Assignment 5

Report File

Question 1: Second Highest Value Transaction in Selected Countries

Observation:

No Of Cores=2   Time Taken=72.718 seconds

```
(env) [cs3304.058@abacus 2023201059_Assignment5]$ sint3 -c 2
salloc: Granted job allocation 1060552
salloc: Waiting for resource configuration
salloc: Nodes node01 are ready for job
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q1.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/11/12 10:53:17 WARN NativeCodeLoader: Unable to load native-hadoop library for your platf
e
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarnin
  FutureWarning
[Stage 0:>                                                              (0
Second highest transacted value is 20000000 which is transacted in GREATER LONDON.
Execution time: 72.71898555755615 seconds
```

No Of Cores=4 Time Taken=44.516 seconds

```
(env) [cs3304.058@abacus 2023201059_Assignment5]$ sint3 -c 4
salloc: Granted job allocation 1060554
salloc: Waiting for resource configuration
salloc: Nodes node01 are ready for job
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q1.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel)
23/11/12 10:55:47 WARN NativeCodeLoader: Unable to load native-hadoop library for your plat
e
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarni
  FutureWarning
Second highest transacted value is 20000000 which is transacted in GREATER LONDON.
Execution time: 44.51680397987366 seconds
[cs3304.058@node01 2023201059_Assignment5]$
```

No Of Cores=6 Time Taken=39.297 seconds

```
salloc: Relinquishing job allocation 1060554
(env) [cs3304.058@abacus 2023201059_Assignment5]$ sint3 -c 6
salloc: Granted job allocation 1060557
salloc: Waiting for resource configuration
salloc: Nodes node01 are ready for job
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q1.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/11/12 10:57:19 WARN NativeCodeLoader: Unable to load native-hadoop library for your platf
e
23/11/12 10:57:20 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarnin
  FutureWarning
Second highest transacted value is 20000000 which is transacted in GREATER LONDON.
Execution time: 39.29767942428589 seconds
[cs3304.058@node01 2023201059_Assignment5]$
```

Change In Execution Time

From 2 to 4 cores: 38.78%

- From 4 to 6 cores: 11.72%

Question 2: Country with the Second Most Transactions

Observation:

No Of Cores=2 Time Taken=25.342 seconds

```
(env) [cs3304.058@abacus 2023201059_Assignment5]$ sint3 -c 2
salloc: Granted job allocation 1060558
salloc: Waiting for resource configuration
salloc: Nodes node01 are ready for job
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q2.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/11/12 10:58:36 WARN NativeCodeLoader: Unable to load native-hadoop library for your platfor
e
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarning:
  FutureWarning
Country with the second most transactions is 'GREATER MANCHESTER'
Execution time: 25.342904090881348 seconds
[cs3304.058@node01 2023201059_Assignment5]$
```

No Of Cores=6 Time Taken=19.797 seconds

```
(env) [cs3304.058@abacus 2023201059_Assignment5]$ sint3 -c 4
salloc: Granted job allocation 1060559
salloc: Waiting for resource configuration
salloc: Nodes node01 are ready for job
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q2.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/11/12 11:00:49 WARN NativeCodeLoader: Unable to load native-hadoop library for your platfor
e
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarning:
  FutureWarning
Country with the second most transactions is 'GREATER MANCHESTER'
Execution time: 19.797455549240112 seconds
[cs3304.058@node01 2023201059_Assignment5]$
```

No Of Cores=6 Time Taken=18.239 seconds

```
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q2.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/11/12 11:03:54 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform.
e
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarning: P
  FutureWarning
Country with the second most transactions is 'GREATER MANCHESTER'
Execution time: 18.239944458007812 seconds
[cs3304.058@node01 2023201059_Assignment5]$
```

- **Change In Execution Time**

  From 2 to 4 cores: 21.88%
- From 4 to 6 cores: 7.86%

Question 3: Number of Transactions for Each Country

Observation:

No Of Cores=2 Time Taken=47.492 seconds

```
(env) [cs3304.058@abacus 2023201059_Assignment5]$ sint3 -c 2
salloc: Granted job allocation 1060562
salloc: Waiting for resource configuration
salloc: Nodes node01 are ready for job
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q3.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/11/12 11:05:11 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform
e
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarning:
  FutureWarning
Execution time: 47.49288558959961 seconds
[cs3304.058@node01 2023201059_Assignment5]$
```

No Of Cores=4 Time Taken=31.010 seconds

```
(env) [cs3304.058@abacus 2023201059_Assignment5]$ sint3 -c 4
salloc: Granted job allocation 1060563
salloc: Waiting for resource configuration
salloc: Nodes node01 are ready for job
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q3.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/11/12 11:06:41 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform
e
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarning:
  FutureWarning
Execution time: 31.010701179504395 seconds
[cs3304.058@node01 2023201059_Assignment5]$
```

No Of Cores=6 Time Taken=23.298 seconds

```
(env) [cs3304.058@abacus 2023201059_Assignment5]$ sint3 -c 6
salloc: Granted job allocation 1060564
salloc: Waiting for resource configuration
salloc: Nodes node01 are ready for job
[cs3304.058@node01 2023201059_Assignment5]$ python3 2023201059_q3.py
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/11/12 11:07:36 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform
e
/home/iiit/cs3304.058/env/lib64/python3.6/site-packages/pyspark/context.py:238: FutureWarning:
  FutureWarning
Execution time: 23.29853343963623 seconds
[cs3304.058@node01 2023201059_Assignment5]$
```

- **Change In Execution Time**

  From 2 to 4 cores: 34.70%
- From 4 to 6 cores: 24.86%

Overall Observations:

- Increasing the number of cores positively influences the execution time for each question.
- The impact of additional cores is more pronounced in scenarios with larger datasets or complex computations.
- While the trend is consistent, the diminishing returns become evident as the core count increases.
- It's crucial to balance the computational power required with the available resources, considering factors like cost and infrastructure constraints