

Thmanyah Real-Time Streaming Assignment

This project simulates a real-time data streaming system using Kafka, Debezium, Redis, BigQuery, and Faust. The architecture ingests user engagement events, enriches them with content metadata, and writes enriched data to Redis, BigQuery, and a mock external system.

Architecture Overview

- PostgreSQL + Debezium → change data capture (CDC) for events.
- Kafka → message broker.
- Faust → real-time -less than 5 sec response- stream processing and enrichment (Also used Flink logic already added).
- Redis → fast storage for enrichment and aggregations.
- Redis Insight → Redis GUI client (for inspection).
- BigQuery → long-term storage and analytics.
- Redpanda Console → Kafka topic inspection (GUI).
- pgAdmin → PostgreSQL GUI client.

High-Level Architecture

Technologies Used

Component	Tool
Streaming Engine	Faust (Python-native, lightweight alternative to Flink for local dev)

Source Connector	PostgreSQL CDC (via Debezium)
Message Broker	Apache Kafka (intermediate pub/sub layer)
Sink Connectors	BigQuery client, Redis client, HTTP client (mock external system)
Infrastructure	Docker Compose
Data Generator	Python script (to simulate engagement & content data into PostgreSQL)

Features

- Real-time event ingestion via Kafka
 - Enrichment with content metadata
 - Track the changes in DB through Debezium for CDC
 - Backfill support for historical PostgreSQL data
 - Redis scoring and n-minute engagement aggregation
 - RedisInsight integration support (manual)
 - BigQuery loading
 - Mock data generation and Dev-friendly tooling
 - Clean modular codebase with app/ structure
 - Makefile for full setup automation
-

Setup Instructions

1. Prerequisites

- `pyenv`
- Docker + Docker Compose
- Python 3.9.x (installed via pyenv)
- Google Cloud credentials (if using BigQuery)
- Redis Insight(Optional)

2. Clone & Prepare

Shell

```
git clone  
https://github.com/your-org/thmanyah-streaming-assignment.git  
cd thmanyah-streaming-assignment  
  
cp .env.example .env # Fill in your local values
```

3. Run All in One

Use the Makefile to set everything up and run the pipeline:

Shell

```
make
```

This will:

- Create a virtual environment
 - Install dependencies
 - Start all Docker services
 - Wait for services to become healthy
 - Register Debezium connector
 - Run Faust streaming app
 - Generate mock data
-

Run Manually (Step-by-Step)

None

```
make setup          # Python venv setup
make up             # Start Docker containers
make wait           # Wait for services (2 mins)
make register-connector # Register Debezium
make run-faust       # Start Faust
make mock-data       # Run mock generator
```

Utilities

None

```
make down           # Stop all Docker containers and delete
them
make clean           # Clean all containers, volumes
make backfill        # Reproduce PostgreSQL data to kafka
topic
```

Project Structure

None

```
├─ app/.            <- Core logic
|   ├─ config.py
|   ├─ models.py
|   └─ faust_app.py
```

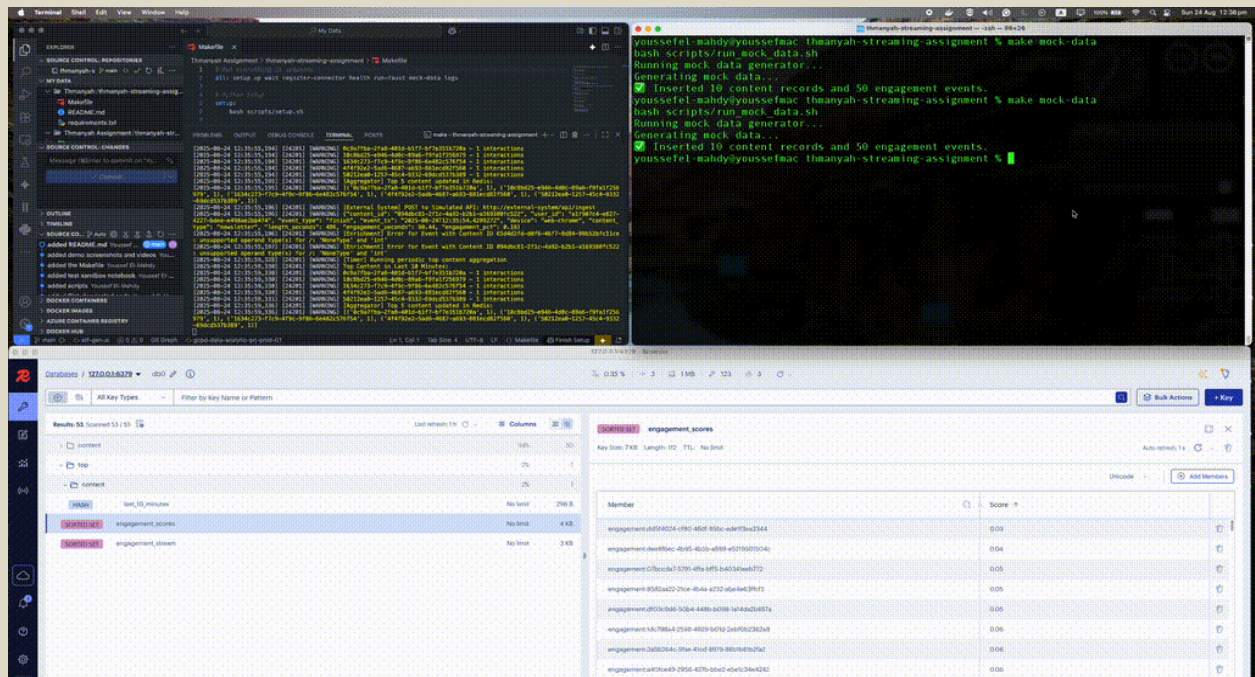
```
|   ├── enrichment.py
|   ├── redis_utils.py
|   └── bigquery_utils.py
├── config/          <- JSON configs (Debezium)
|   └── debezium-source.json
├── docker/          <- Docker compose + SQL init
|   ├── docker-compose.yml
|   └── init-db.sql
├── scripts/         <- Helper scripts (mock, setup, connector,
etc.)
|   ├── setup.sh
|   ├── clean_and_run_faust.sh
|   ├── generate_mock_data.py
|   ├── run_mock_data.sh
|   ├── register_debezium_connector.py
|   └── backfill.py
├── .env.example     <- Environment variables
├── Makefile         <- Task runner
└── requirements.txt <- Python deps
```

Environment Variables

Define all variables in .env (see .env.example for a template):

- Kafka broker URL
 - Redis config
 - BigQuery project + credentials path
 - PostgreSQL and pgAdmin settings
 - External API mock URL
-

Short Demo



Note: More demos and demonstrations for the project can be found with the author upon request. Thanks.

Screenshots

Below are some screenshots demonstrating the project in action:

Docker Running Services

Ask Gordon BETA

Containers

Images

Volumes

Builds

Models BETA

MCP Toolkit BETA

Docker Hub

Docker Scout

Extensions

Containers Give feedback

View all your running containers and applications. [Learn more](#)

Container CPU usage 📉

7.17% / 800% (8 CPUs available)

Container memory usage 📉

1.25GB / 3.74GB

Show charts

Search

Only show running containers

<input type="checkbox"/>	Name	Container ID	Image	Port(s)	CPU (%)	Last started	Actions
<input type="checkbox"/>	jupyter	-	-	-	0%	11 months ago	<div><div></div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/>	thmanyah-streaming-assignment	-	-	-	7.14%	32 seconds ago	<div><div></div><div></div><div></div><div></div></div>
<input type="checkbox"/>	pgadmin	767a90267a2f	dpage/pgadmin4:9	8888:80	0.05%	33 seconds ago	<div><div></div><div></div><div></div><div></div></div>
<input type="checkbox"/>	redis	04115ba19b63	redis:7.2	6379:6379	0.24%	33 seconds ago	<div><div></div><div></div><div></div><div></div></div>
<input type="checkbox"/>	zookeeper	59d87fbc8661	confluentinc/cp-zo		0.13%	33 seconds ago	<div><div></div><div></div><div></div><div></div></div>
<input type="checkbox"/>	postgres	29a61d0f0c5e	debezium/postgres	5432:5432	0.02%	33 seconds ago	<div><div></div><div></div><div></div><div></div></div>
<input type="checkbox"/>	kafka	24376f316b18	confluentinc/cp-k	29092:29092	3.14%	33 seconds ago	<div><div></div><div></div><div></div><div></div></div>
<input type="checkbox"/>	redpanda-console	1e4e7a7eb564	redpandadata/con	9999:8080	0%	32 seconds ago	<div><div></div><div></div><div></div><div></div></div>
<input type="checkbox"/>	connect	d016a222a6eb	debezium/connect	8083:8083	3.56%	32 seconds ago	<div><div></div><div></div><div></div><div></div></div>

Showing 9 items

Engine running | RAM 3.48 GB CPU 2.87% Disk: 17.48 GB used (limit 58.37 GB)

Terminal Update

Kafka Topics

Redpanda

Overview

Topics

Schema Registry

Consumer Groups

Security

Quotas

Connect

Transforms

Reassign Partitions

Cluster

Topics 🔍

5

33

33

Total topicsTotal partitionsTotal replicas

Enter search term/regex

Create topic

Show internal topics

NAME	PARTITIONS	REPLICAS	CLEANUPPOLICY	SIZE
debezium_connect_configs	1	1	compact	4.63 kiB
debezium_connect_offsets	25	1	compact	4 kiB
debezium_connect_statuses	5	1	compact	5.66 kiB
thmanyah.public.content	1	1	delete	49.3 kiB
thmanyah.public.engagement_events	1	1	delete	429 kiB

50 / page

1

Redpanda Console (Platform Version v25.1) (built August 19, 2025) 926B9FC

Kafka Topic: engagement_events

Cluster > Topics

thmanyah.public.engagement_events

Size: 429 kiB | Estimated messages: 500 | Cleanup Policy: Delete | Retention: ~7 days or Infinite

Produce Record | Delete Records

Messages | Consumers | Partitions | Configuration | ACL | Documentation

START OFFSET: Newest - 50 | MAX RESULTS: 50 | Add filter

Filter table content ...

	TIMESTAMP	KEY	VALUE
>	23/08/2025, 17:56:13	Null	{"id":"901","content_id":"dd5f4024-cf80-46df-85bc-ede1f3ea2344","user_id":"807... JSON - 248 B
>	23/08/2025, 17:56:13	Null	{"id":"902","content_id":"6c23d4cb-c7cd-4535-8283-97c8eca64517","user_id":"ae2... JSON - 251 B
>	23/08/2025, 17:56:13	Null	{"id":"903","content_id":"5d8719f3-bf73-4307-b127-fb5adaa13a74","user_id":"447... JSON - 249 B
>	23/08/2025, 17:56:13	Null	{"id":"904","content_id":"5fd5465d-882a-4033-9928-d479891a66a1","user_id":"36c... JSON - 251 B
>	23/08/2025, 17:56:13	Null	{"id":"905","content_id":"b89f4d84-4428-4917-b6bd-a8bce6a68abb","user_id":"3c3... JSON - 251 B
>	23/08/2025, 17:56:13	Null	{"id":"906","content_id":"5fd5465d-882a-4033-9928-d479891a66a1","user_id":"f10... JSON - 251 B

Kafka Topic: content

Cluster > Topics

thmanyah.public.content

Size: 49.3 kiB | Estimated messages: 50 | Cleanup Policy: Delete | Retention: ~7 days or Infinite

Produce Record | Delete Records

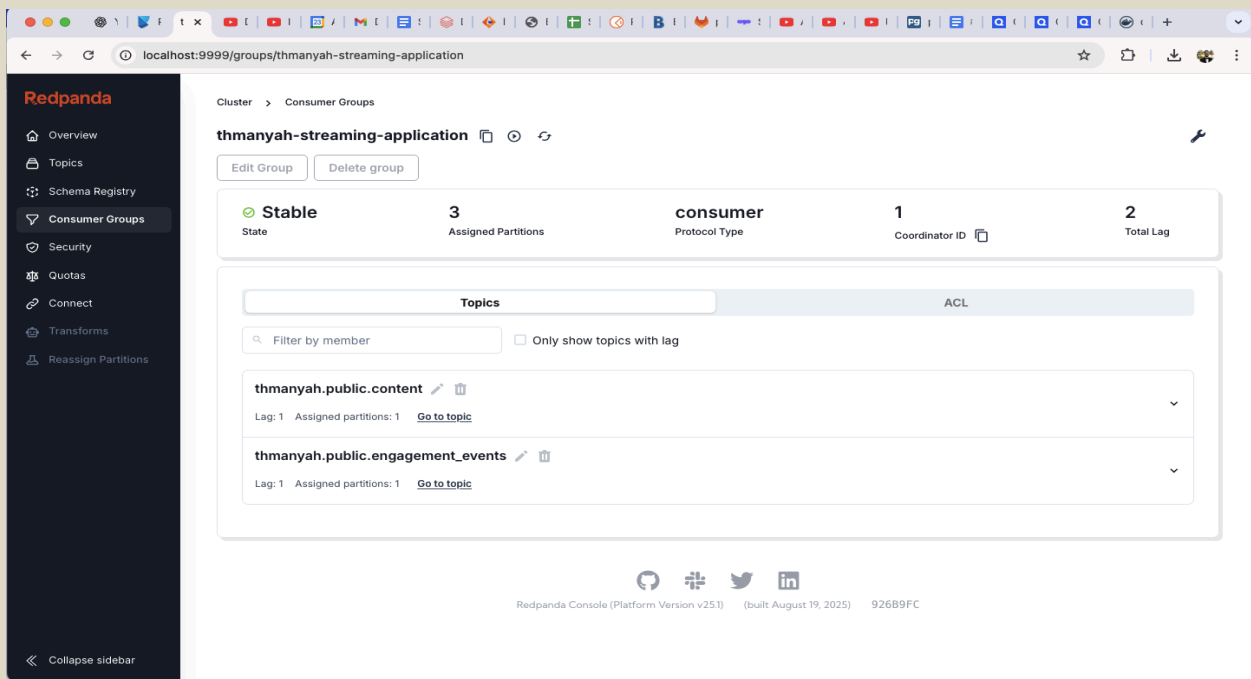
Messages | Consumers | Partitions | Configuration | ACL | Documentation

START OFFSET: Newest - 50 | MAX RESULTS: 50 | Add filter

Filter table content ...

	TIMESTAMP	KEY	VALUE
>	23/08/2025, 16:59:58	{"id":"3242a229-259f-48a0-b529-236b189965... JSON - 45 B	{"id":"3242a229-259f-48a0-b529-236b1899... JSON - 204 B
>	23/08/2025, 16:59:58	{"id":"962d38fe-3b19-4a0a-b7b3-5c576882e3... JSON - 45 B	{"id":"962d38fe-3b19-4a0a-b7b3-5c576882... JSON - 189 B
>	23/08/2025, 16:59:58	{"id":"42503af1-0272-4cb8-8f2b-654f85fda3... JSON - 45 B	{"id":"42503af1-0272-4cb8-8f2b-654f85fd... JSON - 204 B
>	23/08/2025, 16:59:58	{"id":"df88189f-021e-4feb-9e5d-e3062970b8... JSON - 45 B	{"id":"df88189f-021e-4feb-9e5d-e3062970... JSON - 199 B
>	23/08/2025, 16:59:58	{"id":"3a5b264c-5fae-41cd-8979-88b1b81b2f... JSON - 45 B	{"id":"3a5b264c-5fae-41cd-8979-88b1b81b... JSON - 204 B

Kafka Consumer: thmanyah-streaming-application



PgAdmin Thmanyah Database

Table: content

Query: `SELECT * FROM public.content ORDER BY id ASC`

	id [PK] uuid	slug	title	content_type	length_seconds	publish_ts timestamp with time zone
1	21939aa9-cf7e-4c7a-a178-34eac0cde0...	three-serve	Must themselves society somebody look.	video	307	2025-08-18 11:07:27.486975+
2	33a7db65-2a18-427b-959b-1919afa9e...	direction-foot	Throw three southern then before already.	video	74	2025-08-23 11:07:27.487305+
3	56aa2824-7a5a-426b-a2d4-8f1cd99c3...	policy-enter	Chance seat hard three right design.	podcast	724	2025-08-07 11:07:27.487838+
4	59485a83-ae60-4773-be8e-e85057cd...	miss-with-betw...	Whether word camera care.	podcast	289	2025-07-28 11:07:27.486013+
5	66971f17-0aa1-42c4-9869-15c2b93fc...	type-consider-as	Play movement own place crime.	podcast	568	2025-08-19 11:07:27.488164+
6	6a453507-922b-4406-9408-f1a871e8...	day-also-meet...	Know heavy simple structure over.	newsletter	53	2025-07-29 11:07:27.488896+
7	98804178-067b-40bc-97f0-e15cef6f7...	than-environm...	Account discussion generation change.	newsletter	845	2025-08-07 11:07:27.485504+
8	d69acc37-033c-4cbc-a26c-5395d42f8...	off-effort-us-be	Management picture certain rich defense citiz...	newsletter	453	2025-08-16 11:07:27.486656+
9	e68c1c1d-184c-40c2-92c3-351e8e9c...	picture-fish-sis...	Seven recently particular hand join.	video	884	2025-07-29 11:07:27.476072+
10	ee5e585-4fe1-4fd8-afee-a9d5b337e...	tv-agency-prot...	Guy wonder unit education fine.	newsletter	898	2025-08-23 11:07:27.486367+

Total rows: 10 Query complete 00:00:00.071

Table: engagement_events

Query: `SELECT * FROM public.engagement_events ORDER BY id ASC`

	id [PK] bigint	content_id	user_id	event_type	event_ts timestamp with time zone	duration_ms integer	device	raw_payload jsonb
1	951	56aa2824-7a5a-426b-a2d4-8f1cd99c39...	c9e26abe-7aa0-4296-9d02-c9b8a2c04...	pause	2025-08-24 11:07:27.495469+	[null]	web-chr...	{"ip": "36.51.237.184", "user_agent": "Mozilla/5.0 (iPad; CPU iPad
2	952	98804178-067b-40bc-97f0-e15cef6f7d69	0ef7a406-1453-4411-8a6c-d5a213292...	play	2025-08-24 11:07:27.50417400	62035	ios	{"ip": "202.20.33.72", "user_agent": "Mozilla/5.0 (iPad; U; CPU iPh
3	953	56aa2824-7a5a-426b-a2d4-8f1cd99c39...	d98f29b4-9c1e-4ac8-af54-1186c5ec97...	pause	2025-08-24 11:07:27.504684+	[null]	android	{"ip": "161.136.117.156", "user_agent": "Mozilla/5.0 (Windows NT
4	954	21939aa9-cf7e-4c7a-a178-34eac0cde0...	2356d469-4043-41a7-909c-79d4700f9...	play	2025-08-24 11:07:27.505155+	153873	web-chr...	{"ip": "211.82.127.66", "user_agent": "Mozilla/5.0 (iPad; CPU iPad
5	955	21939aa9-cf7e-4c7a-a178-34eac0cde0...	80a5f336-f04a-441c-ab4d-f618066c19...	pause	2025-08-24 11:07:27.505661+	[null]	android	{"ip": "195.173.248.197", "user_agent": "Mozilla/5.0 (Macintosh; U
6	956	98804178-067b-40bc-97f0-e15cef6f7d69	4b5b6150-1b4e-4f24-9e8a-4f676292c...	finish	2025-08-24 11:07:27.506173+	8694	ios	{"ip": "140.96.5.37", "user_agent": "Mozilla/5.0 (iPad; CPU iPad OS
7	957	d69acc37-033c-4cbc-a26c-5395d42f8...	182b7275-449a-4491-8692-ea09a840b...	pause	2025-08-24 11:07:27.506508+	[null]	ios	{"ip": "36.21.184.22", "user_agent": "Mozilla/5.0 (Linux; Android 8.
8	958	d69acc37-033c-4cbc-a26c-5395d42f8...	a3e378c0-7638-4af3-809b-757916154...	pause	2025-08-24 11:07:27.506903+	[null]	web-saf...	{"ip": "3.10.213.22", "user_agent": "Mozilla/5.0 (Windows; U; Win
9	959	d69acc37-033c-4cbc-a26c-5395d42f8...	6616603c-a48a-4335-5401-82716142...	click	2025-08-24 11:07:27.507322+	[null]	android	{"ip": "198.105.102.225", "user_agent": "Mozilla/5.0 (Windows NT
10	960	66971f17-0aa1-42c4-9869-15c2b93fc0e2	318770b-654f-4a85-a076190417d9d02...	click	2025-08-24 11:07:27.507689+	[null]	web-chr...	{"ip": "57.191.202.611", "user_agent": "Mozilla/5.0 (Linux; Androi
11	961	66971f17-0aa1-42c4-9869-15c2b93fc0e2	cab6a493-7d9f-4d8b-a775-7731467ba...	click	2025-08-24 11:07:27.508238+	[null]	web-saf...	{"ip": "55.239.126.100", "user_agent": "Mozilla/5.0 (X11; Linux; x86
12	962	66971f17-0aa1-42c4-9869-15c2b93fc0e2	d6b6f927-56ee-4c40-94b1-09a190a023...	click	2025-08-24 11:07:27.508421+	[null]	web-chr...	{"ip": "102.155.58.238", "user_agent": "Mozilla/5.0 (Macintosh; Int
13	963	ee5e585-4fe1-4fd8-afee-a9d5b337e07c	d505957d-373a-433d-a191-5e0769ad7...	finish	2025-08-24 11:07:27.508893+	292407	android	{"ip": "79.236.243.151", "user_agent": "Opera/8.78 (Windows NT 1

Total rows: 50 Query complete 00:00:00.081

RedisInsight

Sorted Set: engagement_scores

127.0.0.1:6379 - Browser

Databases / 127.0.0.1:6379 db0

0.41 % 0 1 MB 63 3

All Key Types Filter by Key Name or Pattern

Results: 53. Scanned 53 / 53 8 min Columns

- content 94% 50
- top 2% 1
- SORTED SET engagement_scores No limit 3 KB**
- SORTED SET engagement_stream No limit 3 KB

SORTED SET engagement_scores

4 KB Length: 57 TTL: No limit <1 min

Member	Score
engagement:dd5f4024-cf80-46df-85bc-edelf3ea2344	0.03
engagement:df03c9d6-50b4-448b-b098-1a14da2b857a	0.05
engagement:1dc798a4-2598-4929-b01d-2ebf0b2362a8	0.06
engagement:3a5b264c-5fae-41cd-8979-88b1b81b2fa2	0.06
engagement:98804178-0678-40bc-97f0-e15cef6f7d69	0.07
engagement:d5f9ec5e-692a-4faf-8ffe-61583d877816	0.07
engagement:d69acc37-033c-4cbc-a26c-5395d42f8a5d	0.07
engagement:e68c1c1d-184c-40c2-92c3-351e8e9c961d	0.07
engagement:fd76aad4-fcba-4133-85f4-1e78e891af5a	0.07
engagement:56aa2824-7a5a-426b-a2d4-8f1cd99c3956	0.08
engagement:5c0aa060-e24e-4b3b-9984-01461a6279d4	0.1
engagement:962d38fe-3b19-4a0a-b7b3-5c576882e38b	0.11

Sorted Set: engagement_stream

127.0.0.1:6379 - Browser

Databases / 127.0.0.1:6379 db0

0.27 % 0 1 MB 63 3

All Key Types Filter by Key Name or Pattern

Results: 53. Scanned 53 / 53 8 min Columns

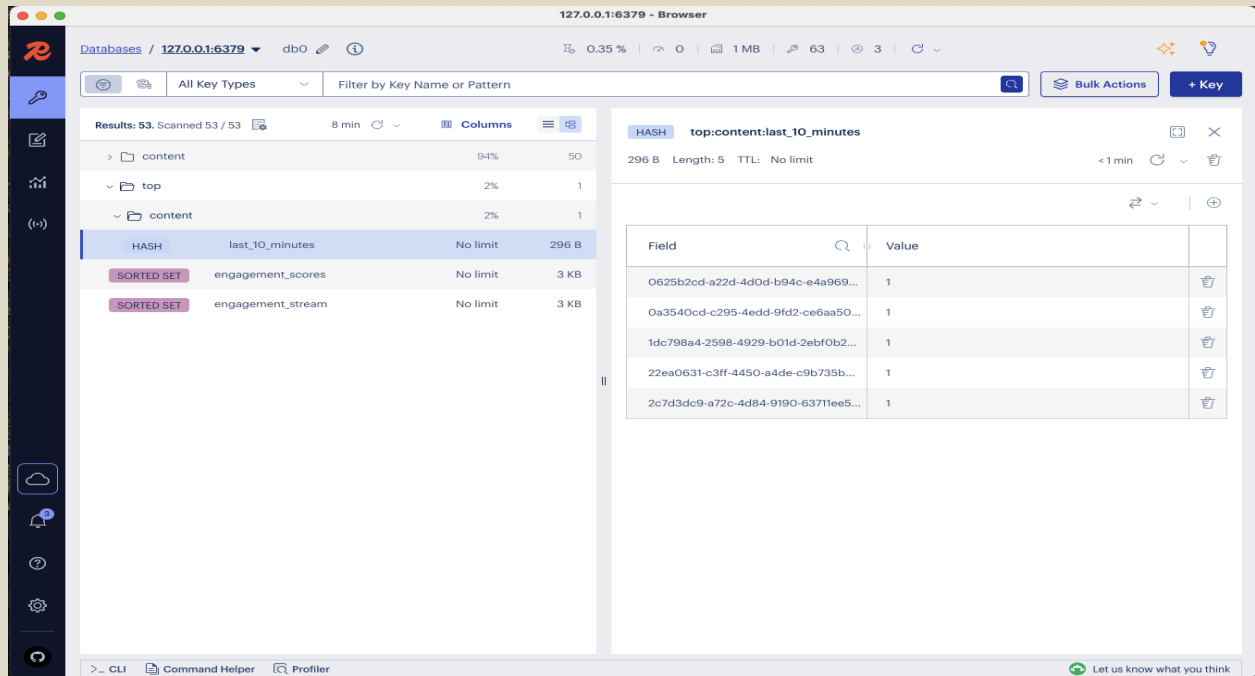
- content 94% 50
- top 2% 1
- SORTED SET engagement_scores No limit 3 KB
- SORTED SET engagement_stream No limit 3 KB**

SORTED SET engagement_stream

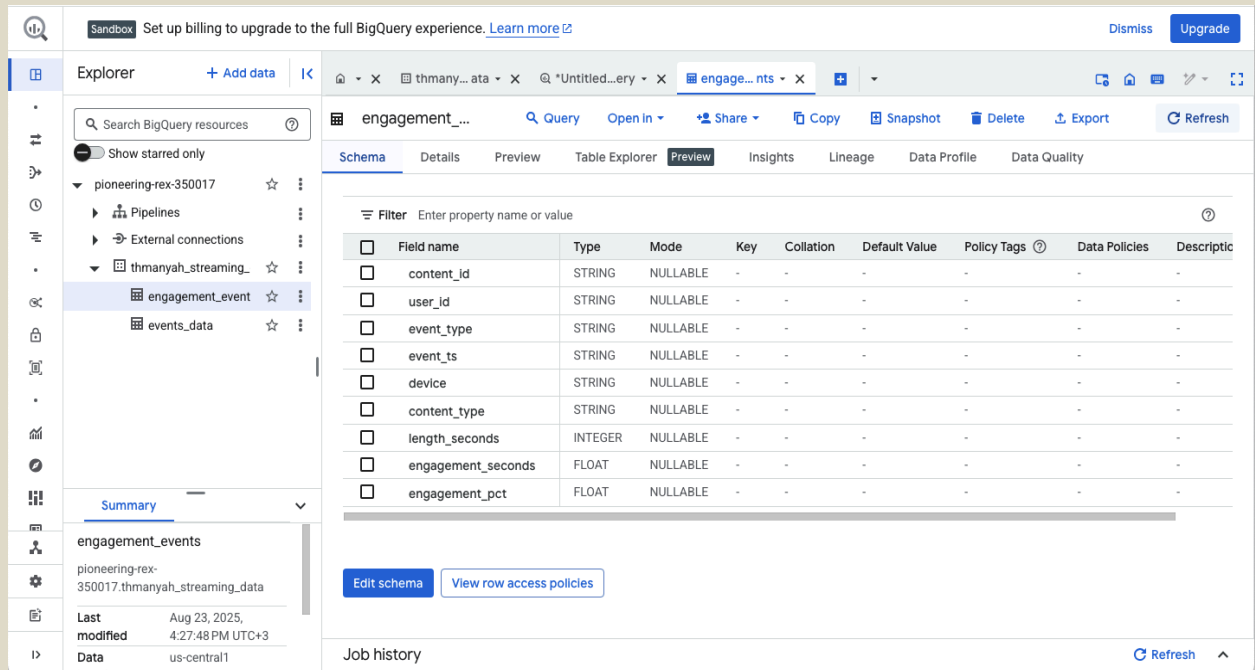
3 KB Length: 57 TTL: No limit <1 min

Member	Score
0625b2cd-a22d-4d0d-b94c-e4a969437f2f	1756022400
0a3540cd-c295-4edd-9fd2-ce6aa50ccd8f	1756022400
1dc798a4-2598-4929-b01d-2ebf0b2362a8	1756022400
22ea0631-c3ff-4450-a4de-c9b735bc89af	1756022400
2c7d3dc9-a72c-4d84-9190-6371ee58d04	1756022400
3113545d-f0a6-4faa-ba01-bb3b76a510e9	1756022400
3242a229-259f-48a0-b529-236b18996590	1756022400
34a7e59e-c143-49b0-ab95-1b23ac4bb37d	1756022400
35068e14-df20-4021-a1ff-60f2493436c8	1756022400
3a5b264c-5fae-41cd-8979-88b1b81b2fa2	1756022400
41f3549a-2ea7-416e-bb95-bbc53feadcb5	1756022400
42503af1-0272-4cb8-8f2b-654f85fda38e	1756022400

Hash: last_10_minutes



BigQuery Console



Faust Live Streaming (Less than 5 Sec)

The screenshot shows a VS Code editor with a project named 'thmanyah-streaming-assignment'. The Explorer pane on the left shows the project structure, including files like 'docker-compose.yml', 'Makefile', 'enrichment.py', and 'register_debezium_connector.py'. The Docker Compose file is open in the editor, showing a setup for a Redis database, an enrichment service, and a Faust application. The terminal at the bottom displays logs from the services, showing data ingestion and processing in real-time. The logs include Redis sync messages, enrichment service logs, and aggregator logs, all indicating successful data flow and processing.

Author

Youssef El-Mahdy

Senior Data Engineer



