

Assignment #1

By

Mohamed Hosny Ahmed

Hisham Gamal El-din

Mohamed Salah Eddin

Mysql

- Create database.
- Create user with specific Ip and privileges .
- Remote Connection.
- Simple Queries.

VM

- Start all services:
 - Start dfs:
 - Namenode:
 - keeps track of where data is physically stored
 - Datanode:
 - HDFS cluster data blocks are replicated across multiple data nodes and access is managed by the NameNode, keep track by rack ID.
 - Start yarn.
 - Job scheduling and tracking.
 - The ResourceManager is a master service and control NodeManager in each of the nodes of a Hadoop cluster

Sqoop

- Tool that offers the capability to extract data from non-Hadoop data stores. Transform the data into a form usable by Hadoop, and then load the data into HDFS. This process is called ETL.
- ✓ **Bulk import:** Sqoop can import individual tables or entire databases into HDFS. The data is stored in the native directories and files in the HDFS file system.
 - ✓ **Direct input:** Sqoop can import and map SQL (relational) databases directly into Hive and HBase.
 - ✓ **Data interaction:** Sqoop can generate Java classes so that you can interact with the data programmatically.
 - ✓ **Data export:** Sqoop can export data directly from HDFS into a relational database using a target table definition based on the specifics of the target database.

Mining Big Data with Hive

- Built on the core elements of Hadoop (HDFS and MapReduce).
- Simple Queries by HiveQL.

HBase

- Start Hbase:
 - Columnar databases.
 - Use hdfs & map-reduce engine for core data storage, real-time read/write data access.
 - “version” attribute, which is nothing more than a timestamp uniquely identifying the cell.
 - Columns in HBase belong to a column family.
 - The schema is defined and created before any data can be stored.

MapReduce

- MapReduce was designed as a generic programming model.
- The real power of MapReduce is the capability to divide and conquer.
- Take a very large problem and break it into smaller, more manageable chunks, operate on each chunk independently, and then pull it all together at the end.