CSEE5590-0004/490-0004: Big Data Programming

Lesson Plan #1

ICP Feedback and Submission Link:

https://forms.gle/xMAmr3zATrtMG5cX7

Lesson Title: Installation/Cloudera / Hue

Lesson Description: *Installation of Cloudera and visualize data with Hue*

Lesson Overview:

- Evaluation Criteria
- Topics to cover in Module 1
    - Cloudera (configuration/Hue)
    - HDFS/ MapReduce and Big Data Applications (2)
    - Hadoop Dependent Query Based No SQL Database Hive
    - Hadoop to SQL Parallel Transfer Engine: Sqoop
    - Parallel Indexing: Solr & Lucene
    - Independent Column Based No SQL Database: Cassandra
- Cloudera Installation
- In class Programming

Lecture and Installation Guideline:

https://umkc.box.com/s/p8dhb8jat5gm7h580wq6kxohj4ixaxdu

In Class Exercise

Dataset1: https://umkc.box.com/s/npk8aaq0wmoz1fttwx4i649qw65cp2x9

Dataset2: https://umkc.box.com/s/z4tfp9yz4uhsrrodemlos8dfxlun7yky

Datasets are also provided in the GitHub repo you clone.

Visualization of Hadoop file with Hue

You are required to follow the steps below to complete your ICP today

- Use the given dataset
- Load it in hadoop hdfs
- Use the second file
- Append it to the first file
- Visualize file with Hue
- View the first and last lines (approximately 5) of merged dataset using appropriate hdfs commands
- Create a new text file and load it into hdfs and try to append all three datasets.

After completion of your ICP fill in the form. Any available TA/instructor will come to you and evaluate ICP.

ICP Guidelines (In-Class Students):
1. ICP Submission is in pairs of four students.
2. Once completed, must be presented to TA or Instructor before the completion of the class
3. Submission after the deadline is considered as late submission. (Check the late submission policy in the syllabus)
4. ICP Code with brief explanation should be pushed to GitHub.
5. Submit a demo video 2-3 min showing your assignment with a voice over explaining your work if you are unable to complete ICP within the deadline due to genuine reason.
6. Provide the video submission link through the submission form
   https://forms.gle/xMAmr3zATrtMG5cX7