Task 1: Explain the below concepts with an example in brief:

* Nosql Databases

A NoSQL database supports processing semi-structured or non-structured data. NoSQL databases support horizontal scaling with ease because it is designed for clustering on commodity servers. Due to horizonal scaling NoSQL databases are ideal to support larges arrays of data that scale to the petabyte range over a cluster of commodity computers.

* Types of Nosql databases

Types of NoSQL databases include key-value store, document-based store, column-based store, and graph-based. The key-value store uses a hash table of keys and values to store the information. Since key-values do not enforce any structure on the data, then it is effective for modeling data that align with the business requirements. The document-based store stores document made of tagged elements. Document-based data store uses documents as an indexed storage structure. The Column-based store is blocks of data stored in only one column. Columns-based databases store the data in a column format. The columns themselves are key-value pairs and can accommodate unstructured data. The graph-based store uses a network representation based on nodes and edges to represent stored data. With a graph-based store the database engine is capable of iterating quickly in any direction through the network of nodes and edges.

* CAP Theorem

The CAP theorem is defined as a triangle connecting C-consistency, A-availability, and P-partition tolerance. Consistency is defined by all users have the same view of the data at any time. Availability is defined by all clients having access to read or write the data at any time. Partition tolerance is defined by the environment continues to function despite the availability of systems on the network. The CAP Theorem indicates that designers must focus on two components of the CAP paradigm C-A, C-P, or A-P.

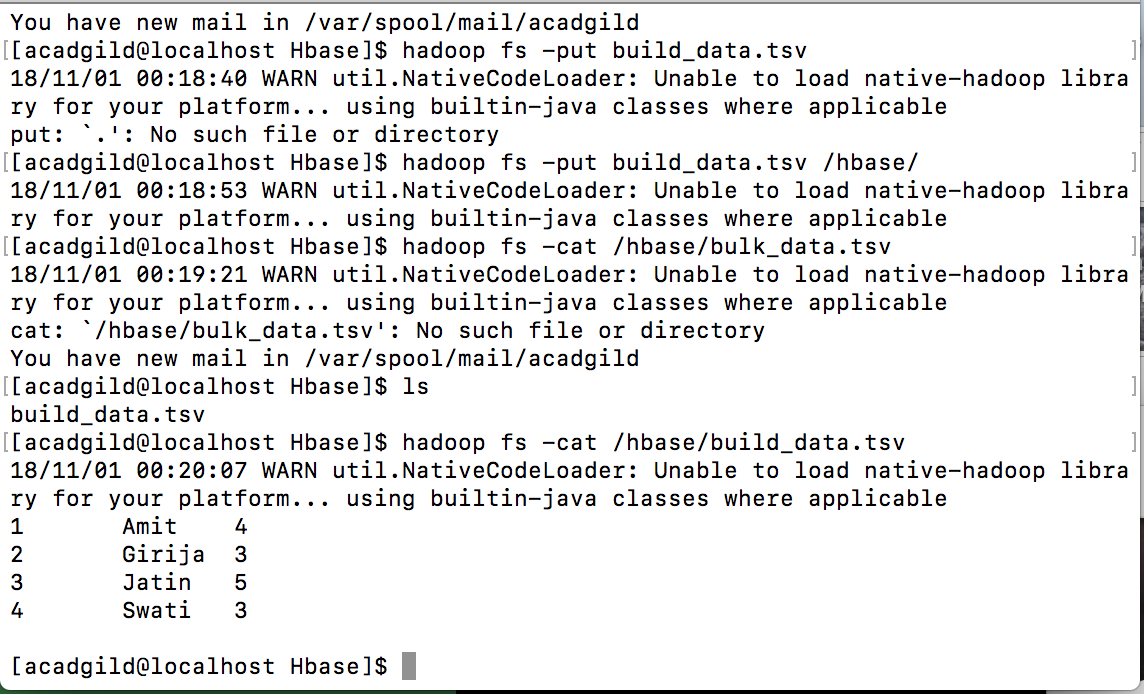
* HBase Architecture

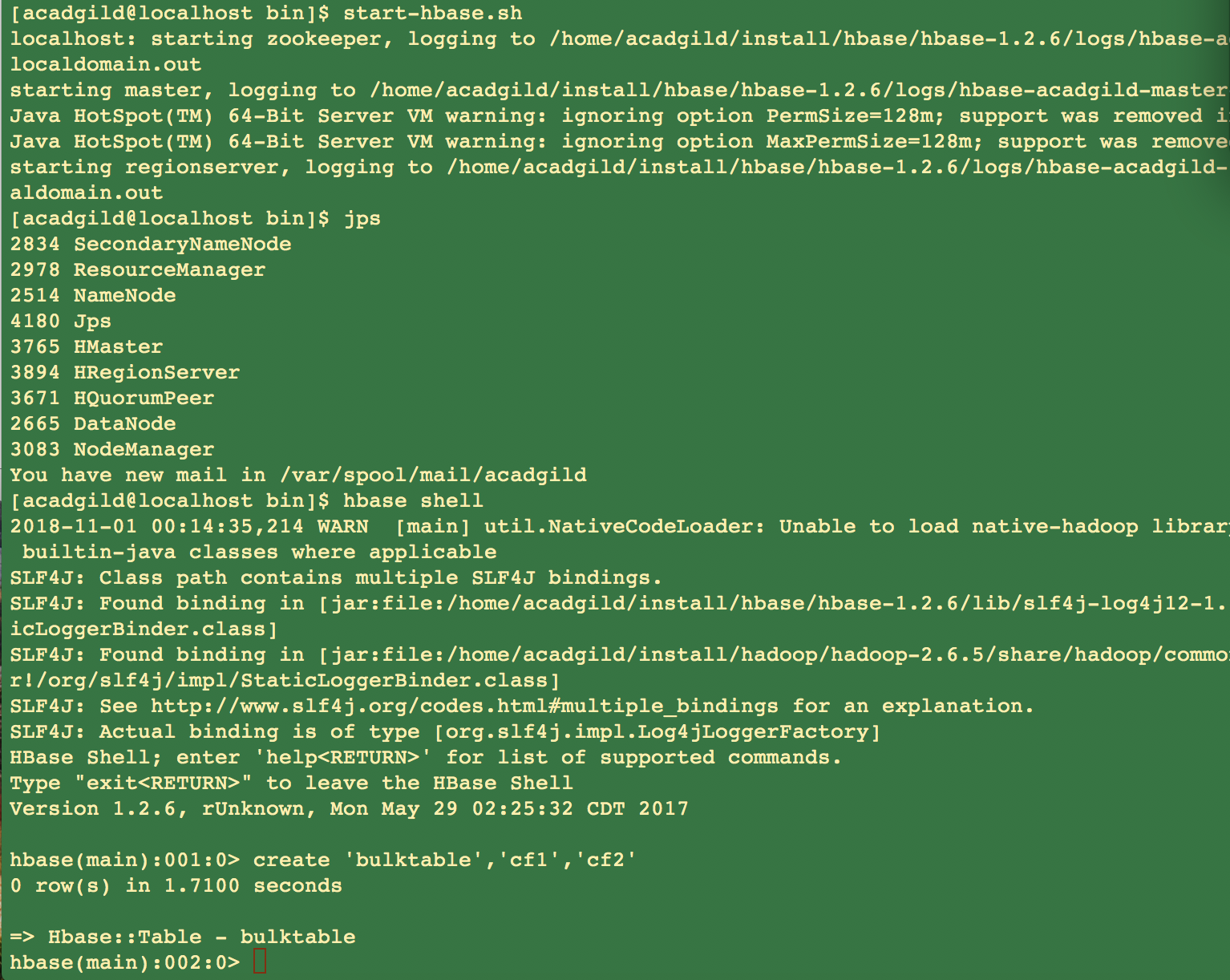
HBase architecture is built on top the HDFS environment. HBase includes region servers that rest on the data node within the Hadoop environment. The Region Server manages the data that is stored on the data node. The data on each data node is stored in HDFS files called the HFile. The Region Server is managed by the HMaster. The HMaster is responsible for identifying and assigning regions on startup, coordinating with Region Servers, and loading balancing between Region Servers. The location of the regions is stored in a HBase Catalog table called the META table. The location of the METE table is Stored in the ZooKeeper. The ZooKeeper is the resource manager for the HBase environment. Clients connect to the ZooKeeper to identify the location of Region Server that hosts the data they require.

* HBase vs RDBMS

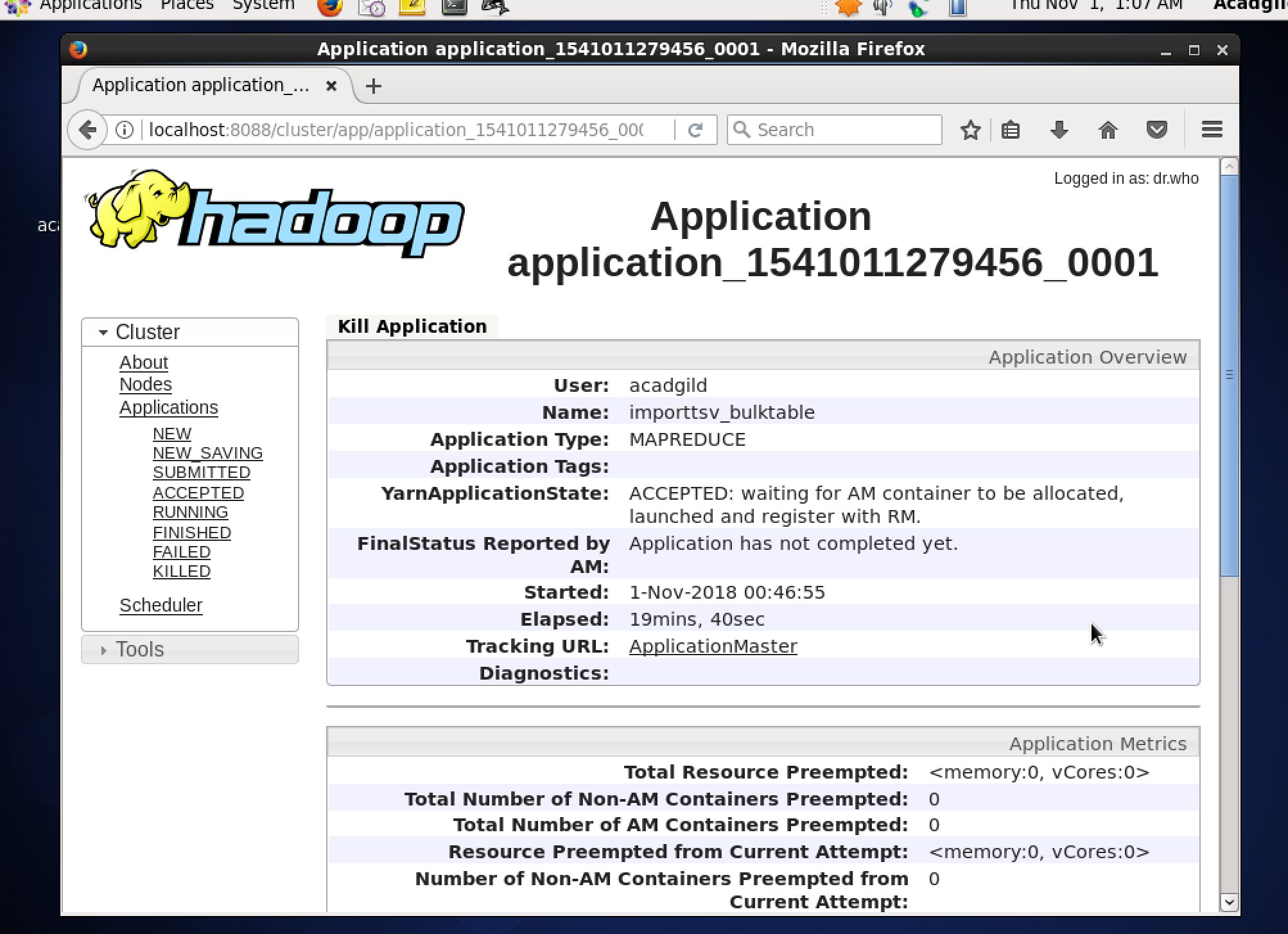
NoSQL databases are different from the typical relational database management system (RDBMS). One difference is NoSQL databases do not require joins to link multiple tables for the desired result whereas a RDBMS database requires the data to be structured in a table format (rows and columns). Tables in a RDBMS database are joined as needed to produce an output that meets the business object. Another difference is horizontal scaling. Horizontal scaling is a problem with RDBMS databases because it requires the data to spread across a number of serves to obtain the breath of scaling. as data becomes less structured and larger in size, RDBMS databases become less efficient. NoSQL overcomes the data size problem by using a clustering design pattern and the non-structured-data problem because of how data is stored within the database. NoSQL databases have the advantage of supporting large volumes of information that might be less structured than RDBMS.

Task 2: Execute blog present in below link









Seems job never completed.