

BUAN 6346 Big Data

Homework 2

Run a YARN job

Step 1: Visit YARN resource Manager(RM)



All Applications

Logged in as: dr.who

Cluster

About Nodes

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

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

Show 20

entries

Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Progress	Tracking UI

No data available in table

Showing 0 to 0 of 0 entries

First

Previous

Next

Last

Step2: Cluster Metrics

Cluster Metrics

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

Step3: Nodes of the Cluster



Nodes of the cluster

Logged in as: dr.who

Cluster

About Nodes

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

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

Show 20 entries

Search:

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Mem Used	Mem Avail	VCores Used	VCores Avail	Version
/default-rack		RUNNING	localhost:53544	localhost:8042	Sun Jun 07 09:56:37 -0700 2020		0	0 B	2 GB	0	8	2.6.0-cdh5.4.3

Showing 1 to 1 of 1 entries

FirstPrevious1NextLast

Step4: Node Manager UI



Logged in as: dr.who

Resource Manager
Node Manager
Node Information
List of Applications
List of Containers
Tools

NodeManager Information

Total Vmem allocated for Containers 4.20 GB

Vmem enforcement enabled false

Total Pmem allocated for Container 2 GB

Pmem enforcement enabled true

Total Vcores allocated for Containers 8

NodeHealthyStatus true

LastNodeHealthTime Sun Jun 07 10:00:37 PDT 2020

NodeHealthReport

Node Manager Version: 2.6.0-cdh5.4.3 from 4cd9f51a3f1ef748d45b8d77d0f211ad44296d4b by jenkins source checksum 8bacc650a3181bc0cfbae816dc81eddd on 2015-06-25T02:41Z

Hadoop Version: 2.6.0-cdh5.4.3 from 4cd9f51a3f1ef748d45b8d77d0f211ad44296d4b by jenkins source checksum 4acea6ac185376e0b48b33695e88e7a7 on 2015-06-25T02:34Z

Step5: Back to Resource Manager Home



Logged in as: dr.who

Cluster
About Nodes
Applications
NEW
NEW SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools

All Applications

Cluster Metrics

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

Show 20 entries Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	Progress	Tracking UI
No data available in table													

Showing 0 to 0 of 0 entries First Previous Next Last

Step6: Change to exercise directory

```
[training@localhost ~]$ cd $DEV1/exercises/yarn
[training@localhost yarn]$ ls -l
total 4
-rwxr-xr-x 1 training training 1320 Aug 27 2015 wordcount.py
[training@localhost yarn]$
```

Step7: Run the example wordcount.py program on the YARN cluster to count the frequency of words

```
[training@localhost yarn]$ spark-submit --master yarn-cluster \
> wordcount.py /loudacre/kb/*
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/zookeeper/lib/slf4j-log4j12-1.7.5.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/flume-ng/lib/slf4j-log4j12-1.7.5.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
20/06/07 10:22:43 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
20/06/07 10:22:46 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0
:8032
20/06/07 10:22:47 INFO yarn.Client: Requesting a new application from cluster wi
th 1 NodeManagers
20/06/07 10:22:47 INFO yarn.Client: Verifying our application has not requested
more than the maximum memory capability of the cluster (1024 MB per container)
20/06/07 10:22:47 INFO yarn.Client: Will allocate AM container, with 896 MB memo
ry including 384 MB overhead
20/06/07 10:22:47 INFO yarn.Client: Setting up container launch context for our
AM
20/06/07 10:22:47 INFO yarn.Client: Preparing resources for our AM container
20/06/07 10:22:49 WARN shortcircuit.DomainSocketFactory: The short-circuit local
reads feature cannot be used because libhadoop cannot be loaded.
20/06/07 10:22:49 INFO yarn.Client: Uploading resource file:/usr/lib/spark/lib/s
park-assembly-1.3.0-cdh5.4.3-hadoop2.6.0-cdh5.4.3.jar -> hdfs://localhost:8020/u
ser/training/.sparkStaging/application_1591548045040_0001/spark-assembly-1.3.0-c
dh5.4.3-hadoop2.6.0-cdh5.4.3.jar
█

20/06/07 10:24:15 INFO yarn.Client:
client token: N/A
diagnostics: N/A
ApplicationMaster host: 192.168.206.128
ApplicationMaster RPC port: 0
queue: root.training
start time: 1591550588022
final status: UNDEFINED
tracking URL: http://localhost:8088/proxy/application_1591548045040_0001/
user: training
20/06/07 10:24:16 INFO yarn.Client: Application report for application_1591548045040_0001 (state: RUNNING)
20/06/07 10:24:17 INFO yarn.Client: Application report for application_1591548045040_0001 (state: RUNNING)
20/06/07 10:24:18 INFO yarn.Client: Application report for application_1591548045040_0001 (state: RUNNING)
```

Step8: Job Browser

The screenshot shows the HUE Job Browser interface. At the top, there's a navigation bar with 'HUE' logo and tabs for 'Query Editors', 'Data Browsers', 'Workflows', and 'Search'. Below this, there's a 'Job Browser' header. A search bar contains 'Username: training' and 'Text: Search for text'. On the right, there are status filters: 'Succeeded', 'Running', 'Failed', and 'Killed'. The main table lists job details:

Logs	ID	Name	Status	User	Maps	Reduces	Queue	Priority	Duration	Submitted	
	1591548045040_0001	wordcount.py	RUNNING	training	00%	00%	root.training	N/A	3m:27s	06/07/20 10:23:08	Kill

At the bottom, it says 'Showing 1 to 1 of 1 entries' with navigation links for 'Previous', '1', and 'Next'.

Step9: Job Browser

This screenshot shows the HUE Job Browser with two job entries. The search filters remain the same. The table now lists two jobs:

Logs	ID	Name	Status	User	Maps	Reduces	Queue	Priority	Duration	Submitted	
	1591548045040_0002	wordcount.py	RUNNING	training	00%	00%	root.training	N/A	20s	06/07/20 10:31:03	Kill
	1591548045040_0001	wordcount.py	SUCCEEDED	training	100%	100%	root.training	N/A	3m:37s	06/07/20 10:23:08	

The bottom status shows 'Showing 1 to 2 of 2 entries'.

Step 10: Running Application

This screenshot shows the detailed view of a job in the HUE Job Browser. On the left, a sidebar contains links for 'APP ID', 'USER', 'STATUS', 'LOGS', 'PROGRESS', 'DURATION', and 'ACTIONS'. The main area displays the job name 'wordcount.py' and a 'Metadata' tab. Below the tab, a table shows job metadata:

Name	Value
Jobs	http://localhost:8089/proxy/application_1591548045040_0002/
Host	http://localhost:8042
Queue Name	root.training
Started	06/07/20 10:31:03
Finished	06/07/20 10:31:52
Pre-empted Resource Vcores	0
VCore seconds	82
Memory seconds	85676
Diagnostics	

The 'ACTIONS' section at the bottom left includes a 'Kill this job' button.

Step 11: YARN RM



All Applications

Logged in as: dr.who

Cluster

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

Tools

Cluster Metrics

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

Show 20 entries

Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Progress	Tracking UI
application_1591548045040_0002	training	wordcount.py	SPARK	root.training	Sun Jun 7 10:31:03 -0700 2020	N/A	RUNNING	UNDEFINED	2	2	2048		ApplicationMaster
application_1591548045040_0001	training	wordcount.py	SPARK	root.training	Sun Jun 7 10:23:08 -0700 2020	Sun Jun 7 10:26:45 -0700 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A		History

Showing 1 to 2 of 2 entries

First Previous 1 Next Last

Step12: Nodes



Nodes of the cluster

Logged in as: dr.who

Cluster

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

Tools

Cluster Metrics

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

Show 20 entries

Search:

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Mem Used	Mem Avail	VCores Used	VCores Avail	Version
/default-rack	RUNNING	localhost:53544	localhost:8042		Sun Jun 07 10:30:37 -0700 2020		2	2 GB	0 B	2	6	2.6.0-cdh5.4.3

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

Step13: Statistics about node



Logged in as: dr.who

ResourceManager

- NodeManager
- Node Information
- List of Applications
- List of Containers

Tools

NodeManager information

Total Vmem allocated for Containers	4.20 GB
Vmem enforcement enabled	false
Total Pmem allocated for Container	2 GB
Pmem enforcement enabled	true
Total VCoers allocated for Containers	8
NodeHealthyStatus	true
LastNodeHealthTime	Sun Jun 07 10:32:37 PDT 2020
NodeHealthReport	
Node Manager Version:	2.6.0-cdh5.4.3 from 4cd9f51a3f1ef748d45b8d77d0f211ad44296d4b by jenkins source checksum 8bacc650a3181bc0cfbae816dc81eddd on 2015-06-25T02:41Z
Hadoop Version:	2.6.0-cdh5.4.3 from 4cd9f51a3f1ef748d45b8d77d0f211ad44296d4b by jenkins source checksum 4acea6ac185376e0b48b33695e88e7a7 on 2015-06-25T02:34Z

Step14: List of Applications**Applications running on this node**

Logged in as: dr:who

▼ ResourceManager	Show: 20 ▾ entries	Search: <input type="text"/>
RM Home	ApplicationId	ApplicationState
► NodeManager	application_1591548045040_0002	RUNNING
► Tools	Showing 1 to 1 of 1 entries	
		First Previous 1 Next Last

Step15: List of Containers

Logged in as: dr:who

▼ ResourceManager

RM Home

► NodeManager

► Tools

Application's information

ApplicationId

application_1591548045040_0002

ApplicationState

RUNNING

User

training

Show 20 ▾ entries

Search:

container_1591548045040_0002_01_000001

container_1591548045040_0002_01_000002

Showing 1 to 2 of 2 entries

First Previous 1 Next Last