

Adil Mubashir Chaudhry 24280007
Hashim Shabbir 24280006

Setting Up Kafka via Docker Compose

```
(env) vboxuser@adil:~/ai601-data-engineering/labs/lab5$ docker-compose -f zk-single-kafka.yml up -d
[+] Running 19/19
 ✓ kafka1 Pulled 298.0s
   ✓ a09bb1026942 Pull complete 52.4s
   ✓ c1b083347de2 Pull complete 52.5s
   ✓ 4d1200200a21 Pull complete 283.8s
   ✓ c57360fb79ac Pull complete 283.9s
   ✓ bc1aa429c05a Pull complete 284.5s
   ✓ 346a1569c408 Pull complete 284.7s
   ✓ 2a8b1ee0cc7c Pull complete 286.0s
   ✓ 485bbb2fee19 Pull complete 286.1s
   ✓ 4b041791a036 Pull complete 286.1s
   ✓ 5040019df3ec Pull complete 286.2s
   ✓ a3749ea76211 Pull complete 286.3s
   ✓ b9fe7d7b902e Pull complete 286.4s
   ✓ 432d3ac62f8b Pull complete 286.5s
   ✓ 7e13fc38f0b3 Pull complete 293.8s
   ✓ 4d66e5b7a7d9 Pull complete 293.9s
 ✓ zoo1 Pulled 298.0s
   ✓ 51601473dc3c Pull complete 293.7s
   ✓ 848d4e00419f Pull complete 293.9s
[+] Running 3/3
 ✓ Network lab5_default Created 0.2s
 ✓ Container zoo1 Started 1.4s
 ✓ Container kafka1 Started 1.6s
(env) vboxuser@adil:~/ai601-data-engineering/labs/lab5$ docker-compose -f zk-single-kafka.yml ps
NAME                IMAGE                                COMMAND                                SERVICE  CREATED      STATUS
kafka1              confluentinc/cp-kafka:7.8.0        "/etc/confluent/dock..."          kafka1   2 minutes ago Up 2 minutes
0.0.0.0:9092->9092/tcp, [::]:9092->9092/tcp, 0.0.0.0:9999->9999/tcp, [::]:9999->9999/tcp, 0.0.0.0:29092->29092/tcp, [::]:29092->29092/tcp
zoo1                confluentinc/cp-zookeeper:7.8.0    "/etc/confluent/dock..."          zoo1     2 minutes ago Up 2 minutes
2888/tcp, 0.0.0.0:2181->2181/tcp, [::]:2181->2181/tcp, 3888/tcp
(env) vboxuser@adil:~/ai601-data-engineering/labs/lab5$
```

Creating Topic

```
(env) vboxuser@adil:~/ai601-data-engineering/labs/lab5$ docker exec -it kafka1 /bin/bash
[appuser@kafka1 ~]$ kafka-topics --version
7.8.0-ccs
[appuser@kafka1 ~]$ kafka-topics --create --topic music_events --bootstrap-server localhost:9092 --partitions 1
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide.
To avoid issues it is best to use either, but not both.
Created topic music_events.
[appuser@kafka1 ~]$ --replication-factor 1
bash: --replication-factor: command not found
[appuser@kafka1 ~]$
```

Music Event Production

```
(env) vboxuser@adil:~/ai601-data-engineering/labs/lab5$ python music_producer.py
Sent event: {'song_id': 303, 'timestamp': 1742199765.6240277, 'region': 'US', 'action': 'play'}
Sent event: {'song_id': 303, 'timestamp': 1742199766.2800322, 'region': 'APAC', 'action': 'play'}
Sent event: {'song_id': 505, 'timestamp': 1742199768.267374, 'region': 'EU', 'action': 'play'}
Sent event: {'song_id': 404, 'timestamp': 1742199769.8471177, 'region': 'EU', 'action': 'play'}
Sent event: {'song_id': 505, 'timestamp': 1742199771.0479708, 'region': 'EU', 'action': 'play'}
Sent event: {'song_id': 101, 'timestamp': 1742199771.6911042, 'region': 'APAC', 'action': 'play'}
Sent event: {'song_id': 202, 'timestamp': 1742199772.5646923, 'region': 'US', 'action': 'play'}
```

Now Trending Script

```
=== Batch: 1 ===
```

window	region	song_id	count	rn
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	EU	505	4	1
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	EU	303	2	2
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	EU	404	2	3
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	APAC	404	3	1
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	APAC	202	2	2
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	APAC	303	2	3
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	US	404	4	1
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	US	101	2	2
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	US	202	2	3

```
=== Batch: 2 ===
```

window	region	song_id	count	rn
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	EU	505	5	1
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	EU	404	3	2
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	EU	101	2	3
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	APAC	101	4	1
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	APAC	202	3	2
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	APAC	303	3	3
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	US	404	5	1
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	US	101	3	2
{2025-03-17 08:25:00, 2025-03-17 08:30:00}	US	202	3	3

Vary the Time Window

```
=== Batch: 1 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|EU|303|4|1|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|EU|101|3|2|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|EU|505|3|3|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|APAC|101|3|1|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|APAC|202|2|2|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|APAC|303|2|3|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|US|101|1|1|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|US|303|1|2|
+-----+-----+-----+-----+

=== Batch: 2 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|EU|505|4|1|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|EU|404|2|2|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|EU|202|2|3|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|APAC|202|3|1|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|APAC|404|2|2|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|US|101|4|1|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|US|404|3|2|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|US|303|2|3|
+-----+-----+-----+-----+

=== Batch: 3 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|EU|101|5|1|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|EU|202|3|2|
|{2025-03-17 08:29:00, 2025-03-17 08:30:00}|APAC|101|4|1|
```

Analysis

The trending songs list changes in every batch, showing that Kafka is recalculating rankings as new data arrives. In the original time window the 505 song id was the top trending in all batches but in the shorter interval every batch has a different trending song. If the window is too short, a song might trend briefly but disappear quickly, making it hard to identify sustained trends.

Add Skip/Like Actions

```
(env) vboxuser@adil:~/ai601-data-engineering/labs/lab5$ python music_producer.py
Sent event: {'song_id': 404, 'timestamp': 1742200391.8650646, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 505, 'timestamp': 1742200393.0354962, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 303, 'timestamp': 1742200394.560335, 'region': 'EU', 'action': 'skip'}
Sent event: {'song_id': 303, 'timestamp': 1742200396.5273533, 'region': 'EU', 'action': 'play'}
Sent event: {'song_id': 202, 'timestamp': 1742200398.4762769, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 404, 'timestamp': 1742200399.9250374, 'region': 'APAC', 'action': 'play'}
Sent event: {'song_id': 505, 'timestamp': 1742200401.4471993, 'region': 'EU', 'action': 'like'}
```

```
+-----+-----+-----+-----+-----+-----+-----+
|window                                     |region|song_id|play_count|skip_count|like_count|skip_ratio|rn |
+-----+-----+-----+-----+-----+-----+-----+
|{2025-03-17 08:33:00, 2025-03-17 08:34:00}|EU    |202    |0        |1        |0        |1.0      |1 |
+-----+-----+-----+-----+-----+-----+-----+

=== Batch: 1 ===
+-----+-----+-----+-----+-----+-----+-----+
|window                                     |region|song_id|play_count|skip_count|like_count|skip_ratio|rn |
+-----+-----+-----+-----+-----+-----+-----+
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|EU    |202    |3        |1        |0        |0.25     |1 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|EU    |101    |2        |2        |0        |0.5      |2 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|EU    |505    |1        |1        |0        |0.5      |3 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|US    |303    |3        |1        |2        |0.16666666666666666|1 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|US    |404    |1        |1        |0        |0.5      |2 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|US    |505    |0        |1        |1        |0.5      |3 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|APAC  |505    |2        |0        |1        |0.0      |1 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|APAC  |303    |1        |1        |0        |0.5      |2 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|APAC  |202    |1        |0        |0        |0.0      |3 |
+-----+-----+-----+-----+-----+-----+-----+

=== Batch: 2 ===
+-----+-----+-----+-----+-----+-----+-----+
|window                                     |region|song_id|play_count|skip_count|like_count|skip_ratio|rn |
+-----+-----+-----+-----+-----+-----+-----+
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|EU    |404    |0        |0        |3        |0.0      |1 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|US    |303    |3        |1        |3        |0.14285714285714285|1 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|US    |202    |1        |0        |0        |0.0      |2 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|US    |101    |0        |2        |0        |1.0      |3 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|APAC  |505    |3        |0        |1        |0.0      |1 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|APAC  |303    |1        |3        |1        |0.6      |2 |
|{2025-03-17 08:34:00, 2025-03-17 08:35:00}|APAC  |202    |1        |1        |1        |0.33333333333333333|3 |
+-----+-----+-----+-----+-----+-----+-----+
```

Set a Different Micro-Batch Interval

```
25/03/17 08:37:18 WARN ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 1000 milliseconds, but spent 31001 milliseconds
=== Batch: 1 ===
+-----+-----+-----+-----+-----+-----+-----+
|window|region|song_id|play_count|skip_count|like_count|skip_ratio|rn|
+-----+-----+-----+-----+-----+-----+-----+
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|EU|404|1|2|0|0.6666666666666666|1|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|EU|202|0|3|0|1.0|2|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|EU|101|0|2|0|1.0|3|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|US|202|1|1|1|0.3333333333333333|1|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|US|505|1|0|1|0.0|2|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|US|101|0|0|2|0.0|3|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|APAC|404|2|0|0|0.0|1|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|APAC|505|1|0|0|0.0|2|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|APAC|202|0|1|0|1.0|3|
+-----+-----+-----+-----+-----+-----+-----+

25/03/17 08:37:45 WARN ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 1000 milliseconds, but spent 27719 milliseconds
=== Batch: 2 ===
+-----+-----+-----+-----+-----+-----+-----+
|window|region|song_id|play_count|skip_count|like_count|skip_ratio|rn|
+-----+-----+-----+-----+-----+-----+-----+
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|EU|202|2|3|1|0.5|1|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|EU|404|2|2|0|0.5|2|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|EU|101|1|4|0|0.8|3|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|US|202|2|1|2|0.2|1|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|US|505|1|1|2|0.25|2|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|US|101|1|0|2|0.0|3|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|APAC|404|4|1|0|0.2|1|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|APAC|101|3|1|1|0.2|2|
|{2025-03-17 08:37:00, 2025-03-17 08:38:00}|APAC|202|0|2|0|1.0|3|
+-----+-----+-----+-----+-----+-----+-----+

25/03/17 08:37:56 WARN ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 1000 milliseconds, but spent 10375 milliseconds
=== Batch: 3 ===
```

Analysis

Changing micro-batch intervals results in faster updates to trending songs, real-time responsiveness. However, as we can see above, the system is struggling to keep up with the 1-second interval. That means that the computation requirements were not being met resulting in the delays.