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

**Lesson Plan #1**

**ICP Feedback and Submission Link :**

**https://forms.gle/wyFHFBL6LzM3xDGr5**

**For Online students:**

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

**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/ums4eav9ms5ejlyo7m5qnruhq917zafr>

Dataset2: <https://umkc.box.com/s/gwha5sfufjat6o1cg2wcbturni9fsvmr>

**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 three/four 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)
4. ICP Code with brief explanation should be pushed to GitHub. Submit Github Link through the Feedback Form (**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 (submit your screenshots as well. The screenshot should have both the code and the output)
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**

***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/)