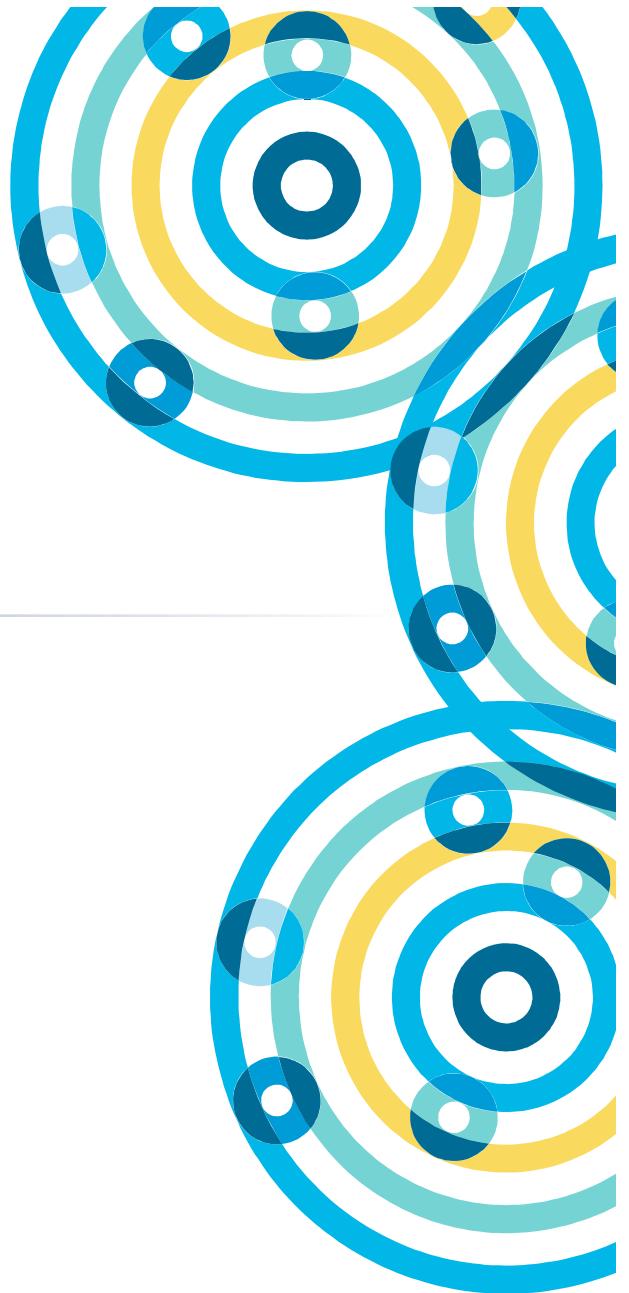


# YARN

---



## What is YARN?

---

YARN is

- **YARN = Yet Another Resource Negotiator**
- **YARN is the Hadoop processing layer that contains**
  - A resource manager
  - A job scheduler
- **YARN allows multiple data processing engines to run on a single Hadoop cluster**
  - Batch programs (e.g. Spark, MapReduce)
  - Interactive SQL (e.g. Impala)
  - Advanced analytics (e.g. Spark, Impala)
  - Streaming (e.g. Spark Streaming)

# YARN Daemons

---

- **Resource Manager (RM)**

- Runs on master node
- Global resource scheduler
- Arbitrates system resources between competing applications
- Has a pluggable scheduler to support different algorithms (capacity, fair scheduler, etc.)

Resource Manager

- **Node Manager (NM)**

- Runs on slave nodes
- Communicates with RM

Node Manager

What are HDFS coun

# Running an Application in YARN

---

## ■ Containers

- Created by the RM upon request
- Allocate a certain amount of resources (memory, CPU) on a slave node
- Applications run in one or more containers

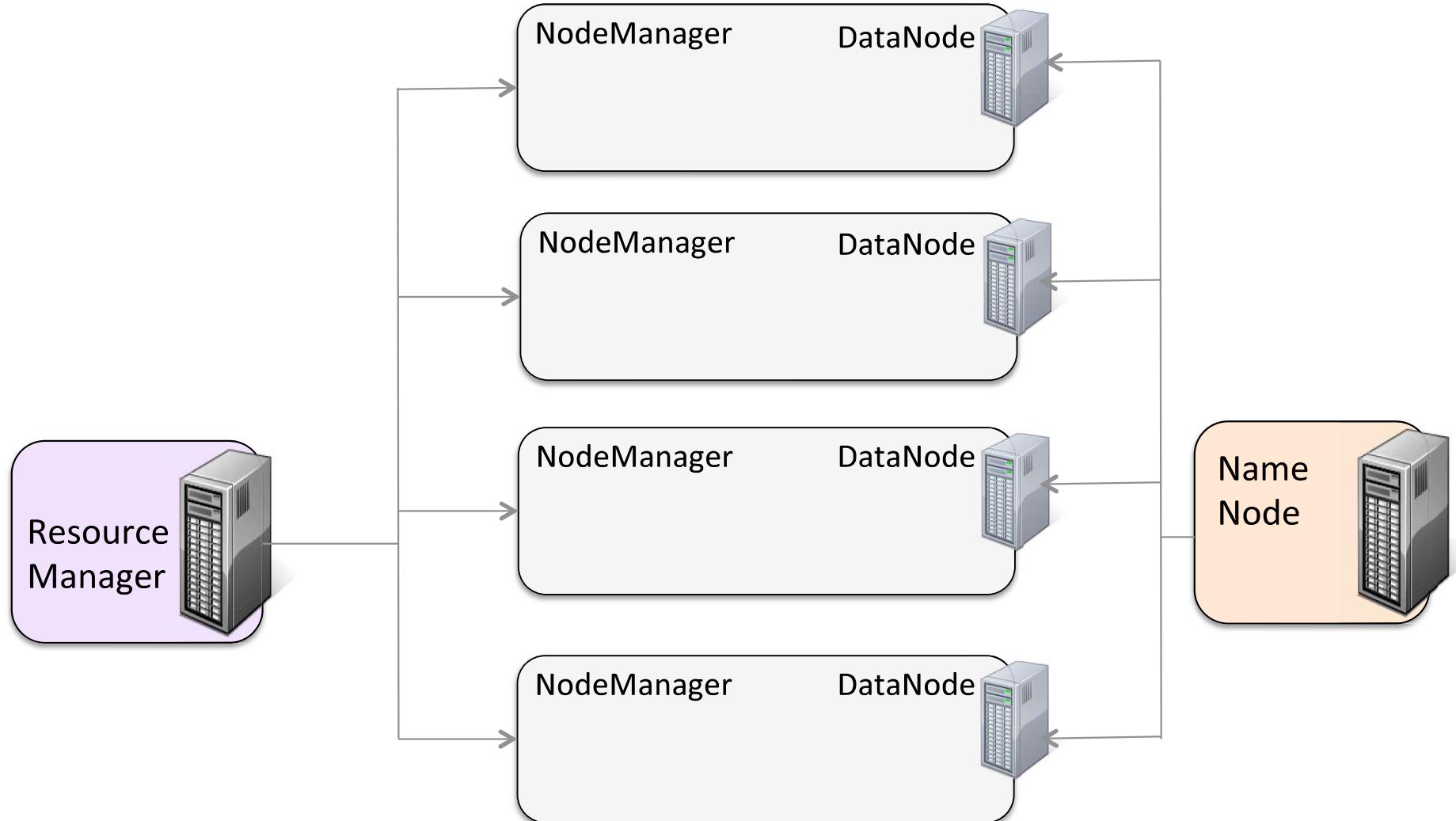


## ■ Application Master (AM)

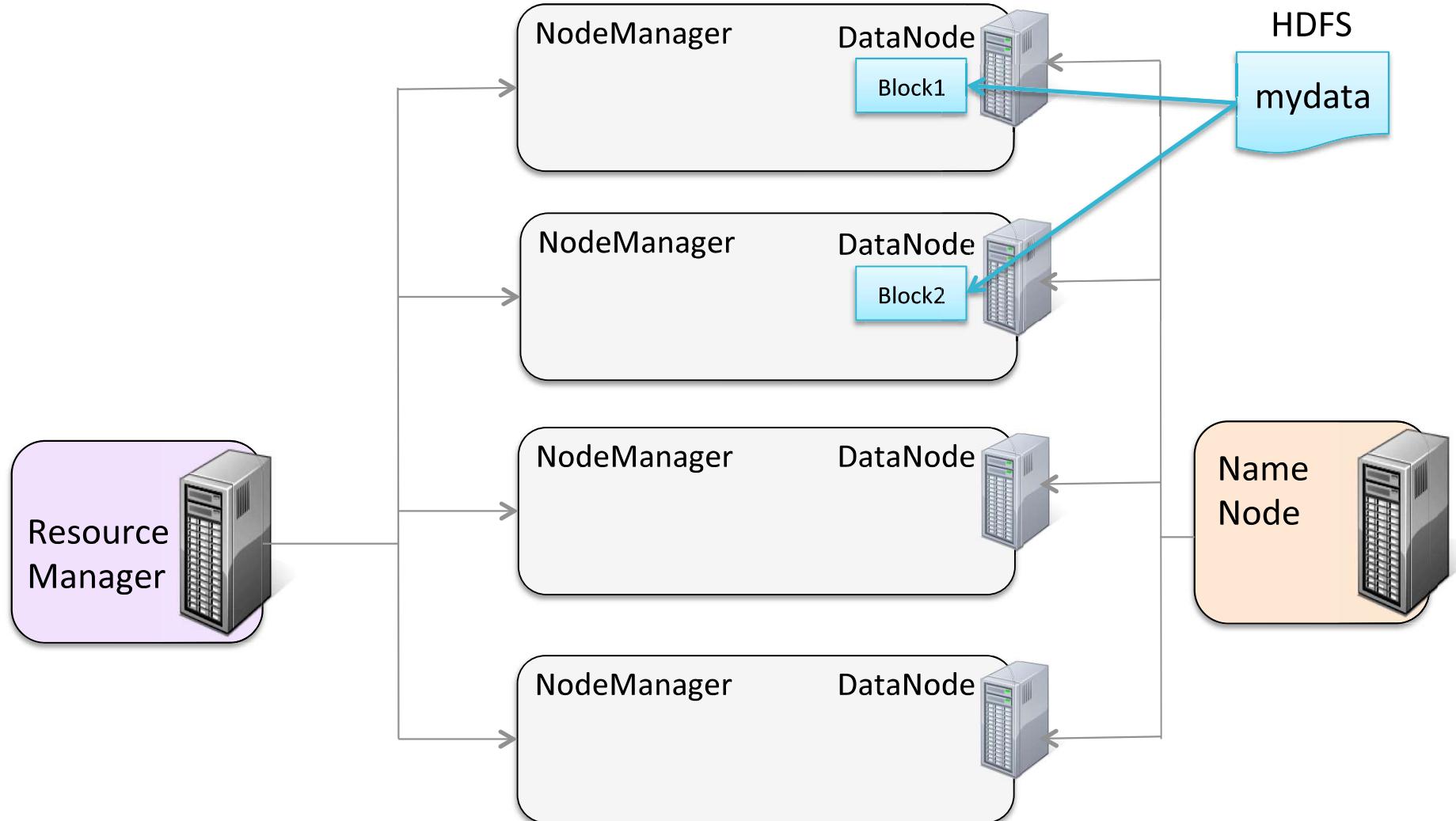
- One per application
- Framework/application specific
- Runs in a container
- Requests more containers to run application tasks



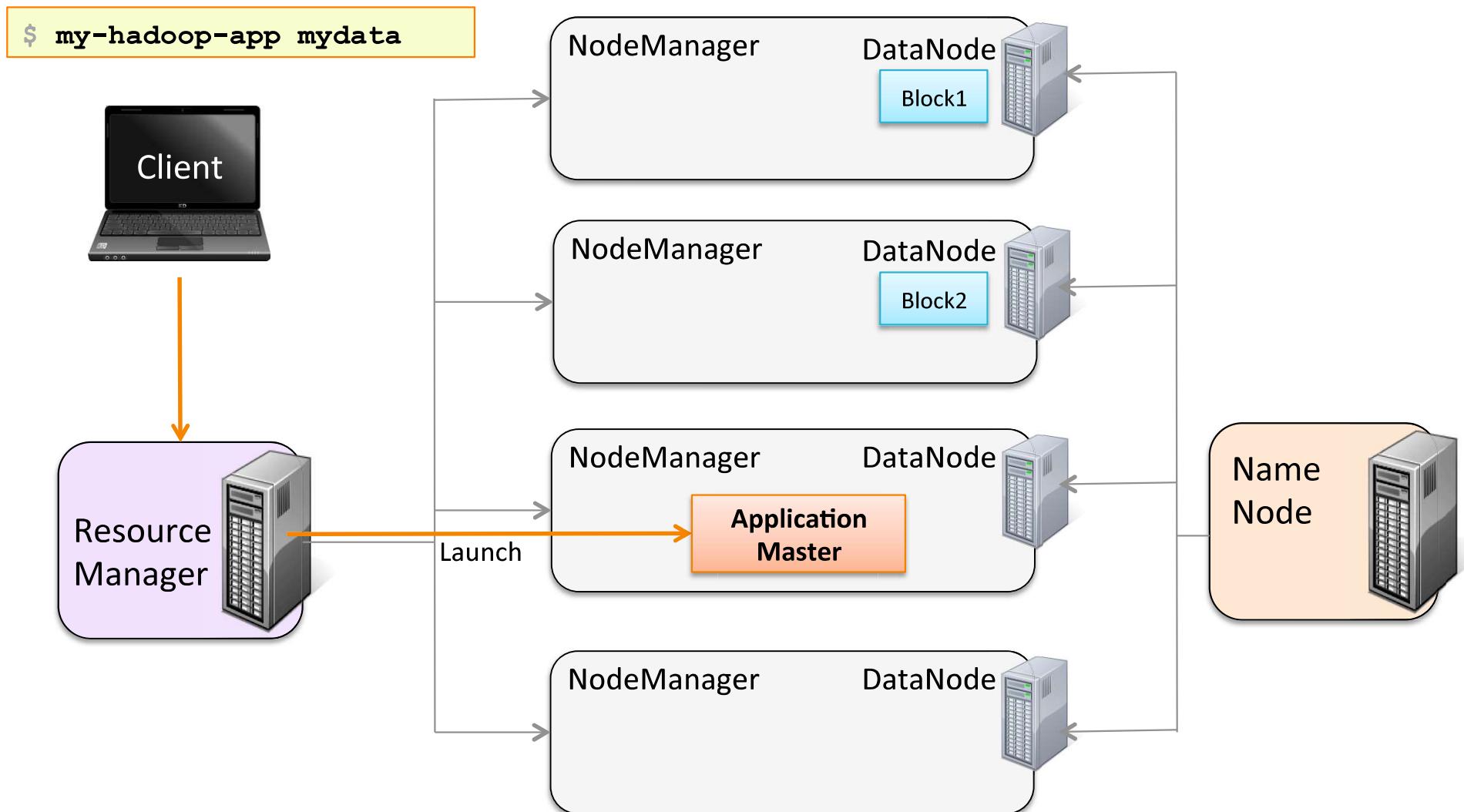
## Running an Application on YARN (1)



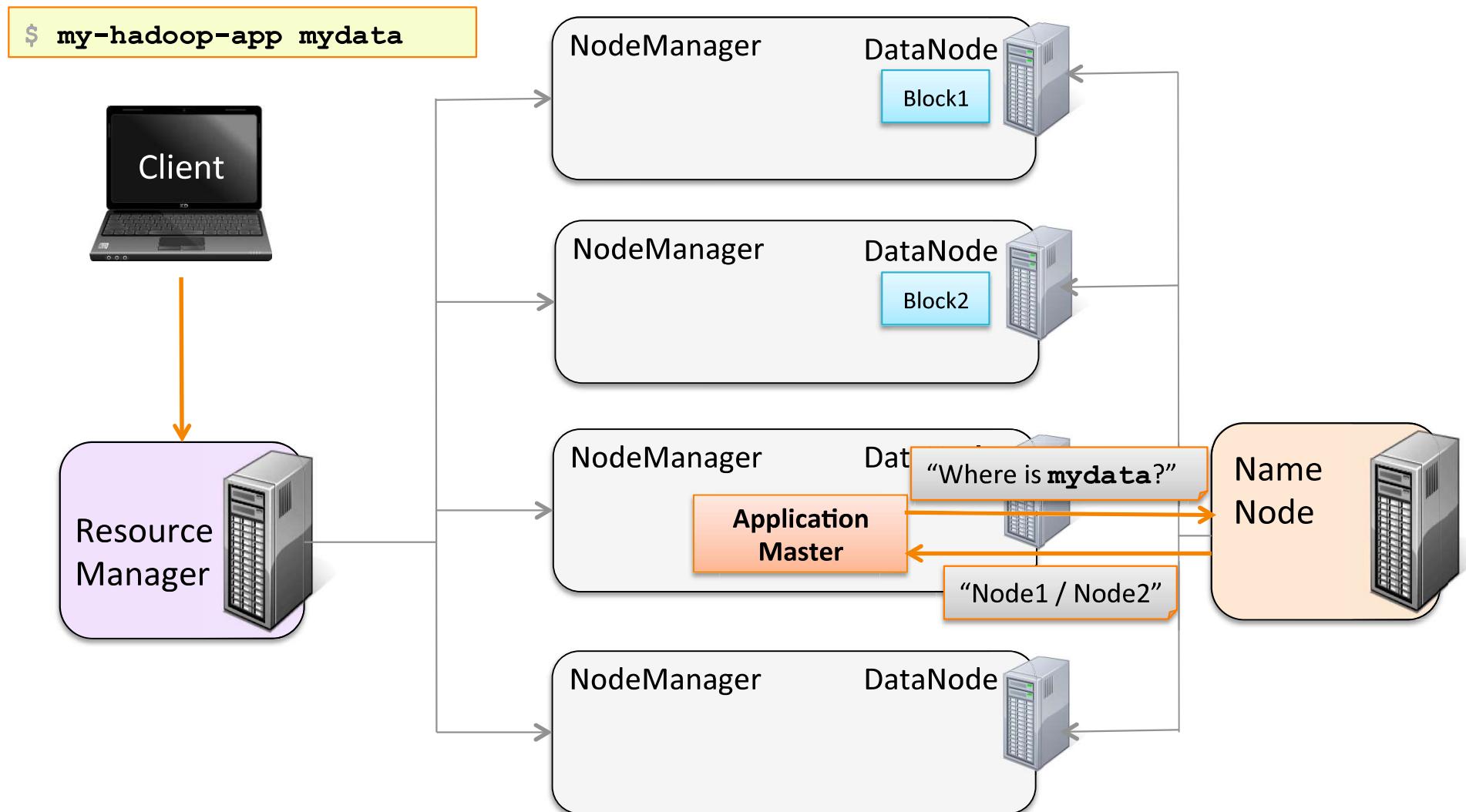
## Running an Application on YARN (2)



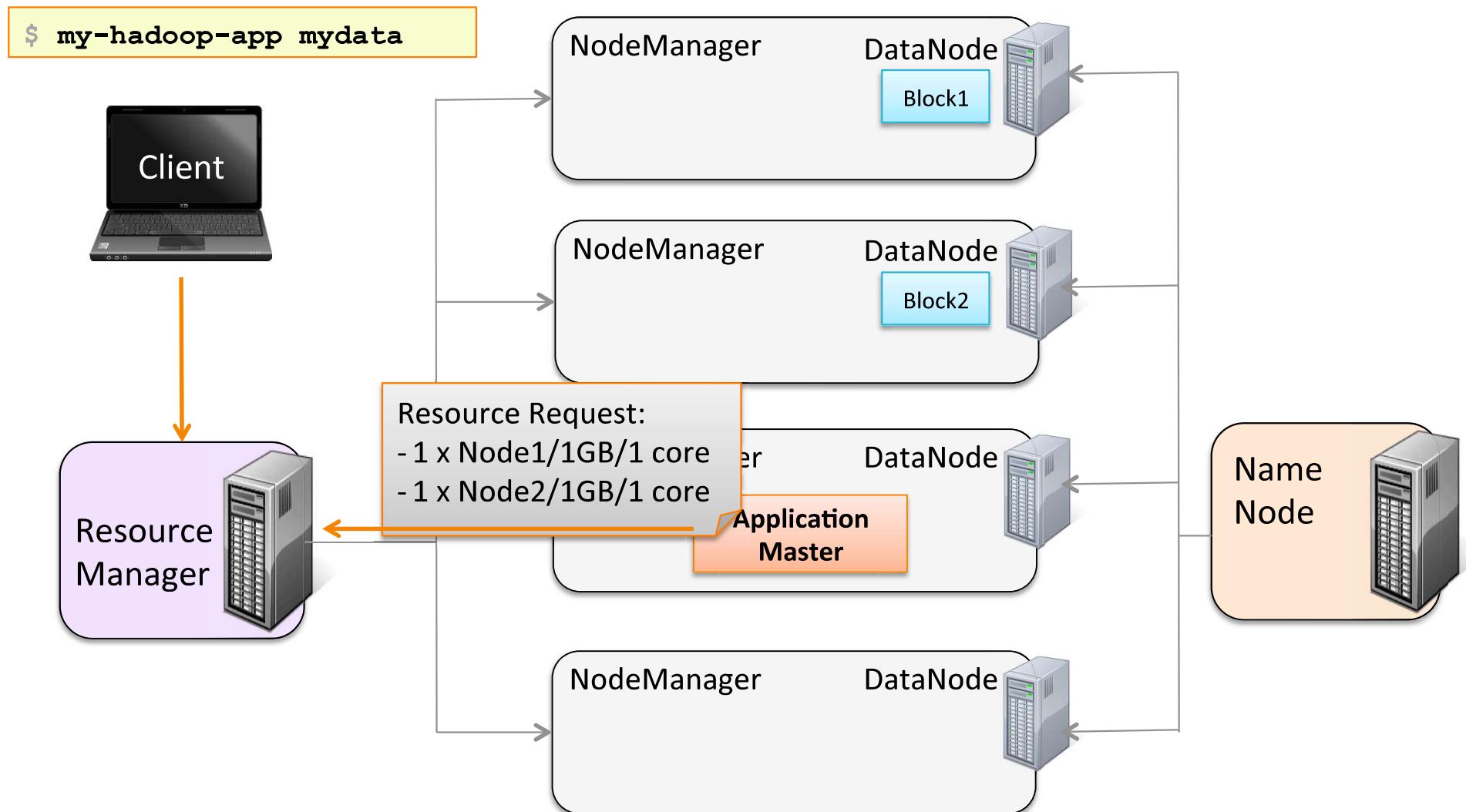
## Running an Application on YARN (3)



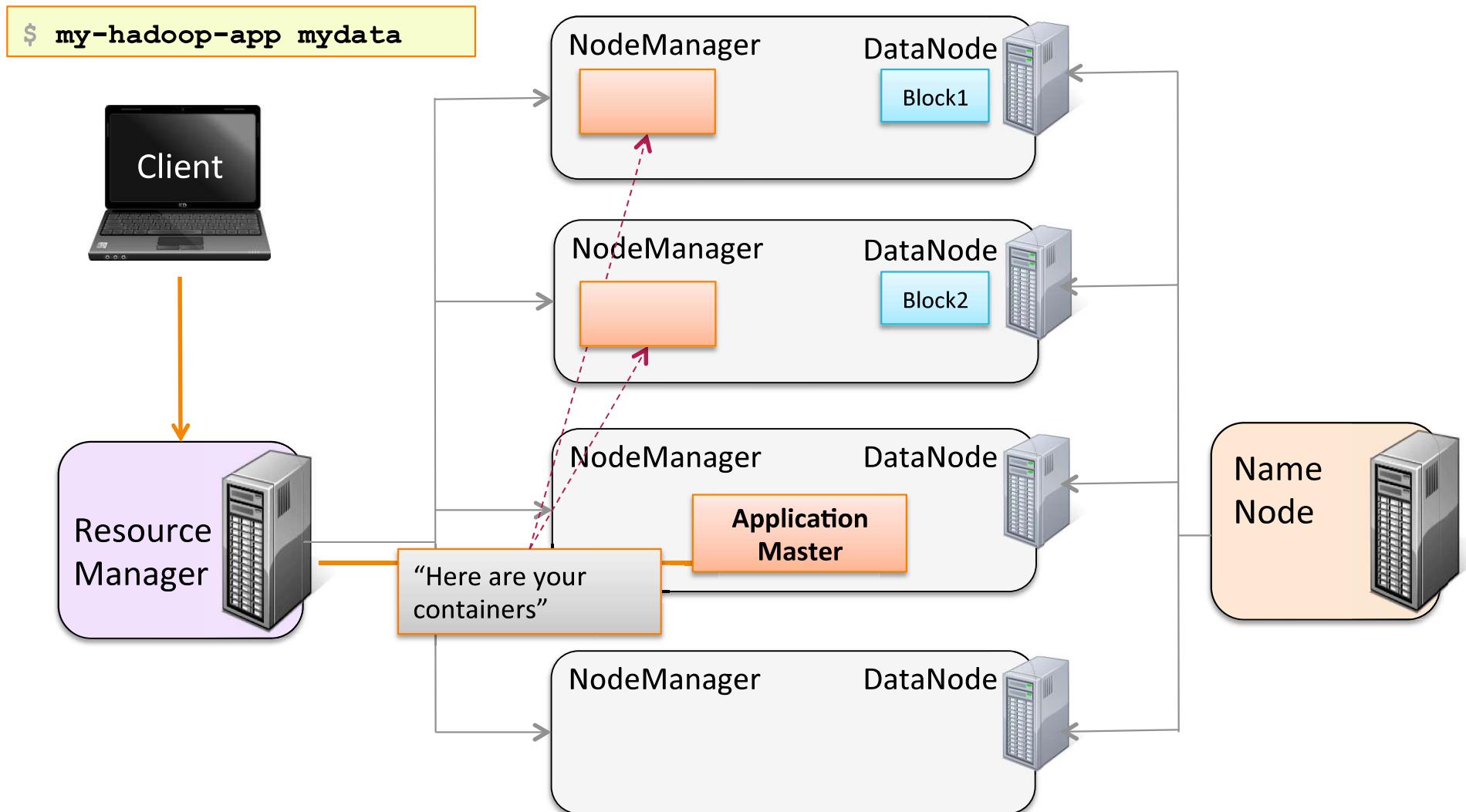
## Running an Application on YARN (4)



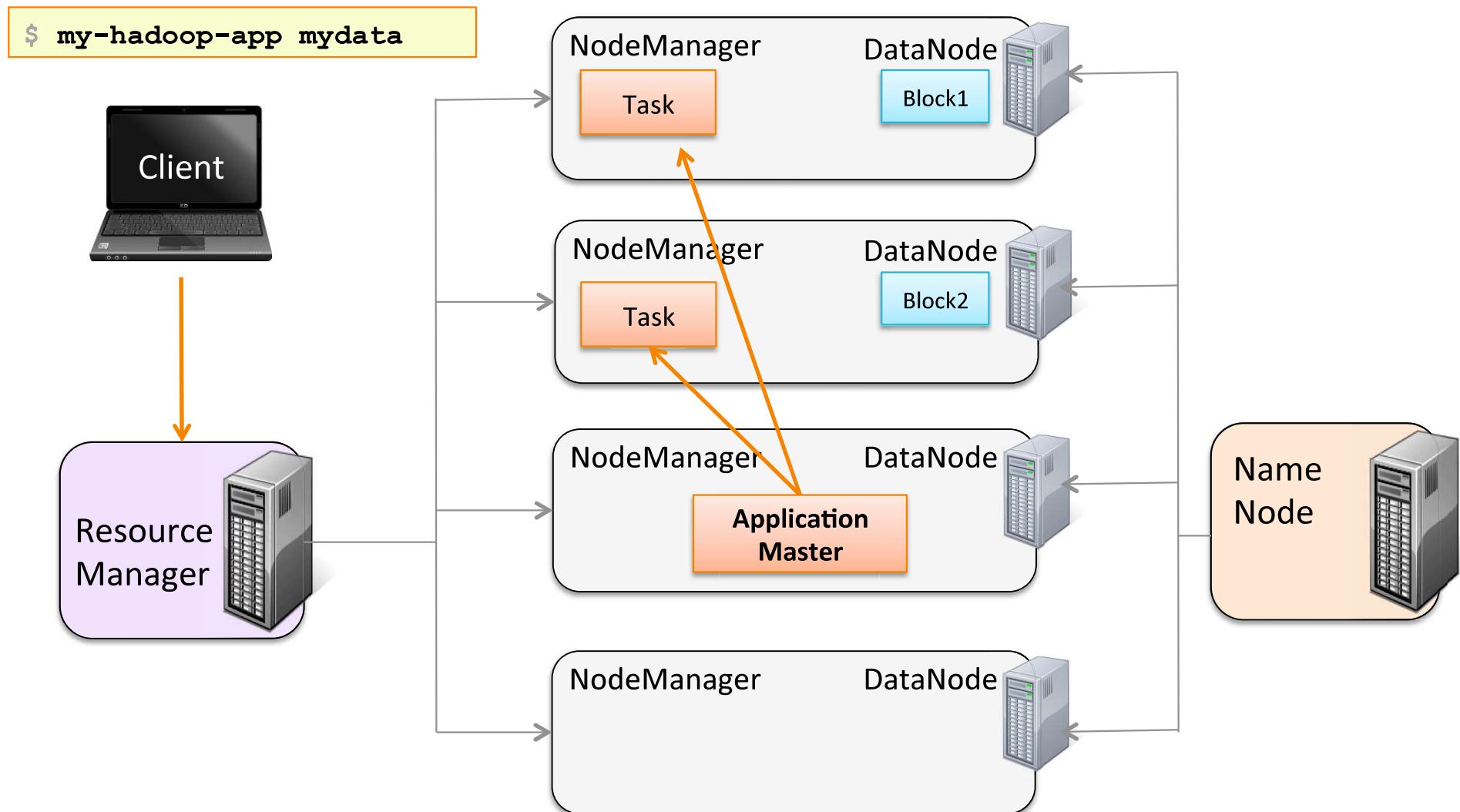
## Running an Application on YARN (5)



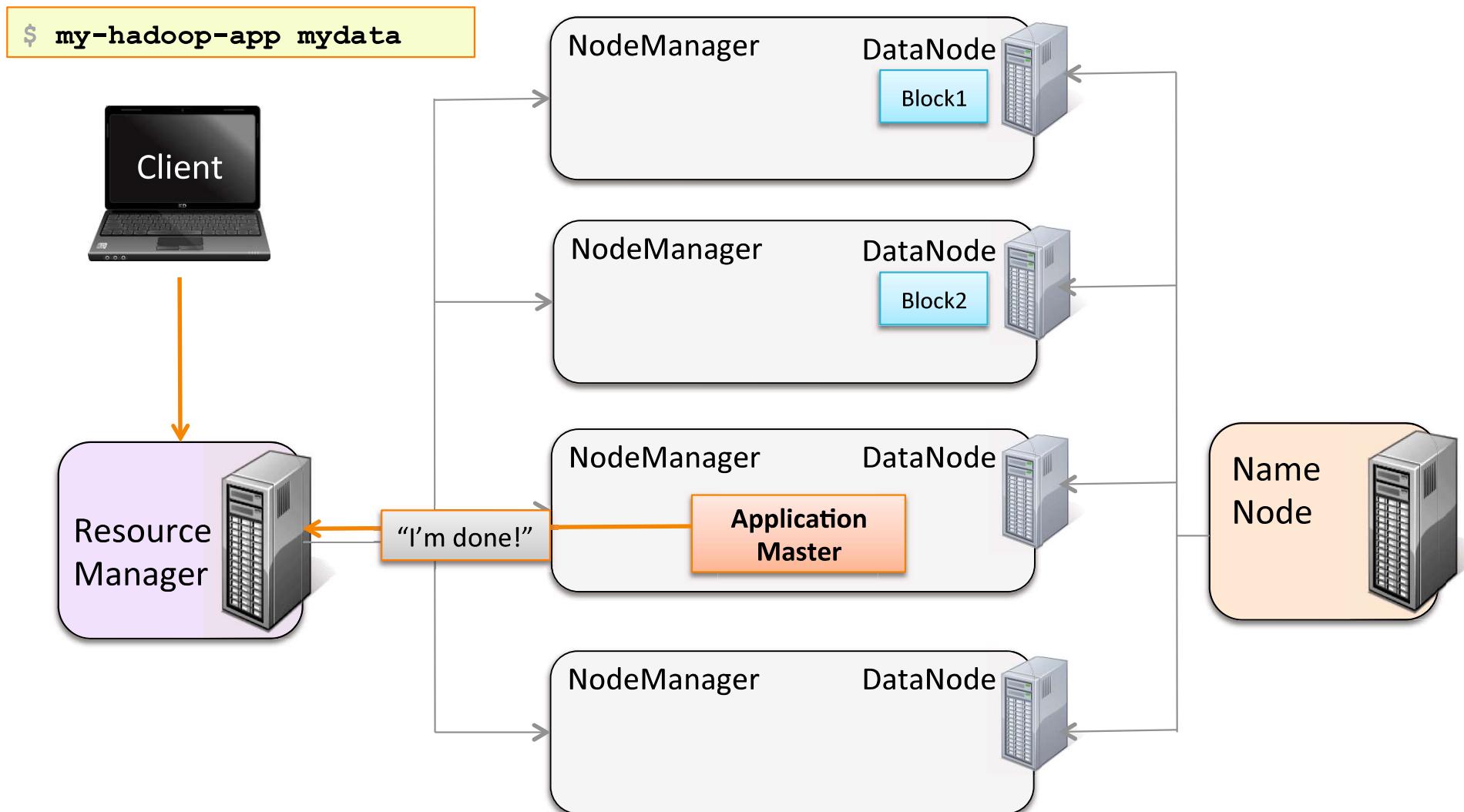
## Running an Application on YARN (6)



## Running an Application on YARN (7)



## Running an Application on YARN (8)



# Working With YARN

---

- **Developers need to be able to**
  - Submit jobs (applications) to run on the YARN cluster
  - Monitor and manage jobs
- **Hadoop includes three major YARN tools for developers**
  - The Hue Job Browser
  - The YARN Web UI
  - The YARN command line
- **YARN administrators can use Cloudera Manager**
  - May also be helpful for developers
  - Included in Cloudera Express and Cloudera Enterprise
  - Not covered in this course

# The Hue Job Browser

- The Hue Job Browser allows you to
  - Monitor the status of a job
  - View the logs
  - Kill a running job

The screenshot shows the Hue Job Browser interface. At the top, there's a navigation bar with links for 'Query Editors', 'Data Browsers', 'Workflows', and 'Search'. To the right of the search bar is a button with a small icon and the number '1', which is highlighted with a red box. Below the navigation bar is the title 'Job Browser' with a blue icon. On the left, there are search filters for 'Username' (set to 'training') and 'Text' (with a placeholder 'Search for text'). To the right of these filters are four colored buttons: green ('Succeeded'), orange ('Running'), red ('Failed'), and dark grey ('Killed'). The main area displays a table of job entries:

Logs	ID	Name	Status	User	Maps	Reduces	Queue	Priority	Duration	Submitted	Action
<a href="#">Logs</a>	<a href="#">1424901249645_0002</a>	webpage.jar	<span>RUNNING</span>	training	<div style="width: 5%;">5%</div>	<div style="width: 5%;">5%</div>	root.training	N/A	18s	03/06/15 11:00:17	<a href="#">Kill</a>
<a href="#">Logs</a>	<a href="#">1424901249645_0001</a>	accounts.jar	<span>SUCCEEDED</span>	training	<div style="width: 100%;">100%</div>	<div style="width: 100%;">100%</div>	root.training	N/A	1m:21s	03/06/15 10:57:48	

At the bottom, it says 'Showing 1 to 2 of 2 entries' and has navigation buttons for 'Previous', 'Next', and page number '1'. A scroll bar is visible on the right side of the table.

## The YARN Web UI

---

- **Resource Manager UI is the main entry point**
  - Runs on the RM host on port 8080 by default
- **Provides more detailed view than Hue**
- **Does not provide any control or configuration**

also try 8088 (in recent versions)

# Resource Manager UI: Nodes

Logged in as: dr.who

## Nodes of the cluster

### Cluster Overview

**Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
4	0	1	3	8	8 GB	8 GB	2 GB	2	0	0	0	0

**User Metrics for dr.who**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved
0	0	1	3	0	0	0	0 B	0 B	0 B

Show 20 entries Search:

Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Mem Used	Mem Avail
/default-rack	RUNNING	qsslave1:8041	qsslave1:8042	21-Nov-2013 13:07:26		4	4 GB	0 B
/default-rack	RUNNING	qsmaster:8041	qsmaster:8042	21-Nov-2013 13:07:17		4	4 GB	0 B

Showing 1 to 2 of 2 entries First Previous 1 Next Last

link to Node Manager UI

List of each node in cluster

# Resource Manager UI: Applications

The screenshot shows the Hadoop Resource Manager UI with the title "All Applications". The left sidebar has sections for Cluster (About, Nodes, Applications, Scheduler, Tools), Cluster Metrics, User Metrics for dr.who, and a table of running applications. A purple callout box points to one of the application rows.

**Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
8	0	1	7	5	6 GB	8 GB	0 B	1	0	0	0	0

**User Metrics for dr.who**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved
0	0	1	7	0	0	0	0 B	0 B	0 B

**Applications**

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
application_1384200217415_0009	training	Process Logs	MAPREDUCE	root.training	Tue, 12 Nov 2013 18:54:38 GMT	N/A	RUNNING	UNDEFINED	<div style="width: 100px; height: 10px;"></div>	ApplicationMaster
application_1384200217415_0008	training	Average Word Length	MAPREDUCE	root.training	Mon, 11 Nov 2013 21:55:21 GMT	Mon, 11 Nov 2013 21:57:30 GMT	FINISHED	SUCCEEDED	<div style="width: 100px; height: 10px;"></div>	History
application_1384200217415_0007	training	Process Logs	MAPREDUCE	root.training	Mon, 11 Nov 2013 21:36:39 GMT	Mon, 11 Nov 2013 21:44:19 GMT	FINISHED	SUCCEEDED	<div style="width: 100px; height: 10px;"></div>	History
application_1384200217415_0006	training	Process Logs	MAPREDUCE	root.training	Mon, 11 Nov 2013	Mon, 11 Nov 2013	FINISHED	SUCCEEDED	<div style="width: 100px; height: 10px;"></div>	History

**Link to Application Details... (next slide)**

**List of running and recent applications**

## Resource Manager UI: Application Detail

Logged in as: dr.who

**hadoop**

**Cluster**

- About
- Nodes
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- REMOVING
- FINISHING
- FINISHED
- FAILED
- KILLED

Scheduler

Tools

**Application Overview**

User:	training
Name:	Process Logs
<b>Application Type:</b>	MAPREDUCE
<b>State:</b>	RUNNING
<b>FinalStatus:</b>	UNDEFINED
<b>Started:</b>	12-Nov-2013 13:54:38
<b>Elapsed:</b>	38sec
<b>Tracking URL:</b>	<a href="#">ApplicationMaster</a>
<b>Diagnostics:</b>	

**ApplicationMaster**

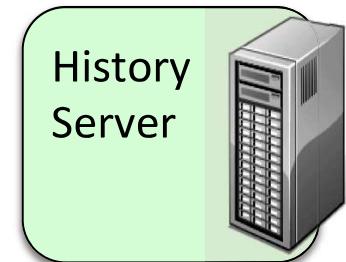
Attempt Number	Start Time	Node	Logs
1	12-Nov-2013 13:54:38	<a href="#">localhost.localdomain:8042</a>	<a href="#">logs</a>

Link to Application Master  
(UI depends on specific framework)

View aggregated log files (optional)

# Job History Server

- YARN does not keep track of job history
- Spark and MapReduce each provide a Job History Server
  - Archives job's metrics and metadata
  - Can be accessed through Job History UI or Hue



Logged in as: dr.who

**JobHistory**

**Retired Jobs**

Show 20 entries Search:

Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
2013.11.21 13:07:38 PST	2013.11.21 13:08:27 PST	job_1385066116114_0004	Process Logs	cloudera	default	SUCCEEDED	4	4	12	12
2013.11.21 13:03:53 PST	2013.11.21 13:04:42 PST	job_1385066116114_0003	Process Logs	cloudera	default	SUCCEEDED	4	4	12	12
2013.11.21 13:01:35 PST	2013.11.21 13:02:28 PST	job_1385066116114_0002	Process Logs	cloudera	default	SUCCEEDED	4	4	12	12
2013.11.21 12:48:00 PST	2013.11.21 12:50:43 PST	job_1385066116114_0001	Word Count	cloudera	default	SUCCEEDED	4	4	1	1
2013.11.21 09:24:45 PST	2013.11.21 09:28:19 PST	job_1385049040288_0003	Word Count	cloudera	default	SUCCEEDED	4	4	1	1

## YARN Command Line

---

- Command to configure and view information about the YARN cluster
  - **yarn <command>**
- Most YARN command line tools are for administrators rather than developers
- Some helpful commands for developers
  - **yarn application**
    - Use **-list** to see running applications
    - Use **-kill** to kill a running application
  - **yarn logs -applicationId <app-id>**
    - View the logs of the specified application

```
$ yarn app -list
```

# Bibliography

---

The following offer more information on topics discussed in this chapter

- ***Hadoop Application Architectures: Designing Real-World Big Data Applications* (published by O'Reilly)**
  - <http://tiny.cloudera.com/archbook>
- **HDFS User Guide**
  - <http://tiny.cloudera.com/hdfsuser>
- **YARN documentation**
  - <http://tiny.cloudera.com/yarndocs>
- **Cloudera Engineering Blog YARN articles**
  - <http://tiny.cloudera.com/yarnblog>