* Hbase is a schema less database, what does it mean?

In traditional RDBMS database the schema of the table is predefined and cannot and be modified easily. The schema is defined while making of the table i.e the datatype of the column. But in case of the HBASE the schema is mentioned at time of loading the data. Also the data stored in HBASE is using the bytes hence it is datatype independent.

In this way even if new data comes into the database HBASE can handle it smoothly without any glitch.

This is one of the pros of HBASE. Also the altering of the records is possible easily with help of the schema less database.

Also considering the data being stored by HBASE is huge amount of unstructured data hence NoSQL feature of the schema less design is very much useful.

Hence HBASE is called schema less design.

* What is the minimum number of column family every Hbase table should have?

Minimum number of column family every HBASE table should is 1. Because HBASE stores the name of the column under the column family only. Also one reason is that HBASE has schema less structure and hence for creating a table it is created with only a column family and while inserting the values the column name is mentioned.

The general flow is that a Client contacts the Zookeeper first to find a particular row key. It does so by retrieving the server name from Zookeeper. With this information it can now query that server to get the server that holds the metatable. Both these details are cached and only looked up once. Lastly, it can query the metaserver and retrieve the server that has the row the client is looking for.

Column family make the above procedure very simple in finding the particular record. Also for each created data in hbase creates the timestamp, value length and other details so it is necessary to create at least the column family.

Hbase creates a spate directory for each and every column so one should not create many column families else the namenode will get overwhelmed so many copies of the directories due to replication factor. Hence column families must be carefully engineered.

* What is the benefit of using connection pool in Hbase?

Database connections are often expensive to create because of the overhead of creating a new connection and initialization. In turn, connection session initialization often requires time consuming processing to perform user authentication, establish transactional contexts and establish other aspects of the session that are required for subsequent database usage. Also the ongoing management of the connections that are established are also one of the limiting factor for creating and destroying connections each time.

Hence the connections are already created and never disconnected and the connections are given to each client in tandem manner.

Also time needed for forming new connection is more hence by connection pooling this time can be saved.