**NAME: Ashish Bisht**

**STUDENT ID: 801168390**

**Cloud Computing for Data Analysis**

**VIDEO CASE 03 : Cloud Tools – Pig, Hive and HBase**

**Video Case Questions:**

• Briefly explain the execution steps followed by Pig.

ANSWER:

Initially Pig uses mappers and reducer techniques to analyse large data sets.

Stages of execution followed by Pig:

1: Loading the data

- The Pig loads the data needed to manipulate the HDFS.

-It can load files or all the files in the directories.

2: Grouping and filtering

- Pig sends the data to the mappers and reducers so that the data is transformed enormously and the output is reduced.

3: Dump and Store

-Finally, load the results into program and store the results in a file system.

• What is the purpose of Hive? Mention some of the advantages of Hive.

ANSWER:

Hive is similar to that used for batch processing in Hadoop. In order to give SQL access to Hadoop data for data analysis, Hive was introduced by Facebook developers. It supports queries which are expressed in hive language also called as HiveQL. HiveQL supported custom MapReduce scripts to plug into queries. It is designed to handle huge amount of data (peta bytes). The main functions handled by the Hive are data summarisation, query and analysis.

Advantages of Hive:

- Multiple users can query data simultaneously.

- Higher level inquiry language and rearranges working with a lot of information.

- Lower expectation than Pig or MapReduce to absorb the information. HiveQL is much closer to SQL than Pig, and more stable than Pig.

- It has a standard based analyser for streamlining consistent plans.

- It supports partitioning of information at the degree of tables to improve execution.

- Metadata store is a major addition in the design which makes the query simple.

• Give some similarities of architectures of HBase and HDFS and MapReduce.

ANSWER:

-HBase ,HDFS and MapReduce have master slave architecture.

- HDFS is a Java-based document framework that can be used to store large collections of information and HBase applications using Java.

- HDFS and Hbase have advanced data storage and management systems.

- All three architectures are able to process structured and unstructured information.