

# **Deploying Spark Apps on a Cluster**



# Objective

Learn the anatomy of a Spark cluster

Start a Spark application directly on a cluster

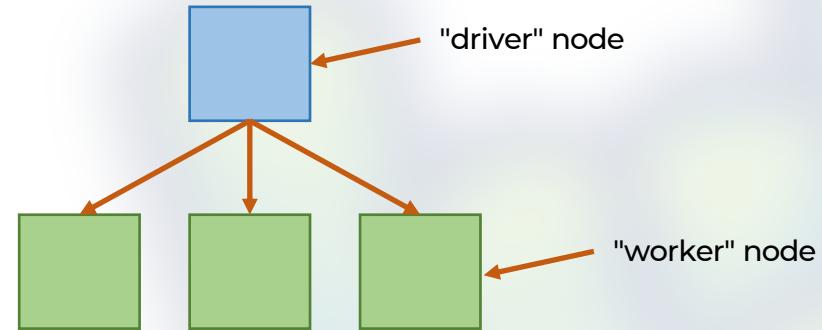


# The Anatomy of a Cluster

## Spark cluster manager

- one node manages the state of the cluster
- the others do the work
- communicate via driver/worker processes

## Standalone, YARN, Mesos

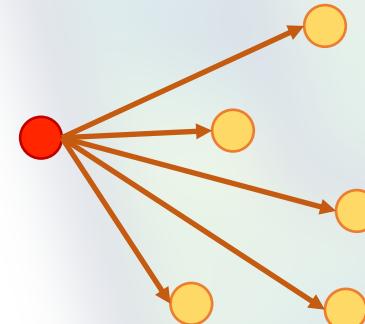


## Spark driver

- manages the state of the stages/tasks of the application
- interfaces with the cluster manager

## Spark executors

- run the tasks assigned by the Spark driver
- report their state and results to the driver



# The Anatomy of a Cluster

## Execution mode

- cluster
- client
- local

## Cluster mode

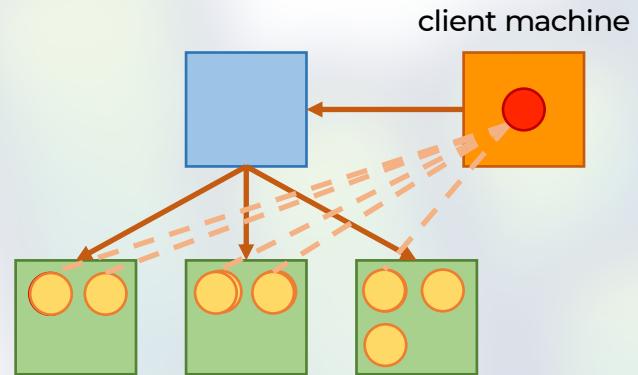
- the Spark driver is launched on a worker node
- the cluster manager is responsible for Spark processes

## Client mode

- the Spark driver is on the client machine
- the client is responsible for the Spark processes and state management

## Local mode

- the entire application runs on the same machine



# Recap

## Spark cluster manager

- driver node + worker nodes
- standalone, YARN, Mesos

## Spark processes

- driver
- executors

## Deploy modes

- cluster: driver + executors launched on cluster
- client: driver launched on client, executors on cluster
- local: everything on the same machine

## Deploy a Spark app

- package a JAR
- ship the JAR + data on the cluster
- submit the JAR to Spark

```
/spark/bin/spark-submit \
--class part6practical.TestApp \
--master spark://(dockerID):7077 \
--deploy-mode client \
--verbose \
--supervise \
spark_playground.jar data/movies.json data/goodMovies
```

# Spark rocks

