1. How are worker, executor and task related to each other?

One Worker node can have multiple executors. Executors are worker nodes' processes in charge of running individual tasks in a given Spark job. They are launched at the beginning of a Spark application and typically run for the entire lifetime of an application. Once they have run the task they send the results to the driver. One executor can run/execute multiple tasks as multiple threads but can execute one thread at a given point of time. Tasks are the actual computations/operations that are to be performed on data.

2. What are the key features of Spark?

* Allows Integration with Hadoop and files included in HDFS.
* Spark has an interactive language shell as it has an independent Scala (the language in which Spark is written) interpreter.
* Spark consists of RDD’s (Resilient Distributed Datasets), which can be cached across computing nodes in a cluster.
* Spark supports multiple analytic tools that are used for interactive query analysis , real-time analysis and graph processing

3. What is Spark Driver?

Spark Driver is the program that runs on the master node of the machine and declares transformations and actions on data RDDs. In simple terms, driver in Spark creates SparkContext, connected to a given Spark Master. The driver also delivers the RDD graphs to Master, where the standalone cluster manager runs.

4. What are the benefits of Spark over MapReduce?

* Spark is easy to program and doesn't require any abstractions.
* spark has interactive mode.
* Spark makes it possible to perform streaming, batch processing and machine learning in the same cluster.
* Spark executes batch processing jobs near about 10 to 100 times faster than map reduce.
* Spark makes it possible to modify the data in real time through Streaming.
* Spark ensures lower latency computations by caching the partial results across its memory.
* Writing spark code is always compact than writing map reduce code as spark provides rich built in APIs.

5. What is Spark Executor?

Executors are Spark processes that run computations and store the data on the worker node. They are launched at the beginning of a Spark application and typically run for the entire lifetime of an application. The final tasks by SparkContext are transferred to executors for their execution.