Lesson 4
Hadoop Yarn

"Big Data Analytics ", Ch.02 L04: Introduction To Hadoop
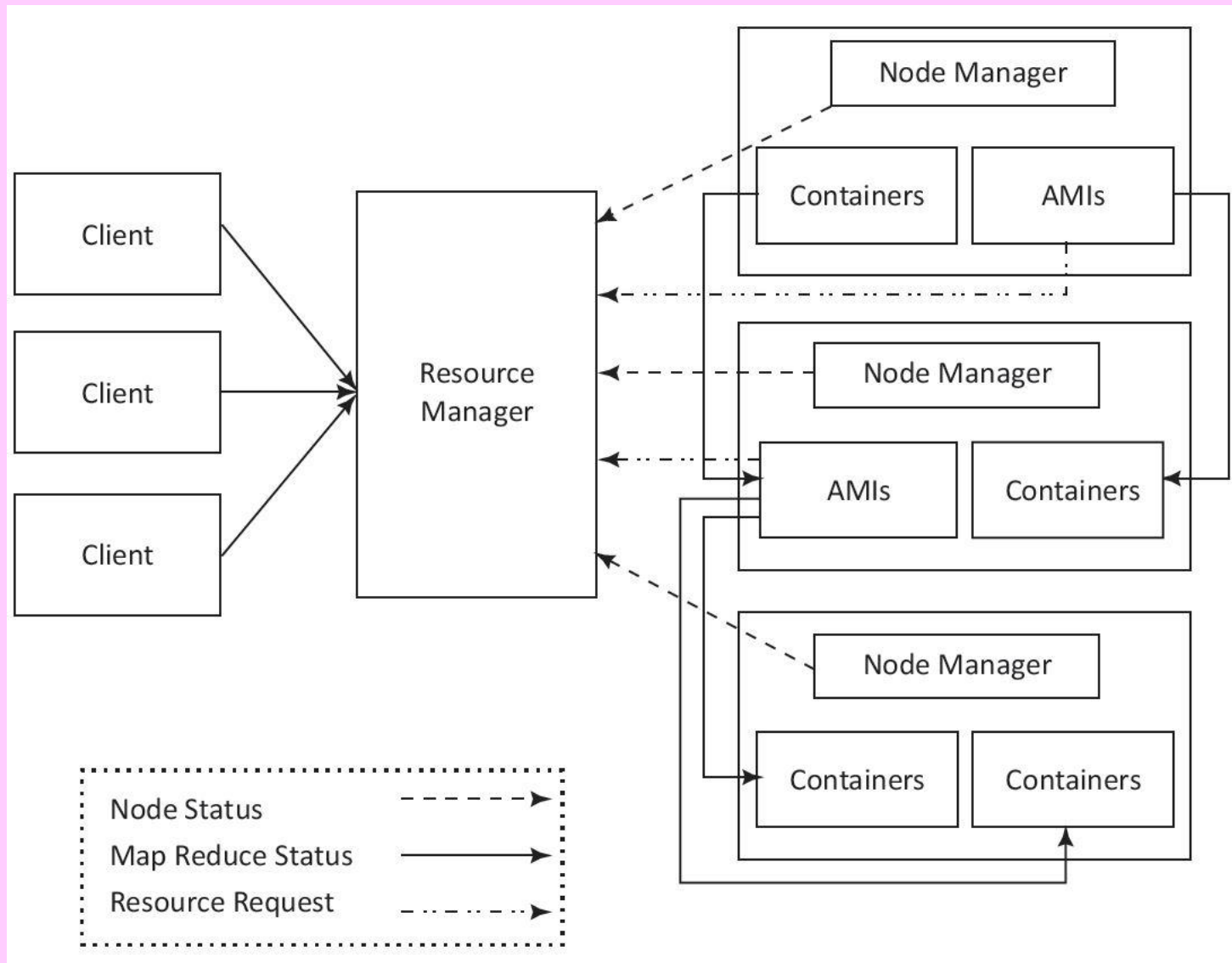Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Hadoop 2

- Hadoop 2 ─ YARN-based system for parallel processing of large datasets and distributed processing of the application tasks

- Hadoop 2 provides the multiple NameNodes. Yarn provides resource Manger and Node Manager

"Big Data Analytics ", Ch.02 L04: Introduction To Hadoop
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

YARN Components

- Client, Resource Manager (RM)

- Node Manager (NM)

- Application Master (AM)

- Containers

Figure 2.5 YARN-based execution model

"Big Data Analytics ", Ch.02 L04: Introduction To Hadoop
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

MasterNode components and Functions

(i) Job History Server and

(ii) Resource Manager(RM)

- A Client Node submits the request of an application to the RM (Master for slaves)

- One RM exists per cluster

- The RM keeps information of all the slave NMs.

"Big Data Analytics ", Ch.02 L04: Introduction To Hadoop
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

RM Functions

- Rendering the Resource Scheduler service that decides how to assign the resources
- Performs resource management as well as scheduling.

Multiple NMs (Slaves) and AMs

- Multiple NMs are at a cluster

- An NM creates an AM instance (AMI) and starts up

- The AMI initializes itself and registers with the RM. Multiple AMIs can be created in an AM.

"Big Data Analytics ", Ch.02 L04: Introduction To Hadoop
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

AMI Functions

- The AMI performs role of an Application Manager (ApplM), that estimates the resources requirement for running an application program or sub-task

- The ApplMs send their requests for the necessary resources to the RM.

NM

- Each NM includes several containers for uses by the subtasks of the application

- All active NMs send the controlling signal periodically to the RM signaling their presence.

NM

- Each NM assigns a container(s) for each AMI

- The container(s) assigned at an instance may be at same NM or another NM

"Big Data Analytics ", Ch.02 L04: Introduction To Hadoop
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

ApplM

- The ApplM at an instance uses the assigned container(s) for running the application sub-task.

"Big Data Analytics ", Ch.02 L04: Introduction To Hadoop
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Summary

We learnt :

- Components of Yarn and their actions

- Resource Manager (RM)

- Node Manager (NM)

- Application Master (AM)

- ApplM

- Containers

"Big Data Analytics ", Ch.02 L04: Introduction To Hadoop
Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

End of Lesson 4 on
Hadoop Yarn