

LAB5:

Running containers:

The Docker Desktop interface displays container management tools. At the top, it shows overall system usage: Container CPU usage at 1.28% (12 CPUs available) and Container memory usage at 368.75MB (3.64GB available). Below this is a search bar and a toggle for 'Only show running containers'. A table lists the containers:

Name	Image	Status	Port(s)	CPU (%)	Last started	Actions
lewis	redis:6.2	Exited	6379:6379	0%	5 days ago	[Play] [Stop] [Delete]
lab5		Running (2/2)		1.41%	1 hour ago	[Play] [Stop] [Delete]
zoo1	confluentinc/cp-zookeeper:7.8.0	Running	2181:2181	0.31%	1 hour ago	[Play] [Stop] [Delete]
kafka1	confluentinc/cp-kafka:7.8.0	Running	29092:29092	1.1%	1 hour ago	[Play] [Stop] [Delete]

At the bottom right, it indicates 'Selected 3 of 5' containers.

Getting data:

A terminal window displays a continuous stream of JSON events. Each event contains fields for song_id, timestamp, region, and action. The actions include 'play', 'like', and 'skip'. The data is being sent from a source identified as 'lewis'.

```
Sent event: {'song_id': 303, 'timestamp': 1742191456.9780357, 'region': 'APAC', 'action': 'play'}
Sent event: {'song_id': 101, 'timestamp': 1742191457.8324306, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 505, 'timestamp': 1742191458.8466344, 'region': 'US', 'action': 'like'}
Sent event: {'song_id': 303, 'timestamp': 1742191460.3902543, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 404, 'timestamp': 1742191461.397085, 'region': 'US', 'action': 'skip'}
Sent event: {'song_id': 404, 'timestamp': 1742191463.0179727, 'region': 'EU', 'action': 'play'}
Sent event: {'song_id': 202, 'timestamp': 1742191463.8237057, 'region': 'US', 'action': 'play'}
Sent event: {'song_id': 404, 'timestamp': 1742191465.6839004, 'region': 'EU', 'action': 'like'}
Sent event: {'song_id': 505, 'timestamp': 1742191467.3752327, 'region': 'US', 'action': 'like'}
Sent event: {'song_id': 101, 'timestamp': 1742191468.7793446, 'region': 'US', 'action': 'like'}
Sent event: {'song_id': 202, 'timestamp': 1742191470.7156336, 'region': 'EU', 'action': 'like'}
Sent event: {'song_id': 202, 'timestamp': 1742191472.1271205, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 303, 'timestamp': 1742191473.859773, 'region': 'US', 'action': 'play'}
Sent event: {'song_id': 505, 'timestamp': 1742191474.6967747, 'region': 'US', 'action': 'like'}
Sent event: {'song_id': 202, 'timestamp': 1742191476.0905564, 'region': 'US', 'action': 'skip'}
Sent event: {'song_id': 303, 'timestamp': 1742191477.929077, 'region': 'APAC', 'action': 'play'}
Sent event: {'song_id': 505, 'timestamp': 1742191479.3853388, 'region': 'US', 'action': 'like'}
Sent event: {'song_id': 404, 'timestamp': 1742191480.0149047, 'region': 'US', 'action': 'skip'}
Sent event: {'song_id': 303, 'timestamp': 1742191481.6061456, 'region': 'US', 'action': 'skip'}
Sent event: {'song_id': 303, 'timestamp': 1742191483.15933, 'region': 'EU', 'action': 'play'}
Sent event: {'song_id': 505, 'timestamp': 1742191484.3845015, 'region': 'US', 'action': 'like'}
Sent event: {'song_id': 505, 'timestamp': 1742191485.2702405, 'region': 'APAC', 'action': 'skip'}
Sent event: {'song_id': 505, 'timestamp': 1742191486.3546112, 'region': 'APAC', 'action': 'skip'}
Sent event: {'song_id': 404, 'timestamp': 1742191487.1881208, 'region': 'EU', 'action': 'skip'}
Sent event: {'song_id': 404, 'timestamp': 1742191487.893762, 'region': 'US', 'action': 'like'}
Sent event: {'song_id': 202, 'timestamp': 1742191488.766699, 'region': 'EU', 'action': 'play'}
Sent event: {'song_id': 505, 'timestamp': 1742191489.2882986, 'region': 'EU', 'action': 'play'}
Sent event: {'song_id': 202, 'timestamp': 1742191489.989904, 'region': 'APAC', 'action': 'skip'}
Sent event: {'song_id': 101, 'timestamp': 1742191490.5321686, 'region': 'APAC', 'action': 'play'}
Sent event: {'song_id': 202, 'timestamp': 1742191491.8114913, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 101, 'timestamp': 1742191492.5328155, 'region': 'US', 'action': 'play'}
Sent event: {'song_id': 303, 'timestamp': 1742191493.5722184, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 303, 'timestamp': 1742191494.0776045, 'region': 'US', 'action': 'play'}
Sent event: {'song_id': 505, 'timestamp': 1742191495.378207, 'region': 'US', 'action': 'like'}
Sent event: {'song_id': 505, 'timestamp': 1742191496.7594373, 'region': 'APAC', 'action': 'skip'}
Sent event: {'song_id': 505, 'timestamp': 1742191497.7278998, 'region': 'APAC', 'action': 'like'}
Sent event: {'song_id': 404, 'timestamp': 1742191499.0470862, 'region': 'APAC', 'action': 'skip'}
Sent event: {'song_id': 404, 'timestamp': 1742191500.9368513, 'region': 'EU', 'action': 'skip'}
```

Introducing new actions -> skip + like

And

Updating the window size to 5 minutes.

```
root@tabdullah:/home/office x Command Prompt x root@tabdullah:/home/office x + v
25/03/17 10:55:25 WARN AdminClientConfig: These configurations '[key.deserializer, value.deserializer, enable.auto.commit, max.poll.records, aut
o.offset.reset]' were supplied but are not used yet.
No data in this batch.
=== Batch: 1 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 10:55:00, 2025-03-17 10:56:00}|US|505|1|1|
|{2025-03-17 10:55:00, 2025-03-17 10:56:00}|APAC|202|2|1|
+-----+-----+-----+-----+

=== Batch: 2 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 10:55:00, 2025-03-17 10:56:00}|EU|202|1|1|
|{2025-03-17 10:55:00, 2025-03-17 10:56:00}|EU|101|1|2|
|{2025-03-17 10:55:00, 2025-03-17 10:56:00}|APAC|303|1|1|
+-----+-----+-----+-----+

No data in this batch.
=== Batch: 4 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 10:56:00, 2025-03-17 10:57:00}|US|404|1|1|
|{2025-03-17 10:56:00, 2025-03-17 10:57:00}|EU|202|1|1|
+-----+-----+-----+-----+

No data in this batch.
=== Batch: 6 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 10:56:00, 2025-03-17 10:57:00}|US|202|1|1|
|{2025-03-17 10:56:00, 2025-03-17 10:57:00}|APAC|404|1|1|
+-----+-----+-----+-----+
```

24°C
Smoke

Search

10:58 am
17/03/2025

WHEN set Batch time to 1 second:

```
root@tabdullah:/home/office x Command Prompt x root@tabdullah:/home/office x + v
25/03/17 11:07:03 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 10.255.255.254, 34103, None)
25/03/17 11:07:03 INFO BlockManagerMasterEndpoint: Registering block manager 10.255.255.254:34103 with 434.4 MiB RAM, BlockManagerId(driver, 10.255.255.254, 34103, None)
25/03/17 11:07:03 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 10.255.255.254, 34103, None)
25/03/17 11:07:03 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 10.255.255.254, 34103, None)
25/03/17 11:07:06 WARN ResolveWriteToStream: Temporary checkpoint location created which is deleted normally when the query didn't fail: /tmp/te
morary-24a868b4-f78d-45b3-b060-229ae819c8fb. If it's required to delete it under any circumstances, please set spark.sql.streaming.forceDeleteT
empCheckpointLocation to true. Important to know deleting temp checkpoint folder is best effort.
25/03/17 11:07:06 WARN ResolveWriteToStream: spark.sql.adaptive.enabled is not supported in streaming DataFrames/Datasets and will be disabled.
25/03/17 11:07:06 WARN AdminClientConfig: These configurations '[key.deserializer, value.deserializer, enable.auto.commit, max.poll.records, aut
o.offset.reset]' were supplied but are not used yet.
No data in this batch.
25/03/17 11:07:21 WARN ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 1000 milliseconds, but spent 14850 milli
seconds
=== Batch: 1 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 11:07:00, 2025-03-17 11:08:00}|APAC|505|1|1|
|{2025-03-17 11:07:00, 2025-03-17 11:08:00}|APAC|404|1|2|
|{2025-03-17 11:07:00, 2025-03-17 11:08:00}|APAC|101|1|3|
|{2025-03-17 11:07:00, 2025-03-17 11:08:00}|EU|101|1|1|
|{2025-03-17 11:07:00, 2025-03-17 11:08:00}|EU|303|1|2|
|{2025-03-17 11:07:00, 2025-03-17 11:08:00}|US|303|1|1|
+-----+-----+-----+-----+
25/03/17 11:07:33 WARN ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 1000 milliseconds, but spent 12158 milli
seconds
=== Batch: 2 ===
+-----+-----+-----+-----+
|window|region|song_id|count|rn|
+-----+-----+-----+-----+
|{2025-03-17 11:07:00, 2025-03-17 11:08:00}|US|101|1|1|
+-----+-----+-----+-----+
25/03/17 11:07:39 WARN ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 1000 milliseconds, but spent 6117 millis
econds
█
```

Observed results:

- With a 1-second interval, the results will update more frequently, but this may increase CPU usage.
- With a 5-second interval, the results will update less frequently, reducing CPU overhead.
- But in reality we need more time interval, we need at least 10-20 seconds interval as we would get warning: "ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 10799 milliseconds"