Spark standalone

the independent mode, with its own complete service, can be deployed separately into a cluster without relying on any other resource management system

It has *masters* and number of *workers* with configured amount of memory and CPU cores.

In Spark standalone cluster mode, Spark allocates resources based on the core.

advantage->

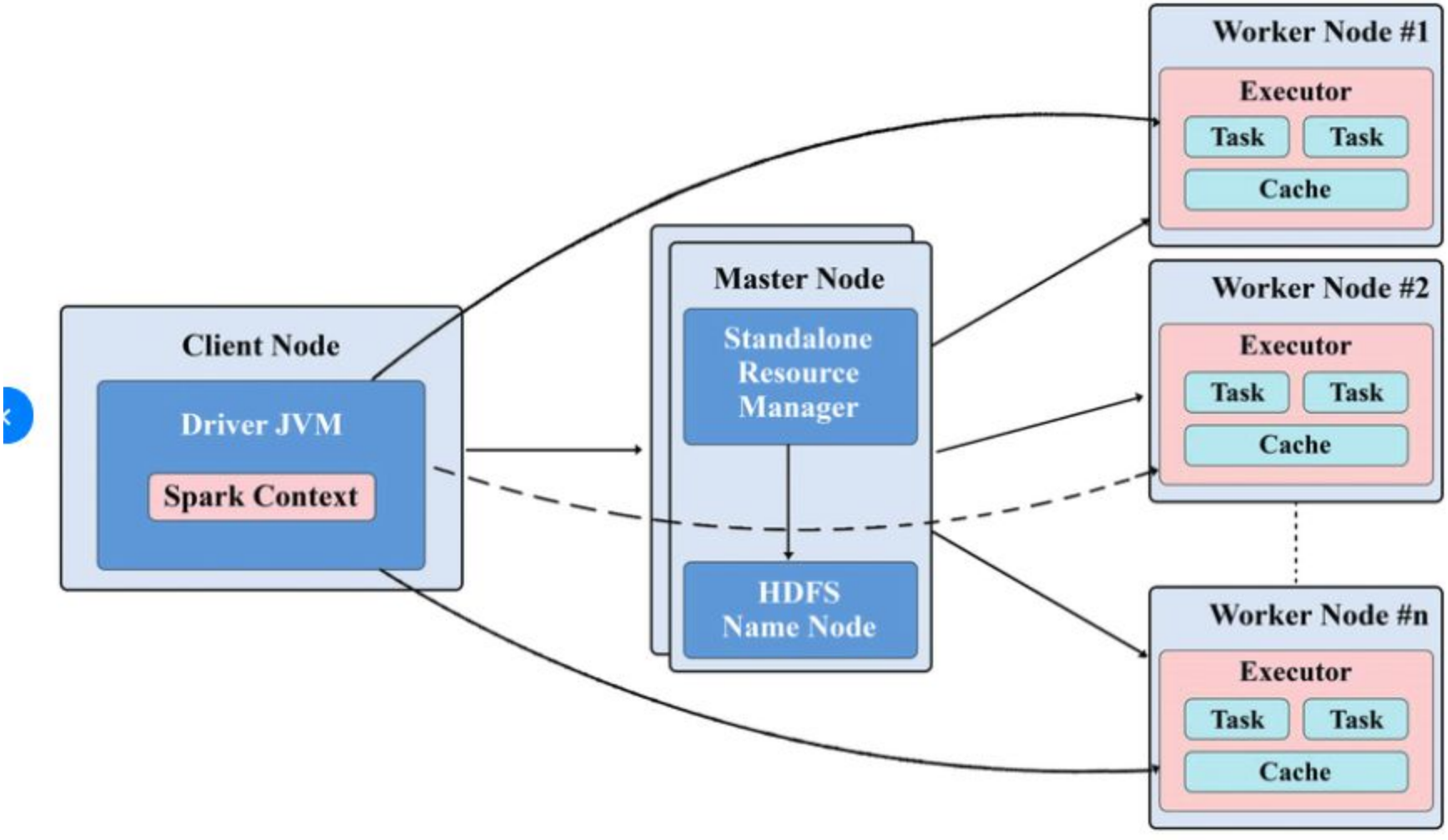
Easy to setup a cluster

Spark itself manages a cluster environment

rapid development

Disadvantage->

Limited in feature


Client Node
Driver JVM
Spark Context
Master Node
Standalone Resource Manager
HDFS Name Node
Worker Node #1
Executor
Task
Task
Cache
Worker Node #2
Executor
Task
Task
Cache
Worker Node #n
Executor
Task
Task
Cache

Spark Local

non-distributed single-JVM deployment mode,

It is the default mode of spark. It does not require any resource manager

Launches executor threads locally for processin

Advantage->

With local mode, we can utilize multiple core of a CPU for processing. Essentially, It is good for parallel computing.

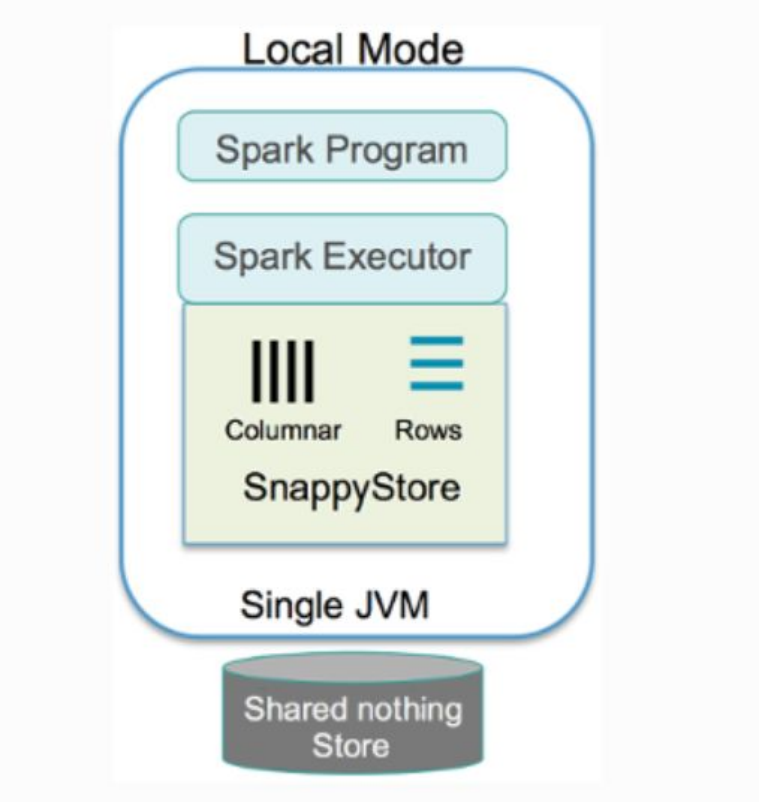
The local mode is also quite useful while testing an spark application.

Easy setup

Disadvantage->

If system fail all data gone.

Local Mode


Local Mode
Spark Program
Spark Executor
Columnar
Rows
SnappyStore
Single JVM
Shared nothing
Store

Spark client mode

here,”driver” component of spark job will run on the machine from which job is submitted. Hence, this spark mode is basically “client mode”.

Advantage->

there is no high network latency of data movement for final result generation between “spark infrastructure” and “driver”

used for interactive and debugging purposes.

Client can also get the information about status of particular job

Disadvantage->

In this mode, the entire application is dependent on the Local machine since the Driver resides in here. In case of any issue in the local machine, the driver will go off

Not suitable for production use cases

Spark cluster mode

spark application master will get started in any of the worker machines

it works with the concept of Fire and Forgets.clien just need to submit the application