

UNIVERSIDAD PRIVADA DEL VALLE

**FACULTAD:** FACULTAD DE INFORMÁTICA Y ELECTRÓNICA

**CARRERA:** INGENIERÍA DE SISTEMAS INFORMÁTICOS

**MATERIA:** DATAWAREHOUSE

PROYECTO FINAL “ANALISIS DE LA INDUSTRIA CINEMATOGRAFICA DE HOLLYWOOD”

**MATERIA:** Datawarehouse

**DOCENTE:** Ing. Fernando Zeballos Orellana

**ESTUDIANTES:** Rojas Cossio Valeria Milka

Andia Espinoza Juan David

Méndez Burgos Manuel Alejandro

Cochabamba – Bolivia

15 / 06 / 2024

Índice

1.Introducción.

2.Objetivos

3.Definición de requisitos del negocio

4.Selección de herramientas y productos

5.Modelo dimensional

6.Diseño físico

7.Diseño de Procesos ETL

8.Desarrollo de herramienta de exploración

9.Conclusiones y Recomendaciones

1. Introducción

En este proyecto, se desarrolla un data warehouse utilizando el dataset "Hollywood Theatrical Market Synopsis 1995 to 2021". El objetivo principal es consolidar los datos relacionados con la industria cinematográfica de Hollywood y proporcionar herramientas de análisis eficientes usando Power BI.

Para el desarrollo del data warehouse, se ha usado la metodología Kimball, ya que es conocida por su enfoque en el modelado dimensional y la creación de data marts orientados al usuario final.

1. Objetivos

- Validar los datos de la industria cinematográfica de Hollywood en un data warehouse.

- Proporcionar una estructura de datos eficiente para el análisis de métricas clave.

- Desarrollar procesos ETL para asegurar la calidad y consistencia de los datos.

- Crear visualizaciones interactivas y dashboards en Power BI para facilitar el análisis de datos.

1. Definición de requisitos del negocio

Los principales requisitos del negocio para este proyecto incluyen:

- Análisis de tendencias de entradas vendidas y total recaudado por año, película, clasificación, género y productora.

- Evaluación del rendimiento financiero de las películas a lo largo del tiempo.

- Comparación de ingresos y asistencia en diferentes años.

1. Selección de herramientas y productos

- Talend: Seleccionado para el proceso de ETL’s ya que su capacidad es apta para manejar grandes volúmenes de datos y su flexibilidad en la transformación de datos.

- Power BI: Elegido para la visualización de datos por su facilidad de uso y sus potentes capacidades de creación de dashboards interactivos.

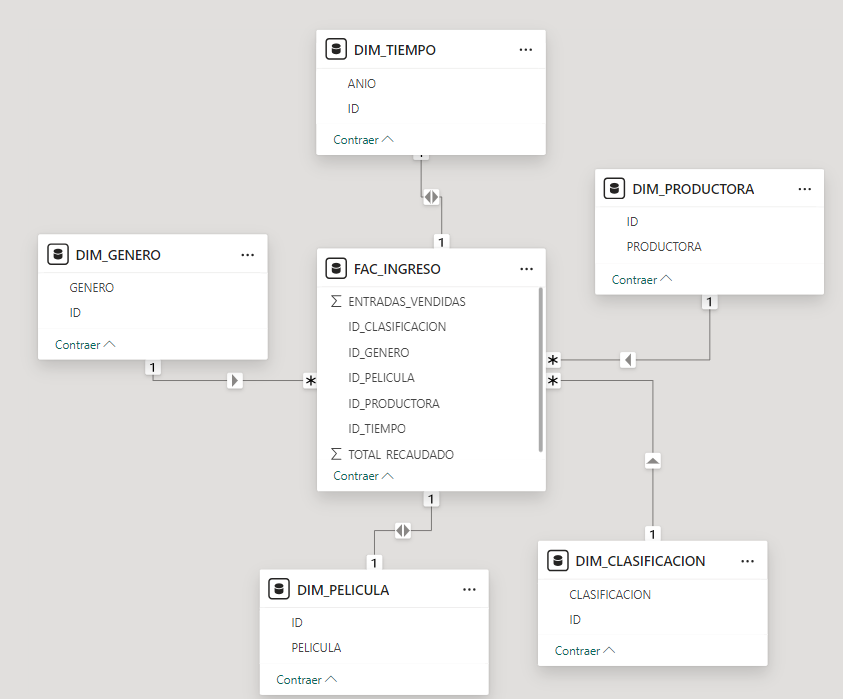
- Base de datos: Se utilizó una base de datos relacional SQLPlus para almacenar los datos transformados.

1. Modelo dimensional

Diagrama

Descripción generada automáticamente

1. Diseño físico



1. Diseño de procesos ETL

Fac\_ingreso

Diagrama

Descripción generada automáticamente

Captura de pantalla con letras

Descripción generada automáticamente

Dim\_pelicula

Imagen que contiene Escala de tiempo

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Dim\_genero

Gráfico

Descripción generada automáticamente con confianza media Texto

Descripción generada automáticamente con confianza media

Dim\_clasificacion

Imagen que contiene Teams

Descripción generada automáticamente Texto

Descripción generada automáticamente

Dim\_productora

Imagen que contiene Escala de tiempo

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Dim\_tiempo

Imagen que contiene Escala de tiempo

Descripción generada automáticamente

Tabla

Descripción generada automáticamente con confianza baja

1. Desarrollo de herramienta de exploración

Se ha usado Power BI para crear varios dashboards interactivos que permiten a los usuarios explorar los datos desde diferentes perspectivas. Entre los dashboards creados se incluyen:

Análisis de entradas vendidas por:

* Año
* Película
* Productora
* Clasificación y genero

Análisis del total recaudado por:

* Año
* Productora
* Película
* Género y clasificación

Comparación de entradas vendidas y total recaudado por:

* Año
* Clasificación y productora
* Película
* Genero

1. Conclusiones y recomendaciones

El proyecto ha permitido consolidar y analizar datos clave de la industria cinematográfica de Hollywood, proporcionando insights valiosos a través de visualizaciones interactivas en Power BI. Para futuras mejoras, nuestras recomendaciones son :

- Integrar más fuentes de datos para un mejor análisis.

- Explorar el uso de técnicas avanzadas de análisis predictivo.

Enlace GitHub del proyecto ETL TALEND y la herramienta de exploración: <https://github.com/ValeRojasC/TalendHollywood.git>