Solution Assignment 2

1. **Draw a top down structure for the various elements in Cassandra?**

   Cluster

   → Key Space(s)

   →Column Family (ies)

   →Row Key(s)

   -->Column(s)→Value and Timestamp.

   In Casandra top down structure, at top it starts with a cluster, which is collection of nodes that form a ring. This ring works on peer-to-peer and gossip protocol.

   Under a cluster, Cassandra has a group of key spaces. To make an understanding, this can be thought of analogous to a database in RDBMS term.

   Then a key space contains group of column family, which is analogous to a table in RDBMS term. This analogy is very loose, though.

   Finally, a column family can contain one or more rows. These rows can contain one or more columns. The column will have a value and a timestamp.

2. **Why is it not a good idea to have multiple applications per key space? If in case you need it when can you have it?**

   Maintaining more than one key spaces is not advised as it requires extra management for Cassandra in order to provide extra maintenance. In addition, there is no functional benefit with regards to the application data model.

Only time you need when you have multiple replication factor and replication strategy for different tables. If you need, you can have it as a cluster can have multiple key spaces.

3. **Please open the web page www.amazon.co.uk and specify what the use cases are where Cassandra can be used?**

Display of inventory and Search

- Amazon may leverage Cassandra to populate its inventory and mange search on the page. The reason is that the page needs to be highly available and needs to deliver higher performance.

- In addition, partitioning and indexing (primary and secondary) can help in higher response time as data from one big table is stored on multiple nodes based on its key. This partitioning-- where records of a table are stored on multiple nodes—help improve both write/read response. Since Cassandra architecture provides default weak consistency --to avoid to get a lock—this also helps in improving overall read performance.

- Cassandra peep-to-peer architecture also helps in delivering better read/or write response as any node in the cluster can take the responsibility of read or write.

4. **Create a model for the column family for any of the use cases created above?**

ItemCategory<<CF>>

<<RowKey>ItemCategoryId
Category_Name
Category_Description

Item<<CF>>

ItemId <<RowKey>>
ItemCategory
ItemName
ItemDescription

ItemPrice
ItemImage
ItemWeight
ItemDimension
ItemRank
ItemReview
ItemLocation
ItemSeller
ItemCount

5. **Please specify any Super columns which could be used in the use cases above?**

   ItemCategory can be used as Super column.

6. **Please specify what Indexes are and let us know where did you use the indexes and the examples created above?**

   Many secondary indexes need to be created to improve the performance. The columns which can be used as a secondary index – item name, item category, item seller.