APACHE SPARK ASSIGNMENT

CASE 2

HARDWARE: 4 Nodes, each has 16 cores and 32GB RAM

Note:

1. 5 Cores = 1 executor
2. 1 Core + 1GB of each node is used by the operating system

Hence, the resources left are:

1 core = 15 cores and 31GB for each node

1. Calculate the number of executors

5 CPU Cores = 1 executor

15 Cores = 15/5

= 3 executors

Total Node = 4

1 Node = 3 executors

4 Nodes = 3x4

= 12 executors

Note:

1 executor shall be used by YARN as Application Master

Therefore,

Num_exe = 12 – 1 = 11 executors

2. Memory Allocation for Executor

1 Node = 3 executors, 31GB

1 executor = 10.3GB

Spark.memory.overhead = 10% of executor memory

Executor.memory = 10.3GB or (10GB)

Overhead => 1.1GB or (1GB)

Total Overhead => 11.4GB or (11GB)

Total num_executor = 11.4 x 3

= 34.2GB or (33GB)

Actual formula for Overhead memory

= max(384MB, 0.07 * executor.memory)

= max(384MB, 0.07 * 10.3GB)

= 720MB or 700MB

Final Executor Memory = 10.3GB – 720MB

Memory for each executor = 9.5GB or (9.3GB)

Final results for 6 Nodes with 16 Cores each and 32GB RAM

5 CPU Cores per executor

11 executors

9 GB RAM