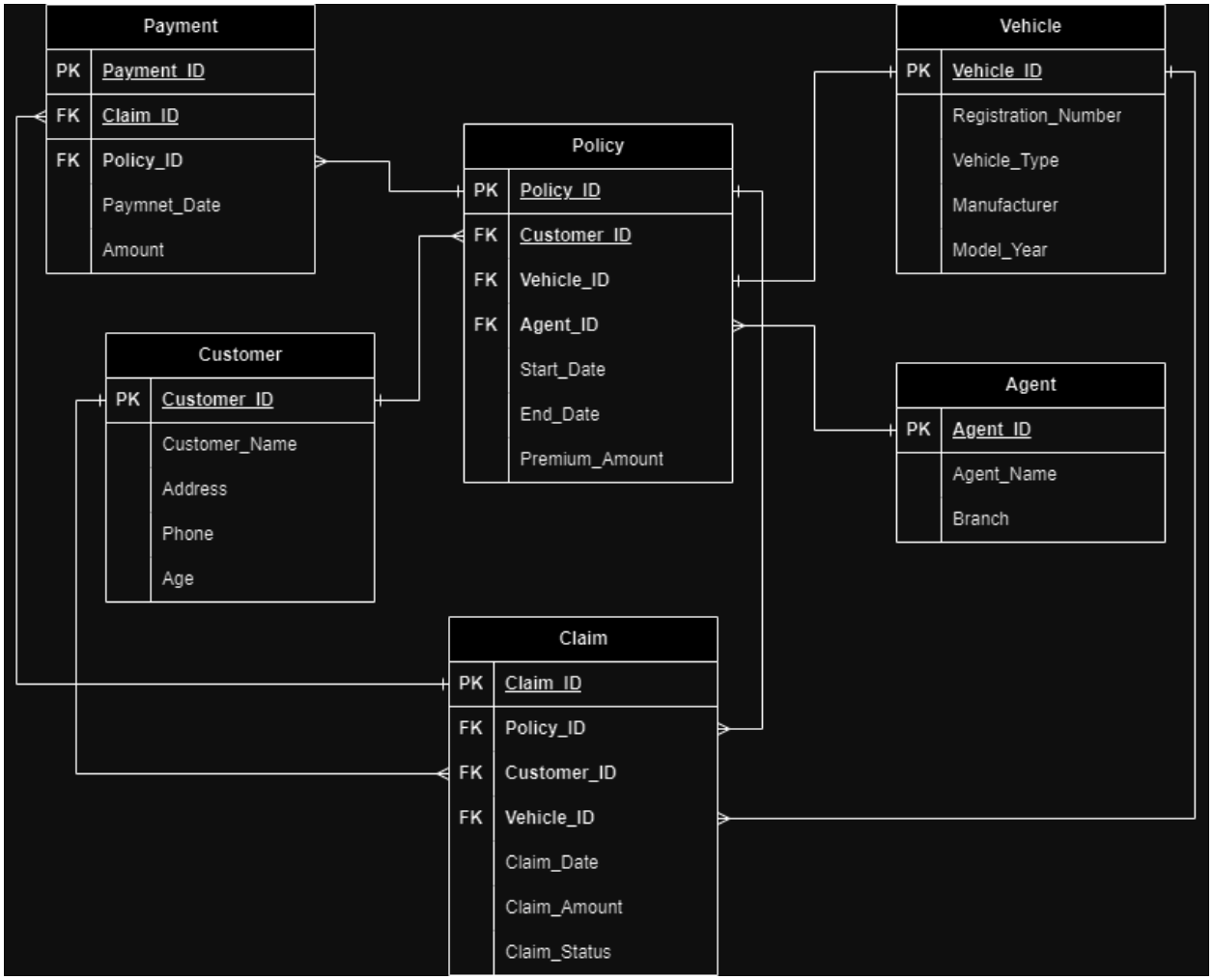
Data Warehouse Design for Multi-Source Databases

1. Motor Insurance Domain

❖ ER Diagram



❖ Names and Types of Dimensions and Facts:

➢ Dimensions:

- **Payment Dimension**: This dimension includes attributes related to payments such as Payment ID, Claim ID, Policy ID, Payment Date, and Amount. It helps in tracking payment details and analyzing financial transactions.
- **Vehicle Dimension**: This dimension includes attributes related to vehicles such as Vehicle ID, Registration Number, Vehicle Type, Manufacturer, and Model Year. It is used to track and analyze vehicle information.
- **Policy Dimension**: This dimension includes attributes related to policies such as Policy ID, Customer ID, Vehicle ID, Agent ID, Start Date, End Date, and Premium Amount. It helps in managing and analyzing insurance policies.
- **Customer Dimension**: This dimension includes attributes related to customers such as Customer ID, Customer Name, Address, Phone, and Age. It helps in analyzing customer details and demographics.

- **Agent Dimension**: This dimension includes attributes related to agents such as Agent ID, Agent Name, and Branch. It is used for tracking agent details and performance.
- **Claim Dimension**: This dimension includes attributes related to claims such as Claim ID, Policy ID, Customer ID, Vehicle ID, Claim Date, Claim Amount, and Claim Status. It helps in managing and analyzing insurance claims.

➢**Facts:**

- **Payment Fact**: This fact table includes attributes such as Payment ID, Claim ID, Policy ID, Payment Date, and Amount. It provides insights into payment transactions and financial details related to claims and policies.
- **Claim Fact**: This fact table includes attributes such as Claim ID, Policy ID, Customer ID, Vehicle ID, Claim Date, Claim Amount, and Claim Status. It helps in analyzing claim details, including amounts and statuses.

- **Type of Data Warehouse and Processing**
  - **Type of Data Warehouse:**
    - **Enterprise Data Warehouse (EDW)**
      - **Reasons:**
        - **Comprehensive Integration**: An EDW integrates data from various sources such as policy management systems, claims processing systems, and customer databases. This provides a unified view of the insurance operations, allowing for a comprehensive analysis of policies, claims, and customer information.
        - **Detailed Analysis and Reporting**: It supports complex queries and detailed reporting, which are essential for assessing policy performance, analyzing claims trends, and understanding customer behavior. This helps in strategic decision-making and regulatory compliance.

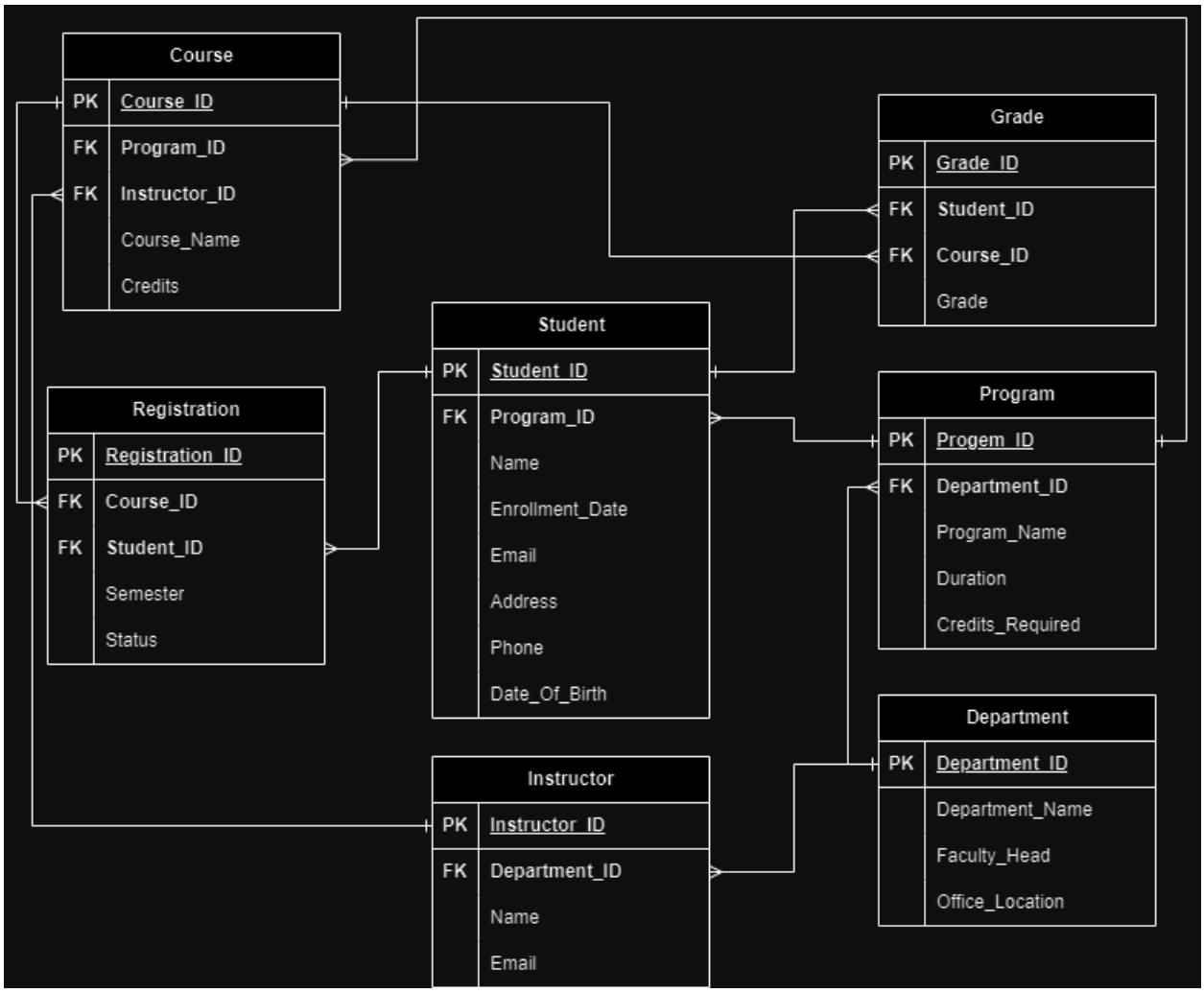
- **Data Consistency and Accuracy**:
  By consolidating data into a single repository, an EDW ensures consistency and accuracy across the organization, which is crucial for maintaining the integrity of insurance records and generating reliable reports.

➢Processing:

- **Batch Processing** is suitable for an EDW in the Motor Insurance domain, as most updates (e.g., new policies, claims processing) occur in bulk and can be handled efficiently in scheduled batches (e.g., daily or weekly). This approach balances performance with the need for timely data updates.

2. University Portal

❖ ER Diagram



❖ Names and Types of Dimensions and Facts:

➢ Dimensions:

- **Course Dimension:** This dimension includes attributes related to courses such as Course ID, Program ID, Instructor ID, Course Name, and Credits. It helps in organizing and analyzing course details.
- **Grade Dimension:** This dimension includes attributes related to grades such as Grade ID, Student ID, Course ID, and Grade. It helps in tracking and analyzing student performance in courses.
- **Student Dimension:** This dimension includes attributes related to students such as Student ID, Program ID, Name, Enrollment Date, Email, Address, Phone, and Date of Birth. It is essential for analyzing student demographics and enrollment details.
- **Program Dimension:** This dimension includes attributes related to programs such as Program ID, Department ID, Program Name, Duration, and Credits Required. It helps in analyzing and organizing academic programs.

- **Department Dimension:** This dimension includes attributes related to departments such as Department ID, Department Name, Faculty Head, and Office Location. It helps in organizing and analyzing department details.
- **Instructor Dimension:** This dimension includes attributes related to instructors such as Instructor ID, Department ID, Name, and Email. It is crucial for tracking instructor details and assignments.

➢**Facts:**

- **Registration Fact:** This fact table includes attributes related to course registrations such as Registration ID, Course ID, Student ID, Semester, and Status. It helps in analyzing registration details and student course enrollment.
- **Grade Fact:** This fact table includes attributes related to grades such as Grade ID, Student ID, Course ID, and Grade. It helps in tracking student performance and course grades.

- **Type of Data Warehouse and Processing:**
  - **Type of Data Warehouse:**
    - **Enterprise Data Warehouse (EDW)**
      - **Reasons:**
        - **Comprehensive Integration:** An EDW integrates data from various university systems, such as student records, course management, faculty details, and financial information. This centralization provides a holistic view of all academic and administrative operations.
        - **Detailed Analysis and Reporting:**
          - It supports extensive analysis and reporting capabilities, allowing for insights into student performance, course effectiveness, and faculty performance. This aids in strategic planning and decision-making.

◆Data Consistency and Accuracy:

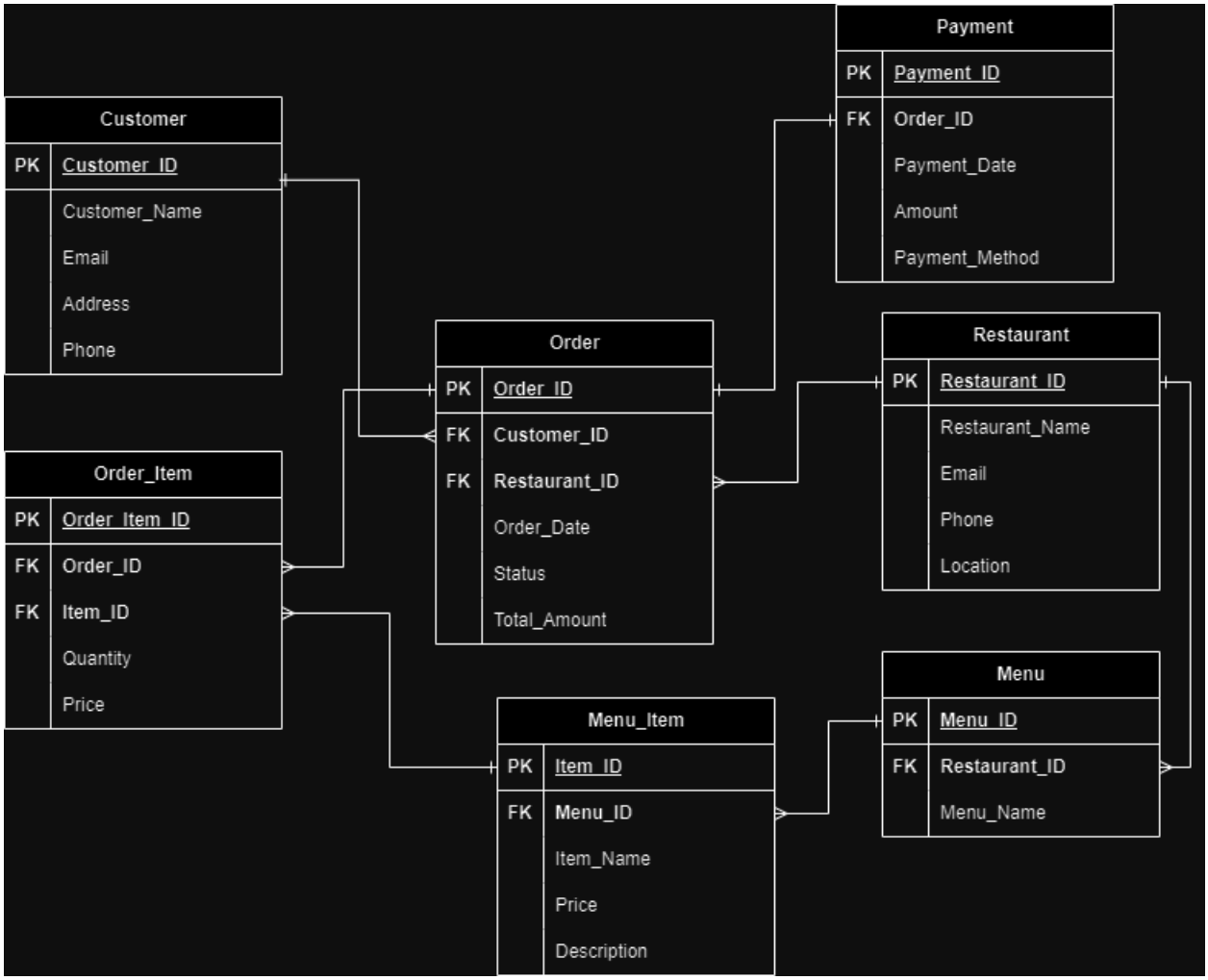
- ➢ By consolidating data from multiple sources, an EDW ensures consistency and accuracy, which is crucial for maintaining reliable academic records and generating comprehensive reports.

➢Processing:

- Batch Processing is appropriate for an EDW in a university portal, as academic data updates (e.g., enrollments, grade submissions) are typically processed in bulk at scheduled intervals (e.g., nightly or weekly). This method ensures efficient data handling and timely updates.

3. Online Food Ordering System

❖ ER Diagram



❖ Names and Types of Dimensions and Facts:

➢ Dimensions:

- **Customer Dimension:** This dimension includes attributes related to customers such as customer-id, name, email, address, and phone. It helps in analyzing customer behavior and preferences.
- **Order Dimension:** This dimension includes attributes related to orders such as order-id, cust-id, rest-id, date, status, and total_amount. It helps in analyzing order patterns and customer purchasing behavior.
- **Payment Dimension:** This dimension includes attributes related to payments such as payment-id, order-id, payment-date, amount, and method. It is crucial for tracking payment methods and amounts.
- **Restaurant Dimension:** This dimension includes attributes related to restaurants such as restaurant-id, name, email, phone, and address. It helps in analyzing restaurant performance and customer preferences.
- **Menu Dimension:** This dimension includes attributes related to menus such as menu-id and name. It helps in tracking the different menus offered by restaurants.

- **Menu Item Dimension:** This dimension includes attributes related to menu items such as item-id, menu-id, item-name, price, and description. It helps in analyzing the popularity and pricing of individual menu items.

➢**Facts:**

- **Order Fact:** This fact table includes metrics such as total_amount and captures order-related details like order-id, cust-id, rest-id, date, status, and total_amount. It helps in analyzing order totals and customer purchase behavior.
- **Payment Fact:** This fact table includes metrics such as amount and records payment details like payment-id, order-id, payment-date, amount, and method. It provides insights into payment methods and financial transactions.
- **Order Item Fact:** This fact table includes metrics such as quantity and price and captures details like order-id, item-id, quantity, and price. It helps in analyzing item sales and understanding order details.

- **Type of Data Warehouse and Processing:**
  - **Type of Data Warehouse**:
    - **Operational Data Store (ODS)**
      - **Reasons:**
        - **Real-Time Data Integration**: An ODS provides real-time or near-real-time integration of data from various operational sources such as order management, inventory systems, and customer interactions. This is crucial for managing live orders, updating inventory levels, and handling customer requests promptly.
        - **Operational Efficiency**: It supports current operational processes by integrating and providing timely data, which helps in optimizing order processing, managing delivery logistics, and ensuring up-to-date customer service.

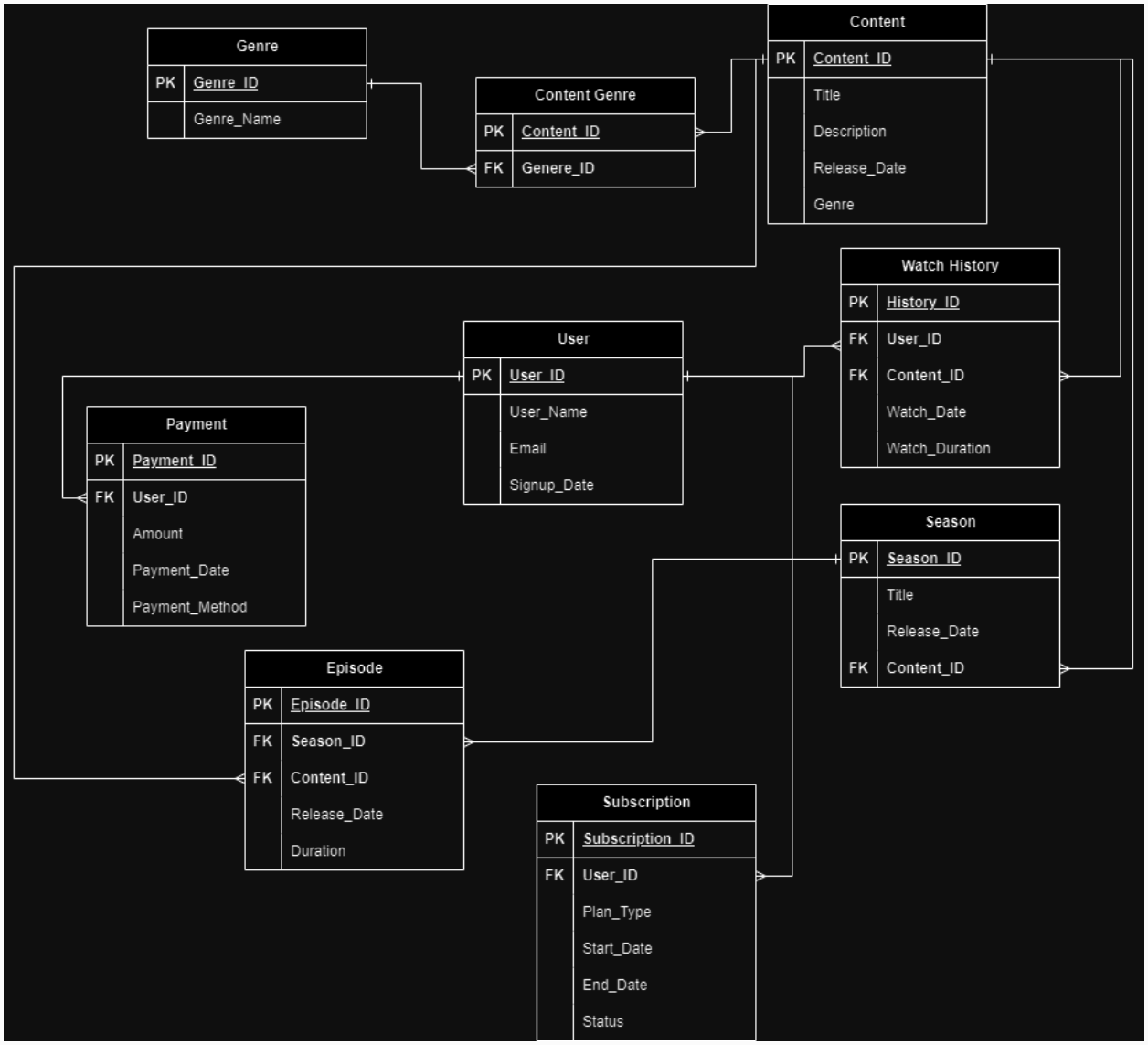
- **Immediate Insights**: By consolidating real-time data, an ODS enables quick access to current operational metrics and statuses, aiding in immediate decision-making and responsiveness to ongoing operational needs.

➢ **Processing**:

- **Real-Time Processing** is best suited for an ODS in this context, as it ensures that order data, inventory levels, and customer interactions are processed and updated immediately, enhancing operational efficiency and customer experience.

4.OTT PLATFORM

❖ER Diagram



❖ Names and Types of Dimensions and Facts:

➢ Dimensions:

- **Genre Dimension**: This dimension includes attributes related to genres such as genre ID, genre name. It categorizes content into various genres like drama, comedy, thriller, etc., which helps in analyzing content preferences and trends.
- **Content Dimension**: This dimension captures details about the content available on the platform, such as content ID, title, type (movie, series), release date, Description, and language. It helps in organizing and tracking the different types of content offered.
- **Episode Dimension**: This dimension focuses on individual episodes within a series and includes attributes like episode ID, episode number, season ID, title, and duration. It helps in analyzing engagement with specific episodes and understanding user preferences within the series.

- **Subscription Dimension**: This dimension stores details about user subscriptions, such as subscription ID, plan type (monthly, yearly), start date, renewal date and subscription status. It's essential for monitoring user subscription behavior and retention.
- **Season Dimension**: This dimension contains details related to the seasons of a series, such as season ID, title, content ID, and release date. It allows tracking the performance of different seasons and analyzing viewership trends.
- **Payment Dimension**: This dimension captures payment-related details, such as payment ID, amount, payment date, method (credit card, PayPal, etc.). It's critical for financial reporting and tracking revenue from subscriptions.
- **User Dimension**: This dimension includes user-related attributes such as user ID, name, email, signup_date. It helps in analyzing user demographics, preferences, and viewing patterns on the platform.

- **Watch History Dimension**: This dimension tracks user activity, including user ID, content ID, watch date, and watch duration. It helps in analyzing user engagement with content and identifying trends in viewing behavior.

➢**Fact**

- **Watch History Fact**: This fact table captures key metrics such as total watch hours, number of views, and engagement for content. It helps in analyzing the popularity and performance of content across different user segments.
- **Subscription Fact**: This fact table records details related to subscriptions, such as the total number of active subscriptions, renewal rates, and revenue from each plan. It helps in tracking subscription growth and user retention trends.
- **Payment Fact**: This fact table stores financial transactions, including total revenue, payment methods, and transaction frequency. It provides insights into revenue generation and payment patterns.

- **Content Performance Fact**: This fact table measures the success of content by tracking metrics such as total views, likes, ratings, and average watch duration per content. It helps in identifying the top-performing content and user preferences.

❖ Type of Data Warehouse and Processing

➢ Type of Data Warehouse:

- **Enterprise Data Warehouse (EDW):**
  - **Reason**: An EDW is suitable as it integrates data from various sources like content management, user activity logs, and financial transactions into a centralized repository. This centralization supports detailed analysis of user engagement and content performance.

➢**Processing**:

- A **Hybrid of Batch and Real-Time Processing** is ideal for an OTT platform.
  - **Batch Processing** can be used for tasks like aggregating watch histories, subscription updates, and generating reports (e.g., daily or weekly summaries).
  - **Real-Time Processing** can handle dynamic recommendations, live content tracking, and immediate user engagement data for features like "continue watching" or personalized recommendations.