Experiment 6

Aim: Streaming Analytics with Kafka and Spark: Set up a data streaming pipeline using Apache Kafka to ingest real-time data and process it using Apache Spark Streaming for real-time analytics.

Starting Services:

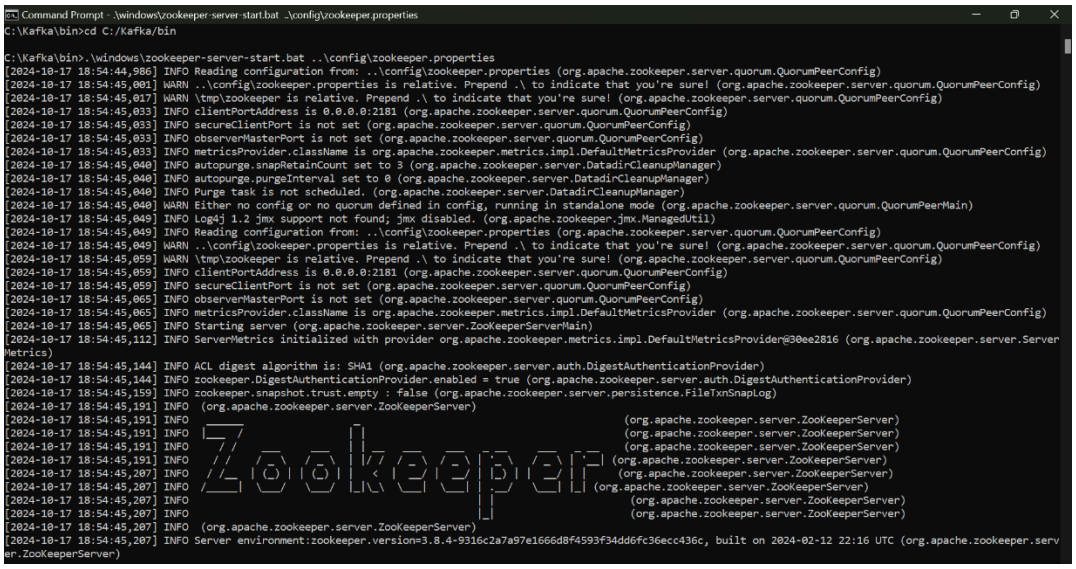

Fig 1: Zookeeper Server

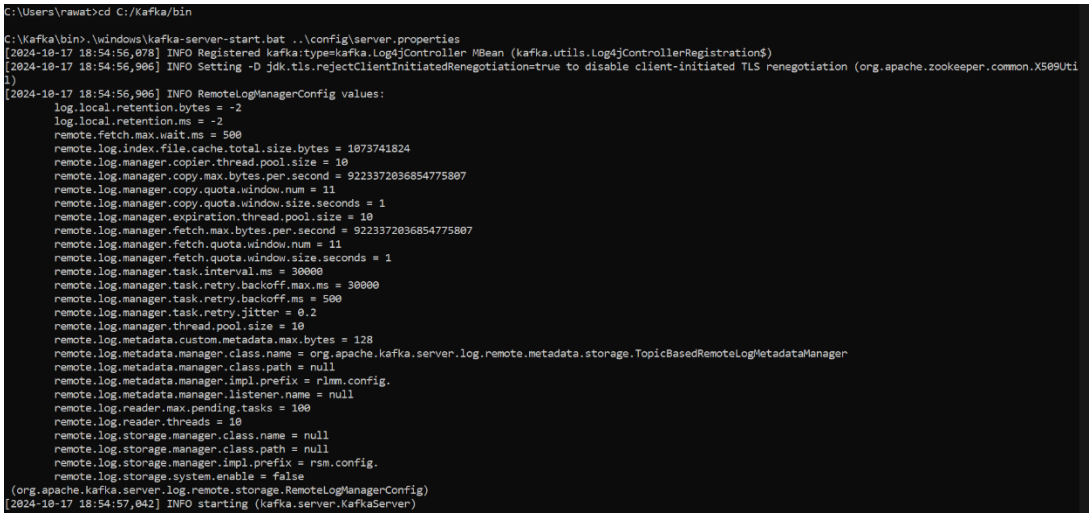

Fig 2: Kafka Server/Broker

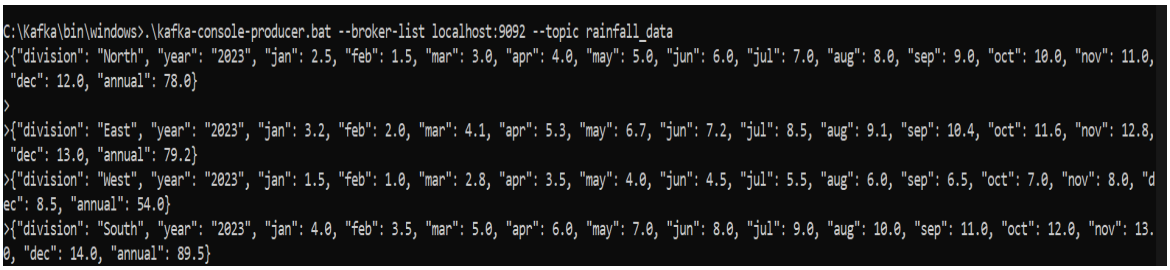

Fig 3: Sending Data to Kafka Streaming using Dictionary line by line

```
C:\Users\rawat\Documents\7 SEMESTER\Big Data Analytics\Lab\Experiment 5>python kafka_producer.py
```

Fig 4: Sending Data to Kafka Streaming using CSV File

Spark Code Execution:

```
C:\Users\rawat\Documents\7 SEMESTER\Big Data Analytics\Lab\Experiment 5>spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.5.2 spark_streaming.py
:: loading settings :: url = jar:file:/C:/spark/jars/ivy-2.5.1.jar!/org/apache/ivy/core/settings/ivysettings.xml
Ivy Default Cache set to: C:\Users\rawat\.ivy2\cache
The jars for the packages stored in: C:\Users\rawat\.ivy2\jars
org.apache.spark#spark-sql-kafka-0-10_2.12 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-c450689e-e0d6-45db-a0f3-cba56eb3e250;1.0
        confs: [default]
        found org.apache.spark#spark-sql-kafka-0-10_2.12;3.5.2 in central
        found org.apache.spark#spark-token-provider-kafka-0-10_2.12;3.5.2 in central
        found org.apache.kafka#kafka-clients;3.4.1 in central
        found org.lz4#lz4-java;1.8.0 in central
        found org.xerial.snappy#snappy-java;1.1.10.5 in central
        found org.slf4j#slf4j-api;2.0.7 in central
        found org.apache.hadoop#hadoop-client-runtime;3.3.4 in central
        found org.apache.hadoop#hadoop-client-api;3.3.4 in central
        found commons-logging#commons-logging;1.1.3 in central
        found com.google.code.findbugs#jsr305;3.0.0 in central
        found org.apache.commons#commons-pool2;2.11.1 in central
:: resolution report :: resolve 708ms :: artifacts dl 110ms
        :: modules in use:
        com.google.code.findbugs#jsr305;3.0.0 from central in [default]
        commons-logging#commons-logging;1.1.3 from central in [default]
        org.apache.commons#commons-pool2;2.11.1 from central in [default]
        org.apache.hadoop#hadoop-client-api;3.3.4 from central in [default]
        org.apache.hadoop#hadoop-client-runtime;3.3.4 from central in [default]
```

Fig 5: Execution of Spark File/Code

```
        org.lz4#lz4-java;1.8.0 from central in [default]
        org.slf4j#slf4j-api;2.0.7 from central in [default]
        org.xerial.snappy#snappy-java;1.1.10.5 from central in [default]
        ---------------------------------------------------------------------
        |                  |            modules            ||   artifacts   |
        |       conf       | number| search|dwnlded|evicted|| number|dwnlded|
        ---------------------------------------------------------------------
        |      default     |   11  |   0   |   0   |   0   ||   11  |   0   |
        ---------------------------------------------------------------------
rieving :: org.apache.spark#spark-submit-parent-0dc1a2df-4075-44ca-aaca-e9eef6e2fd3d
        confs: [default]
        0 artifacts copied, 11 already retrieved (0kB/35ms)
17 19:12:00 INFO SparkContext: Running Spark version 3.5.2
17 19:12:00 INFO SparkContext: OS info Windows 11, 10.0, amd64
```

Fig 6: Output of Execution of Spark File/Code showing installation of all required modules

```
-------------------------------------------
Batch: 1
-------------------------------------------
+--------+-----+------------------+------------------+------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+------------------+
|    NULL|    1|              NULL|              NULL|              NULL|
+--------+-----+------------------+------------------+------------------+

-------------------------------------------
Batch: 2
-------------------------------------------
+--------+-----+------------------+------------------+------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+------------------+
|    NULL|    1|              NULL|              NULL|              NULL|
|   North|    1|              78.0|              78.0|              78.0|
+--------+-----+------------------+------------------+------------------+

-------------------------------------------
Batch: 3
-------------------------------------------
+--------+-----+------------------+------------------+------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+------------------+
|    NULL|    1|              NULL|              NULL|              NULL|
|   North|    2|              78.0|              78.0|              78.0|
+--------+-----+------------------+------------------+------------------+

-------------------------------------------
Batch: 4
-------------------------------------------
+--------+-----+------------------+------------------+------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+------------------+
|    NULL|    2|              NULL|              NULL|              NULL|
|   North|    2|              78.0|              78.0|              78.0|
+--------+-----+------------------+------------------+------------------+
```

Fig 7: Batch Execution of Spark Code showing batch 1-4

```
+--------+-----+------------------+------------------+-------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+-------------------+
|    NULL|    2|              NULL|              NULL|               NULL|
|   North|    2|              78.0|              78.0|               78.0|
+--------+-----+------------------+------------------+-------------------+

----------------------------------------
Batch: 5
----------------------------------------
+--------+-----+------------------+------------------+-------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+-------------------+
|    NULL|    2|              NULL|              NULL|               NULL|
|    East|    1|              79.2|              79.2|   79.19999694824219|
|   North|    2|              78.0|              78.0|               78.0|
+--------+-----+------------------+------------------+-------------------+

----------------------------------------
Batch: 6
----------------------------------------
+--------+-----+------------------+------------------+-------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+-------------------+
|    NULL|    2|              NULL|              NULL|               NULL|
|    East|    1|              79.2|              79.2|   79.19999694824219|
|    West|    1|              54.0|              54.0|               54.0|
|   North|    2|              78.0|              78.0|               78.0|
+--------+-----+------------------+------------------+-------------------+
```

Fig 8: Batch Execution of Spark Code showing batch 5-6

```
----------------------------------------
Batch: 7
----------------------------------------
+--------+-----+------------------+------------------+-------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+-------------------+
|    NULL|    2|              NULL|              NULL|               NULL|
|   South|    2|              89.5|              89.5|               89.5|
| Central|    1|              66.5|              66.5|               66.5|
|    East|    1|              79.2|              79.2|   79.19999694824219|
|    West|    1|              54.0|              54.0|               54.0|
|   North|    2|              78.0|              78.0|               78.0|
+--------+-----+------------------+------------------+-------------------+

----------------------------------------
Batch: 8
----------------------------------------
+--------+-----+------------------+------------------+-------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+-------------------+
|    NULL|    2|              NULL|              NULL|               NULL|
|   South|    2|              89.5|              89.5|               89.5|
| Central|    1|              66.5|              66.5|               66.5|
|    East|    1|              79.2|              79.2|   79.19999694824219|
|    West|    1|              54.0|              54.0|               54.0|
|   North|    3|              78.0|              78.0|               78.0|
+--------+-----+------------------+------------------+-------------------+
```

Fig 9: Batch Execution of Spark Code showing batch 7-8

```
----------------------------------------
Batch: 8
----------------------------------------
+--------+-----+------------------+------------------+-------------------+
|division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------+-----+------------------+------------------+-------------------+
|    NULL|    2|              NULL|              NULL|               NULL|
|   South|    2|              89.5|              89.5|               89.5|
| Central|    1|              66.5|              66.5|               66.5|
|    East|    1|              79.2|              79.2|   79.19999694824219|
|    West|    1|              54.0|              54.0|               54.0|
|   North|    3|              78.0|              78.0|               78.0|
+--------+-----+------------------+------------------+-------------------+

----------------------------------------
Batch: 9
----------------------------------------
+-------------------+-----+------------------+------------------+-------------------+
|           division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+-------------------+-----+------------------+------------------+-------------------+
|               NULL|    2|              NULL|              NULL|               NULL|
|              South|    2|              89.5|              89.5|               89.5|
|ANDAMAN & NICOBAR...|   51|            2352.1|               NaN|                NaN|
|            Central|    1|              66.5|              66.5|               66.5|
|               East|    1|              79.2|              79.2|   79.19999694824219|
|               West|    1|              54.0|              54.0|               54.0|
|              North|    3|              78.0|              78.0|               78.0|
+-------------------+-----+------------------+------------------+-------------------+
```

Fig 10: Batch Execution of Spark Code showing batch 8-9

```
----------------------------------------
Batch: 10
----------------------------------------
+--------------------+-----+------------------+------------------+--------------------+
|           division|count|min_annual_rainfall|max_annual_rainfall|mean_annual_rainfall|
+--------------------+-----+------------------+------------------+--------------------+
|            VIDARBHA|  115|             578.5|            1606.3|  1095.4591303286345|
|NAGA MANI MIZO TR...|  115|            1353.8|            4316.2|   2433.619123641304|
|        CHHATTISGARH|  115|             904.6|            1974.0|  1371.7286875849186|
|                NULL|    2|              NULL|              NULL|                NULL|
|SUB HIMALAYAN WES...|  115|            1988.2|            3655.1|  2752.2173955502717|
|GANGETIC WEST BENGAL|  115|            1015.1|            2099.8|  1490.4878226902174|
|    HIMACHAL PRADESH|  115|             776.1|            1919.2|  1260.3452153744904|
|               BIHAR|  115|             629.2|            1660.4|   1197.63390847911|
|              ORISSA|  115|             987.0|            1945.3|  1458.1695694633152|
|     JAMMU & KASHMIR|  115|             657.0|               NaN|                 NaN|
|   ASSAM & MEGHALAYA|  115|            1743.4|            3403.5|   2580.695658542799|
|               South|    2|              89.5|              89.5|                89.5|
|         LAKSHADWEEP|  114|             992.6|               NaN|                 NaN|
|ANDAMAN & NICOBAR...|  110|            1849.4|               NaN|                 NaN|
|          TAMIL NADU|  115|             318.0|            1365.3|   943.7130437436311|
|             Central|    1|              66.5|              66.5|                66.5|
|NORTH INTERIOR KA...|  115|             470.3|            1095.6|   717.7956508470618|
|  WEST UTTAR PRADESH|  115|             371.9|            1244.2|   827.1147813879925|
|SOUTH INTERIOR KA...|  115|             733.3|            1409.5|  1040.3913027556046|
| EAST MADHYA PRADESH|  115|             653.8|            1747.1|   1205.000001061481|
+--------------------+-----+------------------+------------------+--------------------+
only showing top 20 rows
```

Fig 11: Batch Execution of Spark Code showing batch 10