomas

**Comisión:** 3-A

**Turno:** Mañana

**Año:** 2024

# Documentación

## Introducción

El propósito de este documento es proporcionar una descripción detallada y estructurada del proceso de construcción de un **Data Warehouse (DW)**, abarcando cada uno de los pasos fundamentales que permiten consolidar y optimizar la información para su análisis. Este proceso incluye, en primer lugar, la **extracción de datos desde un sistema OLTP** (Online Transaction Processing), que típicamente contiene datos operacionales y transaccionales. Dichos datos serán trasladados hacia una **Staging Area (SA)**, una zona temporal donde los datos se depuran, limpian y preparan antes de ser cargados en el DW.

La importancia de la SA radica en que es un paso intermedio que permite realizar transformaciones críticas sobre los datos, como la corrección de inconsistencias, la eliminación de duplicados y la normalización de valores. Este paso garantiza que los datos que llegan al DW sean precisos y consistentes. Posteriormente, se efectúa la **carga final al DW**, un repositorio optimizado para consultas y análisis de gran escala, donde los datos estarán organizados de forma que soporten la toma de decisiones estratégicas.

Además de este flujo, se detallarán las **estructuras y herramientas empleadas** para la implementación de este proceso. En este contexto, se hará un enfoque particular en el diseño y desarrollo de los **procesos ETL** (Extract, Transform, Load), que son fundamentales para asegurar la correcta integración y transformación de la información. El ETL no solo se encarga de la extracción de datos desde los sistemas fuente, sino también de su transformación para adaptarse a los requisitos del DW y su posterior carga final, donde estarán listos para su análisis mediante herramientas de inteligencia de negocios (BI).

## Arquitectura del Proyecto

OLTP: La base de datos **Sakila**, un sistema transaccional, es la fuente principal de los datos. En un entorno OLTP (Online Transaction Processing), los datos se generan a través de operaciones diarias como registros de alquileres, ventas o interacciones con los clientes. Este sistema está optimizado para gestionar operaciones frecuentes y rápidas, con un enfoque en la eficiencia del procesamiento de transacciones, pero no está diseñado para el análisis masivo de datos históricos. Por lo tanto, es necesario extraer estos datos de Sakila para trasladarlos a un entorno adecuado para el análisis, como es el caso del Data Warehouse (DW).

**Staging Area (SA)**: Es una zona de almacenamiento temporal crucial para el proceso de ETL. En la **Staging Area**, los datos extraídos desde el OLTP se depuran, transforman y preparan para su inserción final en el DW. En este entorno, las tablas se han diseñado sin restricciones de **NOT NULL** ni claves foráneas, lo cual brinda una mayor flexibilidad durante la carga inicial de datos. Al eliminar estas restricciones, se facilita la limpieza y transformación de los datos sin la rigidez que normalmente implican las reglas de integridad referencial. Aquí se pueden realizar tareas como la eliminación de duplicados, la conversión de formatos y la validación de calidad de los datos. Una vez preparados, los datos están listos para ser migrados al DW.

**Data Warehouse (DW)**: Es el repositorio final donde los datos ya depurados y transformados se almacenan de manera estructurada y optimizada para la generación de consultas y análisis. A diferencia de las bases OLTP, el DW está diseñado para gestionar grandes volúmenes de datos históricos, facilitando consultas complejas y reportes orientados a la toma de decisiones estratégicas. Las tablas en el DW suelen estar modeladas en esquemas dimensionales, como estrella o copo de nieve, para mejorar la eficiencia en las consultas. Las tablas de dimensiones, como **DimCustomer** y **DimFilm**, contienen información descriptiva sobre los clientes y productos, mientras que la tabla de hechos **FactRental** registra las transacciones de alquiler, permitiendo realizar análisis detallados como el comportamiento de los clientes o el rendimiento de los productos.

**Power BI**: Una vez que los datos se encuentran cargados en el DW, herramientas de inteligencia de negocios como **Power BI** juegan un rol fundamental en la visualización y análisis de la información. **Power BI** permite conectarse directamente al DW para generar dashboards interactivos, reportes dinámicos y gráficos visuales que facilitan la interpretación de los datos por parte de los usuarios de negocio. Con Power BI, es posible crear reportes que muestren patrones de comportamiento de los clientes, tendencias de ventas a lo largo del tiempo, así como análisis predictivos basados en los datos históricos almacenados en el DW. Además, Power BI permite actualizar los reportes de manera automática a medida que los datos del DW se actualizan, lo que asegura que las decisiones empresariales estén siempre respaldadas por información actual y relevante.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Diagrama

Descripción generada automáticamente

## Extracción, Transformación y Carga (ETL)

### Carga desde OLTP a Staging Area (SA)

* **Objetivo**: Extraer los datos de las bases OLTP y llevarlos a la SA para limpieza y transformación.
* **Paquete ETL**:
  + **Origen**: Tablas de OLTP (Rental, Customer, Film).
  + **Destino**: Tablas de SA (RentalSA, CustomerSA, FilmSA).

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

### Carga desde Staging Area (SA) al Data Warehouse (DW)

* **Objetivo**: Llevar los datos limpios desde la SA al DW para análisis y reportes.
* **Paquete ETL**:
  + **Origen**: Tablas de SA (RentalSA, CustomerSA, FilmSA).
  + **Destino**: Tablas del DW (DimCustomer, DimFilm, FactRental).

Diagrama, Texto

Descripción generada automáticamente

**Power Bi Reports:**

