**Name: Nikhil Jadhav**

**Student ID: 801075504**

**Cloud Computing for Data Analysis**

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

Watch following videos:

**Video 1:** <https://youtu.be/rxnXHlaSohM>

**Video 2:** <https://youtu.be/uY7Rr7ru9E4>

**Video 3:** <https://youtu.be/kN01ELCAsn8>

Video 1, 2 and 3 gives you a basic knowledge about Pig, Hive and HBase respectively

**Video Case Questions:**

1. Briefly explain the execution steps followed by Pig.
2. What is the purpose of Hive? Mention some of the advantages of Hive.
3. Give some similarities of architectures of HBase and HDFS and MapReduce.

Answers:

1. Following are the execution steps followed by Pig.
   1. Load the data which you want to manipulate from HDFS using the LOAD command.
   2. Run the program using set of transformation (GROUP, FILTER) which you want to apply on the data.
   3. This transformation under the cover are converted into set of Map-Reduce programs.
   4. Finally dump the result into screen or store the result into the file using the DUMP or STORE command.
2. Hive is often used because of its SQL like query language is used as the interface to an Apache Hadoop based data warehouse. Hive is considered friendlier and more familiar to users who are used to using SQL for querying data. It also eliminates the requirement to learn new language as it is similar to SQL querying language.

Following are the advantages of Hive:

* 1. Lower learning curve than Pig or MapReduce – HiveQL is much closer to SQL than Pig – Less trial and error than Pig
  2. Higher level query language – Simplifies working with large amounts of data
  3. Enables serialization and deserialization.
  4. Batch job processing on large datasets.
  5. Supports partitioning of data at the level of tables to improve performance.

1. Similarities between HBase, HDFS and MapReduce is as follows:
   1. HBase is non-relational, non-SQL database that runs on top of HDFS.
   2. HBase, HDFS and MapReduce are all written in Java.
   3. MapReduce and HBase are highly available and accessible.
   4. HDFS and HBase are capable of processing structure and un-structured data.