In this doc file using real world example I have only explained what is Apache Spark and how it works

Apache Spark –

Apache Spark is an open-source , distributedcomputing framework designed for large-scale data processing and analytics.it is known for its speed , ease to use, and ability to handle diverse data processing tasks and spark supports various data processing needs,including batch processing ,real-time stream processing, ml , graph processing.

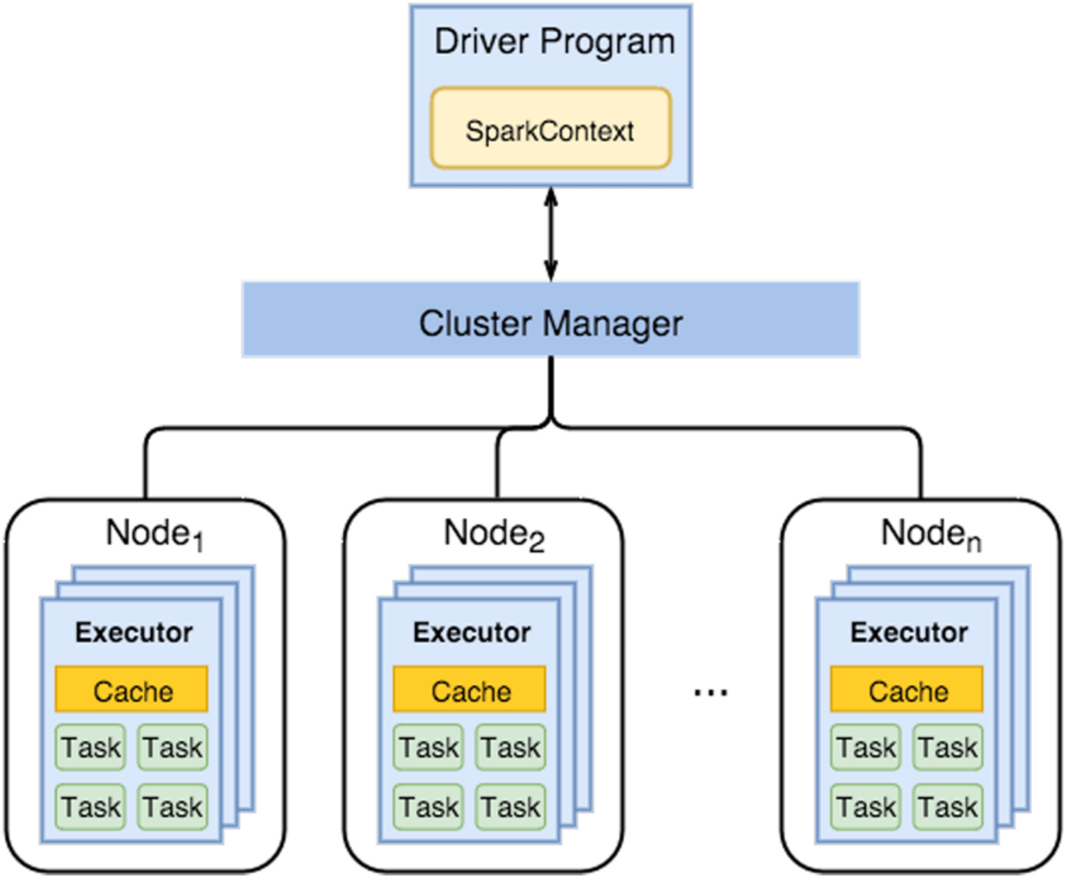
Core Concepts of Apache Spark –

1. **Job -** Think of a job as a task you want to accomplish using Spark, like reading data from a file, processing it, and then saving the results. For example, if you want to count how many times each word appears in a text file, that's a job.
   Let's say you have a file with student grades, and you want to calculate the average grade. This whole like process reading the file, calculating averages, and saving the results is a job.
2. **Stage** - Spark breaks down jobs into smaller steps called stages. Let's say To calculate average grades, you first need to add up all the grades that is one stage, and then divide the sum by the number of students that is another stage. Each of these steps happens in different stages.
3. **Tasks** - A task is a small unit of work that Spark assigns to worker machine, If your data is split into 10 parts (called partitions), Spark will assign 10 tasks to calculate the sum of grades for each part. Each task works on one part of the data.
4. **DAG (Directed Acyclic Graph) –** A roadmap of all the operations Spark will perform to complete the job, If you first filter the grades like remove students with missing grades and then calculate the average, Spark creates a plan DAG showing that it will filter the data first and then do the average calculation. The DAG ensures Spark does the work in the right order without going back and redoing steps.
5. **Executor -** When Spark assigns tasks to workers, the worker runs them inside a process called an executor.
6. **Master** - The master is like the teacher who assigns tasks to students. It tells each worker what part of the job to do and when to do it.
7. **Slave -** The worker machine that follows the master's instructions and runs the executors.

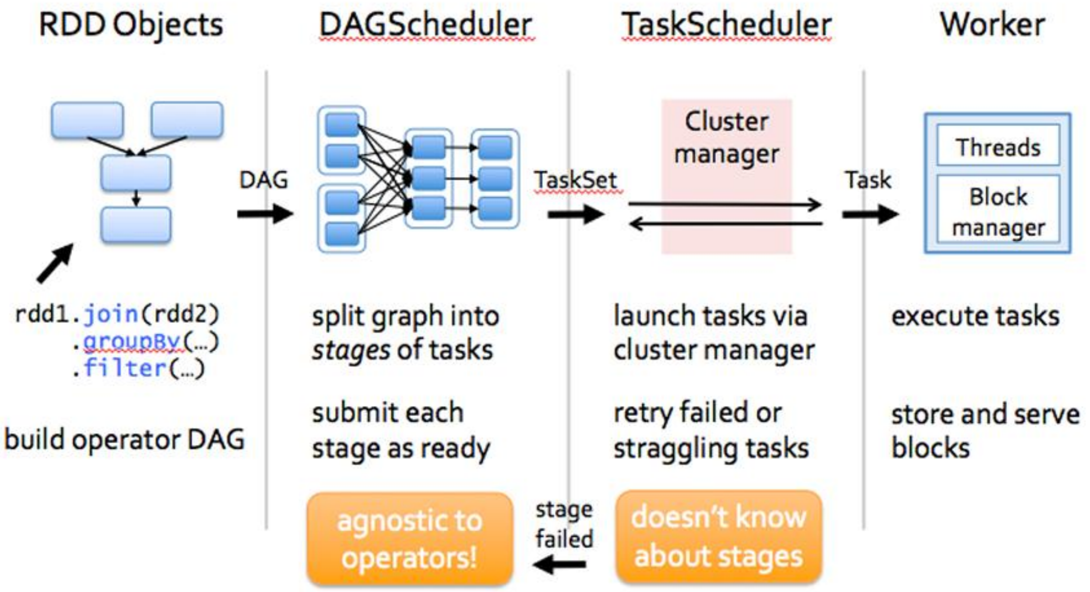
Spark Components—

1. **Spark Driver –** Spark Driver like brain or controller of the Spark application. Imagine you are a project manager in a factory.Spark Driver is like the project manager who creates the work plan, assigns tasks to workers, and monitors their progress.

2. **Cluster Manager --** the Cluster Manager is like the floor manager who assigns workers (executors) to different jobs based on their availability. If a project manager (driver) needs workers, they talk to the floor manager (cluster manager) to request the right number of workers.
3. **Executors --** Executors are like the workers in the factory. The project manager (driver) assigns them specific tasks, and they do the actual work of assembling the product . Once they finish, they either store the result or send it back.



How Apache Spark Works—



Step -1 – RDD (Resilient Distributed Dataset)

Imagine you have a large dataset of customer transactions, and you want to find out the total sales by product. You might create an RDD of transactions and apply transformations and actions.

-- Transformations are operations in Spark that create a new from an existing one. Transformations are lazy, meaning they don't execute right away but instead define a computation pipeline that Spark will execute later, when an action is called.

-- Actions are operations that trigger the execution of the transformations defined When an action is called, Spark runs the computation pipeline and produces a result Actions either return the final value to the driver program or write data to an external storage system.

Step –2 – DAGScheduler

Think of the DAG as a recipe that tells you what to do at each step. If the recipe has several steps, like chopping, cooking, and serving, the DAGScheduler breaks them up and makes sure they happen in the right order

Step –3 – TaskScheduler

Imagine you are running a pizza delivery service. The TaskScheduler is like the dispatcher who sends drivers to deliver pizzas. It doesn't care about the details of the order it only cares about making sure each delivery gets done. If one driver fails to deliver, the dispatcher assigns someone else.

Step—4—Worker

Worker like Executor , Workers execute the tasks, process the data, and send results back to the driver.