**Cloud Computing for Data Analysis**

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

**Student ID: 801104903**

**Student Name: Balasundaram Avudai Nayagam**

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. **Step 1:** The input data is loaded into the HDFS using the ‘Load’ command.

**Step 2:** The input data is processed to extract the required columns. This is done by executing the program as per the map function implementation and they are converted into mapper and reducer tasks.

**Step 3:** With the help of ‘Dump’ command, the output is displayed and the result can be stored using the ‘Store’ command.

1. What is the purpose of Hive? Mention some of the advantages of Hive.
2. Hive is a Hadoop based tool used commonly to perform analysis on the data.

It is less complex to implement as it is easy to write map-reduce programs in it.

**Advantages:**

* Easy to implement and use
* Highly scalable
* Provides external tables for data storage
* Supports data partitioning

1. Give some similarities of architectures of HBase and HDFS and MapReduce.
2. - HBase, MapReduce and HDFS are Java based implementation.

* HDFS and HBase can deal with both structured as well as unstructured data.
* HBase, HDFS and MapReduce all three follow master-slave architecture.