

Environment Setup

- **Cd foldername**: will take us inside the folder
- **python3 -m venv venv**: creates virtual environment (venv) inside the folder
- Manually create folder infra referring to the infrastructure of the docker and created a docker compose yml file inside the folder using command **touch docker-compose**
- that has multiple software using containers and we can run them all at once.
- Open the docker desktop and run **docker compose up -d**

```
[simmi@Anushkas-MacBook-Air Real_time_stock_market % touch docker-compose.yml
[simmi@Anushkas-MacBook-Air Real_time_stock_market % cd infra
[simmi@Anushkas-MacBook-Air infra % docker compose up -d
WARN[0000] /Users/simmi/Real_time_stock_market/infra/docker-compose.yml: the attribute
`version` is obsolete, it will be ignored, please remove it to avoid potential confus
ion
[+] Running 75/76
  ✓ kafka Pulled                                324.0s
  ✓ minio Pulled                                134.5s
  ✓ zookeeper Pulled                            324.0s
  ✓ airflow-webserver Pulled                    327.3s
  ✓ airflow-scheduler Pulled                    327.3s
  ✓ postgres Pulled                             251.4s
  ✓ kafdrop Pulled                             268.2s

[+] Running 8/9
  ✓ Network infra_default Created               0.1s
  ✓ Volume infra_postgres_data Created          0.0s
  ! Container airflow-postgres Starting         1.5s
  ✓ Container zookeeper Started                1.5s
  ✓ Container minio Started                    1.5s
  ✓ Container airflow-scheduler Created         0.2s
  ✓ Container kafka Created                   0.1s
  ✓ Container kafdrop Created                  0.1s
  ✓ Container airflow-daemon Created
Error response from daemon: ports are not available: exposing port TCP 0.0.0.0:5432 ->
127.0.0.1:0: listen tcp 0.0.0.0:5432: bind: address already in use
simmi@Anushkas-MacBook-Air infra %
```

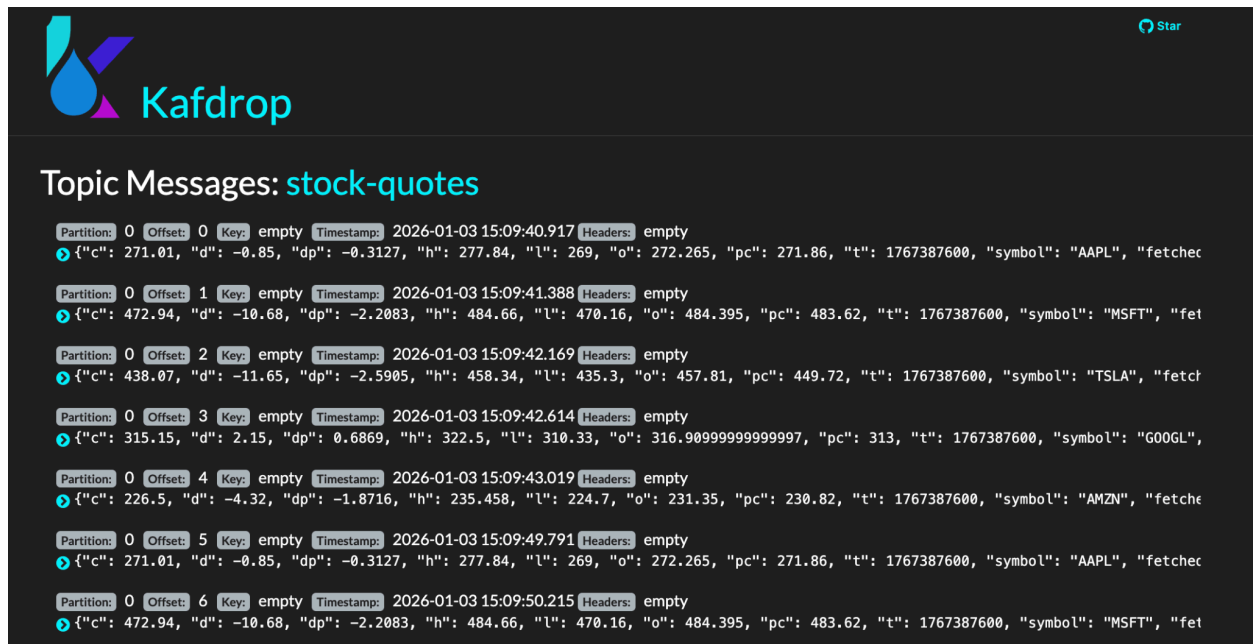
infra	-	-	-	0.3%	1 minute ago	🔍	🔍	🔍	🗑️
airflow-postgres	e262ce5f5984	postgres:15	5432:5432	0%		🔍	▶️	🔍	🗑️
zookeeper	3fec37a8c945	confluentinc/zookeeper:3.8.0	2181:2181	0.19%	1 minute ago	🔍	🔍	🔍	🗑️
minio	6859b4efebf4	minio/minio:RELEASE.2023-08-23T19-17-44Z	9002:9000	0.11%	1 minute ago	🔍	🔍	🔍	🗑️
airflow-webserver	55aa9e920089	apache/airflow:2.5.0	8080:8080	0%		🔍	▶️	🔍	🗑️
kafdrop	871cd46bd7fa	obsidiandry/kafdrop:0.11.0	9000:9000	0%		🔍	▶️	🔍	🗑️
airflow-scheduler	4406776a7c4d	apache/airflow:2.5.0		0%		🔍	▶️	🔍	🗑️
kafka	4d7face23cb5	confluentinc/kafka:7.4.0	29092:29092	0%		🔍	▶️	🔍	🗑️

- Now scheduler and webserver will restart frequently to make them run continuously: **docker-compose run --rm airflow-webserver airflow db init** and **docker-compose up -d airflow-scheduler airflow-webserver**
- Airflow credentials: **docker-compose exec airflow-webserver airflow users create --username simmi --firstname jay --lastname k --role Admin --email anushkak1202@gmail.com --password password123**
- Click on the kafdrop port in the docker, this is kafka UI
- Minio port and credentials are in docker-compose.yml file

- Airflow webserver port with credentials created above.
- Snowflake credentials- username: ,password:
- Create a kafka topic names stock_quotes with 3 partitions
- `source ~/Real_time_stock_market/venv/bin/activate` : Activate virtual environment
- `pip install -r requirements.txt` :will install all the required tools

1.Streaming

- Run producer file and messages can be viewed in kafka

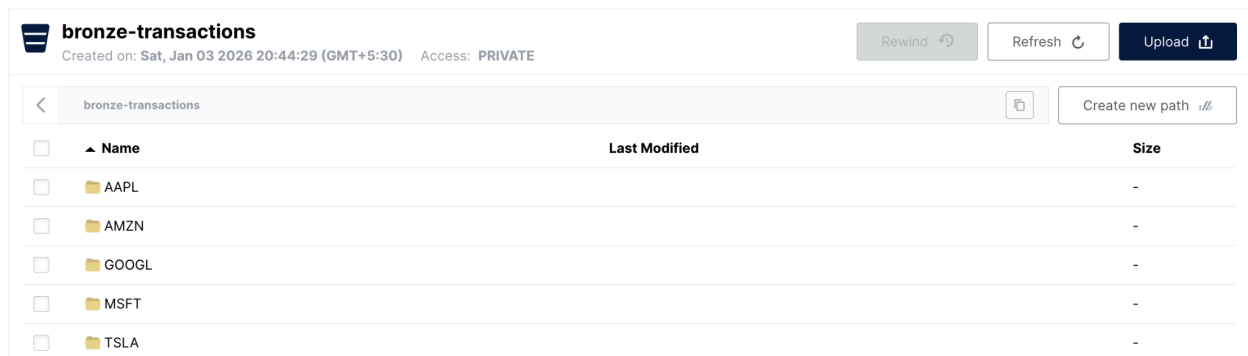


The image shows the Kafdrop web interface for viewing Kafka messages. The title is "Topic Messages: stock-quotes". It displays a list of messages with their metadata and content. Each message entry includes Partition, Offset, Key, Timestamp, Headers, and the message body. The messages are for the "stock-quotes" topic and contain stock price data for AAPL, MSFT, TSLA, and GOOGL.

Partition	Offset	Key	Timestamp	Headers	Message Body
0	0	empty	2026-01-03 15:09:40.917	empty	{"c": 271.01, "d": -0.85, "dp": -0.3127, "h": 277.84, "l": 269, "o": 272.265, "pc": 271.86, "t": 1767387600, "symbol": "AAPL", "fetche
0	1	empty	2026-01-03 15:09:41.388	empty	{"c": 472.94, "d": -10.68, "dp": -2.2083, "h": 484.66, "l": 470.16, "o": 484.395, "pc": 483.62, "t": 1767387600, "symbol": "MSFT", "fet
0	2	empty	2026-01-03 15:09:42.169	empty	{"c": 438.07, "d": -11.65, "dp": -2.5905, "h": 458.34, "l": 435.3, "o": 457.81, "pc": 449.72, "t": 1767387600, "symbol": "TSLA", "fetcl
0	3	empty	2026-01-03 15:09:42.614	empty	{"c": 315.15, "d": 2.15, "dp": 0.6869, "h": 322.5, "l": 310.33, "o": 316.90999999999997, "pc": 313, "t": 1767387600, "symbol": "GOOGL",
0	4	empty	2026-01-03 15:09:43.019	empty	{"c": 226.5, "d": -4.32, "dp": -1.8716, "h": 235.458, "l": 224.7, "o": 231.35, "pc": 230.82, "t": 1767387600, "symbol": "AMZN", "fetche
0	5	empty	2026-01-03 15:09:49.791	empty	{"c": 271.01, "d": -0.85, "dp": -0.3127, "h": 277.84, "l": 269, "o": 272.265, "pc": 271.86, "t": 1767387600, "symbol": "AAPL", "fetche
0	6	empty	2026-01-03 15:09:50.215	empty	{"c": 472.94, "d": -10.68, "dp": -2.2083, "h": 484.66, "l": 470.16, "o": 484.395, "pc": 483.62, "t": 1767387600, "symbol": "MSFT", "fet

2.Storage

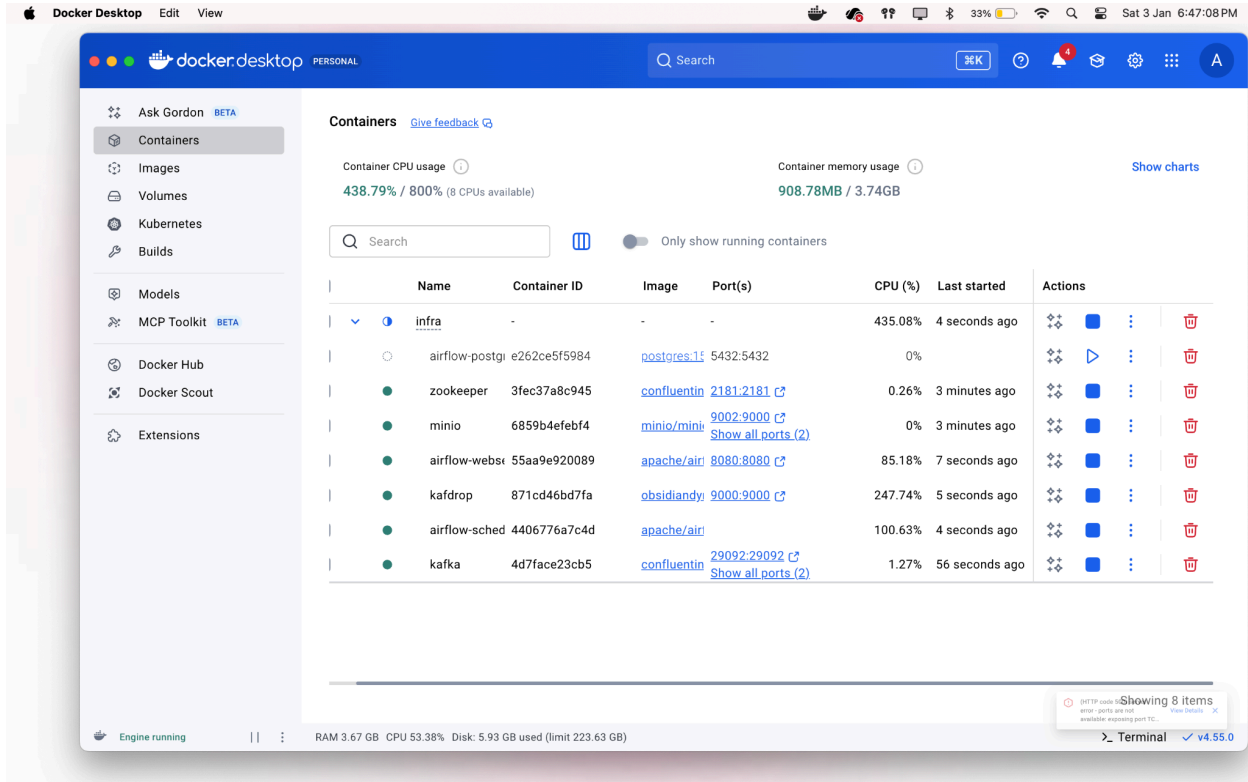
- Similarly, Run consumer file



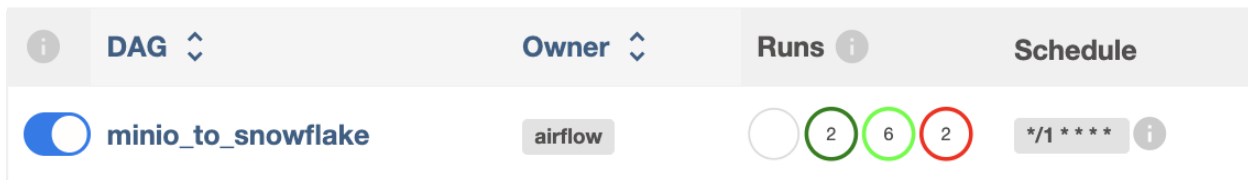
The image shows the "bronze-transactions" storage interface. It displays a list of files and folders. The interface includes a search bar, a "Create new path" button, and a table of files with columns for Name, Last Modified, and Size. The files listed are AAPL, AMZN, GOOGL, MSFT, and TSLA.

Name	Last Modified	Size
AAPL	-	-
AMZN	-	-
GOOGL	-	-
MSFT	-	-
TSLA	-	-



3. Snowflake



- Bronze stage: raw data



- Dbt run: `dbt init dbt_stocks` and `dbt run`
- You will see views of bronze, silver and gold layer in snowflake

Ingestion

Transformation

AI & ML

Monitoring

Marketplace

Horizon Catalog

Catalog

Data sharing

Governance & security

Manage

Compute

Postgres

Admin

\$400 credits left

Trial ends in 30 days

Upgrade

Database Explorer

HORIZON CATALOG

Databases

Search

SNOWFLAKE

SNOWFLAKE_LEARNING_DB

SNOWFLAKE_SAMPLE_DATA

STOCKS_MDS

COMMON

Tables

BRONZE_STOCK_QUOTES_RAW

MY_FIRST_DBT_MODEL

Views

BRONZE_STG_STOCK_QUOTES

GOLD_CANDLESTICK

GOLD_KPI

GOLD_TREECHART

MY_SECOND_DBT_MODEL

SILVER_CLEAN_STOCK_QUOTES

INFORMATION_SCHEMA

PUBLIC

Databases

+ Database

4 Databases

SearchSource All

NAME ↑	SOURCE	OWNER	CREATED	
SNOWFLAKE	Share	—	4 hours ago	...
SNOWFLAKE_LEARNING...	Local	ACCOUNTADMIN	4 hours ago	...
SNOWFLAKE_SAMPLE_D...	Share	ACCOUNTADMIN	4 hours ago	...
STOCKS_MDS	Local	ACCOUNTADMIN	1 hour ago	...