Materialize

Streaming Analytics with dbt: The Fun Parts

Dennis Hume
Developer Experience Engineer

Agenda

- Setup
- Overview of Materialize
- Our dbt Project
- Workshop

Setup

Gitpod

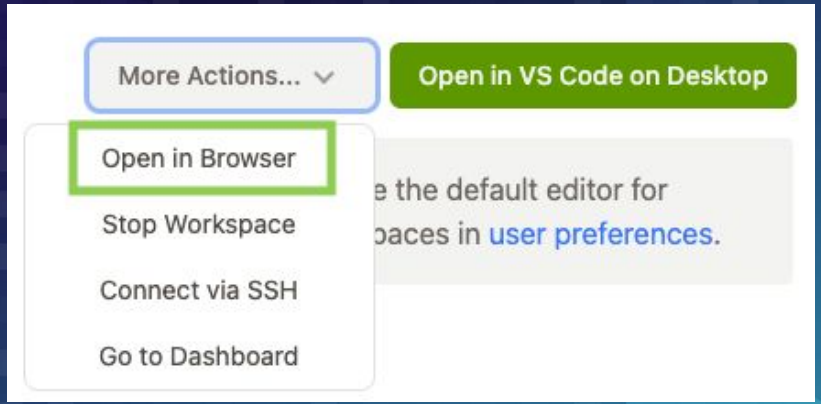
During this workshop we will use Gitpod. This will drop you into a preconfigured instance running in the cloud that has all the dependencies we need:

- **dbt-materialize adapter**
- **psql-client**

To get started use the link in the workshop repo.

Gitpod

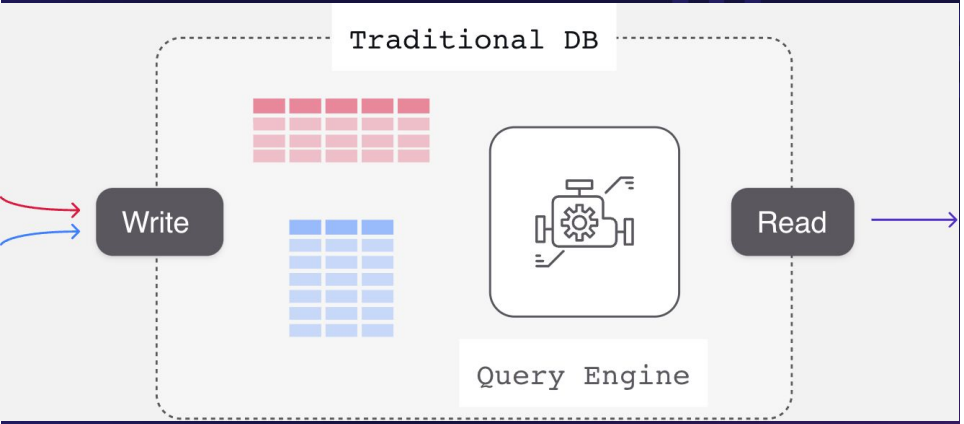
You can chose to open the repo in your browser which will give you an interactive terminal where dbt and psql will be installed.



https://github.com/MaterializeInc/mz-data-council-2023

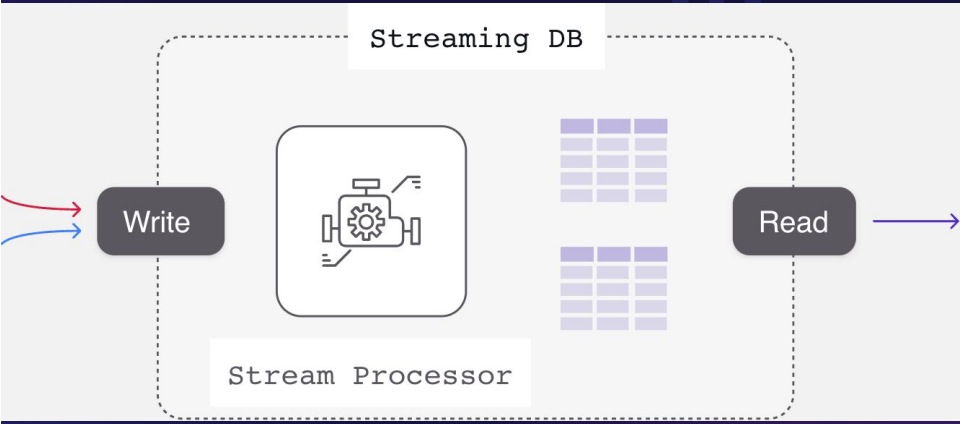
If you would like a trial account
Swing by our booth

Materialize



Traditional Database

In traditional DBs, data is stored in tables matching the structure of the writes (inserts, updates) and all the computation work happens on read queries (selects).



Streaming Database

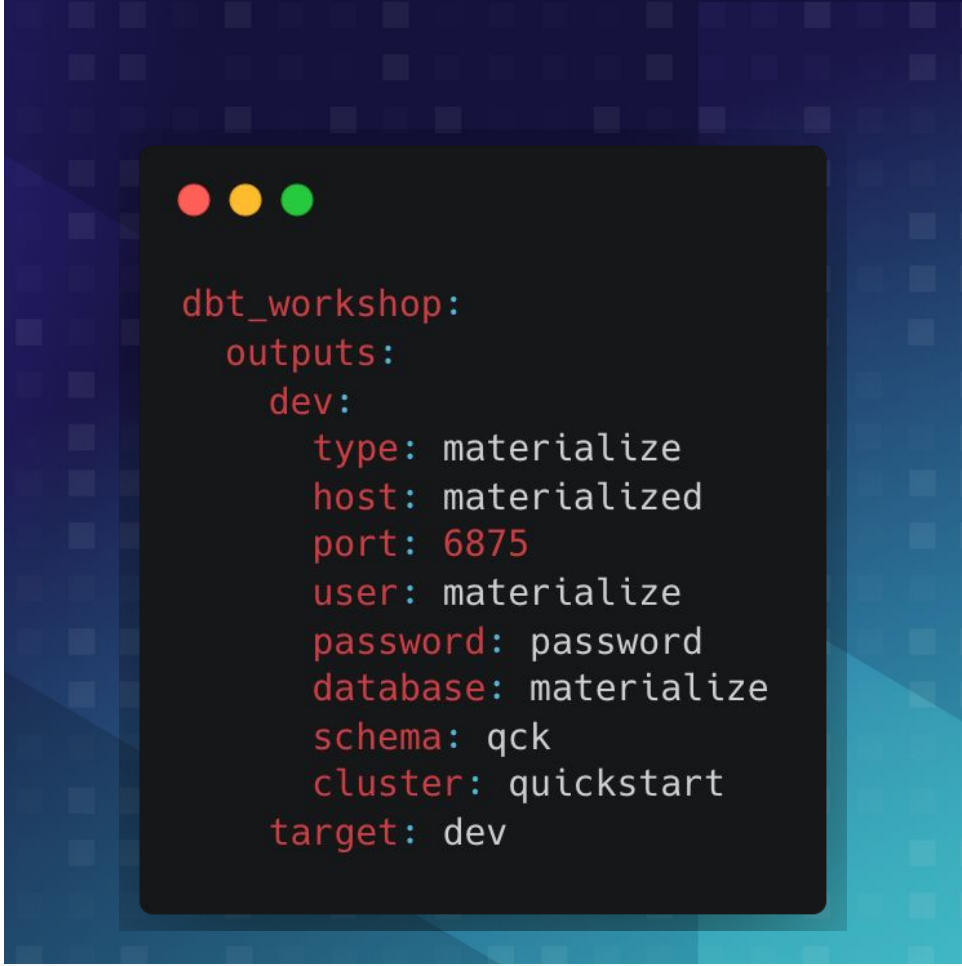
A streaming database moves the work to the write side: Instead of just storing data on writes, a streaming database ask for the queries upfront (in the form of Materialized Views) and incremental updates results as input data arrives.

Our dbt Project

Connection

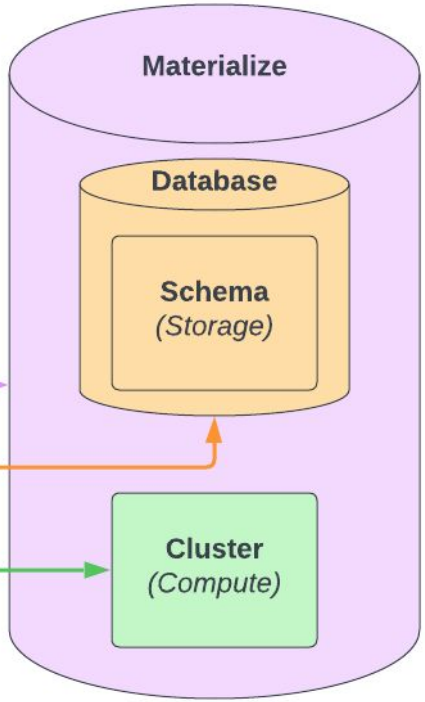
profiles.yml

All our connection details are contained in our *profiles.yml.* This defines how we connect to Materialize and our specific database, schema and cluster.



Connection

```
dbt_workshop:
  outputs:
    dev:
      type: materialize
      host: materialized
      port: 6875
      user: materialize
      password: password
      database: materialize
      schema: qck
      cluster: quickstart
  target: dev
```



Storage and Compute

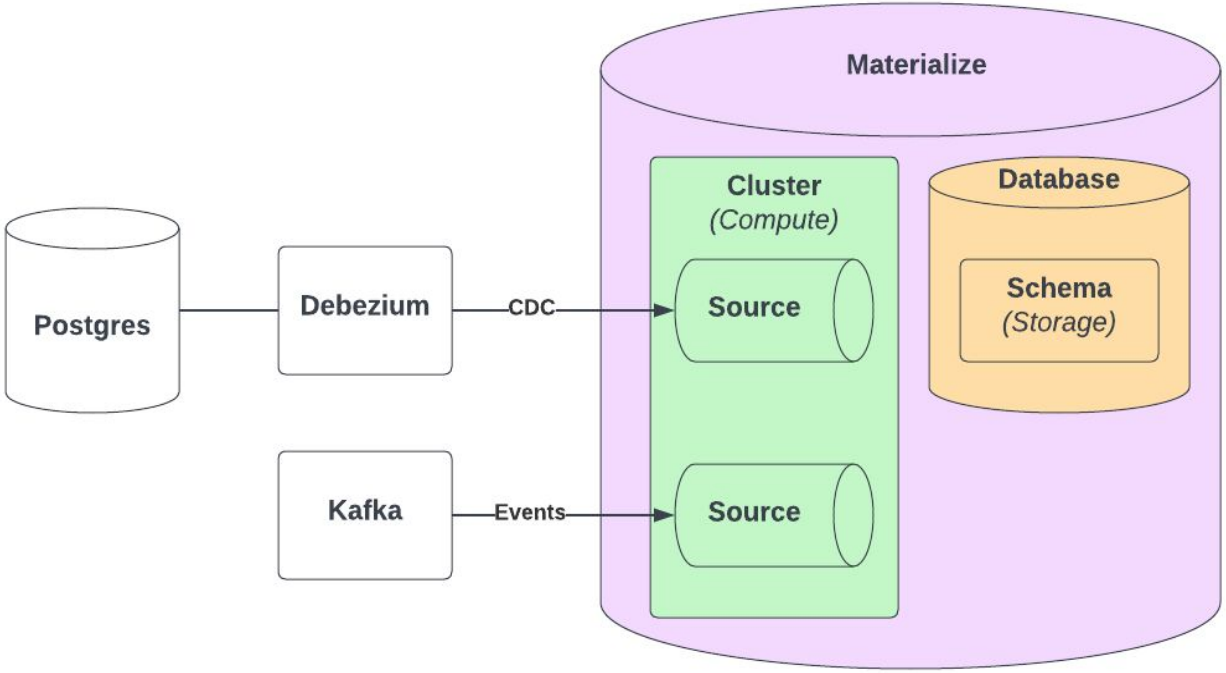
Database and Schema

Database and schema are the namespaces containing our persisted data within the storage layer

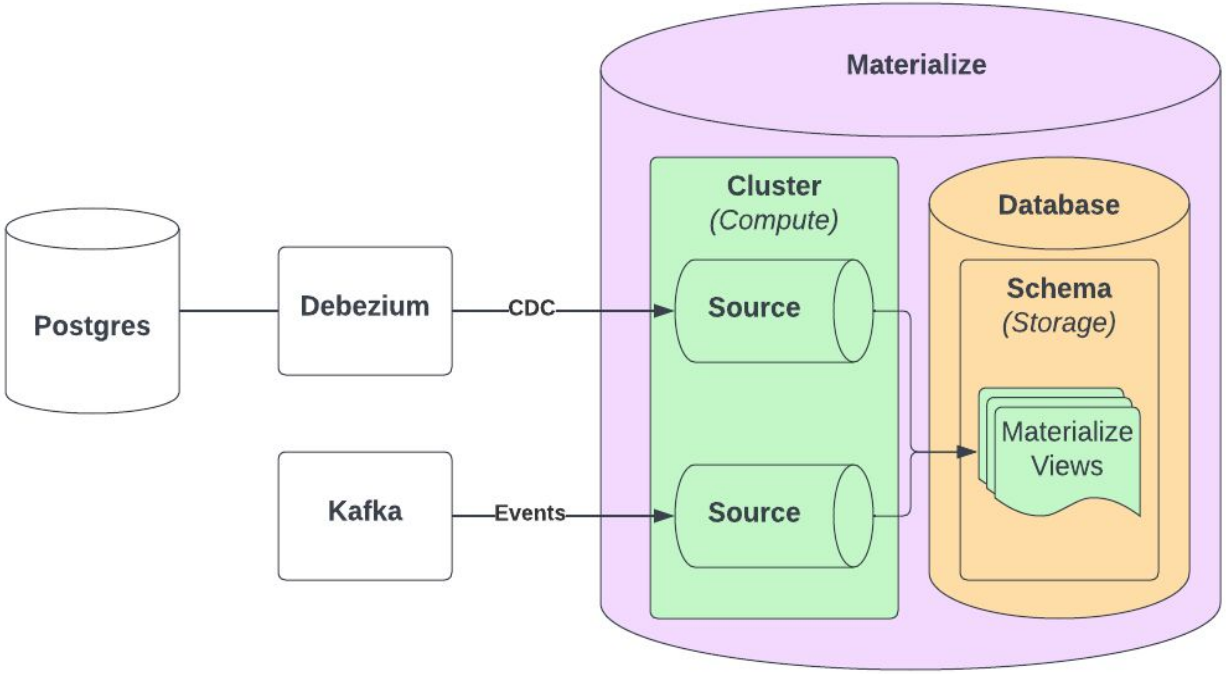
Cluster and Replicas

Clusters describe logical compute resources and replicas allocate physical compute resources

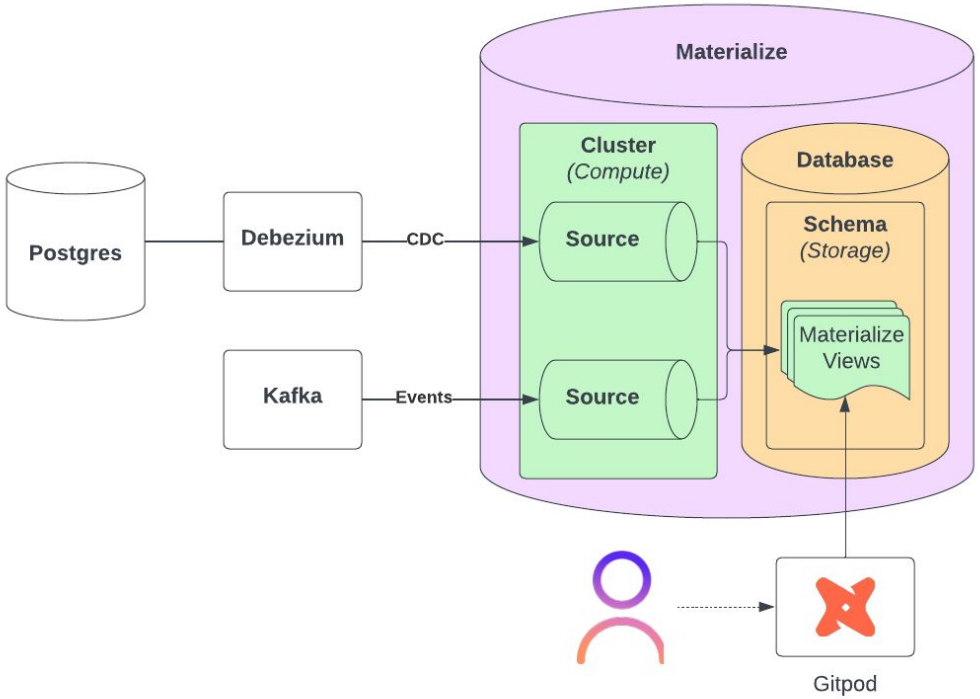
Sources


Materialize
Cluster
(Compute)
Database
Schema
(Storage)
Postgres
Debezium
CDC
Source
Kafka
Events
Source

Materialized Views



Gitpod / dbt


Materialize
Cluster
(Compute)
Database
Schema
(Storage)
Postgres
Debezium
CDC
Source
Materialize
Views
Kafka
Events
Source
Gitpod

Workshop

Questions?

Thank you