**CSEE5590-0001/490-0003: Big Data Programming**

**Lesson Plan # 4**

**ICP Feedback and Submission Link :**

[**https://forms.gle/wyFHFBL6LzM3xDGr5**](https://forms.gle/wyFHFBL6LzM3xDGr5)

**For Online students:**

[**https://forms.gle/wyFHFBL6LzM3xDGr5**](https://forms.gle/wyFHFBL6LzM3xDGr5)

**Lesson Title: *Hive***

**Lesson Description: *Hadoop Dependent Query Based NoSQL Database Hive***

**Lesson Overview:**

*Hive is a data warehousing system to store structured data on Hadoop file system and provides an easy query these data by execution Hadoop MapReduce plans. In this exercise we will learn basics of Hive QL.*

**In Class Exercise:**

1. Create Hive Tables and Perform Queries for Use Case based on Petrol Data. See the Slides for details.

Dataset:

<https://umkc.box.com/s/7umnwj8u6rhtmegrqg13vgf1qs4qdtw0>

1. Create Hive Tables and Perform Queries for Use Case based on Olympics Data. See the Slides for details.

Dataset:

https://umkc.box.com/s/f918eea7k6mw6h7qiwj4b8im97c6hy84

1. Create Hive Tables and Perform Queries for Use Case based on Movielens dataset which has 3 datasets as movies, users and ratings.

Dataset:

<https://umkc.box.com/s/m3i7oabkj00boxuiskv5d4aoklh85w3x>

Perform following tasks:

1. Create 3 tables called movies, ratings and users. Load the data into tables.
2. For movies table:

* List all movies with genre of movie is “Action” and “Drama”

1. For Ratings table:

* List movie ids of all movies with rating equal to 5.

1. Find top 11 average rated "Action" movies with descending order of rating.

( Hint: Need to perform join operation on Movies and Ratings table)

Bonus:

List all the movies with its genre where the movie genre is Action or Drama and the average movie rating is in between 4.4 - 4.9 and only the male users rate the movie.

( Hint: Perform join on three tables users, movies and ratings)

You can refer following document for reference:

<https://umkc.box.com/s/1dcugk08caqzitgqvrthiqe5n6sgznd5>

**ICP Guidelines (In Class Students):**

1. ICP Submission is in pairs of two students.
2. Once completed, must be presented to TA or Instructor before the completion of the class
3. Submission after class is considered as late submission. (Check the late submission policy in the syllabus)

ICP Code with brief explanation in wiki should be pushed to GitHub (submit your screenshots as well. The screenshot should have both the code and the output)Submit GitHub Link through the Feedback Form ([**https://forms.gle/wyFHFBL6LzM3xDGr5**](https://forms.gle/wyFHFBL6LzM3xDGr5)**)**

**Submission Guidelines (for online students):**

1. Submit your source code and documentation to GitHub and represent the work through wiki page properly with detailed explanation (submit your screenshots as well. The screenshot should have both the code and the output). Also submit your wiki to Turnitin where plagiarism should be less than 15%.
2. Comment your code appropriately.
3. Submit a brief demo video 2-3 min showing your assignment with a voice over explaining your work through the Submission Link.
4. Use the following Google link to submit your assignment

(ICP Submission Link#):  
 [**https://forms.gle/wyFHFBL6LzM3xDGr5**](https://forms.gle/wyFHFBL6LzM3xDGr5)

***Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL:*** [*https://catalog.umkc.edu/special-notices/academic-honesty/*](https://catalog.umkc.edu/special-notices/academic-honesty/)