Spark Core

Prerequistes:- Big data terminologies,(mapper Hadoop)

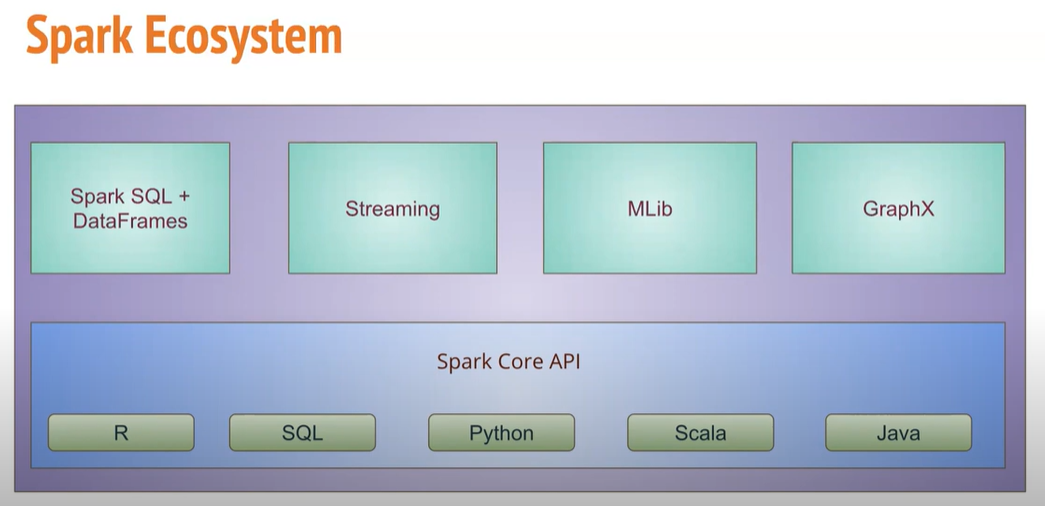
**Spark is a computing engine** .

🡪IN MEMORY processing faster, can connect to specified system.

**Spark unified the data processing**

🡪Same engine can be used for SQL, ML, Graph.

🡪Consistent API



Advantage of Spark:

🡪Evolution of map reduce

🡪parallel processing

🡪low disc I/O operation

Hadoop takes 10 I/O operation for 5 mapper

So how it work

M1 M2 M3 M4 M5

Disc

SO here M1 mapps the program and execute the output and store it in disk that the output of disc is the input for next mapper and so on

Thus disk operation = 2\*5 =10

Where as in spark

M1 M2 M3 M4 M5

Disc

Since it is in memory so M1memory ==M2 memory==M3memory… and only initial and final operation is written on disc

Thus only 2 I/O operations

🡪Active

🡪Compact

🡪Multiple language API

🡪unified data processing