1. The difference between Hbase and Relational Database?

Hbase is non-relational, distributed, fault-tolerant database. In contrast with relational database, hbase works with key-value instead of tabular data.

1. In Hbase what is column families, row key?

Column families are groups of columns of the database that are stored in the same partition of the database. Usually, columns in a family are related between them. Row key is the group of keys of the database.

1. What is the use of ZooKeeper in Hbase?

Is the cluster manager that keeps track of the data partitions and its allocation in disk.

1. What are the optimizations that developer can make while working with spark?

Optimizations in spark could be coalesce operations for a better distribution of the processing within executors. Caching is another way to optimize code by moving to memory data that will be processed. Another way could be to reduce the number of wide transformations to avoid disk read/write operations.

1. Define HRegionServer in Hbase?

Is the data storage of the Hbase system.

1. what are two types of processing in Spark?

Batch processing and streaming. In batch processing the data is already stored in disk and spark loads the data to memory to perform operations. In spark streaming, the data comes to spark from a source and sparks processes the incoming data in memory.

1. How does Spark process in memory?

By using the persist and cache operations to allocate data in memory and perform operations faster.

1. What is Sliding Window in Spark?

It is a sample of data taken from a data stream. Only the data that fall into the window is operated and processed.

1. Explain HBase Architecture?

Hbase has zookeeper as a cluster manager keeping track of the partitions and the column families that are stored in the partitions. It also has the client which make queries and requests to the database and the regionServers that store the data and fulfill the requests.