

Máster en Big Data

Tecnologías de Almacenamiento

6. Hands-On: Ejecución Apache Spark

Presentado por:

José David Angulo y Albert Ripoll

Índice

1. Introducción	3
2. Entorno	3
3. Ejecución Spark	4
4. SparkUI	5

1. Introducción

El objetivo de este Hands-On es aprender a ejecutar un job de Spark mediante un ejemplo precompilado.

2. Entorno

Para la realización de este hands-on, vamos a desplegar un sandbox de MapR en VMWare.

Una vez arrancado el nodo, las credenciales son:

Usuario: *mapr*; Password: *mapr*

Busca la IP de la máquina virtual y accede a HUE <http://<ip>:8888/> y abre una conexión ssh para más comodidad (puedes añadir la ip en hosts con el nombre *maprdemo*).

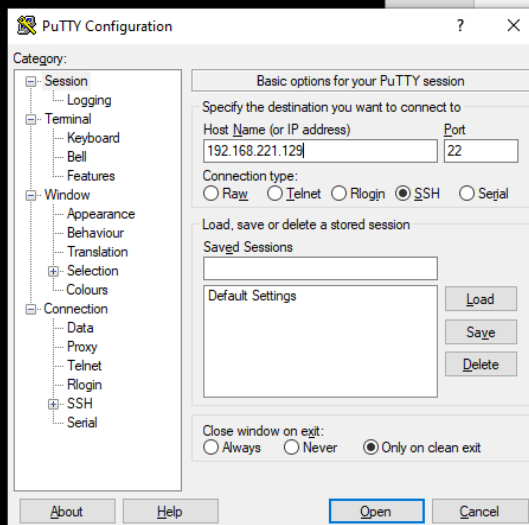
Usuario: *mapr*; Password: *mapr*

ifconfig

```
maprdemo login: mapr
Password:
Last login: Mon Apr 15 11:12:18 on
Welcome to your MapR Demo virtual machine.
[mapr@maprdemo ~]$ ifconfig
ens32: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.221.129 netmask 255.255.255.0 broadcast 192.168.221.255
    inet6 fe80::8fdf:438e:37a3:110c prefixlen 64 scopeid 0x20<link>
    ether 00:0c:29:00:57:e3 txqueuelen 1000 (Ethernet)
    RX packets 66 bytes 7495 (7.3 KiB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 74 bytes 6925 (6.7 KiB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet 127.0.0.1 netmask 255.0.0.0
    inet6 ::1 prefixlen 128 scopeid 0x10<host>
    loop txqueuelen 1 (Local Loopback)
    RX packets 126676 bytes 10879176 (10.3 MiB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 126676 bytes 10879176 (10.3 MiB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

[mapr@maprdemo ~]$
```



Después de ifconfig → la ip de inet la escribimos en host Name or IP address de PuTTY configuration.

```
login as: mapr
Keyboard-interactive authentication prompts from server:
| Password:
End of keyboard-interactive prompts from server
Last login: Mon Apr 15 11:12:58 2024
Welcome to your MapR Demo virtual machine.
```

Volvemos a poner el logging y pass mapr, mapr

3. Ejecución Spark

La ubicación de Spark dentro de la distribución MapR la podemos encontrar en:

`/opt/mapr/spark/spark-2.1.0/`

Se continua después del logging

`cd /opt/mapr/spark/spark-2.1.0/`

```
Last login: Mon Apr 15 11:12:58 2024
Welcome to your MapR Demo virtual machine.
[mapr@maprdemo ~]$ /opt/mapr/spark/spark-2.1.0/
-bash: /opt/mapr/spark/spark-2.1.0/: Is a directory
[mapr@maprdemo ~]$ ls
[mapr@maprdemo ~]$ cd /opt/mapr/spark/spark-2.1.0/
[mapr@maprdemo spark-2.1.0]$
```

Se pide ejecutar el programa SparkPi, contra YARN en modo cliente, con una memoria por ejecutor de 1g.

El jar ya compilado lo podemos encontrar en: `/opt/mapr/spark/spark-2.1.0/examples/jars/spark-examples_2.11-2.1.0-mapr-1710.jar`

Y el package y nombre de la clase es: `org.apache.spark.examples.SparkPi`

El programa acepta por parámetro el número de paralelización que queremos darle, vamos a pasarle, por ejemplo, 4.

Engancha aquí el comando para enviar el Job y una captura de pantalla con la salida.

- En la diapositiva 36 del powerpoint “3. Apache spark” vemos:

```
./bin/spark-submit \
--class org.apache.spark.examples.SparkPi \
--master local[8] \
/path/to/examples.jar \
100
```

```
export HADOOP_CONF_DIR=XXX
./bin/spark-submit \
--class org.apache.spark.examples.SparkPi \
--master yarn \
--deploy-mode cluster \ # can be client for client mode
--executor-memory 20G \
--num-executors 50 \
/path/to/examples.jar \
1000
```

- En la diapositiva 37 del powerpoint “3. Apache spark” vemos:

`spark-submit --jars lib/json-simple-1.1.1.jar -- class com.pragsis.ccma.hbbtv.HbbtvProcess`

`--master yarn-client --executor-memory 15H ccma.jar -i /user/devel/data/pending/ -o /user/devel/data/`

- Por lo que para responder a las condiciones de la práctica se elabora el siguiente código para ejecutar.

`./bin/spark-submit \`

```
--class org.apache.spark.examples.SparkPi \  
--master yarn-client \  
/opt/mapr/spark/spark-2.1.0/examples/jars/spark-examples_2.11-2.1.0-mapr-1710.jar \  
4
```

```
[mapr@maprdemo spark-2.1.0]$ ./bin/spark-submit \  
> --class org.apache.spark.examples.SparkPi \  
> --master yarn-client \  
> /opt/mapr/spark/spark-2.1.0/examples/jars/spark-examples_2.11-2.1.0-mapr-1710.jar \  
> 4  
Warning: Unable to determine $DRILL_HOME  
Warning: Master yarn-client is deprecated since 2.0. Please use master "yarn" with specified deploy mode instead.  
24/04/15 11:39:59 WARN Utils: Your hostname, maprdemo.local resolves to a loopback address: 127.0.0.1; using 192.168.221.129 instead (on interface ens32)  
24/04/15 11:39:59 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address  
24/04/15 11:40:01 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK_HOME.  
Pi is roughly 3.143077857694644  
[mapr@maprdemo spark-2.1.0]$
```

Después se cuelga con la ip original que era 192.168.221.129 y se vuelve a repetir todo otra vez con lo que cambia la ip a 192.168.221.131 y sale

```
[mapr@maprdemo spark-2.1.0]$ ./bin/spark-submit \  
> --class org.apache.spark.examples.SparkPi \  
> --master yarn-client \  
> /opt/mapr/spark/spark-2.1.0/examples/jars/spark-examples_2.11-2.1.0-mapr-1710.jar \  
> 4  
Warning: Unable to determine $DRILL_HOME  
Warning: Master yarn-client is deprecated since 2.0. Please use master "yarn" with specified deploy mode instead.  
24/04/15 12:16:00 WARN Utils: Your hostname, maprdemo.local resolves to a loopback address: 127.0.0.1; using 192.168.221.131 instead (on interface ens32)  
24/04/15 12:16:00 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address  
24/04/15 12:16:02 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK_HOME.  
[Stage 0:> (0 + 0) / 4]  
24/04/15 12:17:03 WARN YarnScheduler: Initial job has not accepted any resources; check your cluster UI to ensure that workers are registered and have sufficient resources  
Pi is roughly 3.142577856444641  
[mapr@maprdemo spark-2.1.0]$
```

Después de colgarse otra vez se repite el proceso y la nueva ip es a 192.168.221.132

Se tiene que repetir para 2, 4 y 8 jobs.

4. SparkUI

Consulta la *SparkUI* después de ejecutar el Job y responde a las siguientes preguntas.

Requisito: No accedas directamente a la interficie del job history de spark, hazlo mediante HUE (acceso user:mapr / pass: mapr)

¿Cuántas tasks se han ejecutado?

¿Cuántos executors se han ejecutado?

Para el job 4: ¿Cuántos RDDs se han creado y que transformaciones los han desencadenado?

Ejecuta el job con 2, 4 y 10 tasks y compara el rendimiento, ¿Que sucede en relación con la performance? ¿Porque se produce este comportamiento?

DETALLES CON Jobs = 2

```
./bin/spark-submit \  
--class org.apache.spark.examples.SparkPi \  
--master yarn-client \  
/opt/mapr/spark/spark-2.1.0/examples/jars/spark-examples_2.11-2.1.0-mapr-1710.jar \  
2
```

```
[mapr@maprdemo ~]$ cd /opt/mapr/spark/spark-2.1.0/  
[mapr@maprdemo spark-2.1.0]$  
[mapr@maprdemo spark-2.1.0]$ ./bin/spark-submit \  
> --class org.apache.spark.examples.SparkPi \  
> --master yarn-client \  
> /opt/mapr/spark/spark-2.1.0/examples/jars/spark-examples_2.11-2.1.0-mapr-1710.  
jar \  
> 2  
Warning: Unable to determine $DRILL_HOME  
Warning: Master yarn-client is deprecated since 2.0. Please use master "yarn" wi  
th specified deploy mode instead.  
24/04/15 12:36:27 WARN Utils: Your hostname, maprdemo.local resolves to a loopba  
ck address: 127.0.0.1; using 192.168.221.132 instead (on interface ens32)  
24/04/15 12:36:27 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another  
address  
24/04/15 12:36:28 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is  
set, falling back to uploading libraries under SPARK_HOME.  
Pi is roughly 3.1460957304786525
```



The screenshot shows the HUE Job Browser interface. At the top, there's a navigation bar with 'HUE' logo and tabs for 'Query Editors', 'Notebooks', 'Data Browsers', and 'Workflows'. Below this, the 'Job Browser' section is active. It includes a search bar with 'Benutzername' set to 'mapr' and a 'Text' search field. A status filter bar shows 'Succeeded' (selected), 'Running', 'Failed', and 'Killed', with a dropdown for 'in the last' set to '7'. Below the filters is a table of jobs.

Logs ID	Name	Application Type	Status	Benutzer	Maps	Reduces	Queue	Priority	Duration	Submitted
1713209692187_0001	Spark Pi	SPARK	SUCCEEDED	mapr	100%	100%	root.mapr	N/A	31s	04/15/24 12:36:32

APP ID

1713209692187_0001

TYPE

SPARK

BENUTZER

mapr

STATUS

SUCCEEDED

LOGS

Logs

PROGRESS

100.0%

DURATION

31s

Spark Pi

Metadata

Name	Value
Jobs	http://maprdemo:18080/history/application_1713209692187_0001/
Host	http://maprdemo:8042
Queue Name	root.mapr
Started	04/15/24 12:36:32
Finished	04/15/24 12:37:04
Pre-empted Resource VCores	0
VCore seconds	57
Memory seconds	110408
Diagnostics	

Para el job 2: ¿Cuántas tasks se han ejecutado? Dos

← → ↻ No es seguro 192.168.221.132:18080/history/application_1713209692187_0001/jobs/ ☆ ⓘ

spark 2.1.0-mapr-1710

Jobs Stages Storage Environment Executors

Spark Pi application UI

Spark Jobs (?)

User: mapr
Total Uptime: 34 s
Scheduling Mode: FIFO
Completed Jobs: 1
▶ Event Timeline

Completed Jobs (1)

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0	reduce at SparkPi.scala:38	2024/04/15 12:36:58	3 s	1/1	2/2

Al darle a reduce at SparkPi.scala:38

Details for Stage 0 (Attempt 0)

Total Time Across All Tasks: 76 ms
Locality Level Summary: Process local: 2
▶ DAG Visualization
▶ Show Additional Metrics
▶ Event Timeline

Summary Metrics for 2 Completed Tasks

Metric	Min	25th percentile	Median	75th percentile	Max
Duration	7 ms	7 ms	69 ms	69 ms	69 ms
GC Time	0 ms	0 ms	16 ms	16 ms	16 ms

▶ Aggregated Metrics by Executor

Executor ID	Address	Task Time	Total Tasks	Failed Tasks	Killed Tasks	Succeeded Tasks
1	maprdemo:34603	2 s	2	0	0	2

Tasks (2)

Index	ID	Attempt	Status	Locality Level	Executor ID / Host	Launch Time	Duration	GC Time	Errors
0	0	0	SUCCESS	PROCESS_LOCAL	1 / maprdemo	2024/04/15 12:36:58	69 ms	16 ms	
1	1	0	SUCCESS	PROCESS_LOCAL	1 / maprdemo	2024/04/15 12:37:00	7 ms		

The screenshot shows the Hadoop Resource Manager web interface. On the left, a sidebar menu includes 'ResourceManager', 'NodeManager', 'Node Information', 'List of Applications', 'List of Containers', and 'Tools'. The main content area displays the following details for a NodeManager:

- Total Vmem allocated for Containers: 10.50 GB
- Vmem enforcement enabled: false
- Total Pmem allocated for Container: 5.00 GB
- Pmem enforcement enabled: true
- Total VCores allocated for Containers: 0
- NodeHealthyStatus: true
- LastNodeHealthTime: Mon Apr 15 12:38:48 PDT 2024
- NodeHealthReport: Node Manager Version: 2.7.0-mapr-1710 from a59f1d36fa1be22772e0e48ce3ee8c9d41530287 by root source checksum d6c4897d604fe570f0cba357c71fd4f on 2017-11-10T02:09Z
- Hadoop Version: 2.7.0-mapr-1710 from a59f1d36fa1be22772e0e48ce3ee8c9d41530287 by root source checksum 89b868a4cd988c71b9b181bcc2a79ce on 2017-11-10T01:59Z

Para el job 2: ¿Cuántos executors se han ejecutado?

The screenshot shows the Hadoop Job History web interface. The breadcrumb trail indicates the path: '192.168.221.132:18080/history/application_1713209692187_0001/executors/'. The 'Executors' tab is selected. A table lists the executors with the following columns: Executor ID, Address, Status, RDD Blocks, Storage Memory, Disk Used, Cores, Active Tasks, Failed Tasks, Complete Tasks, Total Tasks, Task Time (GC Time), Input, Shuffle Read, Shuffle Write, and Logs. Two executors are listed: 'driver' and '1'.

Executor ID	Address	Status	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Logs
driver	192.168.221.132:37543	Active	0	0.0 B / 384.1 MB	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B	
1	mapdemo:34603	Active	0	0.0 B / 956.6 MB	0.0 B	1	0	0	2	2	2 s (16 ms)	0.0 B	0.0 B	0.0 B	stdout stderr

Para el job 2: ¿Cuántos RDDs se han creado y que transformaciones los han desencadenado?

The screenshot shows the Apache Spark web interface. The breadcrumb trail indicates the path: '192.168.221.132:18080/history/application_1713209692187_0001/executors/'. The 'Executors' tab is selected. A 'Summary' table provides an overview of RDDs across different executor states.

	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write
Active(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	2	2	2 s (16 ms)	0.0 B	0.0 B	0.0 B
Dead(0)	0	0.0 B / 0.0 B	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B
Total(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	2	2	2 s (16 ms)	0.0 B	0.0 B	0.0 B

DETALLES CON Jobs = 4

Entonces en el Google Chrome se escribe <http://192.168.221.131:8888/>

Y en job browser vemos el job.

The screenshot shows the Hive Job Browser web interface. The breadcrumb trail indicates the path: '192.168.221.131:8888/jobbrowser/'. The 'Job Browser' tab is selected. A table lists the jobs with the following columns: Logs ID, Name, Application Type, Status, Benutzer, Maps, Reduces, Queue, Priority, Duration, and Submitted. One job is listed with ID '1713207282087_0001'.

Logs ID	Name	Application Type	Status	Benutzer	Maps	Reduces	Queue	Priority	Duration	Submitted
1713207282087_0001	Spark PI	SPARK	SUCCEEDED	mapr	100%	100%	root.mapr	N/A	1m:24s	04/15/24 12:16:06

En view job vemos:

APP ID

1713207282087_0001

TYPE

SPARK

BENUTZER

mapr

STATUS

SUCCEEDED

LOGS

Logs

PROGRESS

100.0%

DURATION

1m:24s

Spark Pi

Metadata

Name	Value
Jobs	http://maprdemo:18080/history/application_1713207282087_0001/
Host	http://maprdemo:8042
Queue Name	root.mapr
Started	04/15/24 12:16:06
Finished	04/15/24 12:17:31
Pre-empted Resource VCores	0
VCores seconds	133
Memory seconds	239991
Diagnostics	

Y su accedemos a los jobs cambiando maprdemo con la ip.

hadoop

ResourceManager

NodeManager

Node Information

List of Applications

List of Containers

Tools

Total Vmem allocated for Containers

Vmem enforcement enabled

Total Pmem allocated for Container

Pmem enforcement enabled

Total VCores allocated for Containers

NodeHealthyStatus

LastNodeHealthTime

NodeHealthReport

Node Manager Version:

Hadoop Version:

192.168.221.131:8042/node

No es seguro

192.168.221.131:18080/history/application_1713207282087_0001/executors/

No es seguro

spark

2.1.0-mapr-1710

Jobs

Stages

Storage

Environment

Executors

Spark Pi applicati

Executors

Summary

	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write
Active(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B
Dead(0)	0	0.0 B / 0.0 B	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B
Total(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B

Executors

Show 20 entries

Search:

Executor ID	Address	Status	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Logs
driver	192.168.221.131:35361	Active	0	0.0 B / 384.1 MB	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B	
1	maprdemo:46568	Active	0	0.0 B / 956.6 MB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B	stdout stderr

Showing 1 to 2 of 2 entries

Previous 1 Next

Para el job 4: ¿Cuántas tasks se han ejecutado? Una

Logs ID	Name	Application Type	Status	Benutzer	Maps	Reduces	Queue	Priority	Duration	Submitted
1713207282087_0001	Spark Pi	SPARK	SUCCEEDED	mapr	100%	100%	root.mapr	N/A	1m:24s	04/15/24 12:16:06

Para el job 4: ¿Cuántos executors se han ejecutado? Se han ejecutado dos. Un driver y un maprdemo

Executor ID	Address	Status	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Logs
driver	192.168.221.131:35361	Active	0	0.0 B / 384.1 MB	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B	
1	maprdemo:46568	Active	0	0.0 B / 956.6 MB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B	stdout stderr

Showing 1 to 2 of 2 entries

Previous 1 Next

Para el job 4: ¿Cuántos RDDs se han creado y que transformaciones los han desencadenado? Se han creado dos que están en activo.

	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write
Active(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B
Dead(0)	0	0.0 B / 0.0 B	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B
Total(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B

DETALLES CON Jobs = 10

```
./bin/spark-submit \
--class org.apache.spark.examples.SparkPi \
--master yarn-client \
/opt/mapr/spark/spark-2.1.0/examples/jars/spark-examples_2.11-2.1.0-mapr-1710.jar \
10
```

```
[mapr@maprdemo spark-2.1.0]$ ./bin/spark-submit \
> --class org.apache.spark.examples.SparkPi \
> --master yarn-client \
> /opt/mapr/spark/spark-2.1.0/examples/jars/spark-examples_2.11-2.1.0-mapr-1710.jar \
> 10
Warning: Unable to determine $DRILL_HOME
Warning: Master yarn-client is deprecated since 2.0. Please use master "yarn" with specified deploy mode instead.
24/04/15 12:52:13 WARN Utils: Your hostname, maprdemo.local resolves to a loopback address: 127.0.0.1; using 192.168.221.132 instead (on interface ens32)
24/04/15 12:52:13 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
24/04/15 12:52:16 WARN Client: Neither spark.yarn.jars nor spark.yarn.archive is set, falling back to uploading libraries under SPARK_HOME.
[Stage 0:> (0 + 0) / 10]
24/04/15 12:53:28 WARN YarnScheduler: Initial job has not accepted any resources; check your cluster UI to ensure that workers are registered and have sufficient resources
Pi is roughly 3.141339141339141
[mapr@maprdemo spark-2.1.0]$
```

← → 🔍 No es seguro 192.168.221.132:8888/jobbrowser/

File Browser Job Browser mapr

Job Browser

Benutzername: mapr Text: Search for text

Succeeded Running Failed Killed in the last 7

Logs ID	Name	Application Type	Status	Benutzer	Maps	Reduces	Queue	Priority	Duration	Submitted
1713209692187_0002	Spark Pi	SPARK	SUCCEEDED	mapr	100%	100%	root.mapr	N/A	1m28s	04/15/24 12:52:21
1713209692187_0001	Spark Pi	SPARK	SUCCEEDED	mapr	100%	100%	root.mapr	N/A	31s	04/15/24 12:36:32

← → ↻ No es seguro 192.168.221.132:8888/jobbrowser/jobs/application_1713209692187_0002

HUE Query Editors ▾ Notebooks Data Browsers ▾ Workflows ▾

Job Browser

APP ID

1713209692187_0002

TYPE

SPARK

BENUTZER

mapr

STATUS

SUCCEEDED

LOGS

Logs

PROGRESS

100.0%

DURATION

1m:28s

Spark Pi

Metadata

Name	Value
Jobs	http://maprdemo:18080/history/application_1713209692187_0002/
Host	http://maprdemo:8042
Queue Name	root.mapr
Started	04/15/24 12:52:21
Finished	04/15/24 12:53:49
Pre-empted Resource VCores	0
VCore seconds	140
Memory seconds	252105
Diagnostics	

Y su acedemos a los jobs cambiando maprdemo con la ip.

← → ↻ No es seguro 192.168.221.132:8042/node

hadoop

ResourceManager
NodeManager
Node Information
List of Applications
List of Containers
Tools

Total Vmem allocated for Containers	10.50 GB
Vmem enforcement enabled	false
Total Pmem allocated for Container	5.00 GB
Pmem enforcement enabled	true
Total VCores allocated for Containers	0
NodeHealthyStatus	true
LastNodeHealthTime	Mon Apr 15 12:56:48 PDT 2024
NodeHealthReport	
Node Manager Version:	2.7.0-mapr-1710 from a59f1d36fa1be22772e0e48ce3ee8c9d41530287 by root source checksum d6c4897d604fe570f0c8a357c71fd4f on 2017-11-10T02:09Z
Hadoop Version:	2.7.0-mapr-1710 from a59f1d36fa1be22772e0e48ce3ee8c9d41530287 by root source checksum 89b68ba4cd988c71b9b181bcc2a79ce on 2017-11-10T01:59Z

← → ↻ No es seguro 192.168.221.131:18080/history/application_1713207282087_0001/executors/

Spark 2.1.0-mapr-1710 Jobs Stages Storage Environment Executors Spark Pi applicati

Executors

Summary

	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write
Active(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B
Dead(0)	0	0.0 B / 0.0 B	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B
Total(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B

Executors

Show 20 entries Search:

Executor ID	Address	Status	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Logs
driver	192.168.221.131:35361	Active	0	0.0 B / 384.1 MB	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B	
1	maprdemo:46568	Active	0	0.0 B / 956.6 MB	0.0 B	1	0	0	4	4	22 s (0.4 s)	0.0 B	0.0 B	0.0 B	stdout stderr

Showing 1 to 2 of 2 entries Previous 1 Next

Para el job 10: ¿Cuántas tasks se han ejecutado? Diez

Spark Jobs (?)

User: mapr
Total Uptime: 1.6 min
Scheduling Mode: FIFO
Completed Jobs: 1
[Event Timeline](#)

Completed Jobs (1)

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0	reduce at SparkPi scala 38	2024/04/15 12:53:09	40 s	1/1	10/10

Detalles

Total Time Across All Tasks: 0.3 s
Locality Level Summary: Process local: 10

[DAG Visualization](#)
[Show Additional Metrics](#)
[Event Timeline](#)

Summary Metrics for 10 Completed Tasks

Metric	Min	25th percentile	Median	75th percentile	Max
Duration	4 ms	5 ms	6 ms	8 ms	0.2 s
GC Time	0 ms	0 ms	0 ms	0 ms	4 s

Aggregated Metrics by Executor

Executor ID	Address	Task Time	Total Tasks	Failed Tasks	Killed Tasks	Succeeded Tasks
1	mapdemo:37147	15 s	10	0	0	10

Tasks (10)

Index	ID	Attempt	Status	Locality Level	Executor ID / Host	Launch Time	Duration	GC Time
0	0	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:34	0.2 s	4 s
1	1	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:48	8 ms	
2	2	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:48	4 ms	
3	3	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:48	4 ms	
4	4	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:48	6 ms	
5	5	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:48	8 ms	
6	6	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:48	5 ms	
7	7	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:48	6 ms	
8	8	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:48	5 ms	
9	9	0	SUCCESS	PROCESS_LOCAL	1 / mapdemo	2024/04/15 12:53:49	5 ms	

Para el job 10: ¿Cuántos executors se han ejecutado?

Executors

Executor ID	Address	Status	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Logs
driver	192.168.221.132:41726	Active	0	0.0 B / 384.1 MB	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B	
1	mapdemo:37147	Active	0	0.0 B / 956.6 MB	0.0 B	1	0	0	10	10	15 s (4 s)	0.0 B	0.0 B	0.0 B	stdout stderr

Para el job 10: ¿Cuántos RDDs se han creado y que transformaciones los han desencadenado?

Executors

Summary

	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write
Active(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	10	10	15 s (4 s)	0.0 B	0.0 B	0.0 B
Dead(0)	0	0.0 B / 0.0 B	0.0 B	0	0	0	0	0	0 ms (0 ms)	0.0 B	0.0 B	0.0 B
Total(2)	0	0.0 B / 1.3 GB	0.0 B	1	0	0	10	10	15 s (4 s)	0.0 B	0.0 B	0.0 B