Hewlett Packard Enterprise

Runbook for Spark 3.1.1 with Livy & JupyterHub

Date Prepared: Mar 2021

Hewlett Packard Enterprise

Document Information

Project Name	**HPE**		
Project Owner		Document Version No	0.1
Quality Review Method			
Prepared By	suhel.shaikh@hpe.com	Preparation Date	23-Mar-2021
Reviewed By	vivek-singh.bhadauriya@hpe.com	Review Date	24-Mar-2021

Table of Contents

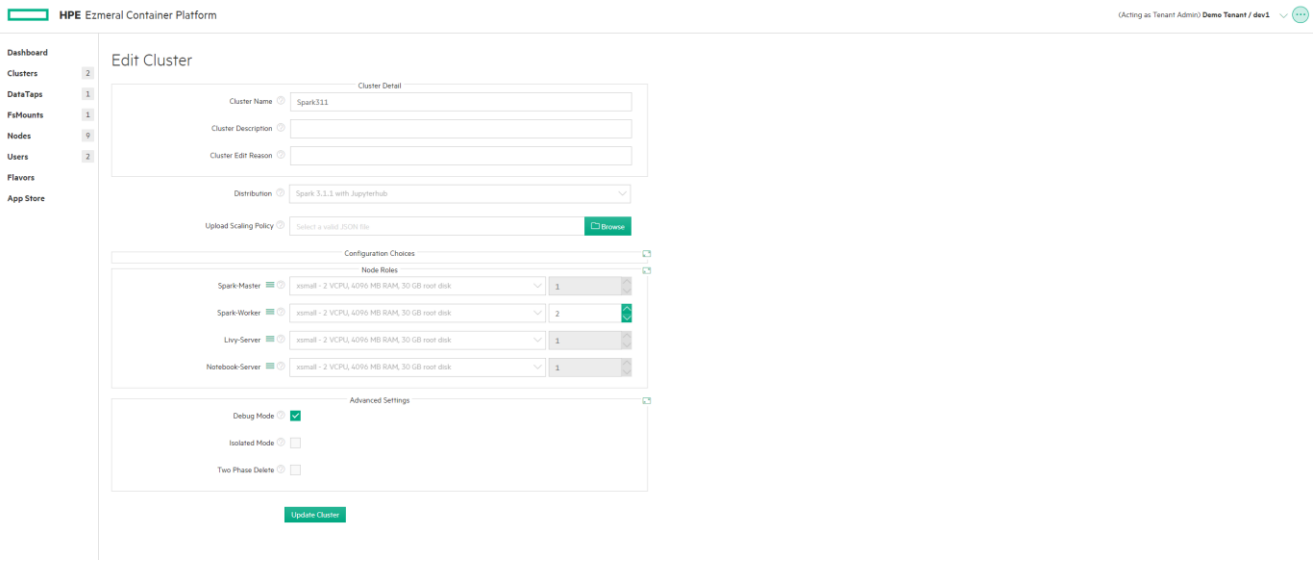
1 SUMMARY

This runbook will demonstrate creating a spark 3.1.1 cluster and different ways to submit Spark jobs.

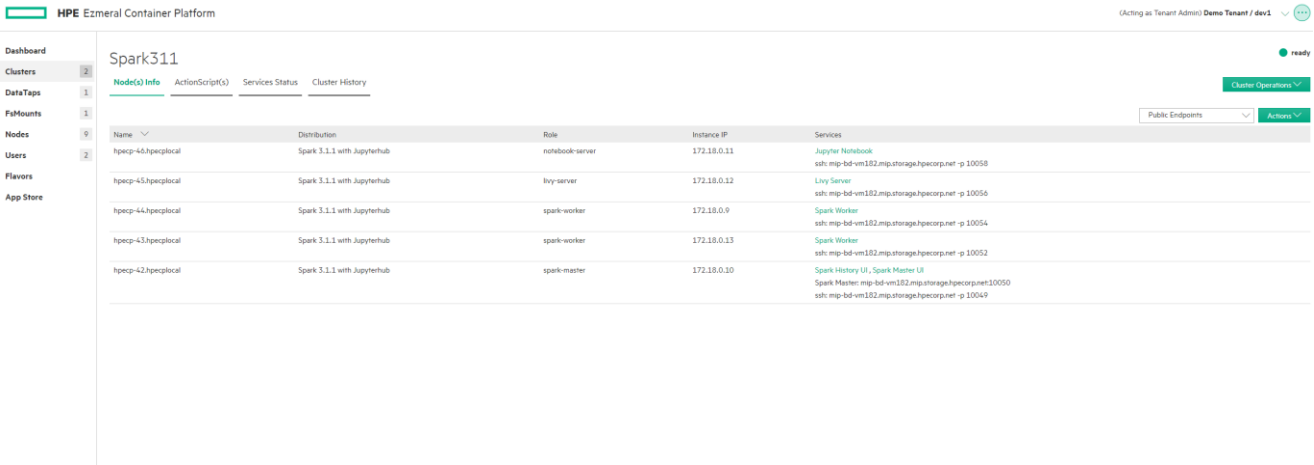
Bin: https://hpecp-engineering.s3.us-east-2.amazonaws.com/Spark311/bdcatalog-centos7-ezmeral-spark311-1.0.11.bin

2 CREATE SPARK 311 CLUSTER

Login to HPECP using any LDAP user and create spark 3.1.1 cluster.



Wait for the cluster state to change to ready.



3 SUBMIT A SPARK JOB FROM LDAP USER

3.1 Submitting Spark-PI Job (Spark PI Jar from Local System)

Login to spark 3.1.1 master node using your LDAP user.

```
[root@mip-bd-vm181 ~]#
[root@mip-bd-vm181 ~]#
[root@mip-bd-vm181 ~]# ssh  dev1@172.18.0.10
dev1@172.18.0.10's password:
Last login: Tue Mar 23 03:27:27 2021 from 172.18.0.2
[dev1@hpecp-42 ~]$
[dev1@hpecp-42 ~]$
[dev1@hpecp-42 ~]$
```

Submit a Spark pi job to spark. Locate your main application file on your local file system.

Command:

```
spark-submit --deploy-mode client --class org.apache.spark.examples.SparkPi /usr/l
ib/spark/spark-3.1.1-bin-hadoop2.7/examples/jars/spark-examples_2.12-3.1.1.jar 100
```

Output:

```
[dev1@hpecp-42 ~]$
[dev1@hpecp-42 ~]$ spark-submit --deploy-mode client --class org.apache.spark.examples.SparkPi /usr/lib/spark/spark-3.1.1-bin-hadoop2.7/examples/jars/spark-examples_2.12-3.1.1.jar 100
21/03/23 03:28:39 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
21/03/23 03:28:39 INFO SparkContext: Running Spark version 3.1.1
21/03/23 03:28:39 INFO ResourceUtils: ==============================================================
21/03/23 03:28:39 INFO ResourceUtils: No custom resources configured for spark.driver.
21/03/23 03:28:39 INFO ResourceUtils: ==============================================================
21/03/23 03:28:39 INFO SparkContext: Submitted application: Spark Pi
21/03/23 03:28:39 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 1, script: , vendor: , memory -> name: memory, amount: 1024, script:
, vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
21/03/23 03:28:39 INFO ResourceProfile: Limiting resource is cpu
21/03/23 03:28:39 INFO ResourceProfileManager: Added ResourceProfile id: 0
21/03/23 03:28:40 INFO SecurityManager: Changing view acls to: dev1
21/03/23 03:28:40 INFO SecurityManager: Changing modify acls to: dev1
21/03/23 03:28:40 INFO SecurityManager: Changing view acls groups to:
21/03/23 03:28:40 INFO SecurityManager: Changing modify acls groups to:
21/03/23 03:28:40 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users  with view permissions: Set(dev1); groups with view permissions: Set(); users  with modif
y permissions: Set(dev1); groups with modify permissions: Set()
21/03/23 03:28:40 INFO Utils: Successfully started service 'sparkDriver' on port 43201.
21/03/23 03:28:40 INFO SparkEnv: Registering MapOutputTracker
21/03/23 03:28:40 INFO SparkEnv: Registering BlockManagerMaster
21/03/23 03:28:40 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
21/03/23 03:28:40 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
21/03/23 03:28:40 INFO SparkEnv: Registering BlockManagerMasterHeartbeat
21/03/23 03:28:40 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-5439164d-5f9f-4dec-970a-273aaf1dfa87
21/03/23 03:28:40 INFO MemoryStore: MemoryStore started with capacity 366.3 MiB
21/03/23 03:28:40 INFO SparkEnv: Registering OutputCommitCoordinator
21/03/23 03:28:40 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
21/03/23 03:28:40 INFO Utils: Successfully started service 'SparkUI' on port 4041.
21/03/23 03:28:40 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://hpecp-42.hpeclocal:4041
21/03/23 03:28:40 INFO SparkContext: Added JAR file:///opt/bluedata/bluedata-dtap.jar at spark://hpecp-42.hpeclocal:43201/jars/bluedata-dtap.jar with timestamp 1616495319869
21/03/23 03:28:40 INFO SparkContext: Added JAR file:/usr/lib/spark/spark-3.1.1-bin-hadoop2.7/examples/jars/spark-examples_2.12-3.1.1.jar at spark://hpecp-42.hpeclocal:43201/jars/spark-examples_2
.12-3.1.1.jar with timestamp 1616495319869
21/03/23 03:28:40 INFO StandaloneAppClient$ClientEndpoint: Connecting to master spark://hpecp-42.hpeclocal:7077...
21/03/23 03:28:40 INFO TransportClientFactory: Successfully created connection to hpecp-42.hpeclocal/172.18.0.10:7077 after 30 ms (0 ms spent in bootstraps)
21/03/23 03:28:40 INFO StandaloneSchedulerBackend: Connected to Spark cluster with app ID app-20210323032840-0019
21/03/23 03:28:40 INFO StandaloneAppClient$ClientEndpoint: Executor added: app-20210323032840-0019/0 on worker-20210319032747-172.18.0.13-42598 (172.18.0.13:42598) with 1 core(s)
21/03/23 03:28:40 INFO StandaloneSchedulerBackend: Granted executor ID app-20210323032840-0019/0 on hostPort 172.18.0.13:42598 with 1 core(s), 1024.0 MiB RAM
21/03/23 03:28:40 INFO StandaloneAppClient$ClientEndpoint: Executor added: app-20210323032840-0019/1 on worker-20210319032746-172.18.0.9-38291 (172.18.0.9:38291) with 1 core(s)
21/03/23 03:28:40 INFO StandaloneSchedulerBackend: Granted executor ID app-20210323032840-0019/1 on hostPort 172.18.0.9:38291 with 1 core(s), 1024.0 MiB RAM
21/03/23 03:28:40 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 33200.
21/03/23 03:28:40 INFO NettyBlockTransferService: Server created on hpecp-42.hpeclocal:33200
21/03/23 03:28:40 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
21/03/23 03:28:40 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, hpecp-42.hpeclocal, 33200, None)
21/03/23 03:28:40 INFO BlockManagerMasterEndpoint: Registering block manager hpecp-42.hpeclocal:33200 with 366.3 MiB RAM, BlockManagerId(driver, hpecp-42.hpeclocal, 33200, None)
21/03/23 03:28:40 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, hpecp-42.hpeclocal, 33200, None)
21/03/23 03:28:40 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, hpecp-42.hpeclocal, 33200, None)
21/03/23 03:28:40 INFO StandaloneAppClient$ClientEndpoint: Executor updated: app-20210323032840-0019/1 is now RUNNING
21/03/23 03:28:40 INFO StandaloneAppClient$ClientEndpoint: Executor updated: app-20210323032840-0019/0 is now RUNNING
21/03/23 03:28:41 INFO SingleEventLogFileWriter: Logging events to dtap://TenantStorage/tmp/spark/app-20210323032840-0019.inprogress
21/03/23 03:28:41 INFO StandaloneSchedulerBackend: SchedulerBackend is ready for scheduling beginning after reached minRegisteredResourcesRatio: 0.0
21/03/23 03:28:41 INFO SparkContext: Starting job: reduce at SparkPi.scala:38
21/03/23 03:28:41 INFO DAGScheduler: Got job 0 (reduce at SparkPi.scala:38) with 100 output partitions
21/03/23 03:28:41 INFO DAGScheduler: Final stage: ResultStage 0 (reduce at SparkPi.scala:38)
21/03/23 03:28:41 INFO DAGScheduler: Parents of final stage: List()
21/03/23 03:28:41 INFO DAGScheduler: Missing parents: List()
21/03/23 03:28:41 INFO DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[1] at map at SparkPi.scala:34), which has no missing parents
```

Spark History Server UI:

3.2 Submitting Spark-PI Job (Spark PI Jar from DTAP)

After logging into the spark master, copy the **spark-examples_2.12-3.1.1.jar** to DTAP.

First we will copy the jar to DTAP.

Command:

```
hadoop fs -put /usr/lib/spark/spark-3.1.1-bin-hadoop2.7/examples/jars/spark-examples_2.12-3.1.1.jar dtap://TenantStorage/
```

```
[dev1@hpecp-42 ~]$ hadoop fs -put /usr/lib/spark/spark-3.1.1-bin-hadoop2.7/examples/jars/spark-examples_2.12-3.1.1.jar dtap://TenantStorage/
[dev1@hpecp-42 ~]$
[dev1@hpecp-42 ~]$
[dev1@hpecp-42 ~]$ hadoop fs -ls dtap://TenantStorage/
Found 8 items
-rw-r--r--   3 dev1 Domain Users    1527168 2021-03-23 03:34 dtap://TenantStorage/code
drwxrwxrwx   - dev1 Domain Users          0 2021-03-22 06:00 dtap://TenantStorage/elkdata
drwxrwxrwx   - dev1 Domain Users          0 2021-03-22 06:00 dtap://TenantStorage/elklogs
drwxrwxrwx   - dev1 Domain Users          0 2021-03-19 03:55 dtap://TenantStorage/input
drwxrwxrwx   - dev1 Domain Users          0 2021-03-19 03:39 dtap://TenantStorage/jar
drwxrwxrwx   - dev1 Domain Users          0 2021-03-19 05:22 dtap://TenantStorage/output
-rw-r--r--   3 dev1 Domain Users    1527168 2021-03-23 03:39 dtap://TenantStorage/spark-examples_2.12-3.1.1.jar
drwxr-xr-x   - dev1 Domain Users          0 2021-03-21 23:08 dtap://TenantStorage/tmp
[dev1@hpecp-42 ~]$
```

Submit a spark pi job to spark where main application file is within DTAP.

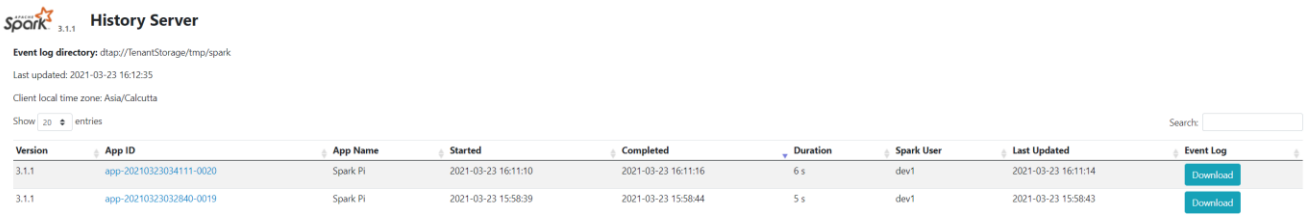
Command:

```
spark-submit --deploy-mode client --class org.apache.spark.examples.SparkPi dtap://TenantStorage/spark-examples_2.12-3.1.1.jar 100
```

Output:

```
[dev1@hpecp-42 ~]$ spark-submit --deploy-mode client --class org.apache.spark.examples.SparkPi dtap://TenantStorage/spark-examples_2.12-3.1.1.jar 100
21/03/23 03:41:09 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
21/03/23 03:41:10 INFO SparkContext: Running Spark version 3.1.1
21/03/23 03:41:10 INFO ResourceUtils: ==============================================================
21/03/23 03:41:10 INFO ResourceUtils: No custom resources configured for spark.driver.
21/03/23 03:41:10 INFO ResourceUtils: ==============================================================
21/03/23 03:41:10 INFO SparkContext: Submitted application: Spark Pi
21/03/23 03:41:10 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 1, script: , vendor: , memory -> name: memory, amount: 1024, script:
, vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
21/03/23 03:41:10 INFO ResourceProfile: Limiting resource is cpu
21/03/23 03:41:10 INFO ResourceProfileManager: Added ResourceProfile id: 0
21/03/23 03:41:10 INFO SecurityManager: Changing view acls to: dev1
21/03/23 03:41:10 INFO SecurityManager: Changing modify acls to: dev1
21/03/23 03:41:10 INFO SecurityManager: Changing view acls groups to:
21/03/23 03:41:10 INFO SecurityManager: Changing modify acls groups to:
21/03/23 03:41:10 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users  with view permissions: Set(dev1); groups with view permissions: Set(); users  with modif
y permissions: Set(dev1); groups with modify permissions: Set()
21/03/23 03:41:10 INFO Utils: Successfully started service 'sparkDriver' on port 39193.
21/03/23 03:41:10 INFO SparkEnv: Registering MapOutputTracker
21/03/23 03:41:10 INFO SparkEnv: Registering BlockManagerMaster
21/03/23 03:41:10 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
21/03/23 03:41:10 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
21/03/23 03:41:10 INFO SparkEnv: Registering BlockManagerMasterHeartbeat
21/03/23 03:41:10 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-8ce76c41-4194-4c47-b123-8852fd7b6610
21/03/23 03:41:10 INFO MemoryStore: MemoryStore started with capacity 366.3 MiB
21/03/23 03:41:10 INFO SparkEnv: Registering OutputCommitCoordinator
21/03/23 03:41:11 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
21/03/23 03:41:11 INFO Utils: Successfully started service 'SparkUI' on port 4041.
21/03/23 03:41:11 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://hpecp-42.hpecplocal:4041
21/03/23 03:41:11 INFO SparkContext: Added JAR file:///opt/bluedata/bluedata-dtap.jar at spark://hpecp-42.hpecplocal:39193/jars/bluedata-dtap.jar with timestamp 1616496070425
21/03/23 03:41:11 INFO SparkContext: Added JAR dtap://TenantStorage/spark-examples_2.12-3.1.1.jar at dtap://TenantStorage/spark-examples_2.12-3.1.1.jar with timestamp 1616496070425
21/03/23 03:41:11 INFO StandaloneAppClient$ClientEndpoint: Connecting to master spark://hpecp-42.hpecplocal:7077...
21/03/23 03:41:11 INFO TransportClientFactory: Successfully created connection to hpecp-42.hpecplocal/172.18.0.10:7077 after 32 ms (0 ms spent in bootstraps)
21/03/23 03:41:11 INFO StandaloneSchedulerBackend: Connected to Spark cluster with app ID app-20210323034111-0020
21/03/23 03:41:11 INFO StandaloneAppClient$ClientEndpoint: Executor added: app-20210323034111-0020/0 on worker-20210319032747-172.18.0.13-42598 (172.18.0.13:42598) with 1 core(s)
21/03/23 03:41:11 INFO StandaloneSchedulerBackend: Granted executor ID app-20210323034111-0020/0 on hostPort 172.18.0.13:42598 with 1 core(s), 1024.0 MiB RAM
21/03/23 03:41:11 INFO StandaloneAppClient$ClientEndpoint: Executor added: app-20210323034111-0020/1 on worker-20210319032746-172.18.0.9-38291 (172.18.0.9:38291) with 1 core(s)
21/03/23 03:41:11 INFO StandaloneSchedulerBackend: Granted executor ID app-20210323034111-0020/1 on hostPort 172.18.0.9:38291 with 1 core(s), 1024.0 MiB RAM
21/03/23 03:41:11 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 44745.
21/03/23 03:41:11 INFO NettyBlockTransferService: Server created on hpecp-42.hpecplocal:44745
21/03/23 03:41:11 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
21/03/23 03:41:11 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, hpecp-42.hpecplocal, 44745, None)
21/03/23 03:41:11 INFO BlockManagerMasterEndpoint: Registering block manager hpecp-42.hpecplocal:44745 with 366.3 MiB RAM, BlockManagerId(driver, hpecp-42.hpecplocal, 44745, None)
21/03/23 03:41:11 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, hpecp-42.hpecplocal, 44745, None)
21/03/23 03:41:11 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, hpecp-42.hpecplocal, 44745, None)
```

Spark History Server UI:

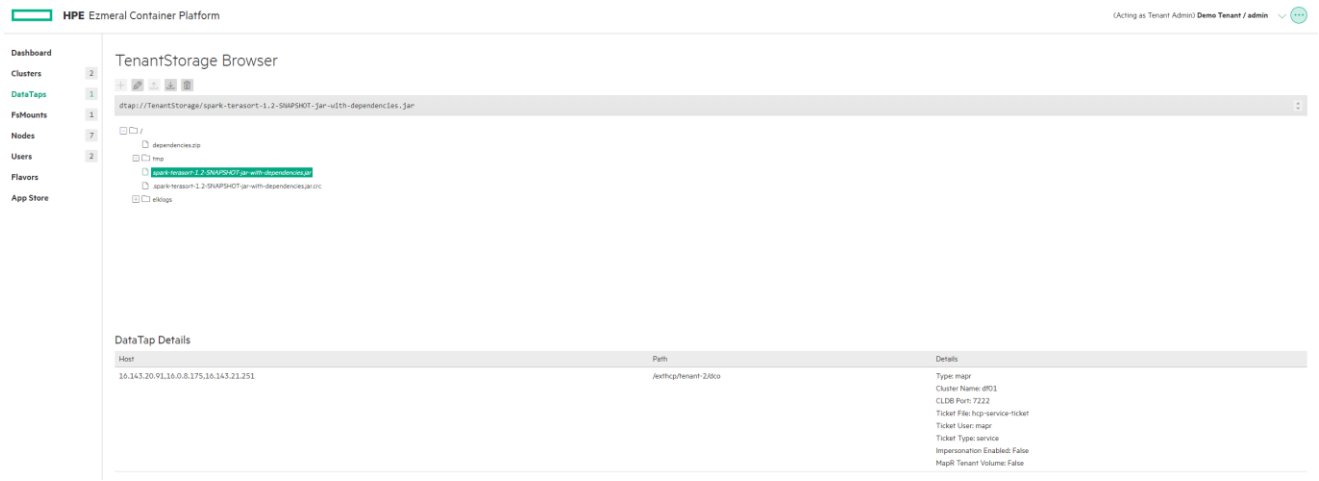


3.3 Submitting Spark-PI Job in Cluster Mode

Upload the **spark-terasort-1.2-SNAPSHOT-jar-with-dependencies.jar** to DTAP from the HPECP Web UI.

You can get the Jar from below link:

Jar: https://hpecp-engineering.s3.us-east-2.amazonaws.com/Spark311/spark-terasort-1.2-SNAPSHOT-jar-with-dependencies.jar



Here we are running TeraGen spark job in cluster mode and our main application will reside in DTAP.

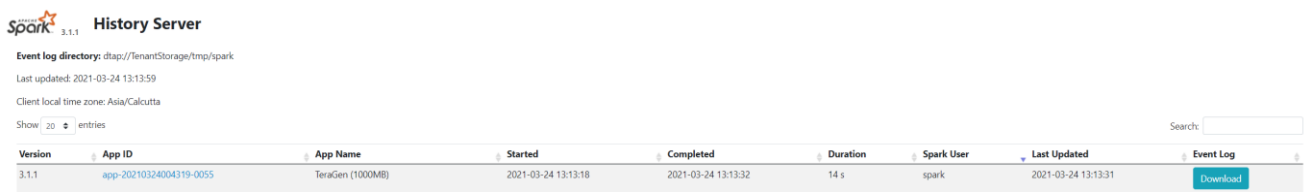
Command:

```
spark-submit --deploy-mode cluster --class com.github.ehiggs.spark.terasort.TeraGe
n dtap://TenantStorage/spark-terasort-1.2-SNAPSHOT-jar-with-dependencies.jar "1G"
 dtap://TenantStorage/tera_gen
```

Output:

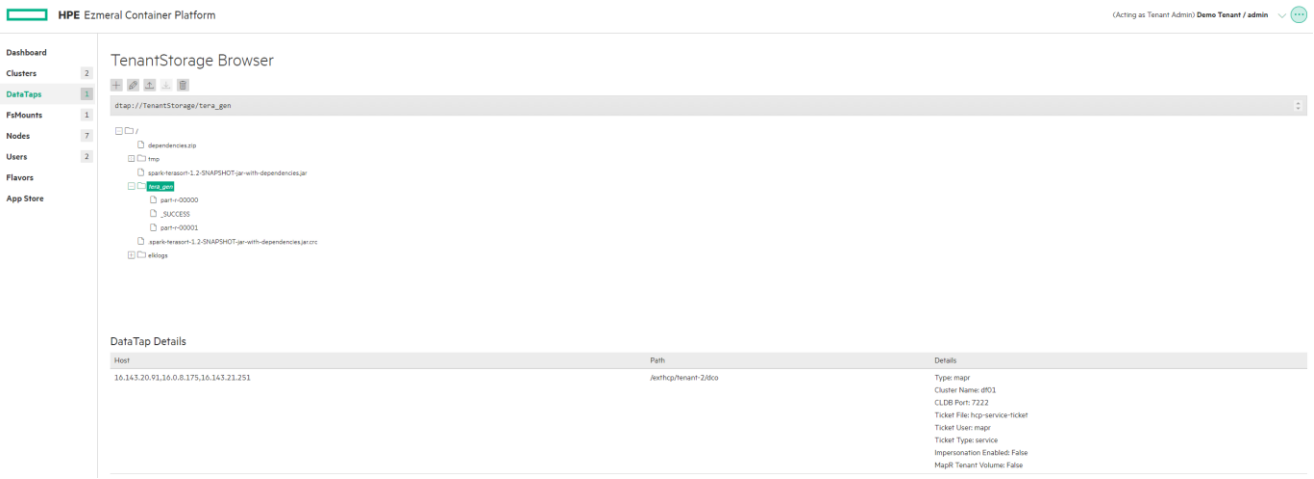
```
[root@hpecp-42 bluedata]#
[root@hpecp-42 bluedata]#
[root@hpecp-42 bluedata]# spark-submit --deploy-mode cluster --class com.github.ehiggs.spark.terasort.TeraGen dtap://TenantStorage/spark-terasort-1.2-SNAPSHOT-jar-with-dependencies.jar "1G"  dta
p://TenantStorage/tera_gen
21/03/24 00:43:15 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
21/03/24 00:43:15 INFO SecurityManager: Changing view acls to: root
21/03/24 00:43:15 INFO SecurityManager: Changing modify acls to: root
21/03/24 00:43:15 INFO SecurityManager: Changing view acls groups to:
21/03/24 00:43:15 INFO SecurityManager: Changing modify acls groups to:
21/03/24 00:43:15 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users  with view permissions: Set(root); groups with view permissions: Set(); users  with modi
fy permissions: Set(root); groups with modify permissions: Set()
21/03/24 00:43:16 INFO Utils: Successfully started service 'driverClient' on port 46231.
21/03/24 00:43:16 INFO TransportClientFactory: Successfully created connection to hpecp-42.hpecplocal/172.18.0.10:7077 after 29 ms (0 ms spent in bootstraps)
21/03/24 00:43:16 INFO ClientEndpoint: ... waiting before polling master for driver state
21/03/24 00:43:16 INFO ClientEndpoint: Driver successfully submitted as driver-20210324004316-0023
21/03/24 00:43:21 INFO ClientEndpoint: State of driver-20210324004316-0023 is RUNNING
21/03/24 00:43:21 INFO ClientEndpoint: Driver running on 172.18.0.9:38291 (worker-20210319032746-172.18.0.9-38291)
21/03/24 00:43:21 INFO ClientEndpoint: spark-submit not configured to wait for completion, exiting spark-submit JVM.
21/03/24 00:43:21 INFO ShutdownHookManager: Shutdown hook called
21/03/24 00:43:21 INFO ShutdownHookManager: Deleting directory /tmp/spark-3fd163a3-4768-4fa4-8087-77388dbed39b
[root@hpecp-42 bluedata]#
```

Spark History Server UI:



Note: When submitting job in cluster mode with any user, the **spark** user will start the driver on worker node.

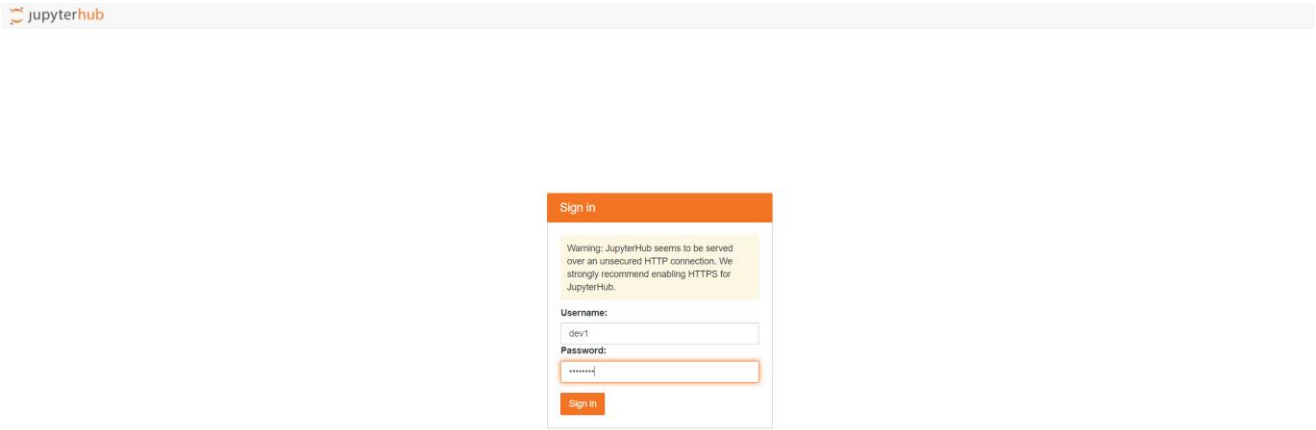
DTAP UI:



4 SPARK INTERECTIVIE SESSION USING LIVY THROUGH NOTEBOOK.

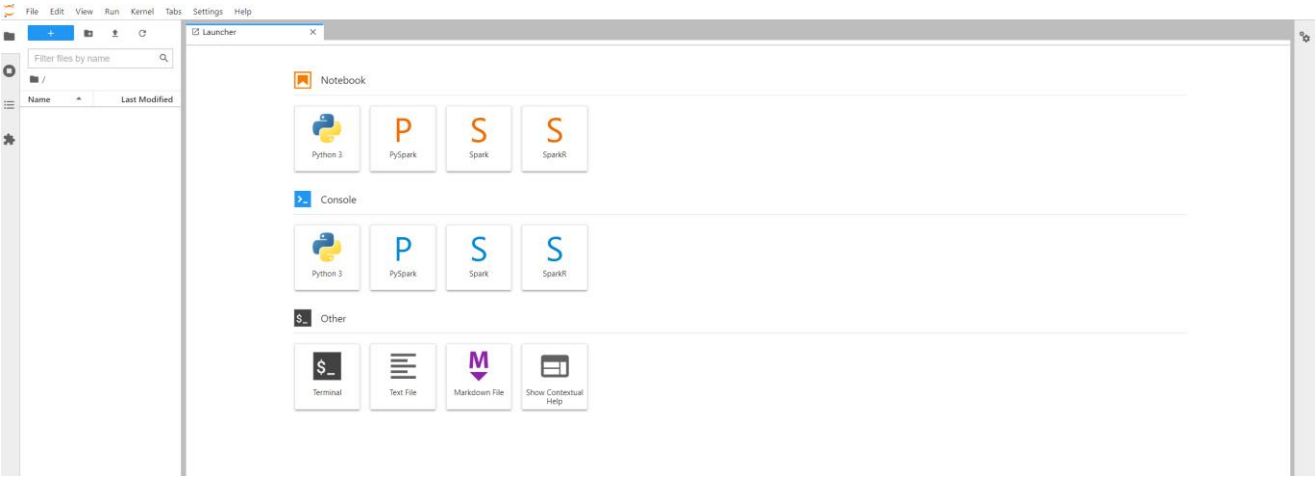
4.1 Login to Jupyter Hub.

Login to JupyterHub service using LDAP user account and password.



4.2 PySpark Notebook

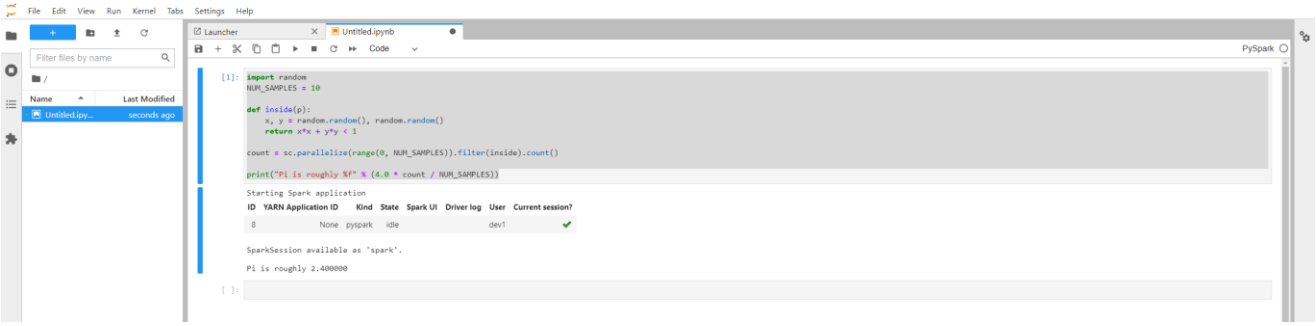
Click on PySpark Notebook and create a notebook.



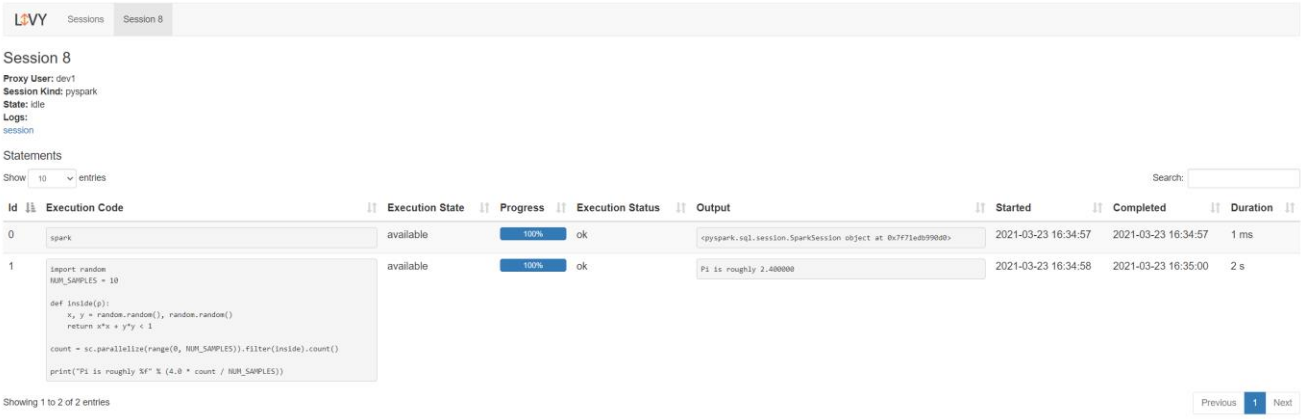
Code:

```
import random

NUM_SAMPLES = 10


def inside(p):

    x, y = random.random(), random.random()

    return x*x + y*y < 1


count = sc.parallelize(range(0, NUM_SAMPLES)).filter(inside).count()


print("Pi is roughly %f" % (4.0 * count / NUM_SAMPLES))
```

Notebook:

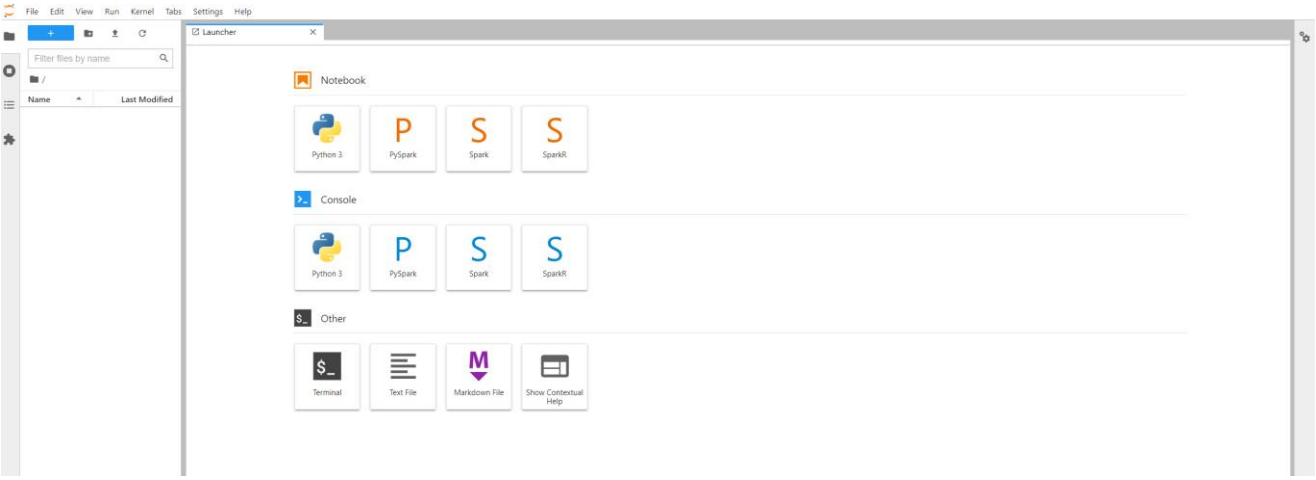


Livy:



4.3 Scala Notebook

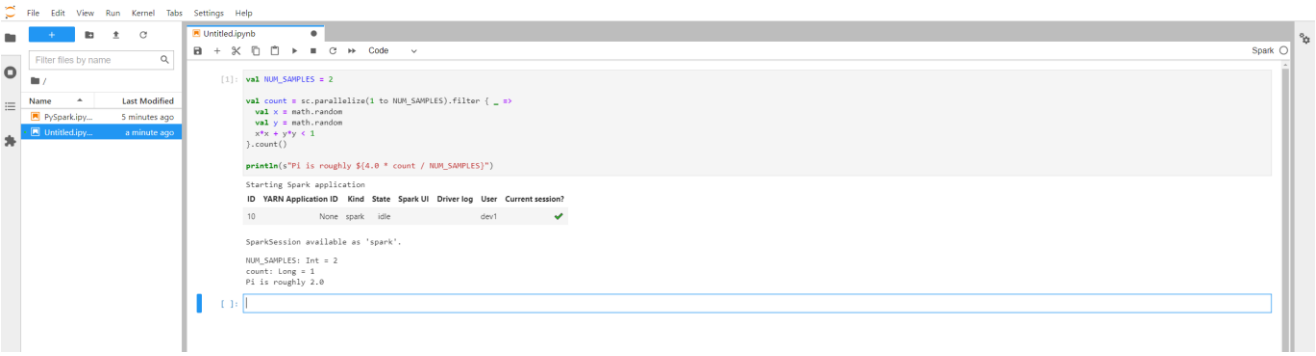
Click on Spark Notebook and create a notebook.



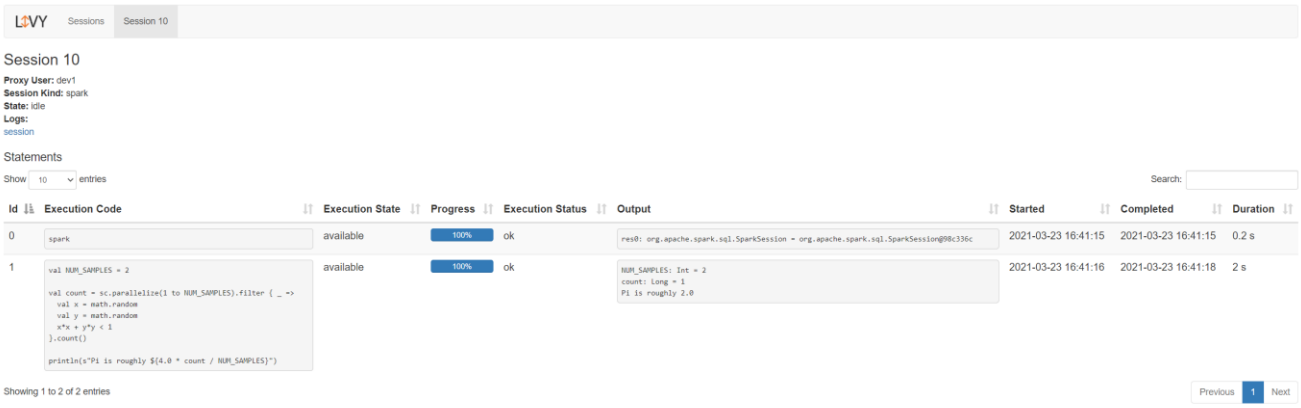
Code:

```
val NUM_SAMPLES = 2


val count = sc.parallelize(1 to NUM_SAMPLES).filter { _ =>
  val x = math.random
  val y = math.random
  x*x + y*y < 1
}.count()


println(s"Pi is roughly ${4.0 * count / NUM_SAMPLES}")
```

Notebook:

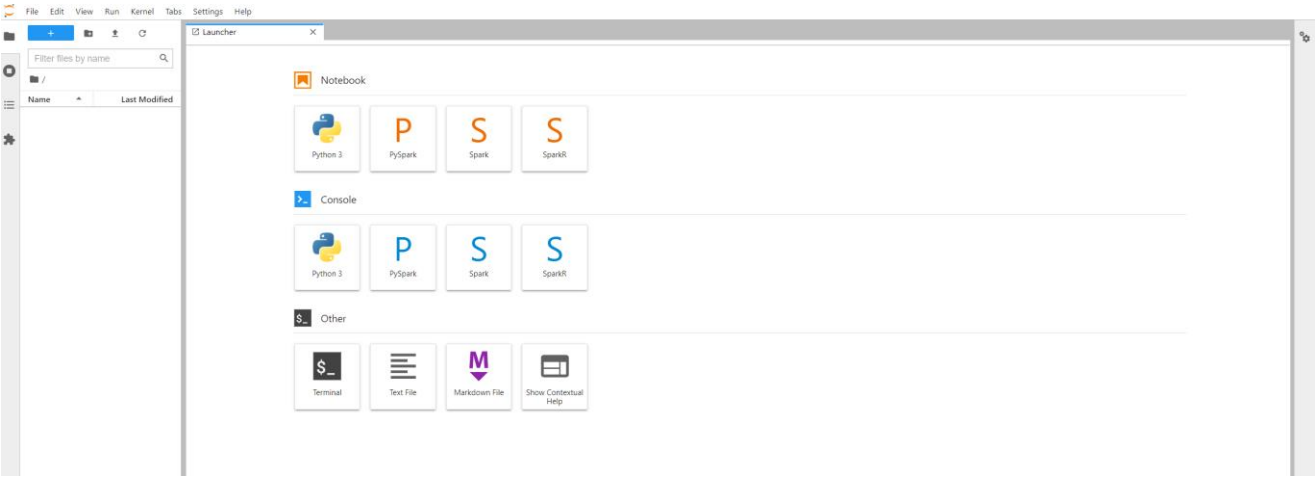


Livy:



4.4 SparkR Notebook

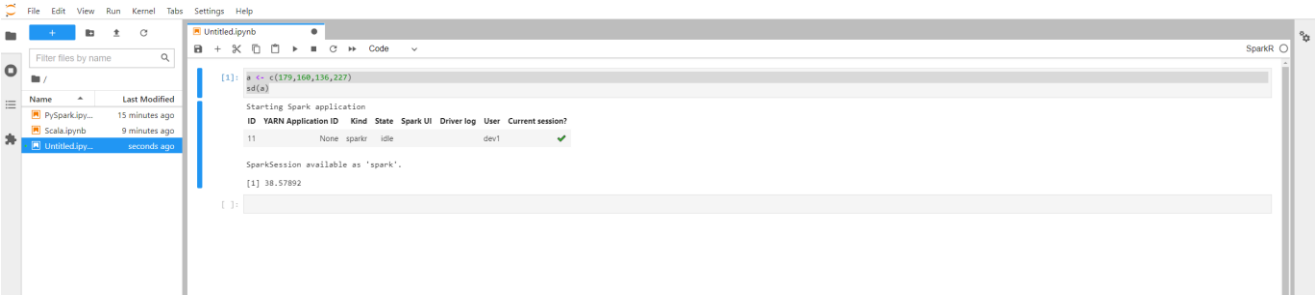
Click on SparkR Notebook and create a notebook.



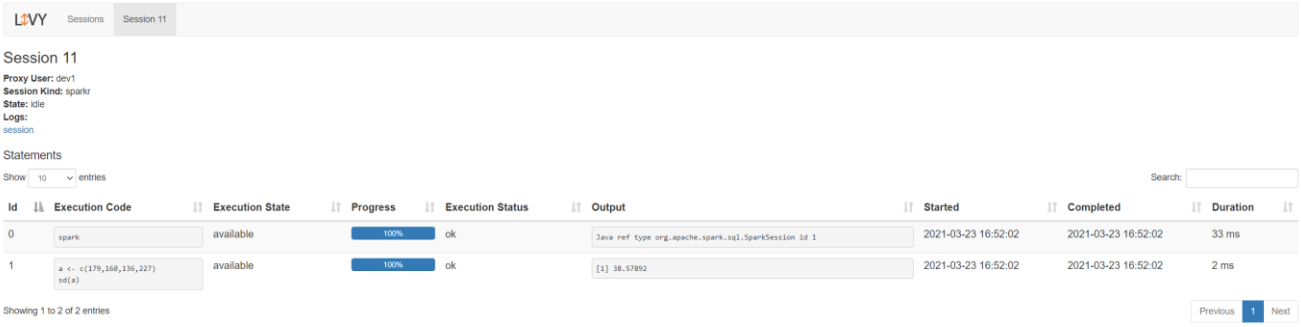
Code:

```
a <- c(179,160,136,227)

sd(a)
```

Notebook:



Livy:


LIVY
Sessions
Session 11
Session 11
Proxy User: dev1
Session Kind: sparkr
State: idle
Logs:
session
Statements
Show 10 entries
Search:
Id
Execution Code
Execution State
Progress
Execution Status
Output
Started
Completed
Duration
0
spark
available
100%
ok
Java ref type org.apache.spark.sql.SparkSession id 1
2021-03-23 16:52:02
2021-03-23 16:52:02
33 ms
1
a <- c(179,160,136,227)
sd(a)
available
100%
ok
[1] 38.57892
2021-03-23 16:52:02
2021-03-23 16:52:02
2 ms
Showing 1 to 2 of 2 entries
Previous
1
Next

5 PASS RUNTIME DEPENDENCIES TO SPARK JOB

5.1 Python Dependencies

Here we are packaging python dependencies and will pass those dependencies to PySpark job at run time.

Below are the commands to package python dependencies in a zip file and upload the zip file to DTAP.

In **requirements.txt** file we just added as single python package openpyxl (python package for reading and writing excel file.)

Note: This work fine with some python package but for some package it will fail.

```
pip3 install -t dependencies -r requirements.txt
cd dependencies/
zip -r ../dependencies.zip .
cd ..
hadoop fs -put dependencies.zip dtap://TenantStorage/
hadoop fs -ls dtap://TenantStorage/
```

PySpark Job:

In the job we are adding dependencies.zip file to our program using **sc.addPyFile()** method. In the logs, it will display all object of openpyxl python package. If you comment line 6 where we are adding dependencies.zip and re execute the job import openpyxl line line will through.

Content of PySparkJob.py

```
from pyspark import SparkContext


sc = SparkContext()


#Adding Dependencies
sc.addPyFile("dtap://TenantStorage/dependencies.zip")



#Loading Python Package
import openpyxl
print(dir(openpyxl))
```

Spark Submit:

```
spark-submit PySparkJob.py
```

Output:

```
[dev1@hpecp-42 ~]$ spark-submit PySparkJob.py
21/03/23 04:46:27 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
21/03/23 04:46:28 INFO SparkContext: Running Spark version 3.1.1
21/03/23 04:46:28 INFO ResourceUtils: ==============================================================
21/03/23 04:46:28 INFO ResourceUtils: No custom resources configured for spark.driver.
21/03/23 04:46:28 INFO ResourceUtils: ==============================================================
21/03/23 04:46:28 INFO SparkContext: Submitted application: PySparkJob.py
21/03/23 04:46:28 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, amount: 1, script: , vendor: , memory -> name: memory, amount: 1024, script:
 , vendor: , offHeap -> name: offHeap, amount: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
21/03/23 04:46:28 INFO ResourceProfile: Limiting resource is cpu
21/03/23 04:46:28 INFO ResourceProfileManager: Added ResourceProfile id: 0
21/03/23 04:46:28 INFO SecurityManager: Changing view acls to: dev1
21/03/23 04:46:28 INFO SecurityManager: Changing modify acls to: dev1
21/03/23 04:46:28 INFO SecurityManager: Changing view acls groups to:
21/03/23 04:46:28 INFO SecurityManager: Changing modify acls groups to:
21/03/23 04:46:28 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users  with view permissions: Set(dev1); groups with view permissions: Set(); users  with modif
y permissions: Set(dev1); groups with modify permissions: Set()
21/03/23 04:46:28 INFO Utils: Successfully started service 'sparkDriver' on port 46343.
21/03/23 04:46:28 INFO SparkEnv: Registering MapOutputTracker
21/03/23 04:46:28 INFO SparkEnv: Registering BlockManagerMaster
21/03/23 04:46:28 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
21/03/23 04:46:28 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
21/03/23 04:46:28 INFO SparkEnv: Registering BlockManagerMasterHeartbeat
21/03/23 04:46:28 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-5e7ec6c9-64bd-407c-b6fc-4d88eaac0111
21/03/23 04:46:28 INFO MemoryStore: MemoryStore started with capacity 366.3 MiB
21/03/23 04:46:28 INFO SparkEnv: Registering OutputCommitCoordinator
21/03/23 04:46:29 INFO Utils: Successfully started service 'SparkUI' on port 4040.
21/03/23 04:46:29 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://hpecp-42.hpecplocal:4040
21/03/23 04:46:29 INFO SparkContext: Added JAR file:///opt/bluedata/bluedata-dtap.jar at spark://hpecp-42.hpecplocal:46343/jars/bluedata-dtap.jar with timestamp 1616499988226
21/03/23 04:46:29 INFO StandaloneAppClient$ClientEndpoint: Connecting to master spark://hpecp-42.hpecplocal:7077...
21/03/23 04:46:29 INFO TransportClientFactory: Successfully created connection to hpecp-42.hpecplocal/172.18.0.10:7077 after 32 ms (0 ms spent in bootstraps)
21/03/23 04:46:29 INFO StandaloneSchedulerBackend: Connected to Spark cluster with app ID app-20210323044629-0031
21/03/23 04:46:29 INFO StandaloneAppClient$ClientEndpoint: Executor added: app-20210323044629-0031/0 on worker-20210319032746-172.18.0.9-38291 (172.18.0.9:38291) with 2 core(s)
21/03/23 04:46:29 INFO StandaloneSchedulerBackend: Granted executor ID app-20210323044629-0031/0 on hostPort 172.18.0.9:38291 with 2 core(s), 1024.0 MiB RAM
21/03/23 04:46:29 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 37948.
21/03/23 04:46:29 INFO NettyBlockTransferService: Server created on hpecp-42.hpecplocal:37948
21/03/23 04:46:29 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
21/03/23 04:46:29 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, hpecp-42.hpecplocal, 37948, None)
21/03/23 04:46:29 INFO BlockManagerMasterEndpoint: Registering block manager hpecp-42.hpecplocal:37948 with 366.3 MiB RAM, BlockManagerId(driver, hpecp-42.hpecplocal, 37948, None)
21/03/23 04:46:29 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, hpecp-42.hpecplocal, 37948, None)
21/03/23 04:46:29 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, hpecp-42.hpecplocal, 37948, None)
21/03/23 04:46:29 INFO StandaloneAppClient$ClientEndpoint: Executor updated: app-20210323044629-0031/0 is now RUNNING
21/03/23 04:46:30 INFO SingleEventLogFileWriter: Logging events to dtap://TenantStorage/tmp/spark/app-20210323044629-0031.inprogress
21/03/23 04:46:30 INFO StandaloneSchedulerBackend: SchedulerBackend is ready for scheduling beginning after reached minRegisteredResourcesRatio: 0.0
21/03/23 04:46:30 INFO SparkContext: Added file dtap://TenantStorage/dependencies.zip at dtap://TenantStorage/dependencies.zip with timestamp 1616499990301
21/03/23 04:46:30 INFO Utils: Fetching dtap://TenantStorage/dependencies.zip to /tmp/spark-710e056c-0eb0-4afc-b6ee-ceccc0d80d87/userFiles-1dc9ffc0-18ca-40a0-9115-a73f5c710b7f/fetchFileTemp7187739
464295905343.tmp
['DEFUSEDXML', 'LXML', 'NUMPY', 'Workbook', '__author__', '__author_email__', '__builtins__', '__cached__', '__doc__', '__file__', '__license__', '__loader__', '__maintainer_email__', '__name__',
 '__package__', '__path__', '__spec__', '__url__', '__version__', '__constants', 'cell', 'chart', 'chartsheet', 'comments', 'compat', 'constants', 'descriptors', 'drawing', 'formatting', 'formula'
, 'load_workbook', 'open', 'packaging', 'pivot', 'reader', 'styles', 'utils', 'workbook', 'worksheet', 'writer', 'xml']
21/03/23 04:46:30 INFO SparkContext: Invoking stop() from shutdown hook
21/03/23 04:46:30 INFO SparkUI: Stopped Spark web UI at http://hpecp-42.hpecplocal:4040
21/03/23 04:46:30 INFO StandaloneSchedulerBackend: Shutting down all executors
21/03/23 04:46:30 INFO CoarseGrainedSchedulerBackend$DriverEndpoint: Asking each executor to shut down
21/03/23 04:46:30 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
21/03/23 04:46:30 INFO MemoryStore: MemoryStore cleared
```

5.2 Jar Dependencies

Working on it.

6 KERBEROS TESTING

Pending

7 SPARK THRIFT SERVER

As of now not include in the bin.