**Group 7:**

**Student ID: 24280052 24280069**

**Student Name: Usmar Haider & Syed Muhammad Mujtaba**

* **This report contains:**

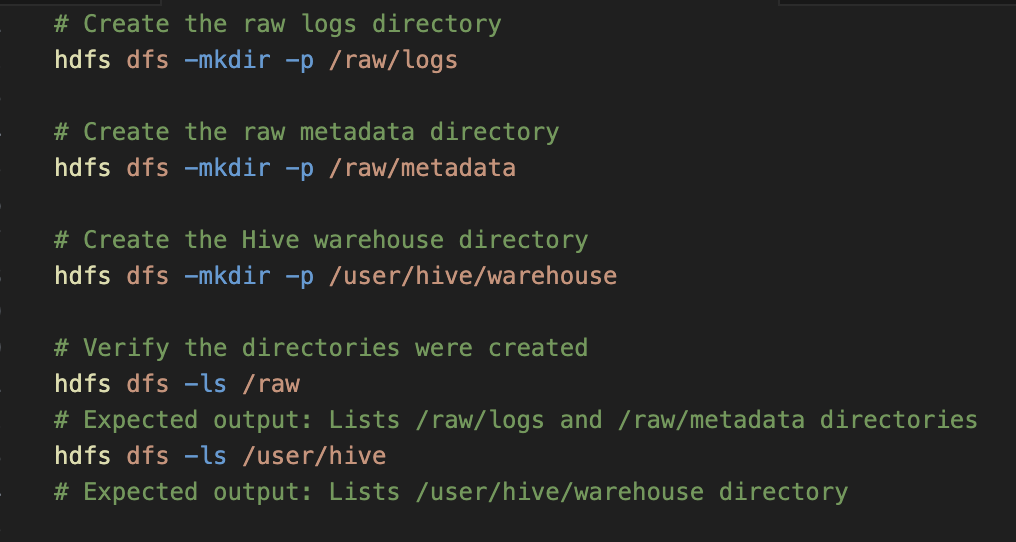
**Scripts:**

* **Step 1: Ingestion Script**
* **Step 2: Raw Tables in Hive**
* **Step 3: Star Schema**
* **Step 4: Transformation**
* **Step 5: Queries**

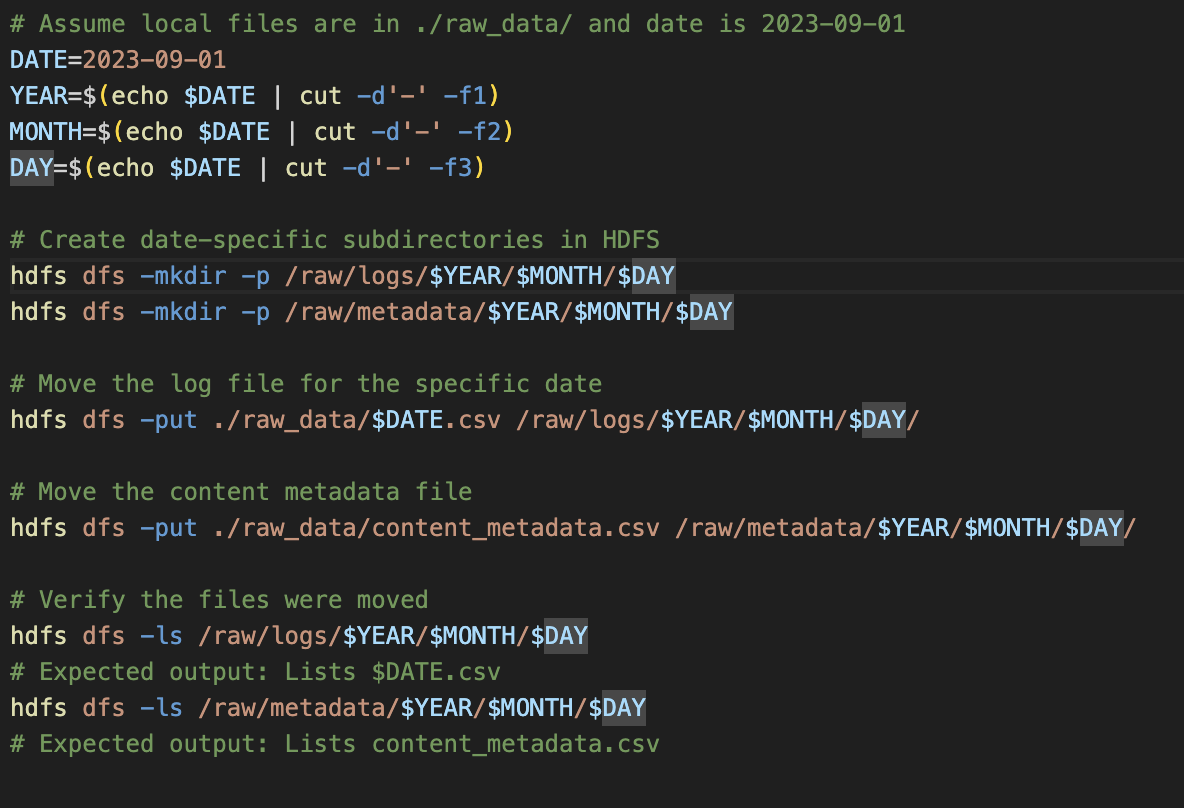
**Step 1: Ingestion Script:**

**Creating Directories in HDFS**

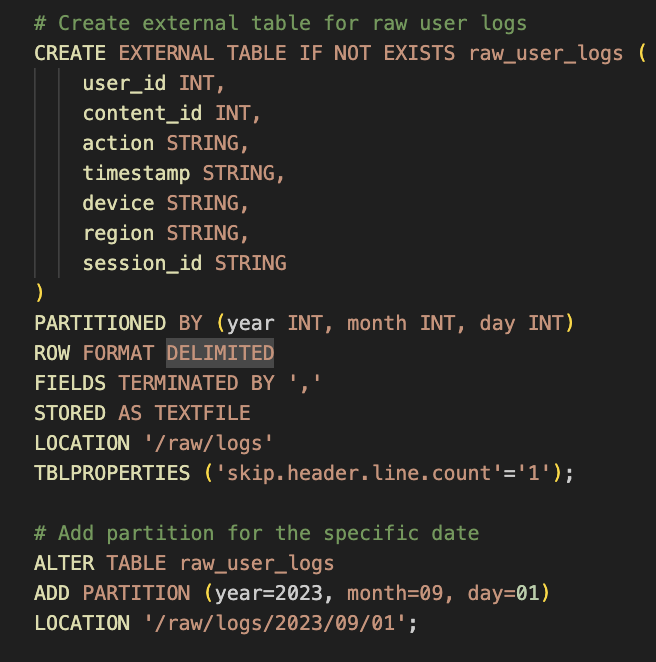
1. This code contains commands to create the necessary directories in HDFS for storing raw logs and metadata, as well as the Hive warehouse.



1. Moving Files into HDFS

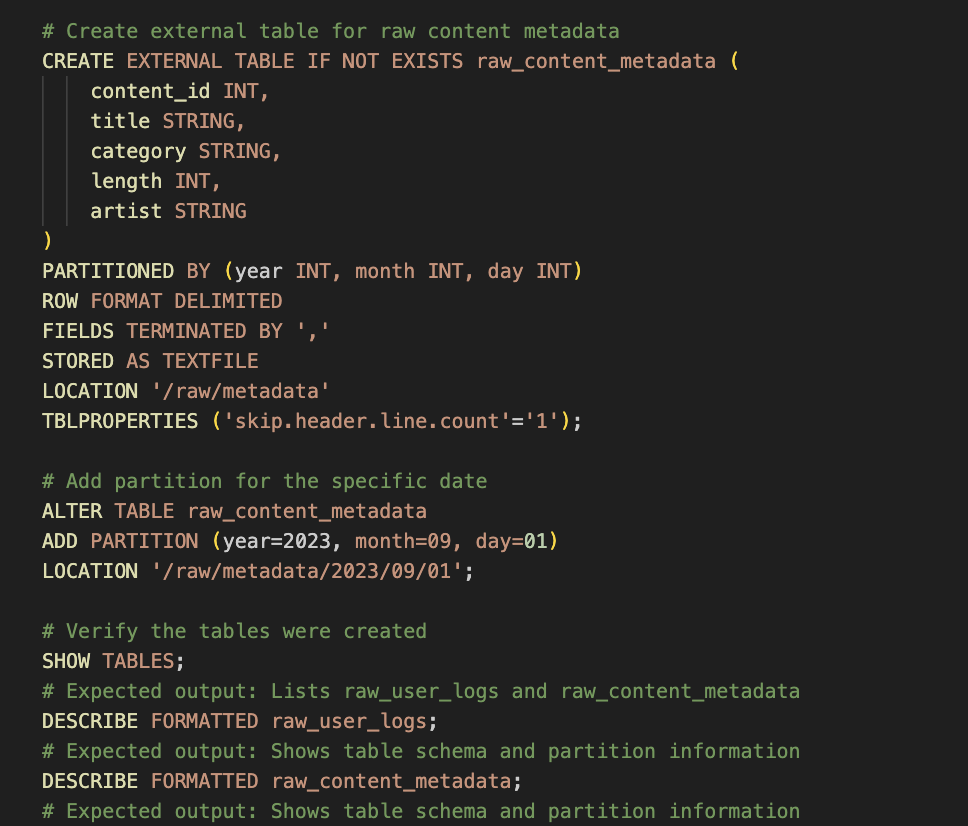


1. Creating Tables in Hive:



**Step 2:**

External Table for raw\_content\_metadata (Partitioned)

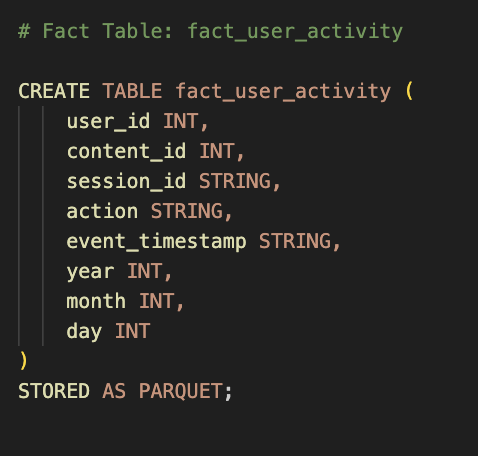


**Step 3: SQL for Tables and Star Schema**

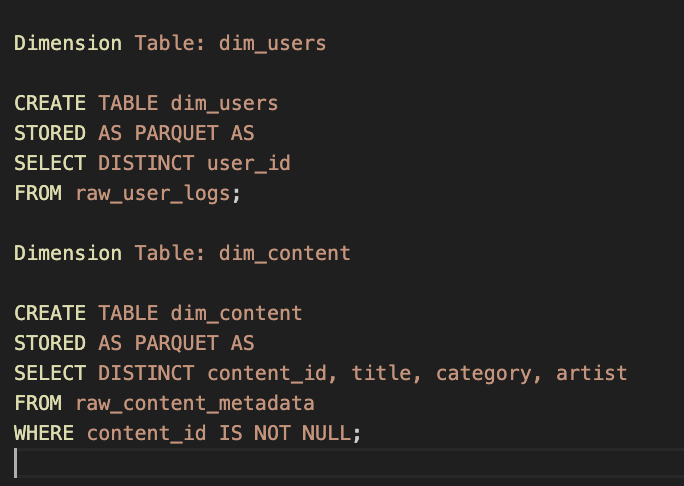
**Star schema as follows:**

* Fact Table: fact\_user\_activity
* Dimension Tables:
  + dim\_users (user-related details)
  + dim\_content (content metadata)
  + dim\_sessions (session-related details)

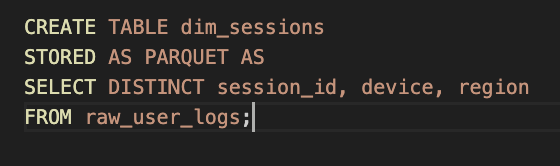
**Fact Table: fact\_user\_activity**



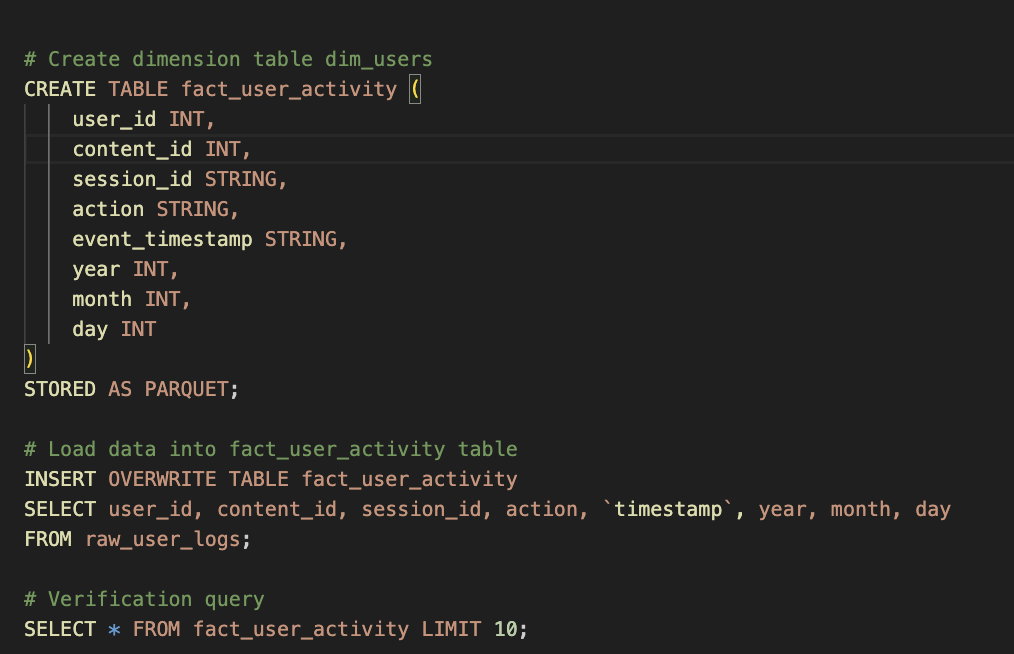
**Dimension Tables:**



**Session Table:**

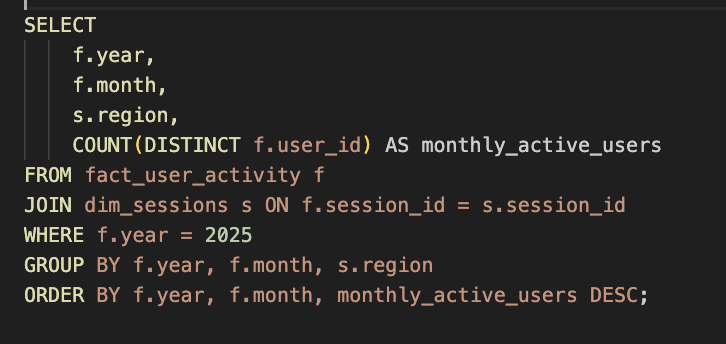


**Step 4: Fact Table & Load Data Using INSERT OVERWRITE**

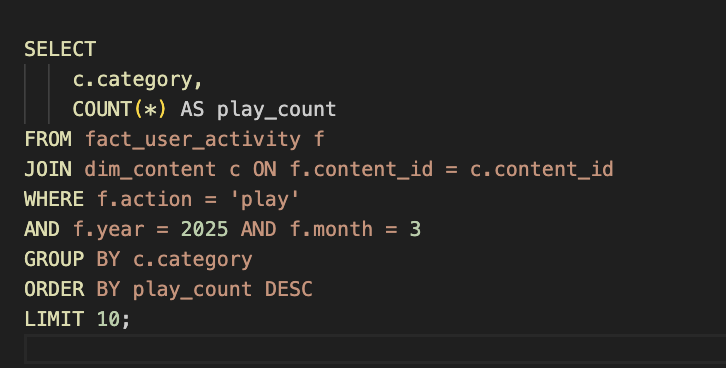


**Step 5: Queries**

**Query 1: Counts Distinct Users Per Region, Per Month**

****

**Query 2: Finds the Most-Played Content Categories**

****

**Query 3: Average Session Length Weekly**

****