

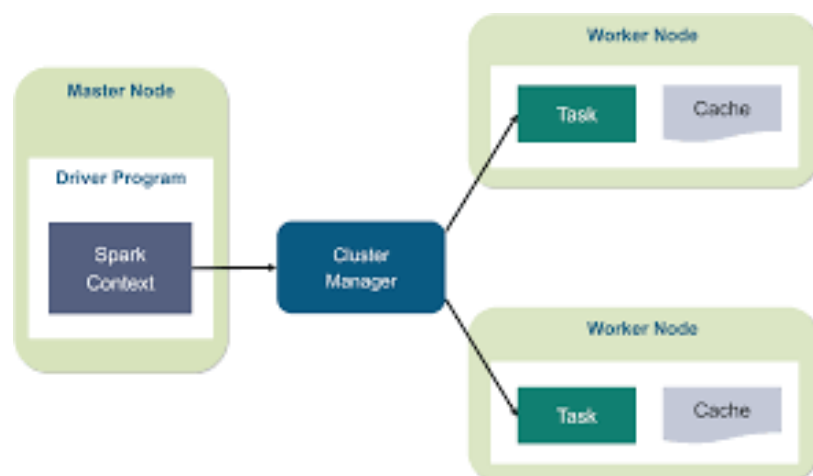


Spark Architecture

Overview:-

Apache Spark follows a master-slave architecture consisting of a Driver, multiple Executors, and distributed Tasks. It is designed for high-performance distributed computing.

sonu vishwakarma





1. Spark Components:-

1.1 Driver Program

- The Driver is the main process that controls the Spark job execution.
- It runs the SparkContext, which is responsible for coordinating tasks across the cluster.
- It translates the Spark application into a DAG (Directed Acyclic Graph) of tasks.
- It allocates resources and schedules tasks on worker nodes.

sonu vishwakarma



1. Spark Components:-

1.2 Executors

- Executors are worker processes that run tasks and store data in-memory for caching.
- Each executor is assigned by the cluster manager.
- They communicate with the driver but do not communicate with each other directly.

sonu vishwakarma



1. Spark Components:-

1.3 Tasks

- A Task is the smallest execution unit in Spark.
- The driver program divides a job into multiple stages and further into tasks.
- Tasks are executed in parallel on different executors.

sonu vishwakarma



2. Spark Execution Process

- User submits a Spark application.
- The Driver creates the SparkContext.
- The Driver requests resources from the cluster manager.
- Executors are launched, and tasks are assigned.
- Executors execute tasks and return results to the Driver.
- The job completes and the executors shut down.
-