

## Group09

### a. Representative Screenshots of the console output when you execute the program.

```
cc@group09-1: /usr/local/spark-1.6.1-bin-hadoop1
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

/home/cc/
Name
SparkHints.txt
search.py
quiz06.py
primes.py
part-00000
NYPD_Complaint_Data_Curr
mpihosts
mpinstall.sh
hwi2-reducer.py
hwi2-mapper.py
helloworld.py
helloSpark.py
HadoopHints.txt
hadoop-install.sh
data
chain.py
cc@10.56.0.148
assign3.py
.xauthority
.viminfo
.sudo_as_admin_successful
.python_history
.profile
.bashrc
.bash_logout
.bash_history
Follow terminal folder

cc@group09-1: /usr/local/spark-1.6.1-bin-hadoop1$ ./bin/spark-submit --master spark://10.56.3.221:7077 /home/cc/assign3.py
21/07/23 09:49:59 INFO spark.SparkContext: Running Spark version 1.6.1
21/07/23 09:49:59 INFO spark.SecurityManager: Changing view acls to: cc
21/07/23 09:49:59 INFO spark.SecurityManager: Changing modify acls to: cc
21/07/23 09:49:59 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(cc); users with modify permissions: Set(cc)
21/07/23 09:50:00 INFO util.Utils: Successfully started service 'sparkDriver' on port 42332.
21/07/23 09:50:00 INFO util.Utils: Slf4jLogger started
21/07/23 09:50:00 INFO Remoting: Starting remoting
21/07/23 09:50:00 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriverActorSystem@10.56.3.221:37364]
21/07/23 09:50:00 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 37364.
21/07/23 09:50:00 INFO spark.SparkEnv: Registering MapOutputTracker
21/07/23 09:50:00 INFO spark.SparkEnv: Registering BlockManagerMaster
21/07/23 09:50:00 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-bc330558-b72c-4d84-abfc-282a38a10439
21/07/23 09:50:00 INFO storage.MemoryStore: MemoryStore started with capacity 511.1 MB
21/07/23 09:50:00 INFO spark.SparkEnv: Registering OutputCommitCoordinator
21/07/23 09:50:00 INFO server.Server: jetty-8.y.z-SNAPSHOT
21/07/23 09:50:00 INFO server.AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
21/07/23 09:50:00 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.
21/07/23 09:50:00 INFO ui.SparkUI: Started SparkUI at http://10.56.3.221:4040
21/07/23 09:50:00 INFO spark.HttpFileServer: HTTP file server directory is /tmp/spark-a6af689d-5cb0-44db-8b21-63b99169e790/httpd-e00d490b-1894-4432-9bfa-3ac58c7553f5
21/07/23 09:50:00 INFO spark.HttpServer: Starting HTTP Server
21/07/23 09:50:00 INFO server.Server: jetty-8.y.z-SNAPSHOT
21/07/23 09:50:00 INFO server.AbstractConnector: Started SocketConnector@0.0.0.0:34821
21/07/23 09:50:00 INFO util.Utils: Successfully started service 'HTTP file server' on port 34821.
21/07/23 09:50:00 INFO util.Utils: Copying /home/cc/assign3.py to /tmp/spark-a6af689d-5cb0-44db-8b21-63b99169e790/userFiles-3385ba89-2351-4a4b-af6-4e5b4c4e7cd/assign3.py
21/07/23 09:50:00 INFO spark.SparkContext: Added file file:/home/cc/assign3.py at http://10.56.3.221:34821/files/assign3.py with timestamp 1627033800499
21/07/23 09:50:00 INFO client.AppClientsClientEndpoint: Connecting to master spark://10.56.3.221:7077...
21/07/23 09:50:00 INFO cluster.SparkDeploySchedulerBackend: Connected to Spark cluster with app ID app-20210723095000-0056
21/07/23 09:50:00 INFO client.AppClientsClientEndpoint: Executor added: app-20210723095000-0056/0 on worker-20210723002628-10.56.0.148-42246 (10.56.0.148:42246) with 2 cores
21/07/23 09:50:00 INFO cluster.SparkDeploySchedulerBackend: Granted executor ID app-20210723095000-0056/0 on hostPort 10.56.0.148:42246 with 2 cores, 1024.0 MB RAM
21/07/23 09:50:00 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 46789.
21/07/23 09:50:00 INFO netty.NettyBlockTransferService: Server created on 46789
21/07/23 09:50:00 INFO storage.BlockManagerMaster: Trying to register BlockManager
21/07/23 09:50:00 INFO storage.BlockManagerMasterEndpoint: Registering block manager 10.56.3.221:46789 with 511.1 MB RAM, BlockManagerId(driver, 10.56.3.221, 46789)
21/07/23 09:50:00 INFO storage.BlockManagerMaster: Registered BlockManager
21/07/23 09:50:00 INFO client.AppClientsClientEndpoint: Executor added: app-20210723095000-0056/1 on worker-20210723002629-10.56.3.221-39073 (10.56.3.221:39073) with 2 cores
21/07/23 09:50:00 INFO cluster.SparkDeploySchedulerBackend: Granted executor ID app-20210723095000-0056/1 on hostPort 10.56.3.221:39073 with 2 cores, 1024.0 MB RAM
21/07/23 09:50:00 INFO client.AppClientsClientEndpoint: Executor updated: app-20210723095000-0056/0 is now RUNNING
21/07/23 09:50:01 INFO client.AppClientsClientEndpoint: Executor updated: app-20210723095000-0056/1 is now RUNNING
21/07/23 09:50:01 INFO cluster.SparkDeploySchedulerBackend: SchedulerBackend is ready for scheduling beginning after reached minRegisteredResourcesRatio: 0.0
What are the top 3 crime types that were reported in the month of July?
PETIT LARCENY 7086
HARRASSMENT 2 5738
ASSAULT 3 & RELATED OFFENSES 4804
How many crimes of type DANGEROUS WEAPONS were reported in the month of July?
DANGEROUS WEAPONS 701
cc@group09-1: /usr/local/spark-1.6.1-bin-hadoop1$ ./bin/spark-submit --master spark://10.56.3.221:7077 /home/cc/assign3.py
```

### b. Output of your Spark program.

What are the top 3 crime types that were reported in the month of July?  
PETIT LARCENY 7086

HARRASSMENT 2 5738

ASSAULT 3 & RELATED OFFENSES 4804

How many crimes of type DANGEROUS WEAPONS were reported in the month of July?  
DANGEROUS WEAPONS 701

c. Describe the contribution of each group member

Sebastian Klemkosky: The project was easy enough that I was able to complete it myself.