UE18CS322: Big Data

D. Sudaroli Vijayakumar

Project Guidelines and Sample Titles

Guidelines:

- Objective of this project is to get exposed to various tools and technologies of Big data Framework.
- Since this course is all about data engineering, it is essential to get exposed to the data pipeline.
- Depending on the problem statement chosen to try to establish the data pipeline.
- You must get exposed to the distributed computation and storage, practically how it's done should be understood as a part of the project.
- Individual and team projects are allowed but I restrict the number of team members as 3.
- As a guideline I am sharing some project titles that could be considered as a Big data Project.

S:No	Project Title	Technologies
1	Simulate a complex real-world data pipeline with Covid dataset.	HDFS, HIVE, Spark, Airflow, Kafka etc.
2	Server log data processing in Hadoop	Flume, Pig,Spark ,Hive etc
3	Track a car in real time using Spark redis	Flume etc
4	Analysing public response of any event with spark streaming	HDFS, HIVE, Spark, Airflow, Kafka etc.
5	Handling unstructured data using spark	Spark
6	Predicting train delays using Spark and Kylin	Hadoop, Spark etc
7	Event data analysis using AWS stack	AWS
8	Analysis of a dataset using Hive	Hive

I also suggest you go through the links below to choose titles:

1. https://www.upgrad.com/blog/big-data-project-ideas-beginners/
2. http://hadoopproject.com/big-data-project-ideas-for-students/
3. https://www.quora.com/What-are-some-mini-projects-in-big-data-that-I-could-do-with-a-team-of-two-or-three-people-in-a-month-or-two-I-am-a-third-year-computer-science-engineering-student-from-India