

# Solves

① 50 nodes, 256 GB size

Containers required?

## — Spark calculations

finding no. of cores, Each executor memory & total executor

when  
cores, RAM, Disk is given

**Number of Executors, Cores ,Memory Calculation in Spark**

**10 - Node Cluster**

	CORES (NUM)	RAM (GB)	DISK (TB)
1	16	64	10
1	16	64	10
1	16	64	10
1	16	64	10
1	16	64	10
1	16	64	10
1	16	64	10
1	16	64	10
1	16	64	10
1	16	64	10
<b>10</b>	<b>160</b>	<b>640</b>	<b>100</b>

Available Cores Per node

Let's assign 5 Cores for each Executor (Num of Executors/Node)

Hadoop/YARN Usage (Cores)	Available Cores Per node	Each Executor Memory	spark.yarn.executor.memoryOverhead (%)	Actual Each Executor Memory (GB)
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
1	(16-1)=15	(15/5)=3	(64/3)=21	(21*7%)=3 (21-3)=18
<b>10</b>	<b>150</b>	<b>30</b>		

Total Executors 30 (Note: Leave 1 Executor for Application Master)  
Each-Executor-Memory 18 GB  
Cores-Per-Executor 5

Cores	Executors	RAM	offheap
$16 - 1 = 15$	5 cores each $15/5 = 3 \text{ exec}$	$64/3 \approx 21$	$21 \times 10\text{GB} = 186 \text{ B}$

- 6 nodes each node have 15 cores  
64GB RAM.  
find no. of concurrent task ?

$16 \times 6 = 96$  total cores

420GB RAM total

task 2.

15 cores  
5 task each executor  
is task ≈ 90 task ?

executors is nothing but JVM  
nodes / 256GB Disk  
executors ?  
can have multiple executors

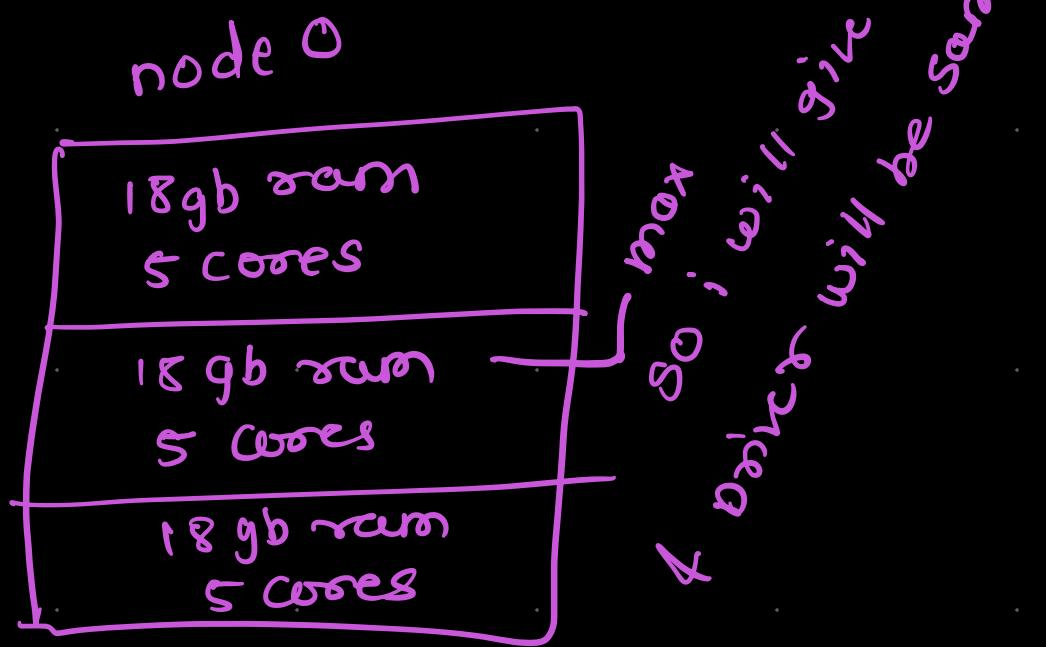
- Some node can have multiple executors
- ①  $\text{--num-executors} = \text{no. of JVM containers}$
- ②  $\text{--executor-memory} = \text{executor-cores} \times \text{min-memory}$
- ③  $\text{--executor-cores} = \text{no. of executors} \times \text{cores per executor}$
- for ex.  $5 \times 10 = 50$  tasks

$\text{Task} = \text{no. of executors} \times \text{cores per executor}$

if you give no. of executors for the cluster not for separate nodes.

④  $\text{--drivers-memory}$

10 nodes each have  
16 cores, 64GB ram  
each executor gets 5 cores  
total executors =  $10 \times 3 = 30$   
Ram usage  
 $64 / 3 = 21.33$  But we have  
to give some to drivers 10-1.  
298 for 10 nodes  
18GB per executor



# Map Reduce!

$$\text{No. of map task} = \frac{\text{Data size}}{\text{block size}}$$

14B into 128MB chunks

so blocks will be

$$\frac{1024}{128} = 8 \text{ blocks}$$

so 8 mappers are required!

— 256 GB data

How many mappers required?

$$256 \times 1024 = 262,144 \text{ MB}$$

$$\frac{262144}{128} \leftarrow \text{block size}$$

2048 ← mappers will be required (128MB)

No. of mappers =

500 nodes 25.6 GB data

Yarn containers

How many mapper4 reducer

$$25.6 \times 1024 = \frac{26,214.4}{256} \text{ block}$$

= 103 mappers

?

# no. of con =  $\min \left[ \begin{array}{l} 2 \times \text{cores}, 1.8 \times \text{disk}, \\ (\text{total} - \text{Reserved}) \end{array} \right]$

min containersize

$$\begin{aligned}
 \min &= \frac{500 \times 4GB}{2000gb} \\
 &= \frac{2000 - 500}{2000 - 500} \\
 &= \frac{1500GB}{1024} \\
 &= 1,536,000 MB
 \end{aligned}$$

$$= 3000$$

$$= 2 \times 500,1.8 \times 25.6 \times 1024, 3000$$

$$= 1000$$

$$\text{Containers} = 1000$$

Reducers =

Reducer =  $0.95 \text{ or } 1.75 \times (\text{no. of nodes} \times$   
- no. of max container  
per node)

$$\begin{aligned} &= 0.95 \times 500 \times 2 \\ &= 950 \end{aligned}$$

Assume -

- = RAM = 4GB
- = 1 core / node

- 1 Shard Key
  - 2 CAP
  - 3 Mongo DB - 1, 2, 3, 10
  - 4 Replica Set
  - 5 Reducer
  - 6 Partition
  - 7 HDFS
  - 8 Map Reduce - 5, 12, 13, 14,
  - 9 Spark ~ 15, 16  
17.
  - 10 Aggregate Framework
  - 11 Executor
  - 12 Resource Manager
  - 13 Node Manager -
  - 14 Application Master -
  - 15 Container
  - 16 Writable Class
- HDFS
- TP

17 Spark Process → 11, 6

18 → Input Split

19 Data node

20 Name node

50 nodes      512 GB

How many mappers, addrees

& how contains

mappers =  $\frac{512 \times 1024}{256 \text{ mb block}}$

=  $\frac{524288}{256}$

= 2048 mappers

no. of containers = 2048

no. of container =  $\min(2 \times \text{cores}, 1.8 \times \text{disk},$

(Total RAM - reserved ram)

min container size

$$= \left[ \frac{(50 \times 4) - 50}{512 \text{ mb}} \right] 1024$$

We assumed 4GB ram each node

$$\Rightarrow \# of C = \min(2 \times 50, 1.8 \times 512, 300)$$
$$= \underline{100}, 1.8 \times 512, 300$$

no. of containers = 100

reducers = no. of nodes  $\times$  no. of max  
containers per node

$\times 0.95$

$$= 50 \times \left( \frac{100}{50} \right) \times 0.95$$

$$= 100 \times 0.95$$

= 95 reducers