

# Exercise 02:

## Spark on YARN



## Understanding the difference between YARN Client and Cluster Modes

**Objective:** To test the consequences of making changes to container sizes, Maximum Application settings, and queue states; configure application preemption and create a set of queues and leaf queues which logically represent the organizations you support and their SLAs

1. Open terminal
2. Change to the `/opt/cloudera/parcels/SPARK3/lib/spark3/examples/jars` directory.

```
cd /opt/cloudera/parcels/SPARK3/lib/spark3/examples/jars
```

3. Open a second terminal window.
4. Once again, change to the

`/opt/cloudera/parcels/SPARK3/lib/spark3/examples/jars` directory.

```
cd /opt/cloudera/parcels/SPARK3/lib/spark3/examples/jars
```

5. Position the two terminal windows so that both are visible on your screen side-by-side.
6. Import sherlock.txt data into HDFS

```
hdfs dfs -mkdir data/  
hdfs dfs -put ~/training/data/sherlock.txt data/
```

7. Type the Spark Submit command to execute wordcount in both terminal windows but do not press Enter to execute it.

### Terminal One

```
spark3-submit --class org.apache.spark.examples.JavaWordCount --master yarn  
--deploy-mode client spark-examples_2.12-3.0.0.2.99.0.0-23.jar  
data/sherlock.txt output1
```

```

felt,: 1
clamped: 2
understand: 23
Wimpole: 1
traced,: 1
faintly: 1
prejudice: 1
parish: 2
later,: 2
sir?: 2
"she: 1
preparations,: 1
innocence.": 1
conclusive.": 1
scratching: 1
whereabouts.: 1
indicated--that: 1
companion: 12
silence,: 7
Upper: 3
absolutely: 24
young: 73
19/10/27 12:28:17 INFO AbstractConnector: Stopped Spark@1c25b8a7(HTTP/1.1,[http/1.1]){0.0.0.0:4041}
19/10/27 12:28:17 INFO SparkUI: Stopped Spark web UI at http://datacouch.training.io:4041
19/10/27 12:28:17 INFO YarnClientSchedulerBackend: Interrupting monitor thread
19/10/27 12:28:17 INFO YarnClientSchedulerBackend: Shutting down all executors

```

## List Application

- While Application is running, open another terminal window and execute below command:

```
$ yarn application -list
```

## Output

```

[training@datacouch ~]$ yarn application -list
19/10/27 12:34:13 INFO client.RMProxy: Connecting to ResourceManager at datacouch.training.io/172.31.0.206:8050
19/10/27 12:34:13 INFO client.AHSProxy: Connecting to Application History server at datacouch.training.io/172.31.0.206:10200
Total number of applications (application-types: [], states: [SUBMITTED, ACCEPTED, RUNNING] and tags: []):1

```

Application-Id	Application-Name	Application-Type	User	Queue
application_1572179256041_0003	JavaWordCount	SPARK	training	default

```


```

To view Logs of the Application, execute below command:

```
$ yarn logs -applicationId <application_id>
```

```
yarn logs -applicationId {your application id}
```

```

End of LogType:prelaunch.err
*****

Container: container_e12_1572179256041_0003_01_000003 on datacouch.training.io_45454_1572179668051
LogAggregationType: AGGREGATED
=====
LogType:prelaunch.out
LogLastModifiedTime:Sun Oct 27 12:34:28 +0000 2019
LogLength:100
LogContents:
Setting up env variables
Setting up job resources
Copying debugging information
Launching container

End of LogType:prelaunch.out
*****

Container: container_e12_1572179256041_0003_01_000003 on datacouch.training.io_45454_1572179668051
LogAggregationType: AGGREGATED
=====
LogType:stderr
LogLastModifiedTime:Sun Oct 27 12:34:28 +0000 2019
LogLength:7145
LogContents:
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/hadoop/yarn/local/filecache/10/spark2-hdp-yarn-archive.tar.gz/slf4j-log4j12-1.7.16.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/3.14.0-315/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

```

## Kill an Application

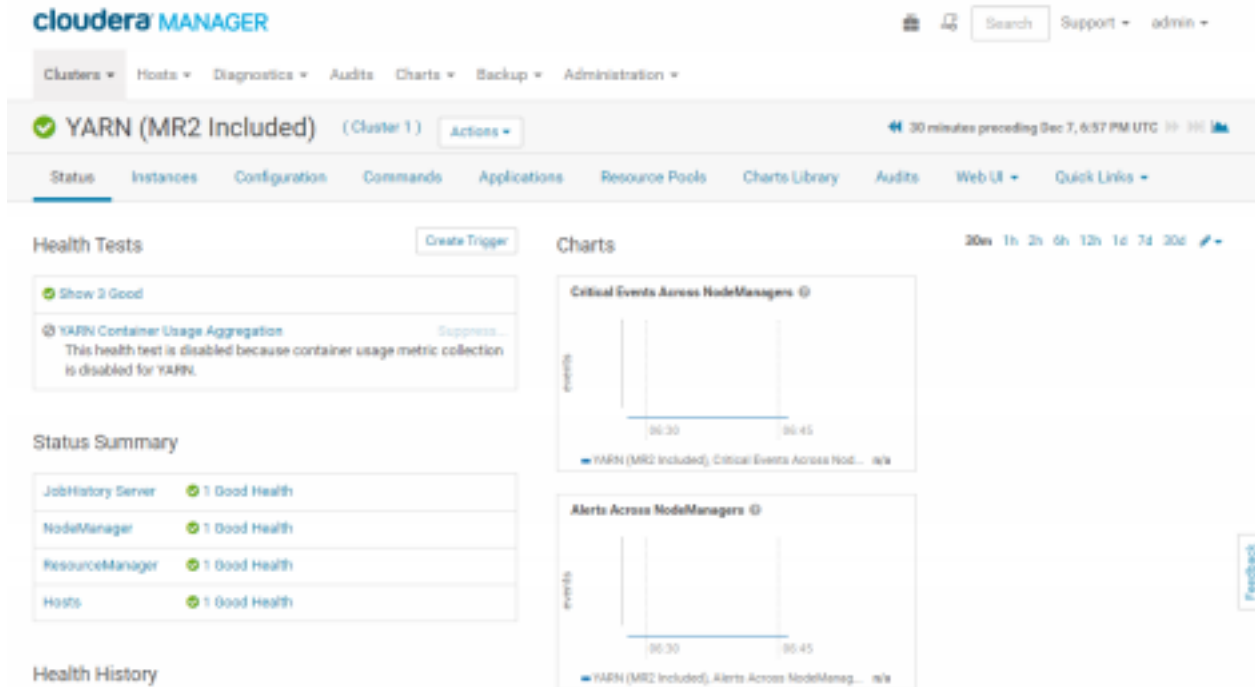
- Below command kills a running application -

```
$ yarn application -kill <application_id>
```

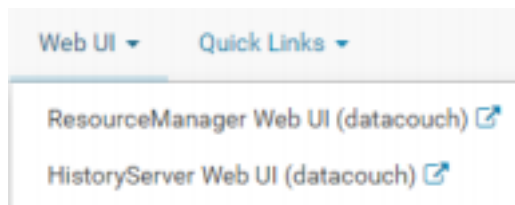
```
yarn application -kill {your application id}
```

## Exploring the YARN Cluster

- Open the browser and connect to the CM at the URL <http://localhost:7180>
- Login to the CM UI using the username admin and the password admin
- Click Services in the CM UI.
- Select the YARN service on the Services page.



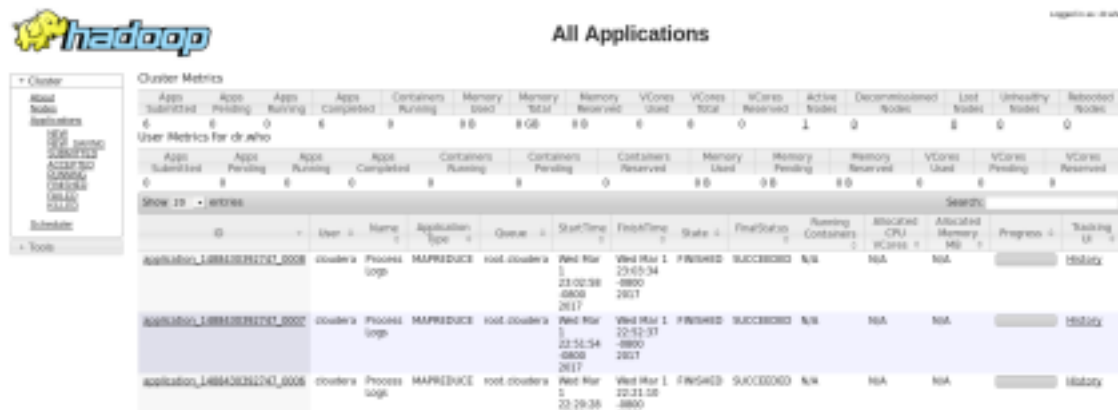
14. Click Web UI and select ResourceManager UI.



The ResourceManager UI Web interface opens in another browser tab.

**NOTE:** If simply clicking the quick link fails to open the ResourceManager UI, replace the default URL in the browser tab with

<http://training.io:8088>



15. This page will be refreshed in a moment once applications are running. Leave this tab open.

## Terminal Two

```
spark3-submit --class org.apache.spark.examples.JavaWordCount
--master yarn --deploy-mode cluster
spark-examples_2.12-3.0.0.2.99.0.0-23.jar data/sherlock.txt output2
```

Open ResourceManager UI Web interface in another browser tab and click on application ID

Cluster Metrics									
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Memory Pending	Memory Reserved
1	0	1	0	2	3 GB	3 GB	0 B	0 B	0 B
Cluster Nodes Metrics									
Active Nodes		Decommissioning Nodes		Decommissioned Nodes		Lost Nodes			
1	0	0	0	0	0	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	0	0	0	0	0	0 B	0 B	0 B
Show 20 entries									
ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers
application_1535000007290_0001	training	org.apache.spark.examples.javaWordCount	SPARK	root.user.training	Mon Dec 9 11:37:05 +0000 2019	N/A	RUNNING	UNDERNED	2
Showing 1 to 1 of 1 entries									

Application Overview	
User:	training
Name:	org.apache.spark.examples.javaWordCount
Application Type:	SPARK
Application Tags:	
State:	FINISHED
FinalStatus:	SUCCEEDED
Started:	Mon Dec 09 11:37:05 +0000 2019
Elapsed:	37sec
Tracking URL:	<a href="#">History</a>
Diagnostics:	
Application Metrics	
Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	71162 MB-seconds, 46 vcore-seconds
ApplicationMaster	
Attempt Number	Start Time
1	Mon Dec 09 11:37:05 +0000 2019
Node	
datacouch.training.io.8042	
Logs	

Log type: stdout  
Log Upload Time: Mon Dec 09 11:37:45 +0000 2019  
Log Length: 173237  
Showing 4096 bytes of 173237 total. Click [here](#) for the full log.

```
: 9
ordered: 9
conspiring.: 1
packet.: 1
other.: 9
newcomers: 2
wait.: 1
London.': 1
Park.': 1
excavating: 1
I.C.: 1
remarked.: 15
breathed: 1
Kate--poor: 1
altar.: 1
boarding-schools.: 1
stock: 3
prisoner: 7
know--faddy: 1
dad.': 1
"Very.: 1
derivative: 3
exert: 1
discoloured.: 1
Between: 3
rate: 1
know.': 1
fairbank.: 1
bow-window: 1
hair: 27
till: 2
timid: 2
disappointment.: 2
ill-dressed: 1
ORANGE: 1
```

## End of Exercise