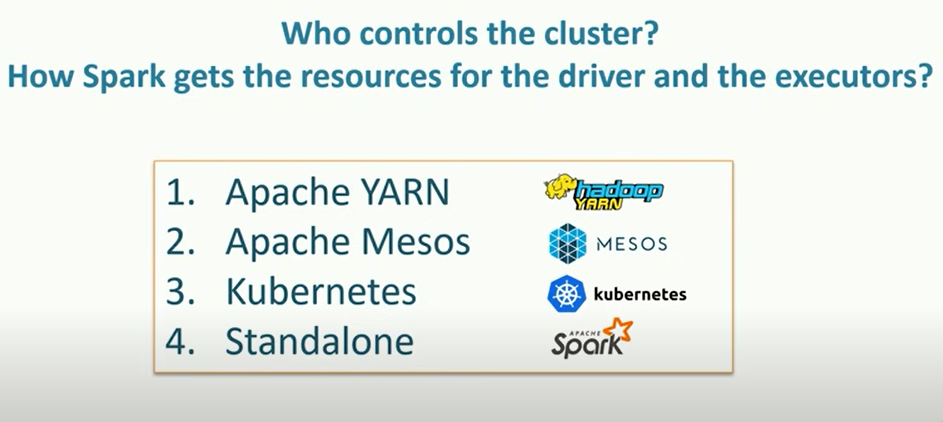
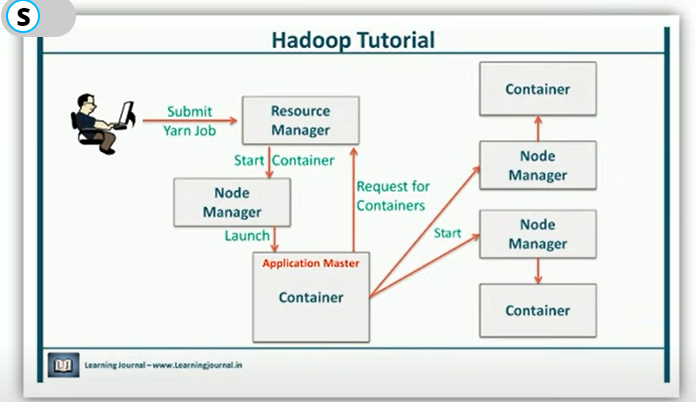


First one is for interactive client and 2nd one is for submit job.



Working of Hadoop YARN



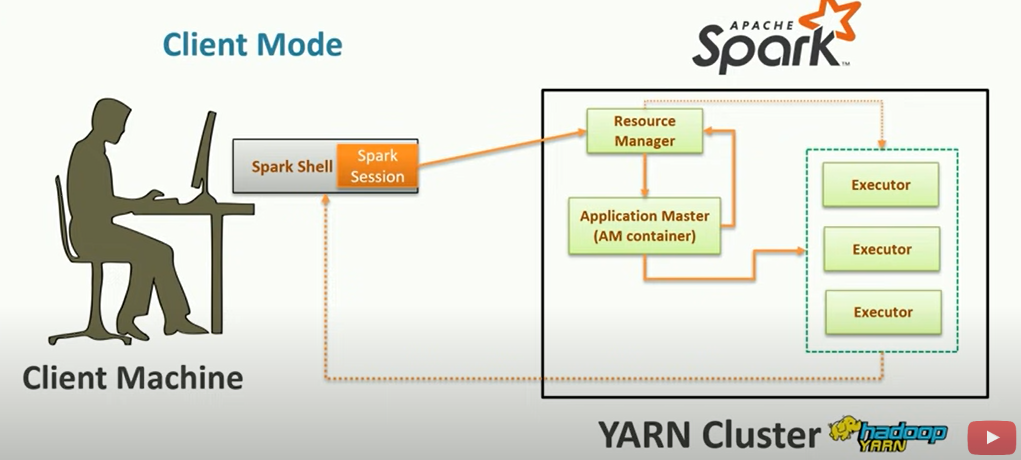
Running Cluster in 3 mode:

1.Client

2.Cluster

3.Local

1.Client Mode



1.Spark session is created by driver automatically in the local machine.

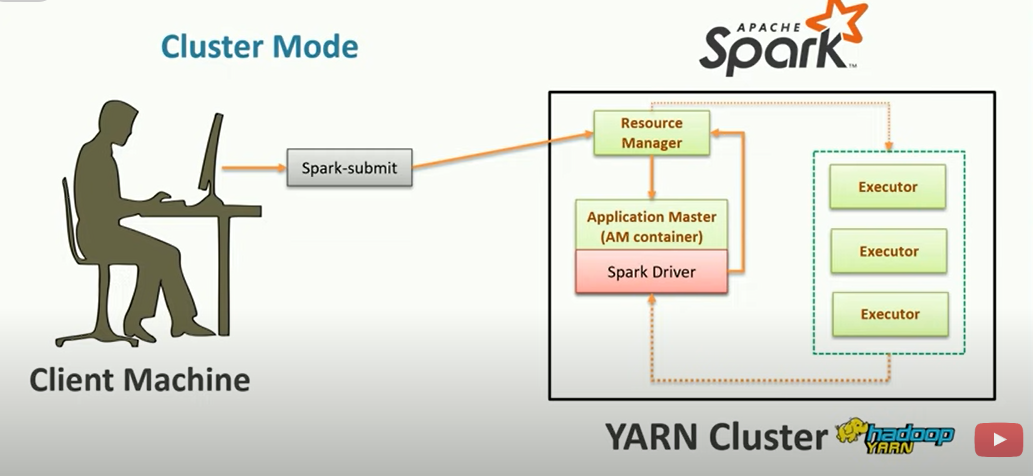
2.Driver sends request to the YARN Resource Manager to create an application.

3. YARN Resource Manager starts an application master.For the client mode the AM acts as executor launcher.

4.AM will reach out to YARN Resource Manager to request for further containers.

5. The YARN Resource Manager will allocate new containers and AM starts the executor in each container.(Each executors are in separate container)

6.These executors directly communicate with driver.



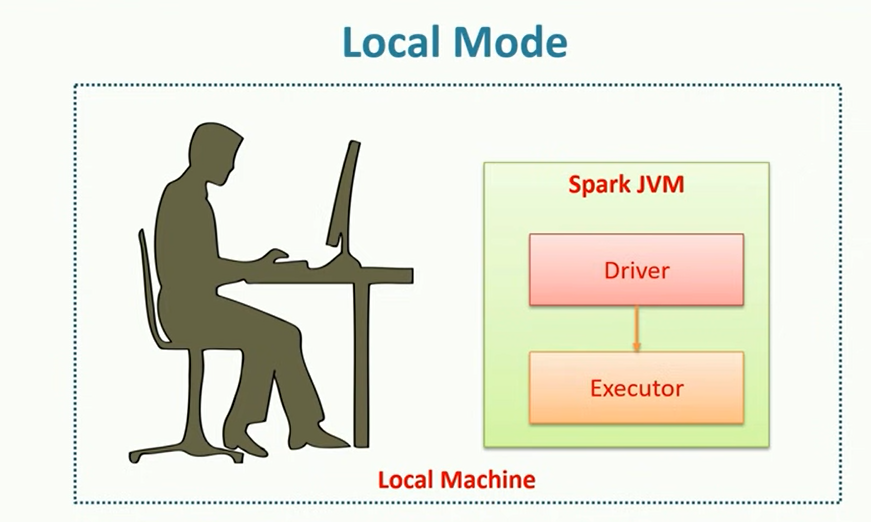
1.Submit the application using SPARK Submit tool.

2.SPARK Submit utility will send spark submit request to YARN Resource manager

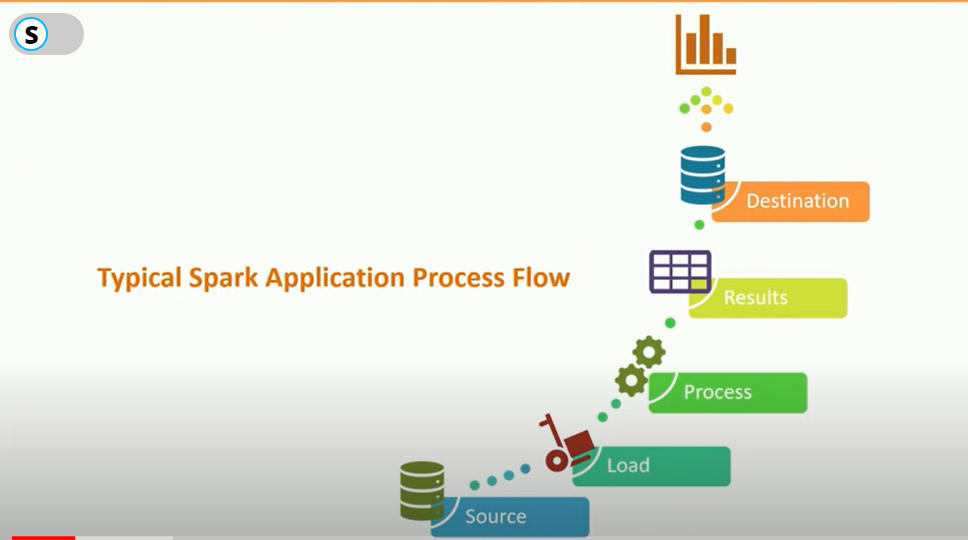
3.The YARN Resource manager starts an Application Manager(AM).

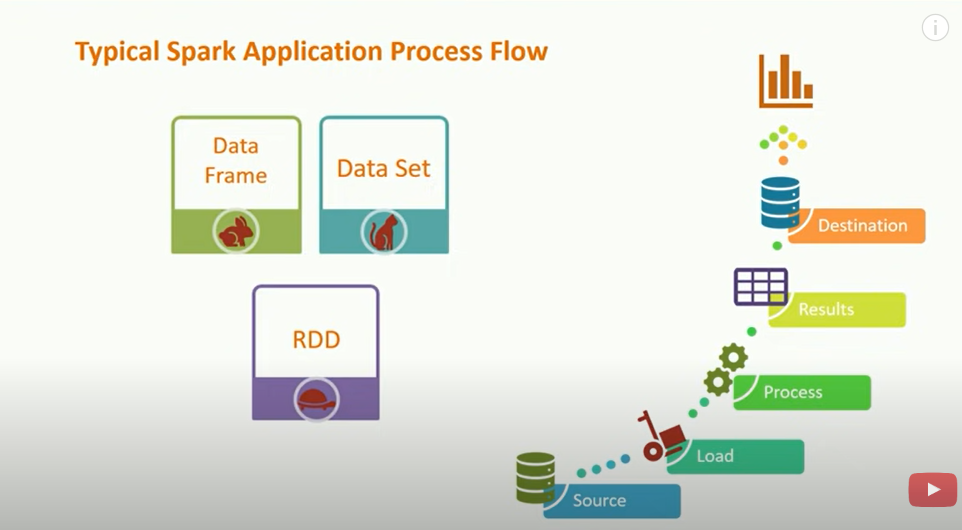
4.Then the driver starts in the AM container and will reach out to YARN Resource Manager to request for further containers.(point of difference in client and cluster mode)

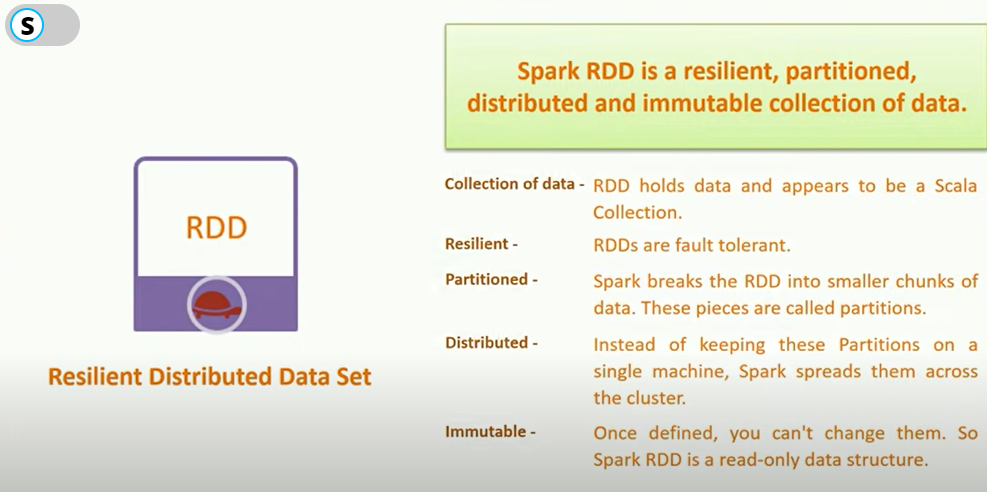
5. The YARN Resource Manager will allocate new containers and driver starts the executor in each container.(Each executors are in separate container)



No cluster required.







Create RDD by following methods

