

Project 1: title of your project

Name (EntryNumber), Name (EntryNumber), Name(EntryNumber)

Due date: April 4, 2021, 11:55pm IST

Please name your report correctly: `< Name1-EntryNumber1-Name2-EntryNumber2-Name3-EntryNumber3>.pdf`
Your report should be structured as follows:

1 Section 1

A description of your project including motivation and how you envision the complete system The ER diagram for your application List of entities and attributes in an easy-to-read table

2 Section 2

1. State the sources for your data.
2. How you downloaded it ("Readymade" vs scraped)
3. Cleanup steps
4. Statistics (Include a table which lists the name of the table, number of tuples in the table, time to load, size of raw dataset, size of raw dataset *after* cleanup.)

3 Section 3

User's view of the system (this should be an itemized list)

1. I need at least a couple of lines of explanation of what the *user* sees. For example: "Movie Listing": User enters the minimum rating and the genre of interest in the search form, the application returns a listing of movies which satisfy these constraints.
2. System view (this should be an itemized list):
 - Couple of lines of explanation for any special functionality that you have. For example: View update triggers: View V1 materializes 5 star movies from the Movies base table, trigger 'five-star' comes into play whenever a 5-star movie is inserted into 'Movies'. Another example: For security purposes, we grant permissions on table 'Movies' only to admins. One more: We built indexes on the following tables/attributes.
 - List of all queries that will potentially be fired (many of these will be based on inputs from the user). This needs to be synchronized with 1. and 2. For each of the item in 1. and 2., write the corresponding SQL query.
3. List of queries you have tested, and sample run times. You can choose what you would like to report here. For example: For Movie Listing: Input by user: 3, Horror, we ran the query on our database and the results were returned in 2ms. Ideally, this will be a table consisting of the query number (from 3.), parameter values, run times.