Spark Internals:

Eg:

```
df1 = spark.read.csv("location")
```

After submitting above code in spark, Spark Context will get created first then code will be submitted to driver.

Driver will understand code and optimize it and then will divided it into stages and tasks.

These tasks will be sent to different executors.

Executor is nothing but worker node, one worker node can have 1 or more than 1 executor based on config.

Once task is assigned to executor then executor will start reading file from source location.

InputFormat API:

When the driver node sends task to an executor, the executor uses the InputFormat API to read the required data from an External file system such as S3,Blob,HDFS.

Its programming interface which defines the logic for dividing data into splits and assigning them to different Spark worker nodes for parallel processing.

Components of InputFormat API:

1) FileInputFormat API: Its responsible to handle different file formats.

2) InputSplit: Its responsible to split data into logically, it will access file from external file system and will divide data.