Aluso Linda

VIDEO STREAMING PLATFORM: STAR SCHEMA DESIGN

Turning Conceptual Design into KPIs

Abstract

This project builds a simple data model to understand how people interact with video content on the Netflix platform. By bringing together details about customers, videos, ratings, time, date and the cast, it helps answer questions such as what is the most popular video, what is the average rating of the videos, how many video trailers were watched but not the actual video. It is a smart way of turning raw video viewing data into insights that can guide recommendation, targeted marketing and audience retention on the platform.

**Star Schema Design for Netflix Analytics**



**Data Catalog**

1. Customer dimension

CustomerKey-Surrogate key that uniquely identifies each customer in the data warehouse

Customer-A natural key that may be used to track the customer data back to the source system

CustomerName- Captures the customer’s name.

Email-Captures the customer’s email.

DateOfBirth-Capture user’s date of birth for appropriate movie recommendation.

MembershipStatus-Capture the user’s status whether active or inactive.

1. Time dimension

TimeKey – Surrogate key that uniquely identifies each time record in the data warehouse.  
TimeValue – Captures the exact time such as 14:35:00 of the video viewing.  
TimeOfDayCategory – Classifies the time into categories like Morning, Afternoon, Evening, or Night for time-based analysis.

1. Date Dimension

DateKey – Surrogate key that uniquely identifies each date in the data warehouse.  
ViewingDate – Captures the actual date a video was viewed.  
DayName – Captures the day of the week for trend analysis.

1. Rating dimension

**RatingKey** – Surrogate key that uniquely identifies each rating entry.  
**RatingId** – Natural key used to track the rating back to the source system.  
**RatingValue** – Captures the rating category assigned to the video content such as 0 for Not rated, 1 for Thumbs down, 2 for Thumbs Up and 3 for Double Thumbs Up.  
**RatingDescription** – Provides details or explanations for the rating such content type.

1. Video dimension

**VideoKey** – Surrogate key that uniquely identifies each video in the data warehouse.  
**VideoID** – Natural key that may be used to track the video back to the source system.  
**VideoName** – Captures the title of the video.  
**RunTimeMinutes** – Captures the total duration of the video in minutes.  
**Genre** – Classifies the video into categories such as Action, Drama, Comedy.  
**ReleaseYear** – Captures the year the video was released.  
**Director** – Captures the name of the director of the video.  
**Producer** – Captures the name of the producer of the video.

1. BridgeVideoCast

**CastKey** – Foreign key to the DimCast table, linking a specific cast member.  
**VideoKey** – Foreign key to the DimVideo table, linking a specific video.

Used to model the many-to-many relationship between videos and cast members

1. Cast dimension

**CastKey** – Surrogate key that uniquely identifies each cast member in the data warehouse.  
**CastID** – Natural key that may be used to track the cast member back to the source system.  
**CastName** – Captures the full name of the cast member.  
**Role** – Captures the role played by the cast member in the video such as Lead, Supporting or Cameo.

1. FactVideoViewing

Viewing Key-Primary Key which uniquely identifies the viewing session. This will be incremental and automatically assigned by the system.

VideoKey – Foreign key to DimVideo which identifies the video that was viewed.

CustomerKey – Foreign key to DimCustomer; identifies who viewed the video.  
RatingKey – Foreign key to DimRating; provides rating context.  
DateKey – Foreign key to DimDate; captures when the video was viewed.  
TimeKey – Foreign key to DimTime; captures time of viewing.  
ViewingStartTime – Captures the exact start time of the viewing.  
ViewingEndTime – Captures the end time of the viewing session.  
VideoWatched (0/1) – Indicates whether the video was completely watched (1 = yes, 0 = no).  
TrailerWatched (0/1) – Indicates whether only the trailer was watched.  
VideoDownloaded (0/1**)** – Indicates if the video was downloaded for offline viewing.  
PercentageWatched – Captures how much of the video was watched, as a percentage

**KPI to Solve Case Questions**

1. Video View Count

Count of records in FactVideoViewing where VideoWatched=1, group by VideoName from DimVideo.

1. Top 10 Videos

Count of records in FactVideoViewing where VideoWatched=1, group by VideoName from DimVideo, ordered by count descending, limited to top 10.

1. Average Rating for Videos

Average RatingValue from DimRating joined FactVideoViewing, group by VideoName from DimVideo.

1. Selected but not watched

Count of records in FactVideoViewing where TrailerWatched=1 AND VideoWatched=0

1. Video Watched with a specific Actor

Count of distinct VideoKey where VideoWatched=1 AND VideoKey exists in DimCast where ActorName= ‘John Dwayne’

**What other metrics do YOU think would be interesting to know?**

The number of directors whose contents were popular among the users by analyzing the average rating value or percentage watched to evaluate the content creator performance.

The average rating of the popular genre to add more of that genre and to market underperforming genre.

**What trends and groupings do you think would be interesting? How would you use the model to determine those trends and groupings?**

Grouping the users by their age group for personalized video recommendation to retain them once they login.

Calculate user age using DateOfBirth from DimCustomer and GetDate (), then group by Age Group and VideoName from DimVideo. This allows for personalized content recommendations based on age demographics.

For trend I would show the pattern of video viewing each day since the release date.

Compute the number of days between ReleaseDate in DimVideo and ViewingDate in DimDate, group by VideoName,ReleaseYear from DimVideo, ViewingDate from DimDate and DaysSinceRelease obtained from the computation. This will help to identify viewing patterns over time since release.