

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`version` is obsolete, it will be ignored, please remove it to avoid potential confus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
✓ kafdrop Pulled
✓ minio Pulled
✓ zookeeper Pulled
✓ airflow-webserver Pulled
✓ airflow-scheduler Pulled
✓ postgres Pulled
✓ kafdrop Pulled
324.0s
134.5s
324.0s
327.3s
327.3s
251.4s
268.2s

[+] Running 8/9
✓ Network infra_default Created
✓ Volume infra_postgres_data Created
✓ Container airflow-postgres Starting
✓ Container zookeeper Started
✓ Container minio Started
✓ Container airflow-scheduler Created
✓ Container kafka Created
✓ Container kafdrop Created
✓ Container airflow-webserver Created
0.1s
0.0s
1.5s
1.5s
1.5s
0.1s
0.1s
0.1s
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	confluentin	2181:2181 ↗	0.19%	1 minute ago	⋮	⋮	⋮
	minio	6859b4efebf4	minio/minio	9002:9000 ↗	0.11%	1 minute ago	⋮	⋮	⋮
	airflow-webserver	55aa9e920089	apache/airflow	8080:8080	0%		⋮	⋮	⋮
	kafdrop	871cd46bd7fa	obsidiandy	9000:9000	0%		⋮	⋮	⋮
	airflow-scheduler	4406776a7c4d	apache/airflow		0%		⋮	⋮	⋮
	kafka	4d7face23cb5	confluentin	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 screenshot shows the Kafdrop interface. At the top, there's a logo and a 'Star' button. Below it, the title 'Topic Messages: stock-quotes' is displayed. The main area lists several Kafka messages in a table format. Each message includes fields: Partition, Offset, Key, Headers, Timestamp, and the actual JSON message content. The messages are as follows:

Partition	Offset	Key	Headers	Timestamp	Message Content
0	0	empty	empty	2026-01-03 15:09:40.917	{"c": 271.01, "d": -0.85, "dp": -0.3127, "h": 277.84, "l": 269, "o": 272.265, "pc": 271.86, "t": 1767387600, "symbol": "AAPL", "fetcher": "kafka"}
0	1	empty	empty	2026-01-03 15:09:41.388	{"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", "fetcher": "kafka"}
0	2	empty	empty	2026-01-03 15:09:42.169	{"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", "fetcher": "kafka"}
0	3	empty	empty	2026-01-03 15:09:42.614	{"c": 315.15, "d": 2.15, "dp": 0.6869, "h": 322.5, "l": 310.33, "o": 316.90999999999997, "pc": 313, "t": 1767387600, "symbol": "GOOGL", "fetcher": "kafka"}
0	4	empty	empty	2026-01-03 15:09:43.019	{"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", "fetcher": "kafka"}
0	5	empty	empty	2026-01-03 15:09:49.791	{"c": 271.01, "d": -0.85, "dp": -0.3127, "h": 277.84, "l": 269, "o": 272.265, "pc": 271.86, "t": 1767387600, "symbol": "AAPL", "fetcher": "kafka"}
0	6	empty	empty	2026-01-03 15:09:50.215	{"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", "fetcher": "kafka"}

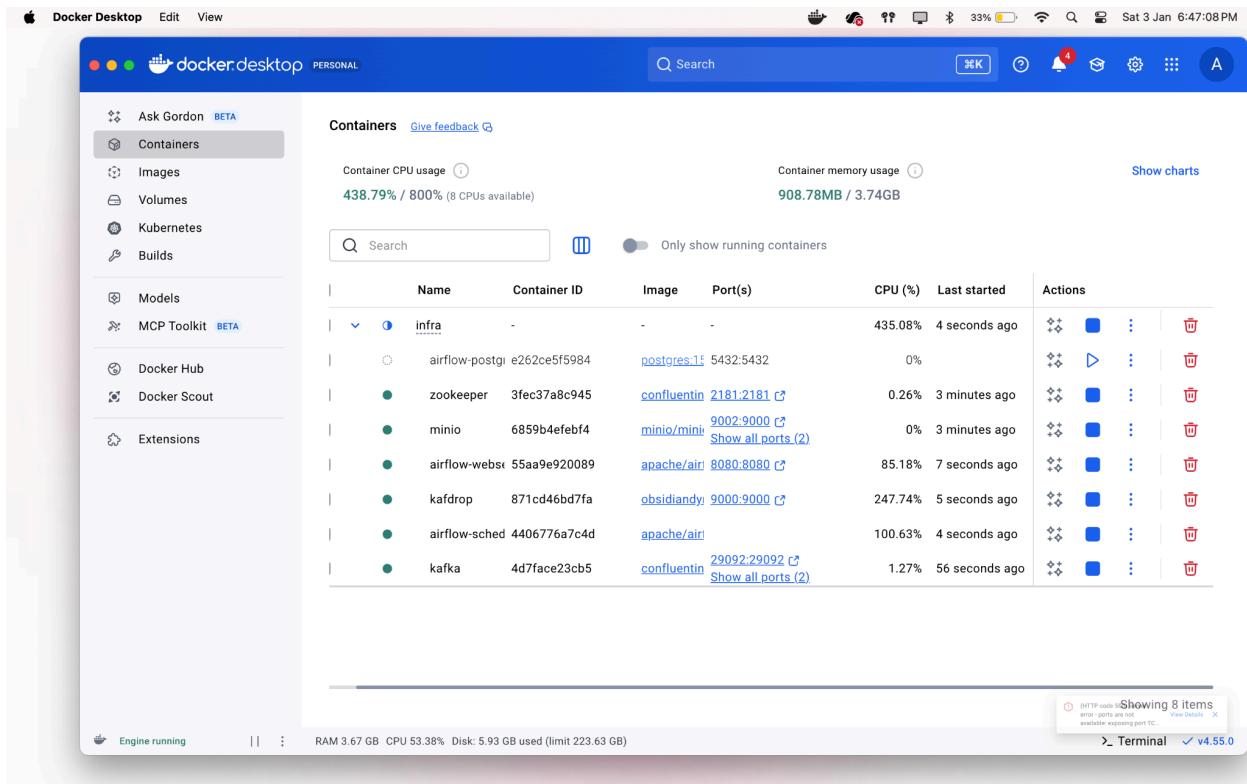
2. Storage

- Similarly, Run consumer file

The screenshot shows a file browser interface. At the top, there's a header bar with a folder icon, the path 'bronze-transactions', a 'Rewind' button, a 'Refresh' button, and an 'Upload' button. Below the header, there's a search bar with the placeholder 'bronze-transactions'. The main area displays a list of files and folders. There are five entries, each represented by a small icon (yellow folder), a checkbox, a name ('AAPL', 'AMZN', 'GOOGL', 'MSFT', 'TSLA'), and a 'Last Modified' column. All entries have a size of '-'.

	Name	Last Modified	Size
<input type="checkbox"/>	AAPL		-
<input type="checkbox"/>	AMZN		-
<input type="checkbox"/>	GOOGL		-
<input type="checkbox"/>	MSFT		-
<input type="checkbox"/>	TSLA		-

3. Snowflake



- Bronze stage: raw data

The screenshot shows the Airflow web interface. At the top, there are dropdown menus for 'DAG' (set to 'minio_to_snowflake'), 'Owner' (set to 'airflow'), 'Runs' (with a help icon), and 'Schedule'. Below this, the DAG 'minio_to_snowflake' is listed with a blue switch icon. To its right are three colored circles: grey (2), green (6), and red (2). Further right are run statistics: '/1 * * * *' and a help icon.

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

The screenshot shows the Database Explorer in the Snowflake Horizon Catalog. The left sidebar includes links for Projects, Ingestion, Transformation, AI & ML, Monitoring, Marketplace, Horizon Catalog, Catalog (which is selected), Data sharing, Governance & security, Manage, Compute, Postgres, and Admin. A message at the bottom indicates \$400 credits left, trial ends in 30 days, and an Upgrade button.

Database Explorer
HORIZON CATALOG

Databases

+ Database

Search Source All ⌂

4 Databases

NAME ↑	SOURCE	OWNER	CREATED
SNOWFLAKE	Share	—	4 hours ago
SNOWFLAKE_LEARNING_DB	Local	ACCOUNTADMIN	4 hours ago
SNOWFLAKE_SAMPLE_DATA	Share	ACCOUNTADMIN	4 hours ago
STOCKS_MDS	Local	ACCOUNTADMIN	1 hour ago

Databases

NAME ↑ SOURCE OWNER CREATED

SNOWFLAKE Share — 4 hours ago

SNOWFLAKE_LEARNING_DB Local ACCOUNTADMIN 4 hours ago

SNOWFLAKE_SAMPLE_DATA Share ACCOUNTADMIN 4 hours ago

STOCKS_MDS Local ACCOUNTADMIN 1 hour ago

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