Q7

Spark Driver

Program which declares the SparkContext. Its role is to split the user program into a series of tasks which can be distributed across the cluster. It also coordinates the execution of tasks and communicates with the Cluster Manager to allocate resources for the application.

➔ Spark Driver is the central coordinating entity of the Spark Application

Spark Executor

Process which task it is to run the tasks in parallel on worker nodes in the cluster ➔The Driver launches the Executor process, which runs the tasks that are assigned to him by the driver.

Responsible for processing the data and executing the code in parallel ➔ Runs the user defined Spark Code and performs the calculations and transformations on the data

Worker Node

Worker nodes are the physical or virtual machines in the Spark cluster that host executors. They communicate with the cluster manager to allocate resources for executors. They track and report the status of executors and tasks back to the driver.