

Assignment Solutions

Task 1:

1) Pi estimation from Spark application:

```
>>> print "Pi is roughly %f" % (4.0 * count / NUM_SAMPLES)
Pi is roughly 3.141940
```

2) Output of one of the executor pods and its description:

```
pujanm@pujanm:~$ kubectl describe po pysparkshell-1551883429969-exec-1
Name:          pysparkshell-1551883429969-exec-1
Namespace:     default
Node:          minikube/192.168.122.216
Start Time:    Wed, 06 Mar 2019 20:13:50 +0530
Labels:        spark-app-selector=spark-application-1551883429892
               spark-exec-id=1
               spark-role=executor
Annotations:   <none>
Status:        Running
IP:            172.17.0.4
Containers:
  executor:
    Container ID:
      docker://259f686a22b94b5a08cd4c176ad980abced9b4a33ccf0bee0af94a460ede4546
    Image:      gcr.io/spark-operator/spark-py:v2.4.0
    Image ID:   docker-pullable://gcr.io/spark-operator/spark-py@sha256:95afcc1f1e0a3ee38db1469ddf147cb696418596d6cb8cc86d5b9cf1fd6e1c02
    Port:       7079/TCP
    Host Port:  0/TCP
    Args:
      executor
    State:      Running
    Started:    Wed, 06 Mar 2019 20:13:54 +0530
    Ready:      True
    Restart Count: 0
    Limits:
      memory: 1408Mi
    Requests:
      cpu:      1
      memory: 1408Mi
    Environment:
      SPARK_DRIVER_URL:
        spark://CoarseGrainedScheduler@192.168.0.102:46359
```

```

SPARK_EXECUTOR_CORES: 1
SPARK_EXECUTOR_MEMORY: 1g
SPARK_APPLICATION_ID: spark-application-1551883429892
SPARK_CONF_DIR: /opt/spark/conf
SPARK_EXECUTOR_ID: 1
SPARK_EXECUTOR_POD_IP: (v1:status.podIP)
SPARK_LOCAL_DIRS:
/var/data/spark-0e7b010d-e169-4836-9a9a-91be3d8b5402
Mounts:
/var/data/spark-0e7b010d-e169-4836-9a9a-91be3d8b5402 from spark-local-dir-1 (rw)
/var/run/secrets/kubernetes.io/serviceaccount from default-token-9mwn4 (ro)
Conditions:
Type      Status
Initialized True
Ready     True
PodScheduled True
Volumes:
spark-local-dir-1:
Type: EmptyDir (a temporary directory that shares a pod's lifetime)
Medium:
default-token-9mwn4:
Type: Secret (a volume populated by a Secret)
SecretName: default-token-9mwn4
Optional: false
QoS Class: Burstable
Node-Selectors: <none>
Tolerations: node.kubernetes.io/not-ready:NoExecute for 300s
              node.kubernetes.io/unreachable:NoExecute for 300s
Events:
Type Reason      Age From          Message
----
Normal Scheduled 108s default-scheduler Successfully assigned
pysparkshell-1551883429969-exec-1 to minikube
Normal SuccessfulMountVolume 108s kubelet, minikube MountVolume.SetUp
succeeded for volume "spark-local-dir-1"
Normal SuccessfulMountVolume 108s kubelet, minikube MountVolume.SetUp
succeeded for volume "default-token-9mwn4"
Normal Pulled 105s kubelet, minikube Container image
"gcr.io/spark-operator/spark-py:v2.4.0" already present on machine
Normal Created 104s kubelet, minikube Created container
Normal Started 104s kubelet, minikube Started container

```

3) Logs of the executor pod:

```

pujanm@pujanm:~$ kubectl logs -f pysparkshell-1551883429969-exec-1
++ id -u

```

```

+ myuid=0
++ id -g
+ mygid=0
+ set +e
++ getent passwd 0
+ uidentry=root:x:0:0:root:/root:/bin/ash
+ set -e
+ '[' -z root:x:0:0:root:/root:/bin/ash ']'
+ SPARK_K8S_CMD=executor
+ case "$SPARK_K8S_CMD" in
+ shift 1
+ SPARK_CLASSPATH='./opt/spark/jars/*'
+ env
+ sed 's/[^=]*=\\(.*)\\1/g'
+ sort -t_ -k4 -n
+ grep SPARK_JAVA_OPT_
+ readarray -t SPARK_EXECUTOR_JAVA_OPTS
+ '[' -n " " ']'
+ '[' -n " " ']'
+ PYSPARK_ARGS=
+ '[' -n " " ']'
+ R_ARGS=
+ '[' -n " " ']'
+ '[' " == 2 ']'
+ '[' " == 3 ']'
+ case "$SPARK_K8S_CMD" in
+ CMD=(${JAVA_HOME}/bin/java "${SPARK_EXECUTOR_JAVA_OPTS[@]}")
-Xms$SPARK_EXECUTOR_MEMORY -Xmx$SPARK_EXECUTOR_MEMORY -cp
"$SPARK_CLASSPATH" org.apache.spark.executor.CoarseGrainedExecutorBackend
--driver-url $SPARK_DRIVER_URL --executor-id $SPARK_EXECUTOR_ID --cores
$SPARK_EXECUTOR_CORES --app-id $SPARK_APPLICATION_ID --hostname
$SPARK_EXECUTOR_POD_IP)
+ exec /sbin/tini -s -- /usr/lib/jvm/java-1.8-openjdk/bin/java -Xms1g -Xmx1g -cp
'/opt/spark/jars/*' org.apache.spark.executor.CoarseGrainedExecutorBackend --driver-url
spark://CoarseGrainedScheduler@192.168.0.102:46359 --executor-id 1 --cores 1 --app-id
spark-application-1551883429892 --hostname 172.17.0.4
2019-03-06 14:43:55 INFO CoarseGrainedExecutorBackend:2566 - Started daemon with
process name: 13@pysparkshell-1551883429969-exec-1
2019-03-06 14:43:55 INFO SignalUtils:54 - Registered signal handler for TERM
2019-03-06 14:43:55 INFO SignalUtils:54 - Registered signal handler for HUP
2019-03-06 14:43:55 INFO SignalUtils:54 - Registered signal handler for INT
2019-03-06 14:43:56 WARN NativeCodeLoader:62 - Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
2019-03-06 14:43:56 INFO SecurityManager:54 - Changing view acls to: root
2019-03-06 14:43:56 INFO SecurityManager:54 - Changing modify acls to: root
2019-03-06 14:43:56 INFO SecurityManager:54 - Changing view acls groups to:

```

2019-03-06 14:43:56 INFO SecurityManager:54 - Changing modify acls groups to:
2019-03-06 14:43:56 INFO SecurityManager:54 - SecurityManager: authentication disabled;
ui acls disabled; users with view permissions: Set(root); groups with view permissions: Set();
users with modify permissions: Set(root); groups with modify permissions: Set()
2019-03-06 14:43:57 INFO TransportClientFactory:267 - Successfully created connection to
/192.168.0.102:46359 after 128 ms (0 ms spent in bootstraps)
2019-03-06 14:43:57 INFO SecurityManager:54 - Changing view acls to: root
2019-03-06 14:43:57 INFO SecurityManager:54 - Changing modify acls to: root
2019-03-06 14:43:57 INFO SecurityManager:54 - Changing view acls groups to:
2019-03-06 14:43:57 INFO SecurityManager:54 - Changing modify acls groups to:
2019-03-06 14:43:57 INFO SecurityManager:54 - SecurityManager: authentication disabled;
ui acls disabled; users with view permissions: Set(root); groups with view permissions: Set();
users with modify permissions: Set(root); groups with modify permissions: Set()
2019-03-06 14:43:58 INFO TransportClientFactory:267 - Successfully created connection to
/192.168.0.102:46359 after 4 ms (0 ms spent in bootstraps)
2019-03-06 14:43:58 INFO DiskBlockManager:54 - Created local directory at
/var/data/spark-0e7b010d-e169-4836-9a9a-91be3d8b5402/blockmgr-3802894e-7a93-4075-
9433-6f904f06f06c
2019-03-06 14:43:58 INFO MemoryStore:54 - MemoryStore started with capacity 413.9 MB
2019-03-06 14:43:58 INFO CoarseGrainedExecutorBackend:54 - Connecting to driver:
spark://CoarseGrainedScheduler@192.168.0.102:46359
2019-03-06 14:43:58 INFO CoarseGrainedExecutorBackend:54 - Successfully registered
with driver
2019-03-06 14:43:58 INFO Executor:54 - Starting executor ID 1 on host 172.17.0.4
2019-03-06 14:43:58 INFO Utils:54 - Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 41849.
2019-03-06 14:43:58 INFO NettyBlockTransferService:54 - Server created on
172.17.0.4:41849
2019-03-06 14:43:58 INFO BlockManager:54 - Using
org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
2019-03-06 14:43:58 INFO BlockManagerMaster:54 - Registering BlockManager
BlockManagerId(1, 172.17.0.4, 41849, None)
2019-03-06 14:43:58 INFO BlockManagerMaster:54 - Registered BlockManager
BlockManagerId(1, 172.17.0.4, 41849, None)
2019-03-06 14:43:58 INFO BlockManager:54 - Initialized BlockManager: BlockManagerId(1,
172.17.0.4, 41849, None)
2019-03-06 14:45:02 INFO CoarseGrainedExecutorBackend:54 - Got assigned task 0
2019-03-06 14:45:02 INFO Executor:54 - Running task 0.0 in stage 0.0 (TID 0)
2019-03-06 14:45:02 INFO TorrentBroadcast:54 - Started reading broadcast variable 0
2019-03-06 14:45:02 INFO TransportClientFactory:267 - Successfully created connection to
/192.168.0.102:35601 after 14 ms (0 ms spent in bootstraps)
2019-03-06 14:45:02 INFO MemoryStore:54 - Block broadcast_0_piece0 stored as bytes in
memory (estimated size 4.2 KB, free 413.9 MB)
2019-03-06 14:45:02 INFO TorrentBroadcast:54 - Reading broadcast variable 0 took 241
ms

2019-03-06 14:45:03 INFO MemoryStore:54 - Block broadcast_0 stored as values in memory (estimated size 6.6 KB, free 413.9 MB)

2019-03-06 14:46:12 INFO PythonRunner:54 - Times: total = 69652, boot = 1437, init = 84, finish = 68131

2019-03-06 14:46:12 INFO Executor:54 - Finished task 0.0 in stage 0.0 (TID 0). 1464 bytes result sent to driver

2019-03-06 14:46:12 INFO CoarseGrainedExecutorBackend:54 - Got assigned task 1

2019-03-06 14:46:12 INFO Executor:54 - Running task 1.0 in stage 0.0 (TID 1)

2019-03-06 14:47:22 INFO PythonRunner:54 - Times: total = 69106, boot = 4, init = 49, finish = 69053

2019-03-06 14:47:22 INFO Executor:54 - Finished task 1.0 in stage 0.0 (TID 1). 1421 bytes result sent to driver