Computer Science Department

University of Computer & Emerging Sciences (FAST-NU)

HOME WORK ASSIGNMENT COVER SHEET

COURSE TITLE DATABASE SYSTEMS COURSE CODE CS219

INSTRUCTORS Teaching Team TYPE [ ] Indiv. [√] Group (Please tick)

ASSIGNMENT NO 2

ASSIGNMENT Understanding "Netflix Movies and TV Shows" Dataset

HAND OUT DATE 16-April-2021 DUE DATE 26-April-2021

Aim and Rules	% Mark
The purpose of this assignment is to improve your skills of database programming, selection of keys, and report writing skills. Submission: Only soft copy in pdf is required. Any mistake or missing material will lead to the reduction of 5% marks. **Group of max 3 students is allowed, individual work and group of 2 is acceptable in extreme situations and with prior written (by email) permission of theory course teacher. Such permission should be sought at least 5 days before the deadline of the assignment.** **Any type of plagiarism will lead to 0% marks of both/all parties.** **Cross Section Groups are not allowed.**	

TO BE COMPLETED BY STUDENT (TEAM LEAD)

Roll # and Name _____

Section _____

GROUP MEMBERS ID

ID	Sec#
ID	Sec#
ID	Sec#

DECLARATION: We/I declare that this Coursework is our group's own work

SIGNATURES (All members)

GRADE/ MARK AWARDED [ ] COMMENTS _____

INSTRUCTOR'S SIGNATURE _____ DATE _____

Detailed Description:

"Netflix Movies and TV Shows" Dataset is available at below URL as a csv file.

https://www.kaggle.com/shivamb/netflix-shows

It contains 12 columns and size of the file is 2.86 MB. The purpose of this assignment is listed as steps below:

1) Read and Understand the details of the dataset on given URL.
2) Plan the entities, their keys and attributes
3) Create a database in Oracle, create tables for all the identified entities having keys and attributes.
4) Download the dataset and load into the created tables. You can use SQL*Loader utility for this purpose. It is a very simple and easy to use utility which can load the data from csv file(s) into Oracle database.
5) Then apply 18 different SQL queries to get various insights of the dataset.

Deliverables:

A report containing following items:

1) Snapshot of the Database Schema designed by your group. The database should contain 3 or more tables. A detailed description of the database schema is required so that a reader could know the idea behind creation of each table.
2) Keys and attributes of each table should be clearly shown along with their datatypes.
3) SQL queries to create database, tables, keys, and attributes.
4) Code of SQL*Loader to load the data into the created tables.
6) SQL queries and their result along with the 2-3 liner description which tell the purpose of each query and tells what information we get from each query and obtained results. You can create queries of your own but they should be your own queries and should not be copied from other groups. The count of 18 queries is divided as below:
    i. 3 Queries which use various aggregation functions
    ii. 3 Queries which use various aggregation functions and involve more than 2 tables (database join)
    iii. 3 Nested queries
    iv. 3 queries involving outer joins and aggregation functions
    v. 3 queries involving set operators.
    vi. 3 queries to compare the difference in execution time of different types of joins like outer join vs equi join etc. The execution time must be obtained from the system and should be displayed along with the output.

Note:

1) A group is likely to get more marks based on nicely written report, well presented material, and technically sound queries.
2) Learn SQL* Loader yourself. It is very easy to learn and use.
3) Individual work or group of 2 is acceptable in extreme situations and with prior written (by email) permission of theory course teacher. Such permission should be sought at least 5 days before the deadline of the assignment. The reply from the course teacher must be placed as a screenshot in the report to show the permission.