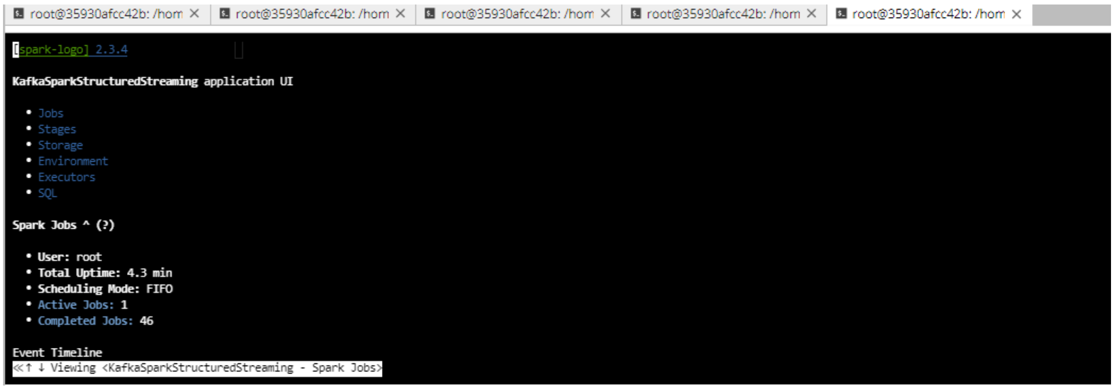
SF Crime Statistics with Spark Streaming Screen Shots

Contents

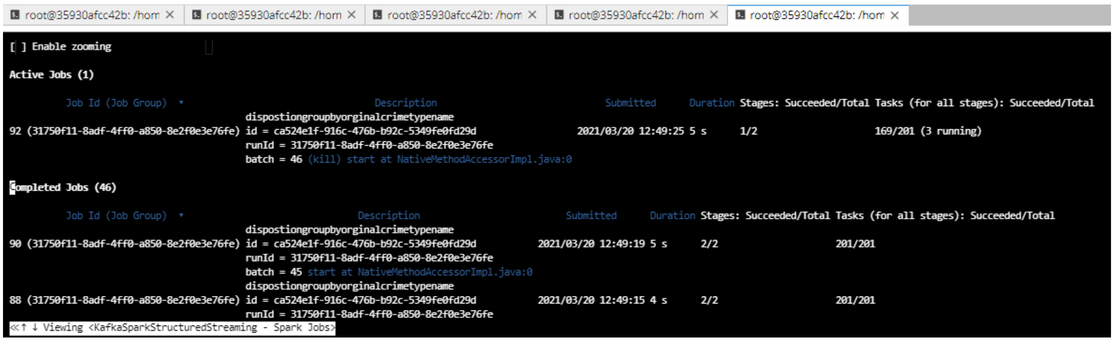
1) kafka-consumer-console output

```
"      \"common_location\": \"Stonestown Galleria, Sf\"\n"

"    },\n"
"   {\n"
"      \"crime_id\": \"183653745\",\n"
"      \"original_crime_type_name\": \"Audible Alarm\",\n"
"      \"report_date\": \"2018-12-31T00:00:00.000\",\n"
"      \"call_date\": \"2018-12-31T00:00:00.000\",\n"
"      \"offense_date\": \"2018-12-31T00:00:00.000\",\n"
"      \"call_time\": \"23:47\",\n"
"      \"call_date_time\": \"2018-12-31T23:47:00.000\",\n"
"      \"disposition\": \"PAS\",\n"
"      \"address\": \"1900 Block Of 18th Av\",\n"
"      \"city\": \"San Francisco\",\n"
"      \"state\": \"CA\",\n"
"      \"agency_id\": \"1\",\n"
"      \"address_type\": \"Premise Address\",\n"
"      \"common_location\": \"\"\n"
"   },\n"
"   {\n"
^CProcessed a total of 66 messages
root@35930afcc42b:/home/workspace#
```

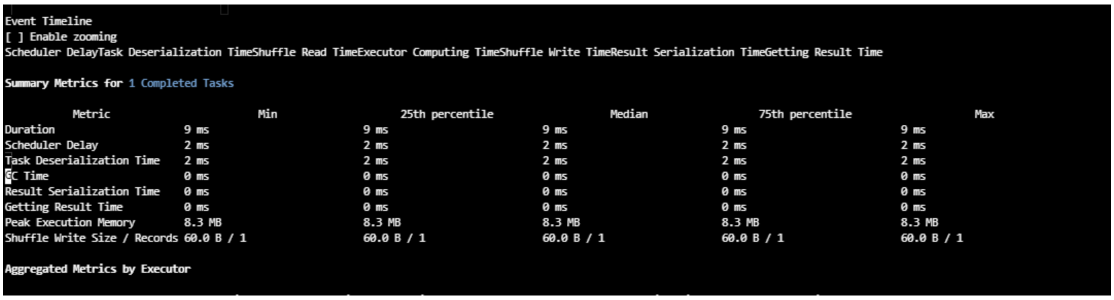
2a) Progress Reporter Screen Shots (only default configuration)



```
[spark-logo] 2.3.4

KafkaSparkStructuredStreaming application UI

  • Jobs
  • Stages
  • Storage
  • Environment
  • Executors
  • SQL

Spark Jobs ^ (?)

  • User: root
  • Total Uptime: 4.3 min
  • Scheduling Mode: FIFO
  • Active Jobs: 1
  • Completed Jobs: 46

Event Timeline
<↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```



```
[ ] Enable zooming

Active Jobs (1)

        Job Id (Job Group)  ▾                          Description                     Submitted      Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                   dispostiongroupbyorginalcrimetypename
92 (31750f11-8adf-4ff0-a850-8e2f0e3e76fe) id = ca524e1f-916c-476b-b92c-5349fe0fd29d            2021/03/20 12:49:25 5 s     1/2                   169/201 (3 running)
                                   runId = 31750f11-8adf-4ff0-a850-8e2f0e3e76fe
                                   batch = 46 (kill) start at NativeMethodAccessorImpl.java:0

Completed Jobs (46)

        Job Id (Job Group)  ▾                        Description                     Submitted      Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                   dispostiongroupbyorginalcrimetypename
90 (31750f11-8adf-4ff0-a850-8e2f0e3e76fe) id = ca524e1f-916c-476b-b92c-5349fe0fd29d         2021/03/20 12:49:19 5 s     2/2                   201/201
                                   runId = 31750f11-8adf-4ff0-a850-8e2f0e3e76fe
                                   batch = 45 start at NativeMethodAccessorImpl.java:0
                                   dispostiongroupbyorginalcrimetypename
88 (31750f11-8adf-4ff0-a850-8e2f0e3e76fe) id = ca524e1f-916c-476b-b92c-5349fe0fd29d         2021/03/20 12:49:15 4 s     2/2                   201/201
                                   runId = 31750f11-8adf-4ff0-a850-8e2f0e3e76fe
<↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```



```
Event Timeline
[ ] Enable zooming
Scheduler DelayTask Deserialization TimeShuffle Read TimeExecutor Computing TimeShuffle Write TimeResult Serialization TimeGetting Result Time

Summary Metrics for 1 Completed Tasks

          Metric                      Min                  25th percentile                Median              75th percentile                Max
Duration                  9 ms                    9 ms                    9 ms                    9 ms                    9 ms
Scheduler Delay           2 ms                    2 ms                    2 ms                    2 ms                    2 ms
Task Deserialization Time 2 ms                    2 ms                    2 ms                    2 ms                    2 ms
GC Time                   0 ms                    0 ms                    0 ms                    0 ms                    0 ms
Result Serialization Time 0 ms                    0 ms                    0 ms                    0 ms                    0 ms
Getting Result Time       0 ms                    0 ms                    0 ms                    0 ms                    0 ms
Peak Execution Memory     8.3 MB                  8.3 MB                  8.3 MB                  8.3 MB                  8.3 MB
Shuffle Write Size / Records 60.0 B / 1           60.0 B / 1              60.0 B / 1              60.0 B / 1              60.0 B / 1

Aggregated Metrics by Executor
```

2b) Spark Streaming UI

2021-03-20 13:09:51 INFO ContextHandler:781 - Started o.s.j.s.ServletContextHandler@e2989ae{/storage/rdd/json,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO ContextHandler:781 - Started o.s.j.s.ServletContextHandler@284f6d5{/environment,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO ContextHandler:781 - Started o.s.j.s.ServletContextHandler@36e9e6{/environment/json,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO ContextHandler:781 - Started o.s.j.s.ServletContextHandler@58766e4e{/executors,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO ContextHandler:781 - Started o.s.j.s.ServletContextHandler@1ad0e817{/executors/json,null,AVAILABLE,@Spark}

```
2021-03-20 13:09:51 INFO  ContextHandler:781 - Started
o.s.j.s.ServletContextHandler@3fc787c6{/executors/threadDump,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO  ContextHandler:781 - Started
o.s.j.s.ServletContextHandler@8892001{/executors/threadDump/json,null,AVAILABLE,@Spa
rk}

2021-03-20 13:09:51 INFO  ContextHandler:781 - Started
o.s.j.s.ServletContextHandler@63e63b40{/static,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO  ContextHandler:781 - Started
o.s.j.s.ServletContextHandler@5e7bf2bc{/,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO  ContextHandler:781 - Started
o.s.j.s.ServletContextHandler@454d1d2e{/api,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO  ContextHandler:781 - Started
o.s.j.s.ServletContextHandler@7a4817f{/jobs/job/kill,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO  ContextHandler:781 - Started
o.s.j.s.ServletContextHandler@63aec600{/stages/stage/kill,null,AVAILABLE,@Spark}

2021-03-20 13:09:51 INFO  SparkUI:54 - Bound SparkUI to 0.0.0.0, and started at
http://35930afcc42b:4040

2021-03-20 13:09:51 INFO  SparkContext:54 - Added JAR
file:///root/.ivy2/jars/org.apache.spark_spark-sql-kafka-0-10_2.11-2.3.4.jar at
spark://35930afcc42b:35823/jars/org.apache.spark_spark-sql-kafka-0-10_2.11-2.3.4.jar with
timestamp 1616245791691

2021-03-20 13:09:51 INFO  SparkContext:54 - Added JAR
file:///root/.ivy2/jars/org.apache.kafka_kafka-clients-0.10.0.1.jar at
spark://35930afcc42b:35823/jars/org.apache.kafka_kafka-clients-0.10.0.1.jar with
timestamp 1616245791695

2021-03-20 13:09:51 INFO  SparkContext:54 - Added JAR file:///root/.ivy2/jars/org.spark-
project.spark_unused-1.0.0.jar at spark://35930afcc42b:35823/jars/org.spark-
project.spark_unused-1.0.0.jar with timestamp 1616245791697

2021-03-20 13:09:51 INFO  SparkContext:54 - Added JAR
file:///root/.ivy2/jars/net.jpountz.lz4_lz4-1.3.0.jar at
spark://35930afcc42b:35823/jars/net.jpountz.lz4_lz4-1.3.0.jar with timestamp
1616245791699

2021-03-20 13:09:51 INFO  SparkContext:54 - Added JAR
file:///root/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.2.6.jar at
spark://35930afcc42b:35823/jars/org.xerial.snappy_snappy-java-1.1.2.6.jar with timestamp
1616245791700
```

```
2021-03-20 13:09:51 INFO  SparkContext:54 - Added JAR file:///root/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar at spark://35930afcc42b:35823/jars/org.slf4j_slf4j-api-1.7.16.jar with timestamp 1616245791703

2021-03-20 13:09:51 INFO  SparkContext:54 - Added file file:/home/workspace/data_stream.py at file:/home/workspace/data_stream.py with timestamp 1616245791748

2021-03-20 13:09:51 INFO  Utils:54 - Copying /home/workspace/data_stream.py to /tmp/spark-d3ee8b74-0f31-4c28-a973-31e8565198bb/userFiles-bdda2a4f-f262-4253-8629-3fb406e16da5/data_stream.py

2021-03-20 13:09:51 INFO  SparkContext:54 - Added file file:///root/.ivy2/jars/org.apache.spark_spark-sql-kafka-0-10_2.11-2.3.4.jar at file:///root/.ivy2/jars/org.apache.spark_spark-sql-kafka-0-10_2.11-2.3.4.jar with timestamp 1616245791769

2021-03-20 13:09:51 INFO  Utils:54 - Copying /root/.ivy2/jars/org.apache.spark_spark-sql-kafka-0-10_2.11-2.3.4.jar to /tmp/spark-d3ee8b74-0f31-4c28-a973-31e8565198bb/userFiles-bdda2a4f-f262-4253-8629-3fb406e16da5/org.apache.spark_spark-sql-kafka-0-10_2.11-2.3.4.jar

2021-03-20 13:09:51 INFO  SparkContext:54 - Added file file:///root/.ivy2/jars/org.apache.kafka_kafka-clients-0.10.0.1.jar at file:///root/.ivy2/jars/org.apache.kafka_kafka-clients-0.10.0.1.jar with timestamp 1616245791782

2021-03-20 13:09:51 INFO  Utils:54 - Copying /root/.ivy2/jars/org.apache.kafka_kafka-clients-0.10.0.1.jar to /tmp/spark-d3ee8b74-0f31-4c28-a973-31e8565198bb/userFiles-bdda2a4f-f262-4253-8629-3fb406e16da5/org.apache.kafka_kafka-clients-0.10.0.1.jar

2021-03-20 13:09:51 INFO  SparkContext:54 - Added file file:///root/.ivy2/jars/org.spark-project.spark_unused-1.0.0.jar at file:///root/.ivy2/jars/org.spark-project.spark_unused-1.0.0.jar with timestamp 1616245791799

2021-03-20 13:09:51 INFO  Utils:54 - Copying /root/.ivy2/jars/org.spark-project.spark_unused-1.0.0.jar to /tmp/spark-d3ee8b74-0f31-4c28-a973-31e8565198bb/userFiles-bdda2a4f-f262-4253-8629-3fb406e16da5/org.spark-project.spark_unused-1.0.0.jar

2021-03-20 13:09:51 INFO  SparkContext:54 - Added file file:///root/.ivy2/jars/net.jpountz.lz4_lz4-1.3.0.jar at file:///root/.ivy2/jars/net.jpountz.lz4_lz4-1.3.0.jar with timestamp 1616245791804

2021-03-20 13:09:51 INFO  Utils:54 - Copying /root/.ivy2/jars/net.jpountz.lz4_lz4-1.3.0.jar to /tmp/spark-d3ee8b74-0f31-4c28-a973-31e8565198bb/userFiles-bdda2a4f-f262-4253-8629-3fb406e16da5/net.jpountz.lz4_lz4-1.3.0.jar
```

```
2021-03-20 13:09:51 INFO  SparkContext:54 - Added file file:///root/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.2.6.jar at file:///root/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.2.6.jar with timestamp 1616245791814

2021-03-20 13:09:51 INFO  Utils:54 - Copying /root/.ivy2/jars/org.xerial.snappy_snappy-java-1.1.2.6.jar to /tmp/spark-d3ee8b74-0f31-4c28-a973-31e8565198bb/userFiles-bdda2a4f-f262-4253-8629-3fb406e16da5/org.xerial.snappy_snappy-java-1.1.2.6.jar

2021-03-20 13:09:51 INFO  SparkContext:54 - Added file file:///root/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar at file:///root/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar with timestamp 1616245791824

2021-03-20 13:09:51 INFO  Utils:54 - Copying /root/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar to /tmp/spark-d3ee8b74-0f31-4c28-a973-31e8565198bb/userFiles-bdda2a4f-f262-4253-8629-3fb406e16da5/org.slf4j_slf4j-api-1.7.16.jar

2021-03-20 13:09:51 INFO  Executor:54 - Starting executor ID driver on host localhost

2021-03-20 13:09:51 INFO  Utils:54 - Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 42769.

2021-03-20 13:09:51 INFO  NettyBlockTransferService:54 - Server created on 35930afcc42b:42769

2021-03-20 13:09:51 INFO  BlockManager:54 - Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

2021-03-20 13:09:51 INFO  BlockManagerMaster:54 - Registering BlockManager BlockManagerId(driver, 35930afcc42b, 42769, None)

2021-03-20 13:09:51 INFO  BlockManagerMasterEndpoint:54 - Registering block manager 35930afcc42b:42769 with 366.3 MB RAM, BlockManagerId(driver, 35930afcc42b, 42769, None)

2021-03-20 13:09:51 INFO  BlockManagerMaster:54 - Registered BlockManager BlockManagerId(driver, 35930afcc42b, 42769, None)

2021-03-20 13:09:51 INFO  BlockManager:54 - Initialized BlockManager: BlockManagerId(driver, 35930afcc42b, 42769, None)

2021-03-20 13:09:52 INFO  ContextHandler:781 - Started o.s.j.s.ServletContextHandler@32cea45a{/metrics/json,null,AVAILABLE,@Spark}

2021-03-20 13:09:52 INFO  SharedState:54 - Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/home/workspace/spark-warehouse/').

2021-03-20 13:09:52 INFO  SharedState:54 - Warehouse path is 'file:/home/workspace/spark-warehouse/'.
```

```
2021-03-20 13:09:52 INFO  ContextHandler:781 - Started o.s.j.s.ServletContextHandler@128b5264{/SQL,null,AVAILABLE,@Spark}
2021-03-20 13:09:52 INFO  ContextHandler:781 - Started o.s.j.s.ServletContextHandler@500787b3{/SQL/json,null,AVAILABLE,@Spark}
2021-03-20 13:09:52 INFO  ContextHandler:781 - Started o.s.j.s.ServletContextHandler@19e76122{/SQL/execution,null,AVAILABLE,@Spark}
2021-03-20 13:09:52 INFO  ContextHandler:781 - Started o.s.j.s.ServletContextHandler@6355ae43{/SQL/execution/json,null,AVAILABLE,@Spark}
2021-03-20 13:09:52 INFO  ContextHandler:781 - Started o.s.j.s.ServletContextHandler@714fee92{/static/sql,null,AVAILABLE,@Spark}
2021-03-20 13:09:53 INFO  StateStoreCoordinatorRef:54 - Registered StateStoreCoordinator endpoint
2021-03-20 13:09:53 INFO  ConsumerConfig:178 - ConsumerConfig values: 
	metric.reporters = []
	metadata.max.age.ms = 300000
	partition.assignment.strategy = [org.apache.kafka.clients.consumer.RangeAssignor]
	reconnect.backoff.ms = 50
	sasl.kerberos.ticket.renew.window.factor = 0.8
	max.partition.fetch.bytes = 1048576
	bootstrap.servers = [localhost:9092]
	ssl.keystore.type = JKS
	enable.auto.commit = false
	sasl.mechanism = GSSAPI
	interceptor.classes = null
	exclude.internal.topics = true
	ssl.truststore.password = null
	client.id = 
	ssl.endpoint.identification.algorithm = null
	max.poll.records = 1
	check.crcs = true
	request.timeout.ms = 40000
	heartbeat.interval.ms = 3000
	auto.commit.interval.ms = 5000
	receive.buffer.bytes = 65536
	ssl.truststore.type = JKS
	ssl.truststore.location = null
	ssl.keystore.password = null
	fetch.min.bytes = 1
	send.buffer.bytes = 131072
	value.deserializer = class org.apache.kafka.common.serialization.ByteArrayDeserializer
	group.id = spark-kafka-source-9e93536e-9209-4586-b28c-206937832ccd--1486384545-driver-0
	retry.backoff.ms = 100
	sasl.kerberos.kinit.cmd = /usr/bin/kinit
	sasl.kerberos.service.name = null
	sasl.kerberos.ticket.renew.jitter = 0.05
	ssl.trustmanager.algorithm = PKIX
	ssl.key.password = null
	fetch.max.wait.ms = 500
	sasl.kerberos.min.time.before.relogin = 60000
	connections.max.idle.ms = 540000
	session.timeout.ms = 30000
	metrics.num.samples = 2
	key.deserializer = class org.apache.kafka.common.serialization.ByteArrayDeserializer
	ssl.protocol = TLS
	ssl.provider = null
	ssl.enabled.protocols = [TLSv1.2, TLSv1.1, TLSv1]
	ssl.keystore.location = null
```

```
    ssl.cipher.suites = null
    security.protocol = PLAINTEXT
    ssl.keymanager.algorithm = SunX509
    metrics.sample.window.ms = 30000
    auto.offset.reset = earliest

2021-03-20 13:09:53 INFO  ConsumerConfig:178 - ConsumerConfig values:
    metric.reporters = []
    metadata.max.age.ms = 300000
    partition.assignment.strategy = [org.apache.kafka.clients.consumer.RangeAssignor]
    reconnect.backoff.ms = 50
    sasl.kerberos.ticket.renew.window.factor = 0.8
    max.partition.fetch.bytes = 1048576
    bootstrap.servers = [localhost:9092]
    ssl.keystore.type = JKS
    enable.auto.commit = false
    sasl.mechanism = GSSAPI
    interceptor.classes = null
    exclude.internal.topics = true
    ssl.truststore.password = null
    client.id = consumer-1
    ssl.endpoint.identification.algorithm = null
    max.poll.records = 1
    check.crcs = true
    request.timeout.ms = 40000
    heartbeat.interval.ms = 3000
    auto.commit.interval.ms = 5000
    receive.buffer.bytes = 65536
    ssl.truststore.type = JKS
    ssl.truststore.location = null
    ssl.keystore.password = null
    fetch.min.bytes = 1
    send.buffer.bytes = 131072
    value.deserializer = class org.apache.kafka.common.serialization.ByteArrayDeserializer
    group.id = spark-kafka-source-9e93536e-9209-4586-b28c-206937832ccd--1486384545-driver-0
    retry.backoff.ms = 100
    sasl.kerberos.kinit.cmd = /usr/bin/kinit
    sasl.kerberos.service.name = null
    sasl.kerberos.ticket.renew.jitter = 0.05
    ssl.trustmanager.algorithm = PKIX
    ssl.key.password = null
    fetch.max.wait.ms = 500
    sasl.kerberos.min.time.before.relogin = 60000
    connections.max.idle.ms = 540000
    session.timeout.ms = 30000
    metrics.num.samples = 2
    key.deserializer = class org.apache.kafka.common.serialization.ByteArrayDeserializer
    ssl.protocol = TLS
    ssl.provider = null
    ssl.enabled.protocols = [TLSv1.2, TLSv1.1, TLSv1]
    ssl.keystore.location = null
    ssl.cipher.suites = null
    security.protocol = PLAINTEXT
    ssl.keymanager.algorithm = SunX509
```

```
        metrics.sample.window.ms = 30000
        auto.offset.reset = earliest

2021-03-20 13:09:53 INFO  AppInfoParser:83 - Kafka version : 0.10.0.1
2021-03-20 13:09:53 INFO  AppInfoParser:84 - Kafka commitId : a7a17cdec9eaa6c5
df print schema
root
 |-- key: binary (nullable = true)
 |-- value: binary (nullable = true)
 |-- topic: string (nullable = true)
 |-- partition: integer (nullable = true)
 |-- offset: long (nullable = true)
 |-- timestamp: timestamp (nullable = true)
 |-- timestampType: integer (nullable = true)

kafka_df printSchema
root
 |-- value: string (nullable = true)

service_table printSchema
root
 |-- crime_id: string (nullable = true)
 |-- original_crime_type_name: string (nullable = true)
 |-- report_date: timestamp (nullable = true)
 |-- call_date: date (nullable = true)
 |-- offense_date: timestamp (nullable = true)
 |-- call_time: string (nullable = true)
 |-- call_date_time: timestamp (nullable = true)
 |-- disposition: string (nullable = true)
 |-- address: string (nullable = true)
 |-- city: string (nullable = true)
 |-- state: string (nullable = true)
 |-- agency_id: string (nullable = true)
 |-- address_type: string (nullable = true)
 |-- common_location: string (nullable = true)
```

distinct_table printSchema

```
root
 |-- original_crime_type_name: string (nullable = true)
 |-- disposition: string (nullable = true)
```

```
root
 |-- key: binary (nullable = true)
 |-- value: binary (nullable = true)
 |-- topic: string (nullable = true)
 |-- partition: integer (nullable = true)
 |-- offset: long (nullable = true)
 |-- timestamp: timestamp (nullable = true)
 |-- timestampType: integer (nullable = true)
```

agg_df printSchema

```
root
 |-- original_crime_type_name: string (nullable = true)
 |-- count: long (nullable = false)
```

```
2021-03-20 13:09:57 INFO  MicroBatchExecution:54 - Starting dispostiongroupbyorginalcrimetypename [id = dceeb911-200c-43dc-b94c-86a95889ce73, runId = 428ef884-e067-45ae-88bb-ab4fbdb05991]. Use file:///tmp/temporary-489d64ca-0fed-460c-b8c4-bfc96081305f to store the query checkpoint.

print query

<pyspark.sql.streaming.StreamingQuery object at 0x7f5d4423c208>

2021-03-20 13:09:57 INFO  ConsumerConfig:178 - ConsumerConfig values:

	metric.reporters = []

	metadata.max.age.ms = 300000

	partition.assignment.strategy = [org.apache.kafka.clients.consumer.RangeAssignor]

	reconnect.backoff.ms = 50

	sasl.kerberos.ticket.renew.window.factor = 0.8

	max.partition.fetch.bytes = 1048576

	bootstrap.servers = [localhost:9092]

	ssl.keystore.type = JKS

	enable.auto.commit = false

	sasl.mechanism = GSSAPI

	interceptor.classes = null

	exclude.internal.topics = true

	ssl.truststore.password = null

	client.id = 

	ssl.endpoint.identification.algorithm = null

	max.poll.records = 1

	check.crcs = true

	request.timeout.ms = 40000

	heartbeat.interval.ms = 3000

	auto.commit.interval.ms = 5000

	receive.buffer.bytes = 65536
	ssl.truststore.type = JKS
	ssl.truststore.location = null
	ssl.keystore.password = null
	fetch.min.bytes = 1
	send.buffer.bytes = 131072
	value.deserializer = class org.apache.kafka.common.serialization.ByteArrayDeserializer
	group.id = spark-kafka-source-66f913df-cd00-42cb-8100-eb11656cec2f--1762755922-driver-0
	retry.backoff.ms = 100
	sasl.kerberos.kinit.cmd = /usr/bin/kinit
	sasl.kerberos.service.name = null
	sasl.kerberos.ticket.renew.jitter = 0.05
	ssl.trustmanager.algorithm = PKIX
	ssl.key.password = null
	fetch.max.wait.ms = 500
	sasl.kerberos.min.time.before.relogin = 60000
	connections.max.idle.ms = 540000
	session.timeout.ms = 30000
	metrics.num.samples = 2
	key.deserializer = class org.apache.kafka.common.serialization.ByteArrayDeserializer
	ssl.protocol = TLS
	ssl.provider = null
	ssl.enabled.protocols = [TLSv1.2, TLSv1.1, TLSv1]
	ssl.keystore.location = null
	ssl.cipher.suites = null
	security.protocol = PLAINTEXT
	ssl.keymanager.algorithm = SunX509
	metrics.sample.window.ms = 30000
	auto.offset.reset = earliest

2021-03-20 13:09:57 INFO  ConsumerConfig:178 - ConsumerConfig values: 
	metric.reporters = []
	metadata.max.age.ms = 300000
	partition.assignment.strategy = [org.apache.kafka.clients.consumer.RangeAssignor]
	reconnect.backoff.ms = 50
	sasl.kerberos.ticket.renew.window.factor = 0.8
	max.partition.fetch.bytes = 1048576
	bootstrap.servers = [localhost:9092]
	ssl.keystore.type = JKS
	enable.auto.commit = false
	sasl.mechanism = GSSAPI
	interceptor.classes = null
	exclude.internal.topics = true
	ssl.truststore.password = null
	client.id = consumer-2
	ssl.endpoint.identification.algorithm = null
	max.poll.records = 1
	check.crcs = true
	request.timeout.ms = 40000
	heartbeat.interval.ms = 3000
	auto.commit.interval.ms = 5000
	receive.buffer.bytes = 65536
	ssl.truststore.type = JKS
	ssl.truststore.location = null
	ssl.keystore.password = null
```

```
    fetch.min.bytes = 1
    send.buffer.bytes = 131072
    value.deserializer = class org.apache.kafka.common.serialization.ByteArrayDeserializer
    group.id = spark-kafka-source-66f913df-cd00-42cb-8100-eb11656cec2f--1762755922-driver-0
    retry.backoff.ms = 100
    sasl.kerberos.kinit.cmd = /usr/bin/kinit
    sasl.kerberos.service.name = null
    sasl.kerberos.ticket.renew.jitter = 0.05
    ssl.trustmanager.algorithm = PKIX
    ssl.key.password = null
    fetch.max.wait.ms = 500
    sasl.kerberos.min.time.before.relogin = 60000
    connections.max.idle.ms = 540000
    session.timeout.ms = 30000
    metrics.num.samples = 2
    key.deserializer = class org.apache.kafka.common.serialization.ByteArrayDeserializer
    ssl.protocol = TLS
    ssl.provider = null
    ssl.enabled.protocols = [TLSv1.2, TLSv1.1, TLSv1]
    ssl.keystore.location = null
    ssl.cipher.suites = null
    security.protocol = PLAINTEXT
    ssl.keymanager.algorithm = SunX509
    metrics.sample.window.ms = 30000
    auto.offset.reset = earliest
```

2c)Batch Logs

```
2021-03-20 13:17:59 INFO  HDFSBackedStateStoreProvider:54 - Aborted version 1 for
HDFSStateStore[id=(op=0,part=42),dir=file:/tmp/temporary-ad577bd0-a924-43c9-9842-
78a78bfcf647/state/0/42]

2021-03-20 13:17:59 INFO  Executor:54 - Finished task 42.0 in stage 1.0 (TID 200). 5576
bytes result sent to driver

2021-03-20 13:17:59 INFO  TaskSetManager:54 - Finished task 42.0 in stage 1.0 (TID 200) in
141 ms on localhost (executor driver) (200/200)

2021-03-20 13:17:59 INFO  TaskSchedulerImpl:54 - Removed TaskSet 1.0, whose tasks have
all completed, from pool

2021-03-20 13:17:59 INFO  DAGScheduler:54 - ResultStage 1 (start at
NativeMethodAccessorImpl.java:0) finished in 10.177 s

2021-03-20 13:17:59 INFO  DAGScheduler:54 - Job 0 finished: start at
NativeMethodAccessorImpl.java:0, took 12.179597 s

2021-03-20 13:17:59 INFO  WriteToDataSourceV2Exec:54 - Data source writer
org.apache.spark.sql.execution.streaming.sources.MicroBatchWriter@1960032 is
committing.

-------------------------------------------

Batch: 0

-------------------------------------------

2021-03-20 13:17:59 INFO  CodeGenerator:54 - Code generated in 20.500517 ms

2021-03-20 13:17:59 INFO  CodeGenerator:54 - Code generated in 12.002051 ms

+------------------------+-----+

|original_crime_type_name|count|

+------------------------+-----+

|                    null| 2496|

+------------------------+-----+



2021-03-20 13:17:59 INFO  WriteToDataSourceV2Exec:54 - Data source writer
org.apache.spark.sql.execution.streaming.sources.MicroBatchWriter@1960032 committed.

2021-03-20 13:17:59 INFO  SparkContext:54 - Starting job: start at
NativeMethodAccessorImpl.java:0
```

```
2021-03-20 13:17:59 INFO  DAGScheduler:54 - Job 1 finished: start at
NativeMethodAccessorImpl.java:0, took 0.000084 s

2021-03-20 13:17:59 INFO  MicroBatchExecution:54 - Streaming query made progress: {
  "id" : "1dd795af-0f34-4027-a041-d083f40f6fbd",
  "runId" : "6db53fab-948b-4b55-9f64-56d6e66e29e2",
  "name" : "dispostiongroupbyorginalcrimetypename",
  "timestamp" : "2021-03-20T13:17:44.428Z",
  "batchId" : 0,
  "numInputRows" : 2496,
  "processedRowsPerSecond" : 161.98325653838666,
  "durationMs" : {
    "addBatch" : 14329,
    "getBatch" : 290,
    "getOffset" : 325,
    "queryPlanning" : 392,
    "triggerExecution" : 15407,
    "walCommit" : 49
  },
  "stateOperators" : [ {
    "numRowsTotal" : 1,
    "numRowsUpdated" : 1,
    "memoryUsedBytes" : 12863
  } ],
  "sources" : [ {
    "description" : "KafkaSource[Subscribe[sf_police_dept_calls_kafka_server]]",
    "startOffset" : null,
    "endOffset" : {
      "sf_police_dept_calls_kafka_server" : {
```

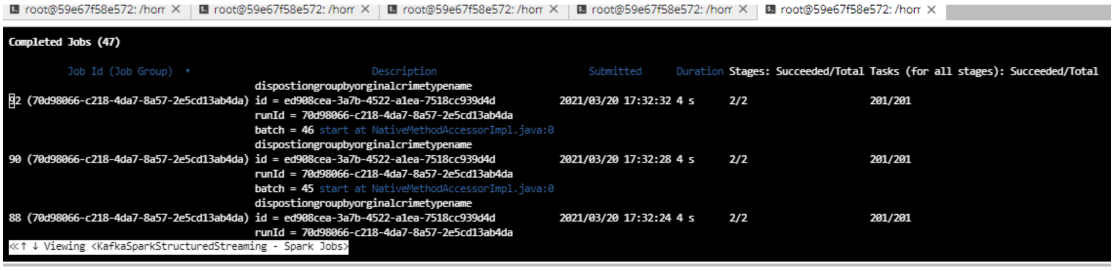
```
      "0" : 2496
    }
  },
  "numInputRows" : 2496,
  "processedRowsPerSecond" : 161.98325653838666
} ],
 "sink" : {
  "description" :
"org.apache.spark.sql.execution.streaming.ConsoleSinkProvider@63b8a109"
 }
}

2021-03-20 13:17:59 INFO  MicroBatchExecution:54 - Committed offsets for batch 1.
Metadata OffsetSeqMetadata(0,1616246279949,Map(spark.sql.shuffle.partitions -> 200,
spark.sql.streaming.stateStore.providerClass ->
org.apache.spark.sql.execution.streaming.state.HDFSBackedStateStoreProvider))

2021-03-20 13:18:00 INFO  KafkaSource:54 - GetBatch called with start =
Some({"sf_police_dept_calls_kafka_server":{"0":2496}}), end =
{"sf_police_dept_calls_kafka_server":{"0":2511}}

2021-03-20 13:18:00 INFO  KafkaSource:54 - Partitions added: Map()

2021-03-20 13:18:00 INFO  KafkaSource:54 - GetBatch generating RDD of offset range:
KafkaSourceRDDOffsetRange(sf_police_dept_calls_kafka_server-0,2496,2511,None)

2021-03-20 13:18:00 INFO  MemoryStore:54 - Block broadcast_4 stored as values in memory
(estimated size 281.0 KB, free 365.4 MB)

2021-03-20 13:18:00 INFO  MemoryStore:54 - Block broadcast_4_piece0 stored as bytes in
memory (estimated size 24.0 KB, free 365.3 MB)

2021-03-20 13:18:00 INFO  BlockManagerInfo:54 - Added broadcast_4_piece0 in memory on
35930afcc42b:38051 (size: 24.0 KB, free: 366.2 MB)

2021-03-20 13:18:00 INFO  SparkContext:54 - Created broadcast 4 from start at
NativeMethodAccessorImpl.java:0

2021-03-20 13:18:00 INFO  MemoryStore:54 - Block broadcast_5 stored as values in memory
(estimated size 281.0 KB, free 365.1 MB)

2021-03-20 13:18:00 INFO  MemoryStore:54 - Block broadcast_5_piece0 stored as bytes in
memory (estimated size 24.0 KB, free 365.0 MB)
```

```
2021-03-20 13:18:00 INFO  BlockManagerInfo:54 - Added broadcast_5_piece0 in memory on 35930afcc42b:38051 (size: 24.0 KB, free: 366.2 MB)

2021-03-20 13:18:00 INFO  SparkContext:54 - Created broadcast 5 from start at NativeMethodAccessorImpl.java:0

2021-03-20 13:18:00 INFO  WriteToDataSourceV2Exec:54 - Start processing data source writer: org.apache.spark.sql.execution.streaming.sources.MicroBatchWriter@64427e8f. The input RDD has 200 partitions.

2021-03-20 13:18:00 INFO  SparkContext:54 - Starting job: start at NativeMethodAccessorImpl.java:0

2021-03-20 13:18:00 INFO  DAGScheduler:54 - Registering RDD 19 (start at NativeMethodAccessorImpl.java:0)

2021-03-20 13:18:00 INFO  DAGScheduler:54 - Got job 2 (start at NativeMethodAccessorImpl.java:0) with 200 output partitions

2021-03-20 13:18:00 INFO  DAGScheduler:54 - Final stage: ResultStage 3 (start at NativeMethodAccessorImpl.java:0)

2021-03-20 13:18:00 INFO  DAGScheduler:54 - Parents of final stage: List(ShuffleMapStage 2)

2021-03-20 13:18:00 INFO  DAGScheduler:54 - Missing parents: List(ShuffleMapStage 2)

2021-03-20 13:18:00 INFO  DAGScheduler:54 - Submitting ShuffleMapStage 2 (MapPartitionsRDD[19] at start at NativeMethodAccessorImpl.java:0), which has no missing parents

2021-03-20 13:18:00 INFO  MemoryStore:54 - Block broadcast_6 stored as values in memory (estimated size 25.4 KB, free 365.0 MB)
```

3a) batch_size = 16384, maxOffsetsPerTrigger = 200, trigger(processingTime= ‘1 second’)

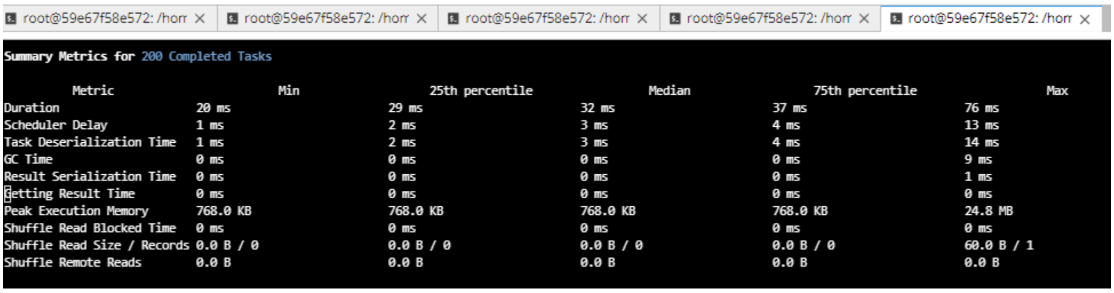
Configuration:

- producer_Server.PY batch_size = not configured
- data_stream.py maxOffsetsPerTrigger = 200
- data_stream.py trigger(processingTime= “1 second”)



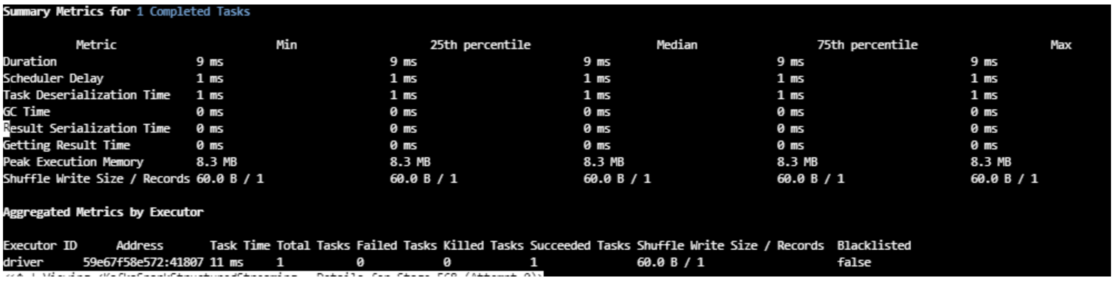
```
Completed Jobs (47)

        Job Id (Job Group)  ▾                         Description                         Submitted      Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                   dispostiongroupbyorginalcrimetypename
92 (70d98066-c218-4da7-8a57-2e5cd13ab4da) id = ed908cea-3a7b-4522-a1ea-7518cc939d4d        2021/03/20 17:32:32 4 s      2/2                 201/201
                                   runId = 70d98066-c218-4da7-8a57-2e5cd13ab4da
                                   batch = 46 start at NativeMethodAccessorImpl.java:0
                                   dispostiongroupbyorginalcrimetypename
90 (70d98066-c218-4da7-8a57-2e5cd13ab4da) id = ed908cea-3a7b-4522-a1ea-7518cc939d4d        2021/03/20 17:32:28 4 s      2/2                 201/201
                                   runId = 70d98066-c218-4da7-8a57-2e5cd13ab4da
                                   batch = 45 start at NativeMethodAccessorImpl.java:0
                                   dispostiongroupbyorginalcrimetypename
88 (70d98066-c218-4da7-8a57-2e5cd13ab4da) id = ed908cea-3a7b-4522-a1ea-7518cc939d4d        2021/03/20 17:32:24 4 s      2/2                 201/201
                                   runId = 70d98066-c218-4da7-8a57-2e5cd13ab4da
<<↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```

One batch had a very high max time



```
Summary Metrics for 200 Completed Tasks

         Metric                  Min                 25th percentile            Median             75th percentile              Max
Duration                20 ms               29 ms               32 ms               37 ms               76 ms
Scheduler Delay         1 ms                2 ms                3 ms                4 ms                13 ms
Task Deserialization Time 1 ms              2 ms                3 ms                4 ms                14 ms
GC Time                 0 ms                0 ms                0 ms                0 ms                9 ms
Result Serialization Time 0 ms              0 ms                0 ms                0 ms                1 ms
Getting Result Time     0 ms                0 ms                0 ms                0 ms                0 ms
Peak Execution Memory   768.0 KB            768.0 KB            768.0 KB            768.0 KB            24.8 MB
Shuffle Read Blocked Time 0 ms              0 ms                0 ms                0 ms                0 ms
Shuffle Read Size / Records 0.0 B / 0       0.0 B / 0           0.0 B / 0           0.0 B / 0           60.0 B / 1
Shuffle Remote Reads    0.0 B               0.0 B               0.0 B               0.0 B               0.0 B
```

Another sample batch was evenly distributed

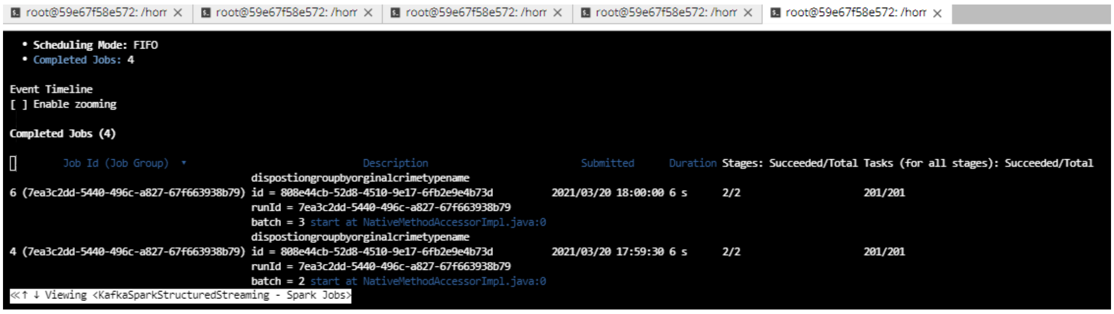


```
Summary Metrics for 1 Completed Tasks

         Metric                  Min                 25th percentile             Median              75th percentile              Max
Duration                9 ms                9 ms                9 ms                9 ms                9 ms
Scheduler Delay         1 ms                1 ms                1 ms                1 ms                1 ms
Task Deserialization Time 1 ms              1 ms                1 ms                1 ms                1 ms
GC Time                 0 ms                0 ms                0 ms                0 ms                0 ms
Result Serialization Time 0 ms              0 ms                0 ms                0 ms                0 ms
Getting Result Time     0 ms                0 ms                0 ms                0 ms                0 ms
Peak Execution Memory   8.3 MB              8.3 MB              8.3 MB              8.3 MB              8.3 MB
Shuffle Write Size / Records 60.0 B / 1     60.0 B / 1          60.0 B / 1          60.0 B / 1          60.0 B / 1

Aggregated Metrics by Executor

Executor ID      Address        Task Time Total Tasks Failed Tasks Killed Tasks Succeeded Tasks Shuffle Write Size / Records  Blacklisted
driver      59e67f58e572:41807 11 ms     1            0            0            1               60.0 B / 1                    false
```

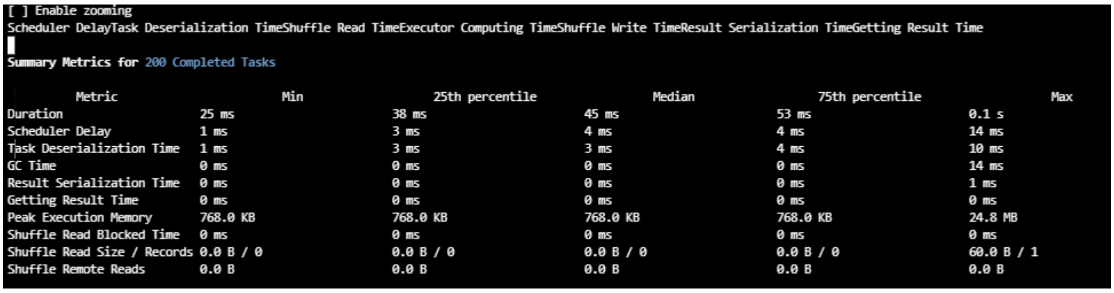
3b) batch_size = 16384, maxOffsetsPerTrigger = default, trigger(processingTime= ‘30 seconds’)

Configuration:

- producer_Server.PY batch_size = not configured
- data_stream.py maxOffsetsPerTrigger = 200
- data_stream.py trigger(processingTime= “30 seconds”)



```
 • Scheduling Mode: FIFO
 • Completed Jobs: 4

Event Timeline
[ ] Enable zooming

Completed Jobs (4)

▯       Job Id (Job Group)  ▾                         Description                      Submitted       Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                        dispostiongroupbyorginalcrimetypename
6 (7ea3c2dd-5440-496c-a827-67f663938b79) id = 808e44cb-52d8-4510-9e17-6fb2e9e4b73d        2021/03/20 18:00:00 6 s     2/2                    201/201
                                        runId = 7ea3c2dd-5440-496c-a827-67f663938b79
                                        batch = 3 start at NativeMethodAccessorImpl.java:0
                                        dispostiongroupbyorginalcrimetypename
4 (7ea3c2dd-5440-496c-a827-67f663938b79) id = 808e44cb-52d8-4510-9e17-6fb2e9e4b73d        2021/03/20 17:59:30 6 s     2/2                    201/201
                                        runId = 7ea3c2dd-5440-496c-a827-67f663938b79
                                        batch = 2 start at NativeMethodAccessorImpl.java:0
«↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```



```
[ ] Enable zooming
Scheduler DelayTask Deserialization TimeShuffle Read TimeExecutor Computing TimeShuffle Write TimeResult Serialization TimeGetting Result Time
█
Summary Metrics for 200 Completed Tasks

          Metric                    Min                     25th percentile              Median                   75th percentile                Max
Duration                  25 ms                  38 ms                  45 ms                  53 ms                  0.1 s
Scheduler Delay           1 ms                   3 ms                   4 ms                   4 ms                   14 ms
Task Deserialization Time 1 ms                   3 ms                   3 ms                   4 ms                   10 ms
GC Time                   0 ms                   0 ms                   0 ms                   0 ms                   14 ms
Result Serialization Time 0 ms                   0 ms                   0 ms                   0 ms                   1 ms
Getting Result Time       0 ms                   0 ms                   0 ms                   0 ms                   0 ms
Peak Execution Memory     768.0 KB               768.0 KB               768.0 KB               768.0 KB               24.8 MB
Shuffle Read Blocked Time 0 ms                   0 ms                   0 ms                   0 ms                   0 ms
Shuffle Read Size / Records 0.0 B / 0            0.0 B / 0              0.0 B / 0              0.0 B / 0              60.0 B / 1
Shuffle Remote Reads      0.0 B                  0.0 B                  0.0 B                  0.0 B                  0.0 B
```

3b Batch Report

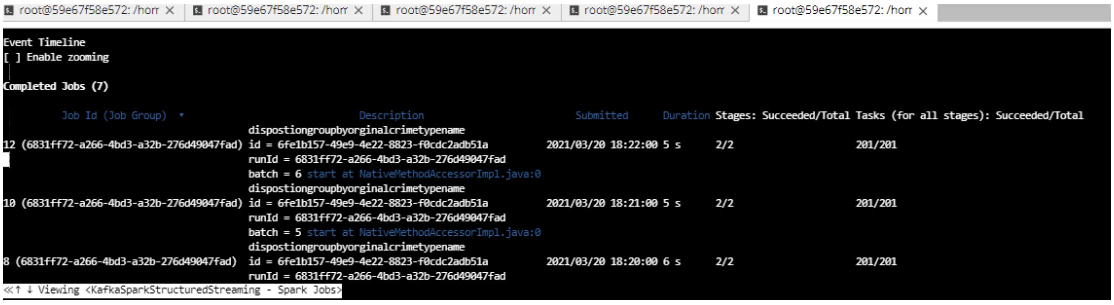
```
-------------------------------------------

Batch: 26

-------------------------------------------

+------------------------+-----+

|original_crime_type_name|count|

+------------------------+-----+

|                    null| 4105|

+------------------------+-----+
```

```
2021-03-20 18:11:34 INFO  WriteToDataSourceV2Exec:54 - Data source writer org.apache.spark.sql.execution.streaming.sources.MicroBatchWriter@3b427d37 committed.

2021-03-20 18:11:34 INFO  SparkContext:54 - Starting job: start at NativeMethodAccessorImpl.java:0

2021-03-20 18:11:34 INFO  DAGScheduler:54 - Job 53 finished: start at NativeMethodAccessorImpl.java:0, took 0.000039 s

2021-03-20 18:11:34 INFO  MicroBatchExecution:54 - Streaming query made progress: {

  "id" : "808e44cb-52d8-4510-9e17-6fb2e9e4b73d",
"runId" : "7ea3c2dd-5440-496c-a827-67f663938b79",
"name" : "dispostiongroupbyorginalcrimetypename",
"timestamp" : "2021-03-20T18:11:30.000Z",
"batchId" : 26,
"numInputRows" : 30,
"inputRowsPerSecond" : 1.0,
"processedRowsPerSecond" : 6.106248727864848,
"durationMs" : {
  "addBatch" : 4876,
  "getBatch" : 4,
  "getOffset" : 3,
  "queryPlanning" : 15,
  "triggerExecution" : 4912,
  "walCommit" : 14
},
"stateOperators" : [ {
  "numRowsTotal" : 1,
  "numRowsUpdated" : 1,
  "memoryUsedBytes" : 17583
} ],
"sources" : [ {
  "description" : "KafkaSource[Subscribe[sf_police_dept_calls_kafka_server]]",
  "startOffset" : {
    "sf_police_dept_calls_kafka_server" : {
      "0" : 4075
    }
  },
    "endOffset" : {
      "sf_police_dept_calls_kafka_server" : {
        "0" : 4105
      }
    },
    "numInputRows" : 30,
    "inputRowsPerSecond" : 1.0,
    "processedRowsPerSecond" : 6.106248727864848
  } ],
  "sink" : {
    "description" :
"org.apache.spark.sql.execution.streaming.ConsoleSinkProvider@11b11ffe"
  }
}
```

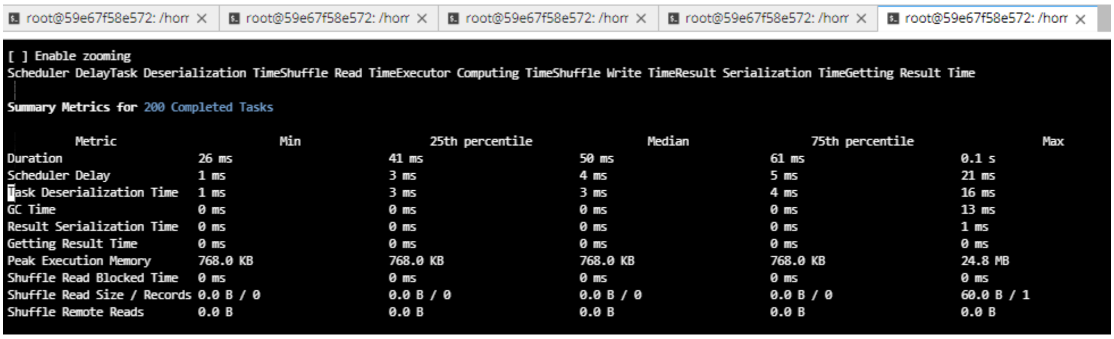
3c) batch_size = 16384, maxOffsetsPerTrigger = default, trigger(processingTime= ‘60 seconds’)

Configuration:

- producer_Server.PY batch_size = not configured
- data_stream.py maxOffsetsPerTrigger = 200
- data_stream.py trigger(processingTime= “60 seconds”)



```
Event Timeline
[ ] Enable zooming

Completed Jobs (7)

        Job Id (Job Group)  ▾                      Description                      Submitted      Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                   dispostiongroupbyorginalcrimetypename
12 (6831ff72-a266-4bd3-a32b-276d49047fad) id = 6fe1b157-49e9-4e22-8823-f0cdc2adb51a      2021/03/20 18:22:00 5 s     2/2                    201/201
                                   runId = 6831ff72-a266-4bd3-a32b-276d49047fad
                                   batch = 6 start at NativeMethodAccessorImpl.java:0
                                   dispostiongroupbyorginalcrimetypename
10 (6831ff72-a266-4bd3-a32b-276d49047fad) id = 6fe1b157-49e9-4e22-8823-f0cdc2adb51a      2021/03/20 18:21:00 5 s     2/2                    201/201
                                   runId = 6831ff72-a266-4bd3-a32b-276d49047fad
                                   batch = 5 start at NativeMethodAccessorImpl.java:0
                                   dispostiongroupbyorginalcrimetypename
8 (6831ff72-a266-4bd3-a32b-276d49047fad)  id = 6fe1b157-49e9-4e22-8823-f0cdc2adb51a      2021/03/20 18:20:00 6 s     2/2                    201/201
                                   runId = 6831ff72-a266-4bd3-a32b-276d49047fad
≪↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```



```
[ ] Enable zooming
Scheduler DelayTask Deserialization TimeShuffle Read TimeExecutor Computing TimeShuffle Write TimeResult Serialization TimeGetting Result Time

Summary Metrics for 200 Completed Tasks

          Metric                    Min                 25th percentile             Median              75th percentile              Max
Duration                 26 ms                 41 ms                 50 ms                 61 ms                 0.1 s
Scheduler Delay          1 ms                  3 ms                  4 ms                  5 ms                  21 ms
Task Deserialization Time 1 ms                 3 ms                  3 ms                  4 ms                  16 ms
GC Time                  0 ms                  0 ms                  0 ms                  0 ms                  13 ms
Result Serialization Time 0 ms                 0 ms                  0 ms                  0 ms                  1 ms
Getting Result Time      0 ms                  0 ms                  0 ms                  0 ms                  0 ms
Peak Execution Memory    768.0 KB              768.0 KB              768.0 KB              768.0 KB              24.8 MB
Shuffle Read Blocked Time 0 ms                 0 ms                  0 ms                  0 ms                  0 ms
Shuffle Read Size / Records 0.0 B / 0          0.0 B / 0             0.0 B / 0             0.0 B / 0             60.0 B / 1
Shuffle Remote Reads     0.0 B                 0.0 B                 0.0 B                 0.0 B                 0.0 B
```

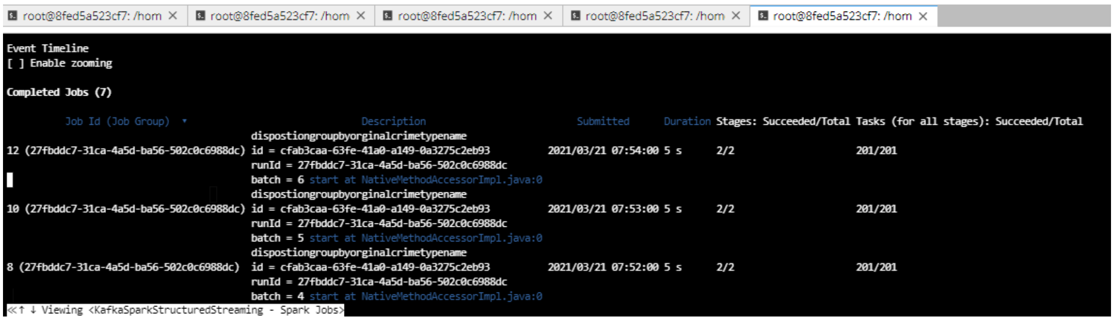
3c Batch Report

```
-------------------------------------------

Batch: 9

-------------------------------------------

+------------------------+-----+

|original_crime_type_name|count|

+------------------------+-----+

|                    null| 2000|

+------------------------+-----+



2021-03-20 18:25:05 INFO  WriteToDataSourceV2Exec:54 - Data source writer org.apache.spark.sql.execution.streaming.sources.MicroBatchWriter@1963c7f2 committed.

2021-03-20 18:25:05 INFO  SparkContext:54 - Starting job: start at NativeMethodAccessorImpl.java:0

2021-03-20 18:25:05 INFO  DAGScheduler:54 - Job 19 finished: start at NativeMethodAccessorImpl.java:0, took 0.000060 s

2021-03-20 18:25:05 INFO  MicroBatchExecution:54 - Streaming query made progress: {

  "id" : "6fe1b157-49e9-4e22-8823-f0cdc2adb51a",

  "runId" : "6831ff72-a266-4bd3-a32b-276d49047fad",

  "name" : "dispostiongroupbyorginalcrimetypename",

  "timestamp" : "2021-03-20T18:25:00.000Z",

  "batchId" : 9,

  "numInputRows" : 200,

  "inputRowsPerSecond" : 3.3333333333333335,
  "processedRowsPerSecond" : 38.18980332251289,
  "durationMs" : {
    "addBatch" : 5187,
    "getBatch" : 6,
    "getOffset" : 3,
    "queryPlanning" : 24,
    "triggerExecution" : 5237,
    "walCommit" : 16
  },
  "stateOperators" : [ {
    "numRowsTotal" : 1,
    "numRowsUpdated" : 1,
    "memoryUsedBytes" : 17583
  } ],
  "sources" : [ {
    "description" : "KafkaSource[Subscribe[sf_police_dept_calls_kafka_server]]",
    "startOffset" : {
      "sf_police_dept_calls_kafka_server" : {
        "0" : 1800
      }
    },
    "endOffset" : {
      "sf_police_dept_calls_kafka_server" : {
        "0" : 2000
      }
    },
    "numInputRows" : 200,
    "inputRowsPerSecond" : 3.3333333333333335,
    "processedRowsPerSecond" : 38.18980332251289
  } ],
  "sink" : {
    "description" : "org.apache.spark.sql.execution.streaming.ConsoleSinkProvider@282c0a93"
  }
}
```

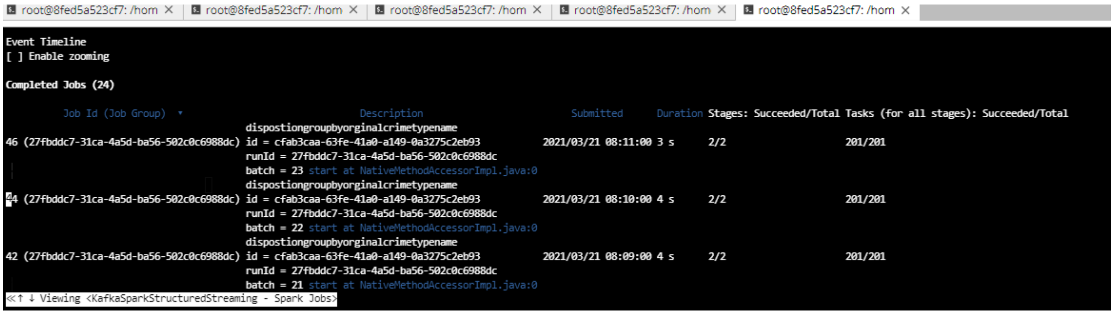
3d) batch_size = 16384, maxOffsetsPerTrigger = 400, trigger(processingTime= ‘60 seconds’)

Configuration:

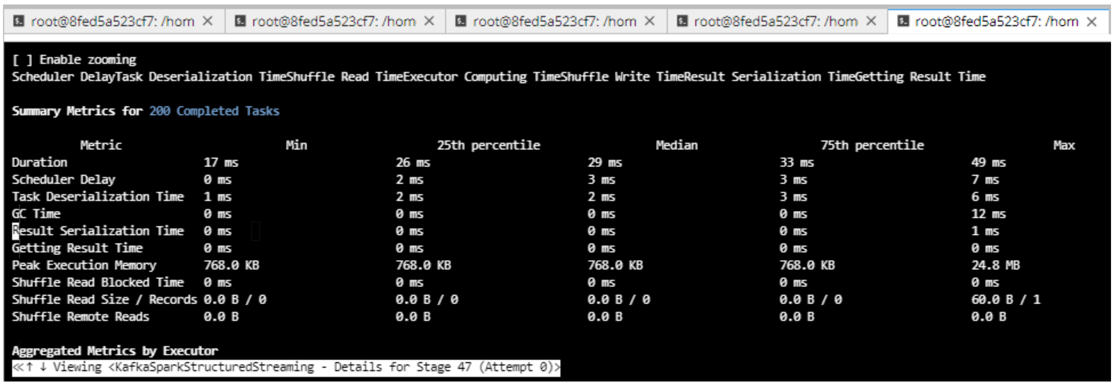
- producer_Server.PY batch_size = not configured
- data_stream.py maxOffsetsPerTrigger = 400
- data_stream.py trigger(processingTime= “60 seconds”)



```
Event Timeline
[ ] Enable zooming

Completed Jobs (7)

        Job Id (Job Group)  ▾                       Description                      Submitted     Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                          dispostiongroupbyorginalcrimetypename
12 (27fbddc7-31ca-4a5d-ba56-502c0c6988dc) id = cfab3caa-63fe-41a0-a149-0a3275c2eb93         2021/03/21 07:54:00 5 s      2/2                    201/201
                                          runId = 27fbddc7-31ca-4a5d-ba56-502c0c6988dc
                                          batch = 6 start at NativeMethodAccessorImpl.java:0
                                          dispostiongroupbyorginalcrimetypename
10 (27fbddc7-31ca-4a5d-ba56-502c0c6988dc) id = cfab3caa-63fe-41a0-a149-0a3275c2eb93         2021/03/21 07:53:00 5 s      2/2                    201/201
                                          runId = 27fbddc7-31ca-4a5d-ba56-502c0c6988dc
                                          batch = 5 start at NativeMethodAccessorImpl.java:0
                                          dispostiongroupbyorginalcrimetypename
8 (27fbddc7-31ca-4a5d-ba56-502c0c6988dc)  id = cfab3caa-63fe-41a0-a149-0a3275c2eb93         2021/03/21 07:52:00 5 s      2/2                    201/201
                                          runId = 27fbddc7-31ca-4a5d-ba56-502c0c6988dc
                                          batch = 4 start at NativeMethodAccessorImpl.java:0
«↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```



```
Event Timeline
[ ] Enable zooming

Completed Jobs (24)

        Job Id (Job Group)  ▾                       Description                      Submitted     Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                          dispostiongroupbyorginalcrimetypename
46 (27fbddc7-31ca-4a5d-ba56-502c0c6988dc) id = cfab3caa-63fe-41a0-a149-0a3275c2eb93         2021/03/21 08:11:00 3 s      2/2                    201/201
                                          runId = 27fbddc7-31ca-4a5d-ba56-502c0c6988dc
                                          batch = 23 start at NativeMethodAccessorImpl.java:0
                                          dispostiongroupbyorginalcrimetypename
44 (27fbddc7-31ca-4a5d-ba56-502c0c6988dc) id = cfab3caa-63fe-41a0-a149-0a3275c2eb93         2021/03/21 08:10:00 4 s      2/2                    201/201
                                          runId = 27fbddc7-31ca-4a5d-ba56-502c0c6988dc
                                          batch = 22 start at NativeMethodAccessorImpl.java:0
                                          dispostiongroupbyorginalcrimetypename
42 (27fbddc7-31ca-4a5d-ba56-502c0c6988dc) id = cfab3caa-63fe-41a0-a149-0a3275c2eb93         2021/03/21 08:09:00 4 s      2/2                    201/201
                                          runId = 27fbddc7-31ca-4a5d-ba56-502c0c6988dc
                                          batch = 21 start at NativeMethodAccessorImpl.java:0
«↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```



```
[ ] Enable zooming
Scheduler DelayTask Deserialization TimeShuffle Read TimeExecutor Computing TimeShuffle Write TimeResult Serialization TimeGetting Result Time

Summary Metrics for 200 Completed Tasks

         Metric                   Min                   25th percentile            Median                75th percentile             Max
Duration                  17 ms                 26 ms                 29 ms                 33 ms                 49 ms
Scheduler Delay           0 ms                  2 ms                  3 ms                  3 ms                  7 ms
Task Deserialization Time 1 ms                  2 ms                  2 ms                  3 ms                  6 ms
GC Time                   0 ms                  0 ms                  0 ms                  0 ms                  12 ms
Result Serialization Time 0 ms                  0 ms                  0 ms                  0 ms                  1 ms
Getting Result Time       0 ms                  0 ms                  0 ms                  0 ms                  0 ms
Peak Execution Memory     768.0 KB              768.0 KB              768.0 KB              768.0 KB              24.8 MB
Shuffle Read Blocked Time 0 ms                  0 ms                  0 ms                  0 ms                  0 ms
Shuffle Read Size / Records 0.0 B / 0           0.0 B / 0             0.0 B / 0             0.0 B / 0             60.0 B / 1
Shuffle Remote Reads      0.0 B                 0.0 B                 0.0 B                 0.0 B                 0.0 B

Aggregated Metrics by Executor
<↑ ↓ Viewing <KafkaSparkStructuredStreaming - Details for Stage 47 (Attempt 0)>
```

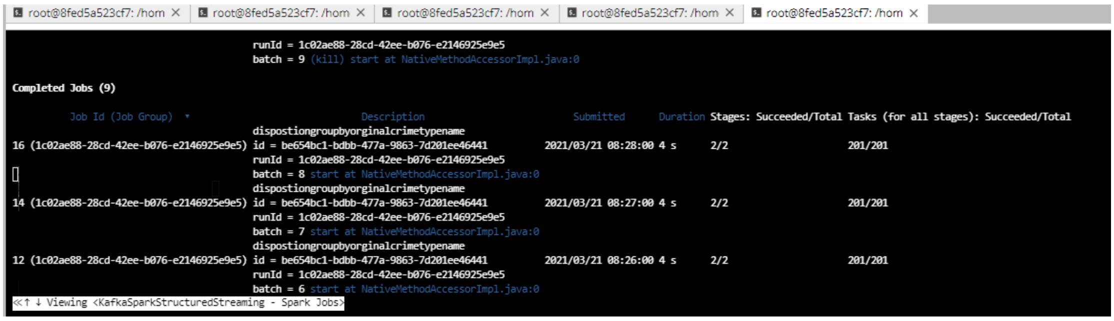
3d Batch report

```
-------------------------------------------

Batch: 27

-------------------------------------------

+-----------------------+-----+

|original_crime_type_name|count|

+-----------------------+-----+

|                   null| 2166|

+-----------------------+-----+



2021-03-21 08:15:04 INFO  WriteToDataSourceV2Exec:54 - Data source writer org.apache.spark.sql.execution.streaming.sources.MicroBatchWriter@67b16ef4 committed.

2021-03-21 08:15:04 INFO  SparkContext:54 - Starting job: start at NativeMethodAccessorImpl.java:0

2021-03-21 08:15:04 INFO  DAGScheduler:54 - Job 55 finished: start at NativeMethodAccessorImpl.java:0, took 0.000050 s

2021-03-21 08:15:04 INFO  MicroBatchExecution:54 - Streaming query made progress: {

  "id" : "cfab3caa-63fe-41a0-a149-0a3275c2eb93",

  "runId" : "27fbddc7-31ca-4a5d-ba56-502c0c6988dc",

  "name" : "dispostiongroupbyorginalcrimetypename",

  "timestamp" : "2021-03-21T08:15:00.000Z",

  "batchId" : 27,
"numInputRows" : 60,
"inputRowsPerSecond" : 1.0,
"processedRowsPerSecond" : 12.523481527864746,
"durationMs" : {
  "addBatch" : 4744,
  "getBatch" : 7,
  "getOffset" : 8,
  "queryPlanning" : 15,
  "triggerExecution" : 4791,
  "walCommit" : 16
},
"stateOperators" : [ {
  "numRowsTotal" : 1,
  "numRowsUpdated" : 1,
  "memoryUsedBytes" : 17583
} ],
"sources" : [ {
  "description" : "KafkaSource[Subscribe[sf_police_dept_calls_kafka_server]]",
  "startOffset" : {
    "sf_police_dept_calls_kafka_server" : {
      "0" : 2106
    }
  },
  "endOffset" : {
    "sf_police_dept_calls_kafka_server" : {
      "0" : 2166
    }
    },
    "numInputRows" : 60,
    "inputRowsPerSecond" : 1.0,
    "processedRowsPerSecond" : 12.523481527864746
  } ],
  "sink" : {
    "description" : "org.apache.spark.sql.execution.streaming.ConsoleSinkProvider@f65c223"
  }
}
```

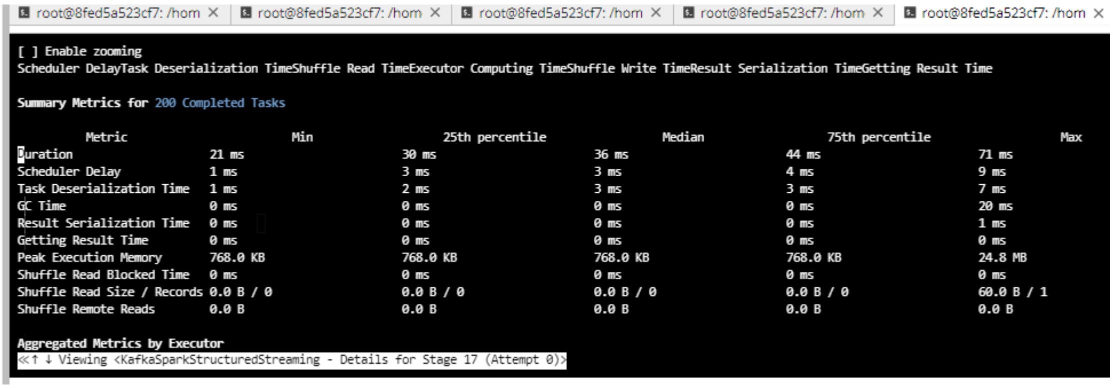
3e) batch_size = 16384, maxOffsetsPerTrigger = 1000, trigger(processingTime= ‘60 seconds’)

Configuration:

- producer_Server.PY batch_size = not configured
- data_stream.py maxOffsetsPerTrigger = 1000
- data_stream.py trigger(processingTime= “60 seconds”)



```
                                        runId = 1c02ae88-28cd-42ee-b076-e2146925e9e5
                                        batch = 9 (kill) start at NativeMethodAccessorImpl.java:0

Completed Jobs (9)

         Job Id (Job Group)  ▾                      Description                  Submitted      Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                        dispostiongroupbyorginalcrimetypename
16 (1c02ae88-28cd-42ee-b076-e2146925e9e5) id = be654bc1-bdbb-477a-9863-7d201ee46441       2021/03/21 08:28:00 4 s      2/2                    201/201
                                        runId = 1c02ae88-28cd-42ee-b076-e2146925e9e5
                                        batch = 8 start at NativeMethodAccessorImpl.java:0
                                        dispostiongroupbyorginalcrimetypename
14 (1c02ae88-28cd-42ee-b076-e2146925e9e5) id = be654bc1-bdbb-477a-9863-7d201ee46441       2021/03/21 08:27:00 4 s      2/2                    201/201
                                        runId = 1c02ae88-28cd-42ee-b076-e2146925e9e5
                                        batch = 7 start at NativeMethodAccessorImpl.java:0
                                        dispostiongroupbyorginalcrimetypename
12 (1c02ae88-28cd-42ee-b076-e2146925e9e5) id = be654bc1-bdbb-477a-9863-7d201ee46441       2021/03/21 08:26:00 4 s      2/2                    201/201
                                        runId = 1c02ae88-28cd-42ee-b076-e2146925e9e5
                                        batch = 6 start at NativeMethodAccessorImpl.java:0
<<↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```



```
[ ] Enable zooming
Scheduler DelayTask Deserialization TimeShuffle Read TimeExecutor Computing TimeShuffle Write TimeResult Serialization TimeGetting Result Time

Summary Metrics for 200 Completed Tasks

        Metric                  Min                   25th percentile              Median                 75th percentile              Max
Duration               21 ms                    30 ms                    36 ms                    44 ms                    71 ms
Scheduler Delay        1 ms                     3 ms                     3 ms                     4 ms                     9 ms
Task Deserialization Time  1 ms                 2 ms                     3 ms                     3 ms                     7 ms
GC Time                0 ms                     0 ms                     0 ms                     0 ms                     20 ms
Result Serialization Time  0 ms                 0 ms                     0 ms                     0 ms                     1 ms
Getting Result Time    0 ms                     0 ms                     0 ms                     0 ms                     0 ms
Peak Execution Memory  768.0 KB                 768.0 KB                 768.0 KB                 768.0 KB                 24.8 MB
Shuffle Read Blocked Time  0 ms                 0 ms                     0 ms                     0 ms                     0 ms
Shuffle Read Size / Records 0.0 B / 0           0.0 B / 0                0.0 B / 0                0.0 B / 0                60.0 B / 1
Shuffle Remote Reads   0.0 B                    0.0 B                    0.0 B                    0.0 B                    0.0 B

Aggregated Metrics by Executor
<<↑ ↓ Viewing <KafkaSparkStructuredStreaming - Details for Stage 17 (Attempt 0)>
```

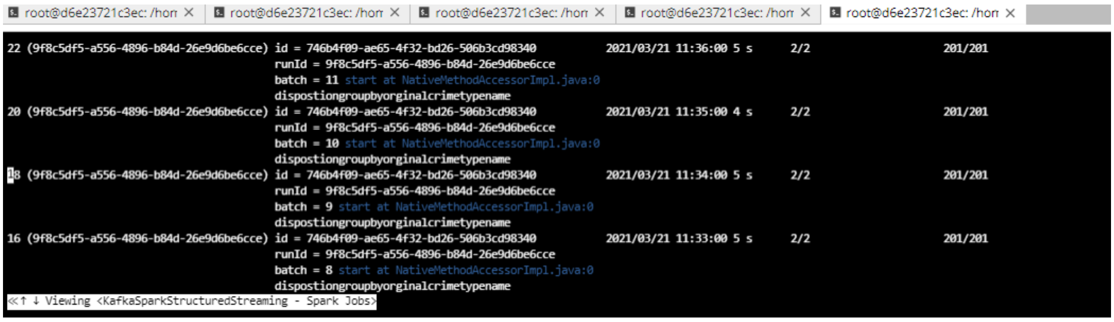
3e Batch report

```
Batch: 13
-------------------------------------------
+------------------------+-----+
|original_crime_type_name|count|
+------------------------+-----+
|                    null| 3244|
+------------------------+-----+


2021-03-21 08:33:05 INFO  WriteToDataSourceV2Exec:54 - Data source writer org.apache.spark.sql.execution.streaming.sources.MicroBatchWriter@6e87402b committed.
2021-03-21 08:33:05 INFO  SparkContext:54 - Starting job: start at NativeMethodAccessorImpl.java:0
2021-03-21 08:33:05 INFO  DAGScheduler:54 - Job 27 finished: start at NativeMethodAccessorImpl.java:0, took 0.000608 s
2021-03-21 08:33:05 INFO  MicroBatchExecution:54 - Streaming query made progress: {
  "id" : "be654bc1-bdbb-477a-9863-7d201ee46441",
  "runId" : "1c02ae88-28cd-42ee-b076-e2146925e9e5",
  "name" : "dispostiongroupbyorginalcrimetypename",
  "timestamp" : "2021-03-21T08:33:00.000Z",
  "batchId" : 13,
  "numInputRows" : 60,
  "inputRowsPerSecond" : 1.0,
  "processedRowsPerSecond" : 11.342155009451796,
  "durationMs" : {
    "addBatch" : 5242,
    "getBatch" : 4,
  "getOffset" : 7,
  "queryPlanning" : 15,
  "triggerExecution" : 5290,
  "walCommit" : 21
},
"stateOperators" : [ {
  "numRowsTotal" : 1,
  "numRowsUpdated" : 1,
  "memoryUsedBytes" : 17583
} ],
"sources" : [ {
  "description" : "KafkaSource[Subscribe[sf_police_dept_calls_kafka_server]]",
  "startOffset" : {
    "sf_police_dept_calls_kafka_server" : {
      "0" : 3184
    }
  },
  "endOffset" : {
    "sf_police_dept_calls_kafka_server" : {
      "0" : 3244
    }
  },
  "numInputRows" : 60,
  "inputRowsPerSecond" : 1.0,
  "processedRowsPerSecond" : 11.342155009451796
} ],
"sink" : {
    "description" :
"org.apache.spark.sql.execution.streaming.ConsoleSinkProvider@5b281be"

  }

}

^CTraceback (most recent call last):

  File "/home/workspace/data_stream.py", line 163, in <module>

    run_spark_job(spark)
```

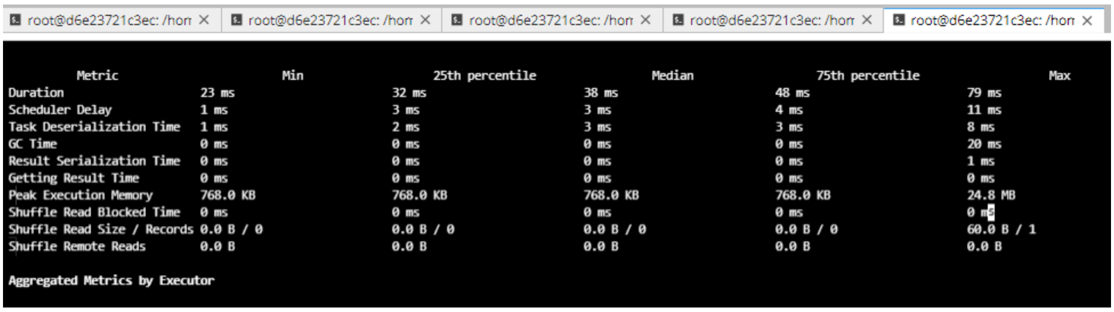
3f) batch_size = 35000, maxOffsetsPerTrigger = 1000, trigger(processingTime= ‘60 seconds’)

Configuration:

- producer_Server.PY batch_size = 35000
- data_stream.py maxOffsetsPerTrigger = 1000
- data_stream.py trigger(processingTime= “60 seconds”)


root@d6e23721c3ec: /hom ×
root@d6e23721c3ec: /hom ×
root@d6e23721c3ec: /hom ×
root@d6e23721c3ec: /hom ×
root@d6e23721c3ec: /hom ×

```
22 (9f8c5df5-a556-4896-b84d-26e9d6be6cce) id = 746b4f09-ae65-4f32-bd26-506b3cd98340        2021/03/21 11:36:00 5 s     2/2                 201/201
                                          runId = 9f8c5df5-a556-4896-b84d-26e9d6be6cce
                                          batch = 11 start at NativeMethodAccessorImpl.java:0
                                          dispostiongroupbyorginalcrimetypename
20 (9f8c5df5-a556-4896-b84d-26e9d6be6cce) id = 746b4f09-ae65-4f32-bd26-506b3cd98340        2021/03/21 11:35:00 4 s     2/2                 201/201
                                          runId = 9f8c5df5-a556-4896-b84d-26e9d6be6cce
                                          batch = 10 start at NativeMethodAccessorImpl.java:0
                                          dispostiongroupbyorginalcrimetypename
18 (9f8c5df5-a556-4896-b84d-26e9d6be6cce) id = 746b4f09-ae65-4f32-bd26-506b3cd98340        2021/03/21 11:34:00 5 s     2/2                 201/201
                                          runId = 9f8c5df5-a556-4896-b84d-26e9d6be6cce
                                          batch = 9 start at NativeMethodAccessorImpl.java:0
                                          dispostiongroupbyorginalcrimetypename
16 (9f8c5df5-a556-4896-b84d-26e9d6be6cce) id = 746b4f09-ae65-4f32-bd26-506b3cd98340        2021/03/21 11:33:00 5 s     2/2                 201/201
                                          runId = 9f8c5df5-a556-4896-b84d-26e9d6be6cce
                                          batch = 8 start at NativeMethodAccessorImpl.java:0
                                          dispostiongroupbyorginalcrimetypename
«↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```


root@d6e23721c3ec: /hom ×
root@d6e23721c3ec: /hom ×
root@d6e23721c3ec: /hom ×
root@d6e23721c3ec: /hom ×
root@d6e23721c3ec: /hom ×

```
        Metric                     Min                25th percentile            Median             75th percentile            Max
Duration                    23 ms                 32 ms                 38 ms                 48 ms                 79 ms
Scheduler Delay             1 ms                  3 ms                  3 ms                  4 ms                  11 ms
Task Deserialization Time   1 ms                  2 ms                  3 ms                  3 ms                  8 ms
GC Time                     0 ms                  0 ms                  0 ms                  0 ms                  20 ms
Result Serialization Time   0 ms                  0 ms                  0 ms                  0 ms                  1 ms
Getting Result Time         0 ms                  0 ms                  0 ms                  0 ms                  0 ms
Peak Execution Memory       768.0 KB              768.0 KB              768.0 KB              768.0 KB              24.8 MB
Shuffle Read Blocked Time   0 ms                  0 ms                  0 ms                  0 ms                  0 ms
Shuffle Read Size / Records 0.0 B / 0             0.0 B / 0             0.0 B / 0             0.0 B / 0             60.0 B / 1
Shuffle Remote Reads        0.0 B                 0.0 B                 0.0 B                 0.0 B                 0.0 B

Aggregated Metrics by Executor
```

3f Batch report

Batch: 18

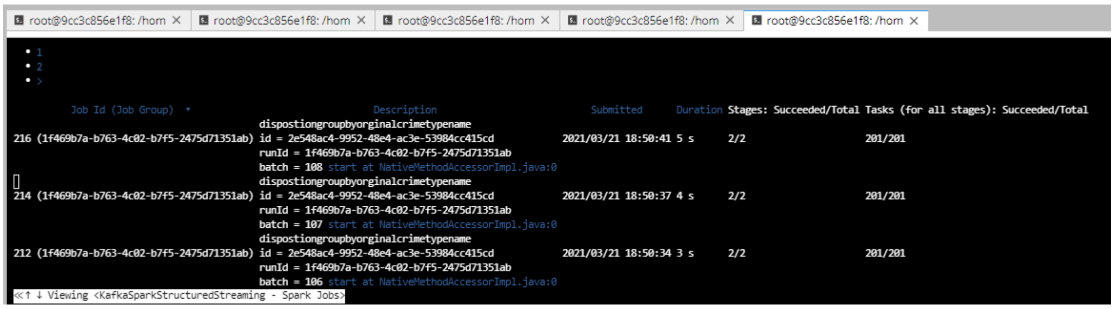
```
+------------------------+-----+
|original_crime_type_name|count|
+------------------------+-----+
|                    null| 1520|
+------------------------+-----+


2021-03-21 11:43:04 INFO  WriteToDataSourceV2Exec:54 - Data source writer org.apache.spark.sql.execution.streaming.sources.MicroBatchWriter@59a6e75d committed.
2021-03-21 11:43:04 INFO  SparkContext:54 - Starting job: start at NativeMethodAccessorImpl.java:0
2021-03-21 11:43:04 INFO  DAGScheduler:54 - Job 37 finished: start at NativeMethodAccessorImpl.java:0, took 0.000069 s
2021-03-21 11:43:04 INFO  MicroBatchExecution:54 - Streaming query made progress: {
  "id" : "746b4f09-ae65-4f32-bd26-506b3cd98340",
  "runId" : "9f8c5df5-a556-4896-b84d-26e9d6be6cce",
  "name" : "dispostiongroupbyorginalcrimetypename",
  "timestamp" : "2021-03-21T11:43:00.001Z",
  "batchId" : 18,
  "numInputRows" : 60,
  "inputRowsPerSecond" : 0.9999833336111065,
  "processedRowsPerSecond" : 13.917884481558803,
  "durationMs" : {
    "addBatch" : 4268,
    "getBatch" : 4,
    "getOffset" : 3,
    "queryPlanning" : 20,
    "triggerExecution" : 4311,
    "walCommit" : 15
  },
  "stateOperators" : [ {
    "numRowsTotal" : 1,
    "numRowsUpdated" : 1,
    "memoryUsedBytes" : 17583
  } ],
  "sources" : [ {
    "description" : "KafkaSource[Subscribe[sf_police_dept_calls_kafka_server]]",
    "startOffset" : {
      "sf_police_dept_calls_kafka_server" : {
        "0" : 1460
      }
    },
    "endOffset" : {
      "sf_police_dept_calls_kafka_server" : {
        "0" : 1520
      }
    },
    "numInputRows" : 60,
    "inputRowsPerSecond" : 0.9999833336111065,
    "processedRowsPerSecond" : 13.917884481558803
  } ],
  "sink" : {
    "description" :
"org.apache.spark.sql.execution.streaming.ConsoleSinkProvider@255998d3"
  }
}
^CTraceback (most recent call last):
```

3g) batch_size = default, maxOffsetsPerTrigger = default, trigger(processingTime= default)

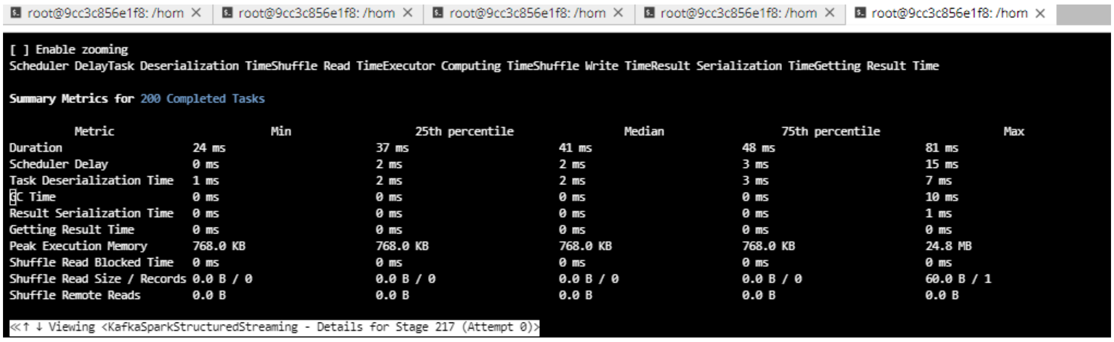
Configuration:

- producer_Server.PY batch_size = no configuration
- data_stream.py maxOffsetsPerTrigger = no configuration
- data_stream.py trigger(processingTime= no configuration

First sample :

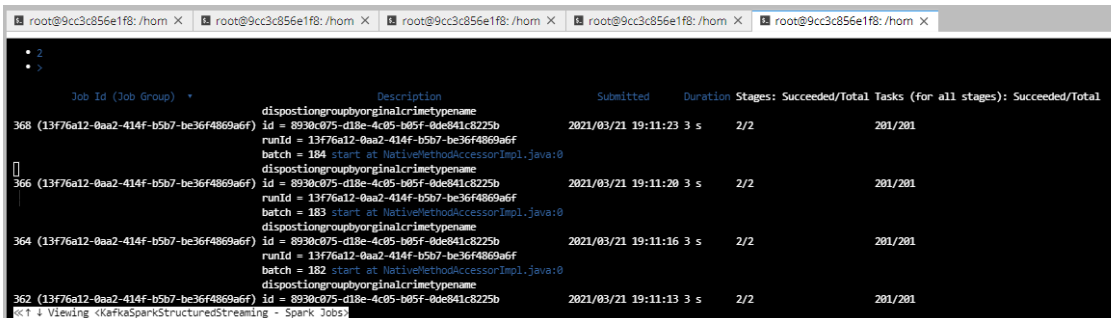


```
  • 1
  • 2
  • >

       Job Id (Job Group)  ▾                         Description                       Submitted      Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                               dispostiongroupbyorginalcrimetypename
216 (1f469b7a-b763-4c02-b7f5-2475d71351ab) id = 2e548ac4-9952-48e4-ac3e-53984cc415cd         2021/03/21 18:50:41 5 s     2/2                     201/201
                                               runId = 1f469b7a-b763-4c02-b7f5-2475d71351ab
                                               batch = 108 start at NativeMethodAccessorImpl.java:0
▯                                              dispostiongroupbyorginalcrimetypename
214 (1f469b7a-b763-4c02-b7f5-2475d71351ab) id = 2e548ac4-9952-48e4-ac3e-53984cc415cd         2021/03/21 18:50:37 4 s     2/2                     201/201
                                               runId = 1f469b7a-b763-4c02-b7f5-2475d71351ab
                                               batch = 107 start at NativeMethodAccessorImpl.java:0
                                               dispostiongroupbyorginalcrimetypename
212 (1f469b7a-b763-4c02-b7f5-2475d71351ab) id = 2e548ac4-9952-48e4-ac3e-53984cc415cd         2021/03/21 18:50:34 3 s     2/2                     201/201
                                               runId = 1f469b7a-b763-4c02-b7f5-2475d71351ab
                                               batch = 106 start at NativeMethodAccessorImpl.java:0
«↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```

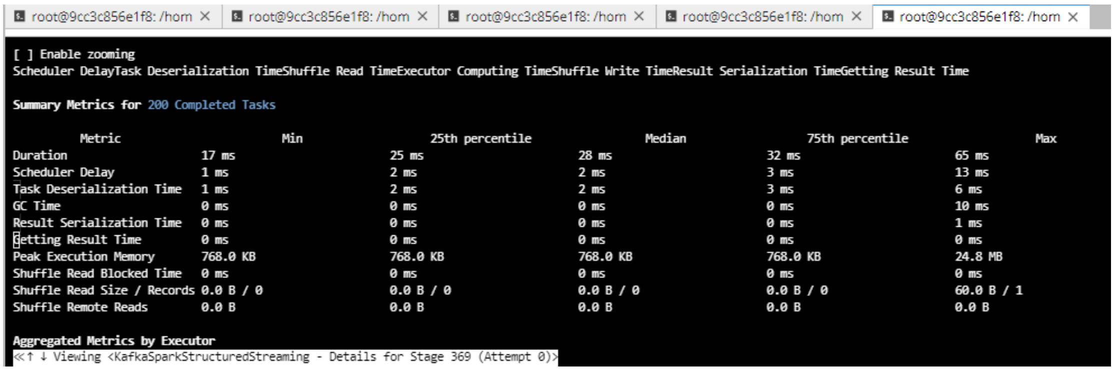


```
[ ] Enable zooming
Scheduler DelayTask Deserialization TimeShuffle Read TimeExecutor Computing TimeShuffle Write TimeResult Serialization TimeGetting Result Time

Summary Metrics for 200 Completed Tasks

        Metric                    Min                 25th percentile               Median               75th percentile                 Max
Duration                   24 ms                 37 ms                 41 ms                 48 ms                 81 ms
Scheduler Delay            0 ms                  2 ms                  2 ms                  3 ms                  15 ms
Task Deserialization Time  1 ms                  2 ms                  2 ms                  3 ms                  7 ms
GC Time                    0 ms                  0 ms                  0 ms                  0 ms                  10 ms
Result Serialization Time  0 ms                  0 ms                  0 ms                  0 ms                  1 ms
Getting Result Time        0 ms                  0 ms                  0 ms                  0 ms                  0 ms
Peak Execution Memory      768.0 KB              768.0 KB              768.0 KB              768.0 KB              24.8 MB
Shuffle Read Blocked Time  0 ms                  0 ms                  0 ms                  0 ms                  0 ms
Shuffle Read Size / Records 0.0 B / 0            0.0 B / 0             0.0 B / 0             0.0 B / 0             60.0 B / 1
Shuffle Remote Reads       0.0 B                 0.0 B                 0.0 B                 0.0 B                 0.0 B

«↑ ↓ Viewing <KafkaSparkStructuredStreaming - Details for Stage 217 (Attempt 0)>
```

Second Sample had faster times:



```
  • 2
  • >

       Job Id (Job Group)  ▾                         Description                       Submitted      Duration Stages: Succeeded/Total Tasks (for all stages): Succeeded/Total
                                               dispostiongroupbyorginalcrimetypename
368 (13f76a12-0aa2-414f-b5b7-be36f4869a6f) id = 8930c075-d18e-4c05-b05f-0de841c8225b         2021/03/21 19:11:23 3 s     2/2                     201/201
                                               runId = 13f76a12-0aa2-414f-b5b7-be36f4869a6f
                                               batch = 184 start at NativeMethodAccessorImpl.java:0
▯                                              dispostiongroupbyorginalcrimetypename
366 (13f76a12-0aa2-414f-b5b7-be36f4869a6f) id = 8930c075-d18e-4c05-b05f-0de841c8225b         2021/03/21 19:11:20 3 s     2/2                     201/201
                                               runId = 13f76a12-0aa2-414f-b5b7-be36f4869a6f
                                               batch = 183 start at NativeMethodAccessorImpl.java:0
                                               dispostiongroupbyorginalcrimetypename
364 (13f76a12-0aa2-414f-b5b7-be36f4869a6f) id = 8930c075-d18e-4c05-b05f-0de841c8225b         2021/03/21 19:11:16 3 s     2/2                     201/201
                                               runId = 13f76a12-0aa2-414f-b5b7-be36f4869a6f
                                               batch = 182 start at NativeMethodAccessorImpl.java:0
                                               dispostiongroupbyorginalcrimetypename
362 (13f76a12-0aa2-414f-b5b7-be36f4869a6f) id = 8930c075-d18e-4c05-b05f-0de841c8225b         2021/03/21 19:11:13 3 s     2/2                     201/201
«↑ ↓ Viewing <KafkaSparkStructuredStreaming - Spark Jobs>
```



```
[ ] Enable zooming
Scheduler DelayTask Deserialization TimeShuffle Read TimeExecutor Computing TimeShuffle Write TimeResult Serialization TimeGetting Result Time

Summary Metrics for 200 Completed Tasks

          Metric                        Min                   25th percentile                 Median                  75th percentile                   Max
Duration                    17 ms                       25 ms                       28 ms                       32 ms                       65 ms
Scheduler Delay             1 ms                        2 ms                        2 ms                        3 ms                        13 ms
Task Deserialization Time   1 ms                        2 ms                        2 ms                        3 ms                        6 ms
GC Time                     0 ms                        0 ms                        0 ms                        0 ms                        10 ms
Result Serialization Time   0 ms                        0 ms                        0 ms                        0 ms                        1 ms
Getting Result Time         0 ms                        0 ms                        0 ms                        0 ms                        0 ms
Peak Execution Memory       768.0 KB                    768.0 KB                    768.0 KB                    768.0 KB                    24.8 MB
Shuffle Read Blocked Time   0 ms                        0 ms                        0 ms                        0 ms                        0 ms
Shuffle Read Size / Records 0.0 B / 0                   0.0 B / 0                   0.0 B / 0                   0.0 B / 0                   60.0 B / 1
Shuffle Remote Reads        0.0 B                       0.0 B                       0.0 B                       0.0 B                       0.0 B

Aggregated Metrics by Executor
<<↑ ↓ Viewing <KafkaSparkStructuredStreaming - Details for Stage 369 (Attempt 0)>
```