

DATA ENGINEERING PLATFORMS (MSCA 31012)

sbharadwaj@uchicago.edu | jchan530@uchicago.edu

Submissions

- Team : One Submission/per team (PPT) sent to sbharadwaj@uchicago.com
- Individual : Submission on Canvas.
 - A word document containing
 - Screenshot of the EER diagram
 - SQL (DDL/DML) queries used along with any assumptions & outputs (first few rows) for each question
 - A tableau workbook with reports and dashboard.

Note: Save your brand new workbook as a .twbx before you make any data connections and then set your data connections to Extract instead of Live when you connect to data, then Tableau will automatically create the extracts for the data connections

Part A (Team): Business use case , datasets and tools relating to the final project

1. Add 3 – 4 slides to the final project presentation work completed as part of assignment 2 which addresses the following **– { 0 Points }**
 - Design considerations (include data preparation steps / platform considerations)
 - Enhanced Entity Relationship (EER) model or Dimensional model for your project

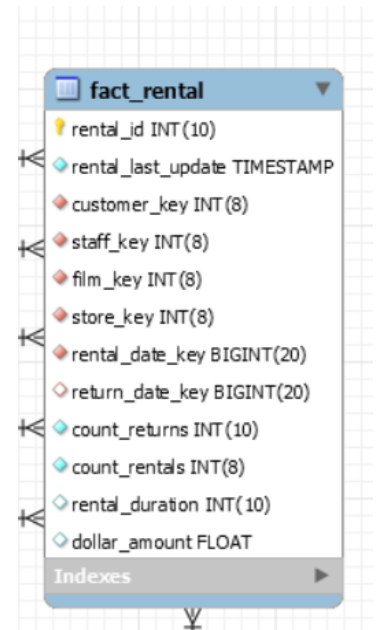
Part B (Individual): Dimensional Modeling(Sakila) , Insights

Data (Sakila dataset)

- We will use the Sakila database schema which can be found at:
<http://dev.mysql.com/doc/index-other.html/>
- Full documentation:
<http://dev.mysql.com/doc/sakila/en/>

2. Download Sakila Snowflake dimensional data model (DDL) from canvas —{ 40 Points }

- Run the DDL & DML provided for the Sakila dimension model
- For the missing Fact table **fact_rental** create the DDL and the DML scripts to pull data from the Sakila ER model. Modify/improve the fact table script if you need more metrics for the Tableau report below (For example to include the dollar amount for each rental use rental rate * number of days rented).
- Generate the EER diagram for the final dimensional model.
- Identify any scripts errors and areas of improvement in the data model.



Note :

- For “count_returns” and “count_rentals” the value in the fact_rental table would be 1. This is so that you add up all the ones if you wanted to see the total number of returns and rentals. The only case it would be a 0 is when there is no return or rental date. You can treat it as a NULL or a Zero depending on your assumptions
- The calculation for the dollar amount can be simple or complex depending on your assumptions. If you want to take a simple approach just use the formula: dollar amount = (number of days between rental date and return date) * rental rate
- If you are facing issues in inserting data into the fact_rental table, it might be because the query is timing out. The way to increase the SQL timeout settings through the Workbench.
 - In the application menu, select Edit > Preferences > SQL Editor.
 - Look for MySQL Session section and increase the DBMS connection read time out value.
 - Save the settings, quit MySQL Workbench and reopen the connection.

3. Using Tableau and referencing the snowflake dimensional model (Sakila data warehouse) create 5 different reports that provides insights into the sakila dataset. You may use any of the below use cases or you can create your own reporting use cases/Insights:— { 40 Points }

- Create a report that helps visualize list of movies titles grouped by categories
- Create a report that provides the list of overdue rentals grouped by customers

- c. Create a report that list the Labels and language for all DVDs
 - d. Create visualization with payments for each customer grouped by city
 - e. Create charts that gives insight into the number of customers, ranged by number of rentals made
 - f. Create charts that gives insight into the number of rentals for each store and name of store manager
 - g. Create table that lists which 3 actors' films have the highest demand for rentals
 - h. Build charts that gives insight into the number of rentals per month for each store
2. Create a dashboard based on the report set that you have created (For example: Store level dashboard that shows inventory, number of customers per store, revenue, etc.)
- Note: Make sure the context within the dashboard can be switched using filters.

– { 20 Points }