

Activity 3 - Monitoring user experience in near real-time

Technical Specification: Event Aggregation Logic

Scenario: Continuous reporting of critical crash events

Given a stream of incoming event logs
When a log entry has a `content` field containing the substring "crash"
And the `severity` value is either "High" or "Critical"
And logs are grouped by the `user_id` field, such as the crash count per `user_id` **And** the system aggregates these occurrences in **10-second intervals** based strictly on the event `timestamp` field
Then the system must output the aggregated results for each interval as they complete **When** the crash count of a given `user_id` is higher than 2 per interval.

Implementation Notes:

- Ensure the search for "crash" handles case sensitivity according to project standards.
- The 10-second interval logic must be tied to the record's metadata (`timestamp`), not the system arrival time.

Example of output

Batch: 12

Interval	user_id	crash_count
{2026-01-11 14:42:50, 2026-01-11 14:43:00}	user_1836	5
{2026-01-11 14:42:50, 2026-01-11 14:43:00}	user_1184	3
{2026-01-11 14:42:50, 2026-01-11 14:43:00}	user_1946	3
{2026-01-11 14:42:50, 2026-01-11 14:43:00}	user_1551	3
{2026-01-11 14:42:50, 2026-01-11 14:43:00}	user_1841	3
{2026-01-11 14:42:50, 2026-01-11 14:43:00}	user_1287	3
{2026-01-11 14:42:50, 2026-01-11 14:43:00}	user_1028	3
{2026-01-11 14:42:50, 2026-01-11 14:43:00}	user_1288	3

2. Non-Functional Requirements

- **Scalability:** The architecture must support horizontal scaling, allowing the logic to be distributed across a cluster of multiple machines.
- **Fault Tolerance:** The system must support recovery in the event of infrastructure failure of the worker nodes.

Deliverables for Activity 3

You are required to submit your application source code accompanied by a technical discussion. This discussion must explain how your specific implementation satisfies the requirements, including a discussion on your solution could handle the scenario of late-arriving records that can appear after a 10-second interval has concluded. Furthermore, you must provide a performance and scalability report that evaluates the performance and efficiency of your solution and discuss its ability to execute effectively across a multi-machine environment. Submit via Moodle until 27.01.2026

Delete the topic where the log records were produced:

```
docker exec -it kafka kafka-topics.sh \
  --bootstrap-server localhost:9092 \
  --delete \
  --topic logs
```

Step 1: Prepare Kafka Topic

```
docker exec -it kafka kafka-topics.sh \
  --bootstrap-server localhost:9092 \
  --create \
  --topic logs \
  --partitions 4 \
  --replication-factor 1
```

Step 2: Ensure load generator is started

```
cd load-generator
docker compose up -d
```

Step 3: Run the Spark Application

Configuration Used:

Parameter	Value	Purpose
num-executors	1	Single-node deployment
executor-cores	8	Parallel task execution
executor-memory	4G	State storage + processing
Kafka partitions	4	Parallel ingestion
shuffle.partitions	16 (default)	2× executor cores

Key Parameters for the Python Script:

Parameter	Default	Purpose
WINDOW_DURATION	10 seconds	Tumbling window size
WATERMARK_DELAY	30 seconds	Late arrival tolerance
startingOffsets	latest	Kafka read position
outputMode	update	Stream output mode

Run the script inside the spark client (connect to container via VS Code):

```
spark-submit \
  --master spark://spark-master:7077 \
  --packages org.apache.spark:spark-sql-kafka-0-10_2.13:4.0.0 \
  --num-executors 1 \
  --executor-cores 8 \
  --executor-memory 4G \
  /opt/spark-apps/spark_crash_monitoring.py
```

Step 4: Monitor the Application

Spark Web UI

- **Driver:** http://localhost:4040

Key Tabs to Monitor:

1. **Structured Streaming** - Batch processing times, input/output rates
2. **SQL** - Query DAG and execution plan
3. **Executors** - Resource utilization

Step 5: Interpret the Output

Output:

```
-----
Batch: 0
-----
+-----+-----+-----+
|Interval|user_id|crash_count|
+-----+-----+-----+
+-----+-----+-----+

-----
Batch: 1
-----
+-----+-----+-----+
```

Interval	user_id	crash_count
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1361	3
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1424	6
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1631	6
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1987	4
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1411	3
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1281	3
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1497	3
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1568	3
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1588	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1886	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1356	6
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1083	5
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1086	5
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1696	9
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1412	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1526	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1038	3
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1728	3
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1690	3
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1282	3
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1840	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1889	4
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1458	6
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1266	8
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1404	5
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1353	3
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1562	3
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1891	3
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1138	4
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1878	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1585	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1189	8
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1172	6
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1207	3
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1704	4
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1668	3
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1709	4
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1391	4
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1740	4
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1522	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1180	3
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1671	7
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1080	5
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1153	5
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1853	3
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1137	3
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1791	3
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1096	4
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1316	6
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1843	8
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1097	5
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1587	3

{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1889 5	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1036 3	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1725 3	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1037 3	
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1614 3	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1213 3	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1495 5	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1030 5	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1602 11	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1592 3	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1856 3	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1733 3	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1637 6	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1418 5	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1286 3	
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1440 4	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1678 3	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1107 7	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1447 6	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1591 5	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1971 6	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1339 3	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1689 3	
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1933 3	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1687 4	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1887 4	
{2026-01-20 13:53:00, 2026-01-20 13:53:10}	user_1291 3	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1318 3	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1376 5	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1686 6	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1563 4	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1915 8	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1397 5	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1031 3	
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1225 6	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1543 4	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1734 5	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1572 6	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1391 3	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1456 3	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1574 3	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1092 4	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1026 4	
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1088 3	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1041 4	
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1243 4	
{2026-01-20 13:52:50, 2026-01-20 13:53:00}	user_1084 3	
{2026-01-20 13:53:10, 2026-01-20 13:53:20}	user_1701 3	

+-----+-----+-----+
only showing top 100 rows

Batch: 2

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

Interval	user_id	crash_count
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1411	5
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1976	3
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1602	3
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1300	6
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1269	4
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1480	5
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1443	3
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1462	6
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1885	4
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1022	6
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1696	6
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1649	4
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1112	5
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1053	3
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1042	4
{2026-01-20 13:53:50, 2026-01-20 13:54:00}	user_1223	5
{2026-01-20 13:53:50, 2026-01-20 13:54:00}	user_1774	3
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1100	5
{2026-01-20 13:53:50, 2026-01-20 13:54:00}	user_1693	4
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1045	4
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1425	5
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1032	5
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1187	5
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1483	6
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1789	3
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1698	6
{2026-01-20 13:53:50, 2026-01-20 13:54:00}	user_1981	3
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1875	3
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1898	3
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1888	3
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1354	4
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1699	4
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1282	4
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1003	4
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1393	5
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1568	4
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1206	5
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1010	3
{2026-01-20 13:53:50, 2026-01-20 13:54:00}	user_1104	5
{2026-01-20 13:53:50, 2026-01-20 13:54:00}	user_1465	5
{2026-01-20 13:53:50, 2026-01-20 13:54:00}	user_1268	3
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1603	4
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1008	3
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1425	3
{2026-01-20 13:53:20, 2026-01-20 13:53:30}	user_1404	6
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1974	7
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1529	3
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1515	4
{2026-01-20 13:53:30, 2026-01-20 13:53:40}	user_1029	7
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1758	7
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1989	5
{2026-01-20 13:53:40, 2026-01-20 13:53:50}	user_1658	5

{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1088 4	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1466 6	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1050 4	
{2026-01-20 13:53:50, 2026-01-20 13:54:00} user_1519 3	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1353 4	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1498 5	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1562 6	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1704 5	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1552 3	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1066 6	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1434 5	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1596 4	
{2026-01-20 13:53:50, 2026-01-20 13:54:00} user_1733 4	
{2026-01-20 13:53:50, 2026-01-20 13:54:00} user_1068 6	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1873 6	
{2026-01-20 13:53:50, 2026-01-20 13:54:00} user_1598 3	
{2026-01-20 13:53:50, 2026-01-20 13:54:00} user_1565 3	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1172 7	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1207 4	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1373 3	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1589 9	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1247 7	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1895 5	
{2026-01-20 13:53:50, 2026-01-20 13:54:00} user_1704 8	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1340 4	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1513 8	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1667 4	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1309 5	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1340 5	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1390 3	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1840 5	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1395 5	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1053 3	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1357 6	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1109 5	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1068 3	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1540 4	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1565 4	
{2026-01-20 13:53:20, 2026-01-20 13:53:30} user_1418 9	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1838 5	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1937 5	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1198 3	
{2026-01-20 13:53:30, 2026-01-20 13:53:40} user_1677 5	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1224 6	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1295 3	
{2026-01-20 13:53:40, 2026-01-20 13:53:50} user_1851 4	
{2026-01-20 13:53:50, 2026-01-20 13:54:00} user_1639 7	
{2026-01-20 13:53:50, 2026-01-20 13:54:00} user_1459 5	

+-----+-----+-----+
only showing top 100 rows

Batch: 3

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

Interval	user_id	crash_count
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1204	8
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1469	6
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1434	5
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1617	5
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1221	10
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1269	4
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1593	5
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1533	3
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1562	4
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1288	5
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1392	6
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1247	7
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1010	3
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1161	5
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1747	11
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1468	5
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1988	7
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1904	5
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1565	3
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1390	5
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1231	4
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1836	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1633	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1412	8
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1361	9
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1937	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1777	3
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1500	3
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1520	6
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1145	5
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1757	7
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1693	7
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1929	4
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1548	5
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1908	8
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1220	6
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1830	7
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1748	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1050	4
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1710	3
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1783	4
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1330	3
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1302	4
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1970	3
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1417	4
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1634	4
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1981	9
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1477	10
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1883	8
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1207	7
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1296	6
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1779	3

{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1859	7
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1080	7
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1768	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1456	6
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1989	3
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1177	4
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1754	5
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1889	5
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1844	3
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1762	4
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1227	6
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1143	3
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1968	5
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1675	3
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1669	3
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1457	3
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1720	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1993	6
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1603	7
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1284	5
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1482	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1381	3
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1801	4
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1235	4
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1075	8
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1197	8
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1128	3
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1195	5
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1248	4
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1639	5
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1746	5
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1847	4
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1904	5
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1019	6
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1353	9
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1824	6
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1316	5
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1694	6
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1520	4
{2026-01-20 13:54:00, 2026-01-20 13:54:10}	user_1601	3
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1945	4
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1357	8
{2026-01-20 13:54:10, 2026-01-20 13:54:20}	user_1964	9
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1480	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1435	3
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1653	4
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1643	8
{2026-01-20 13:54:20, 2026-01-20 13:54:30}	user_1512	3

+-----+-----+-----+
only showing top 100 rows

Batch: 4

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

Interval	user_id	crash_count
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1572	6
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1417	5
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1549	8
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1200	5
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1310	4
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1940	6
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1356	6
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1173	4
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1295	4
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1833	4
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1788	4
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1473	3
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1946	3
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1945	6
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1878	7
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1510	4
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1441	4
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1276	3
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1439	5
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1862	3
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1514	7
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1806	4
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1710	4
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1891	4
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1319	4
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1150	11
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1810	6
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1374	4
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1711	3
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1762	4
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1330	4
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1783	6
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1117	6
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1608	5
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1292	5
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1969	5
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1261	4
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1077	4
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1745	7
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1929	7
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1504	6
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1986	4
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1337	4
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1177	4
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1595	5
{2026-01-20 13:54:30, 2026-01-20 13:54:40}	user_1500	3
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1414	5
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1455	6
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1920	4
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1258	6
{2026-01-20 13:54:40, 2026-01-20 13:54:50}	user_1517	5
{2026-01-20 13:54:50, 2026-01-20 13:55:00}	user_1743	3

```

|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1812|7|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1726|6|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1551|3|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1482|5|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1269|3|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1901|6|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1606|5|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1477|5|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1446|4|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1622|3|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1423|4|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1458|3|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1711|4|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1507|4|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1008|4|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1715|5|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1225|6|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1649|3|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1020|5|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1288|3|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1652|4|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1471|5|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1232|3|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1407|7|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1958|5|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1472|6|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1840|5|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1813|4|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1512|7|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1933|8|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1180|4|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1990|5|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1074|3|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1161|3|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1204|6|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1678|4|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1550|5|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1688|4|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1273|3|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1329|5|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1512|3|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1820|6|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1715|8|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1294|3|
|{2026-01-20 13:54:50, 2026-01-20 13:55:00}|user_1019|5|
|{2026-01-20 13:54:30, 2026-01-20 13:54:40}|user_1992|6|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1941|6|
|{2026-01-20 13:54:40, 2026-01-20 13:54:50}|user_1979|3|
+-----+-----+-----+

```

All requirements satisfied

- Processes high-velocity streams with stateful aggregations

- Handles late data within 30s watermark
- Near real-time detection (<30s latency)
- Horizontally scalable with checkpointing for fault tolerance

Batch Statistics:

- 100+ critical users per batch (top 100 displayed)
- 4 consecutive windows per batch
- Crash range: 3-11 per user per window

Step 6: Testing and Validation

Test 1: Verify Filtering

Expected: Only "crash" logs with "High"/"Critical" severity appear Result: only "High"/"Critical" severity appear

Test 2: Verify Windowing

Expected: Results grouped in 10-second intervals based on event timestamp Result: grouping by 10-second interval works

Test 3: Verify Threshold

Expected: Users with <=2 crashes do not appear Result: no numbers smaller than 3

Test 4: Late Data Handling

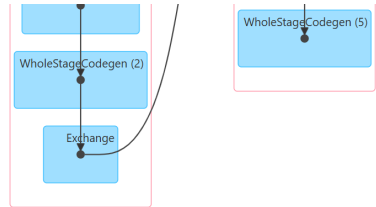
Expected: Events arriving up to 30 seconds late are processed Result: processed if 30 seconds late or less; ignored if older than 30 seconds

Step 7: Performance Testing

Overall performance:

Job Id (Job Group) ▾	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
5 (bc4d9f50-c272-4f21-9d1c-ec3809e7d929)	id = 58c6e398-9318-445d-a500-805bc0e63e2c runId = bc4d9f50-c272-4f21-9d1c-ec3809e7d929 batc... start at <unknown>-0	2026/01/20 14:40:28	5 s	2/2	204/204
4 (bc4d9f50-c272-4f21-9d1c-ec3809e7d929)	id = 58c6e398-9318-445d-a500-805bc0e63e2c runId = bc4d9f50-c272-4f21-9d1c-ec3809e7d929 batc... start at <unknown>-0	2026/01/20 14:40:22	6 s	2/2	204/204
3 (bc4d9f50-c272-4f21-9d1c-ec3809e7d929)	id = 58c6e398-9318-445d-a500-805bc0e63e2c runId = bc4d9f50-c272-4f21-9d1c-ec3809e7d929 batc... start at <unknown>-0	2026/01/20 14:40:15	7 s	2/2	204/204
2 (bc4d9f50-c272-4f21-9d1c-ec3809e7d929)	id = 58c6e398-9318-445d-a500-805bc0e63e2c runId = bc4d9f50-c272-4f21-9d1c-ec3809e7d929 batc... start at <unknown>-0	2026/01/20 14:40:03	11 s	2/2	204/204
1 (bc4d9f50-c272-4f21-9d1c-ec3809e7d929)	id = 58c6e398-9318-445d-a500-805bc0e63e2c runId = bc4d9f50-c272-4f21-9d1c-ec3809e7d929 batc... start at <unknown>-0	2026/01/20 14:39:45	17 s	2/2	204/204
0 (bc4d9f50-c272-4f21-9d1c-ec3809e7d929)	id = 58c6e398-9318-445d-a500-805bc0e63e2c runId = bc4d9f50-c272-4f21-9d1c-ec3809e7d929 batc... start at <unknown>-0	2026/01/20 14:39:17	27 s	2/2	204/204

Time of a single run:



Completed Stages (2)

age: 1

1 Pages. Jump to 1. Show 100 items in a page. Go

Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
11	id = 58c6e398-9318-445d-a500-805bc0e63e2c runId = bc4d9f50-c272-4f21-9d1c-ec3809e7d929 batch = 5 start at <unknown>	2026/01/20 14:40:30	3 s	200/200			91.0 KiB	
10	id = 58c6e398-9318-445d-a500-805bc0e63e2c runId = bc4d9f50-c272-4f21-9d1c-ec3809e7d929 batch = 5 start at <unknown>	2026/01/20 14:40:28	2 s	4/4				91.0 KiB

Processing speed (input vs processed):

Streaming Queries (1)

age: 1

1 Pages. Jump to 1. Show 100 items in a page. Go

Name	Status	ID	Run ID	Start Time	Duration	Avg Input /sec	Avg Process /sec	Latest Batch
<no name>	RUNNING	58c6e398-9318-445d-a500-805bc0e63e2c	bc4d9f50-c272-4f21-9d1c-ec3809e7d929	2026/01/20 14:39:10	2 minutes 26 seconds	97916.24	121072.83	13

age: 1

1 Pages. Jump to 1. Show 100 items in a page. Go

Resource usage:

Executors

Show 20 entries

Executor ID	Address	Status	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Local
driver	b2af8b4106a3:40455	Active	0	6.8 MiB / 434.4 MiB	0.0 B	0	0	0	0	0	7.8 min (1 s)	0.0 B	0.0 B	0.0 B	
0	172.22.0.5:35657	Active	0	6.8 MiB / 2.2 GiB	0.0 B	8	9	0	5394	5403	18 min (16 s)	0.0 B	2.2 MiB	2.3 MiB	st

Showing 1 to 2 of 2 entries

Observations

- Single executor with 8 cores handles moderate load effectively
- State size fits comfortably in 4G memory
- 30-second watermark balances late data handling and memory usage

Step 8: Possible Scaling

Scaling

- Increase executors to distribute load across machines
- Kafka partitions distributed to Spark tasks
- State partitioned by (window, user_id) keys
- **Best practice:** num_kafka_partitions ≥ num_executor_cores

Clean up the load-generator folder under logs-processing.

```
docker compose down -v
```