Assignment 12.2

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

Workers are the slave daemon in Spark responsible for computation. Workers hold many executors, for many applications. Workers contain executors for driving out the computations. 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.

2. **What are the key features of Spark?**

RDD – Resilient Distributed Datasets

Data Frame

Rich Set of API

Data Caching

Real Time Stream Processing

Strong ecosystem Tool Support

Unified Platform

3. **What is Spark Driver?**

The spark driver is the program that creates the SparkContext, connecting to a given Spark Master.

4. **What are the benefits of Spark over MapReduce?**

Apache Spark

Easy to program and does not require any abstractions.

Programmers can perform streaming, batch processing and machine learning ,all in the same cluster.

Has in-built interactive mode.

Executes jobs 10 to 100 times faster than Hadoop MapReduce.

Programmers can modify the data in real-time through Spark streaming.

Apache Hadoop

Difficult to program and requires abstractions.

It is used for generating reports that help find answers to historical queries.

No in-built interactive mode except tools like Pig and Hive.

Hadoop MapReduce does not leverage the memory of the hadoop cluster to the maximum.

Allows you to just process a batch of stored data.

5. What is Spark Executor?

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.