BigData Analytics

**Introduction:**

This is a computer-based practical paper for BigData Analytics subject. The practical will be conducted leveraging Hadoop environment In Ubuntu.

**Instructions:**

1. **All questions are compulsory**
2. **All names used as part of these questions are fictional and have no relevance or correlation to any real-world person or objects.**
3. **Save your shell command code/Queries in a text file as well.**

**Activity:**

**Question 1: (15 marks)**

There is a text file containing information of about audit informations. The audit log contains (audit\_date, audit\_time, auditor\_id, auditor\_signatory, audit\_observation, audit\_category, audit\_severity, audit\_status) and auditor details contains (auditor\_id, auditor\_name, auditor\_designation, auditor\_phone, auditor\_supervisor\_id).

Note : Auditor\_Signatory is none but the audit supervisors. Audit status can be “open”, “closed”, “accepted”,”rejected”, “inprogress”. Audit severity is between 1 and 4, 1 being most critical and 4 being normal.

Implement the following queries in Hive.

1. Write queries to load the audit and auditor details.

2. Write a query to identify all auditors who have not done any audit.

3. Write a query to find out the number of audit observations per category and severity.

4. List all audit\_date, audit\_observation, auditor\_name, audior\_supervisor\_name whose audit status is closed.

5. List all the audit information done by auditor whose name contains “AK” or “MAK”

**Question 2: (15 marks)**

There is a flight information file available to you. You are required to act on the same using Pig

1. Load flight data into Pig.

2. Find out the airline, flight number, source and destination airport for flights on day 5.

3. Find out all flights whose flight number contains “NB” or “JB”

4. Get the count of flights based on the source airport

5. Get the flights which has departed before time.