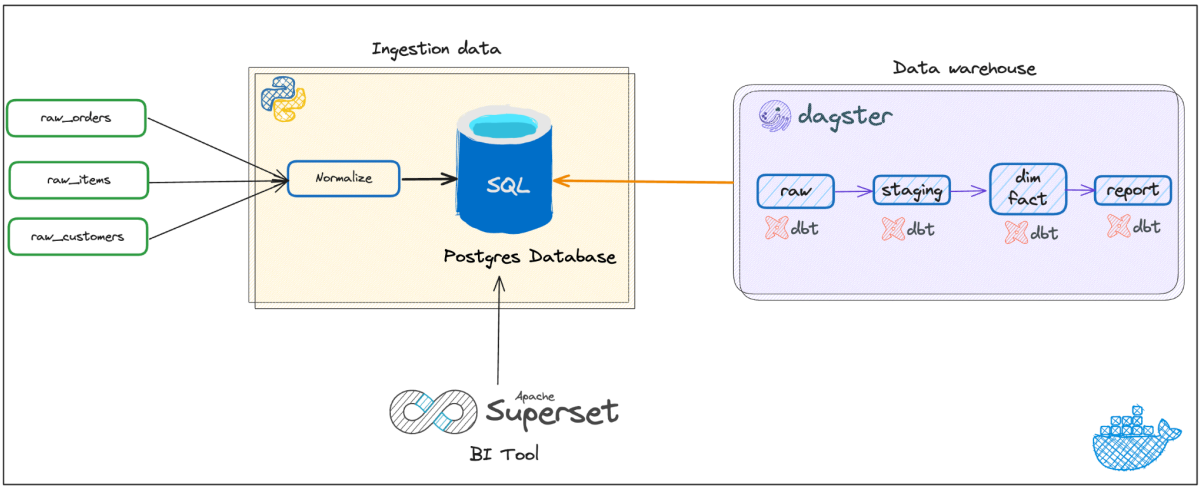
Solution - ViTran

Architecture



- **Tech stack**:
  - Python
  - Dagster
  - DBT
  - Postgres
  - Docker && Docker-compose

- **Pipeline**:
  - Ingestion raw data to Postgres by Python.
  - Use dbt to extract raw data to build data warehouse and report.
  - Manage dbt dags by Dagster.
  - Use Superset to visualize data.
- **How to run?** Please checkout README.md file.
- **Why choose Dagster to manage dbt and Superset for visualizing?**

Choosing Dagster for managing dbt and Superset for visualization can offer several advantages:

1. Modular Data Pipeline Management: Dagster provides a framework for building modular data pipelines, allowing you to define, schedule, and monitor your dbt runs alongside other data processing tasks. With Dagster, you can easily integrate dbt into your broader data pipeline, orchestrating dependencies and ensuring proper execution.
2. Data Lineage and Dependency Management: Dagster tracks data lineage and dependencies between different components of your data pipeline. This means you can understand how changes to your dbt models impact downstream processes and make informed decisions about data transformations and analysis.

On the visualization side, Superset offers several benefits:

1. Rich Visualizations: Superset provides a wide range of visualization options, including charts, graphs, and dashboards, allowing you to create rich and interactive visualizations to explore and analyze your data.
2. Self-Service Analytics: Superset empowers users to explore and analyze data on their own, without relying on data engineers or analysts to generate reports. Users can create their own dashboards and visualizations, democratizing access to data across the organization.
3. Integration with Data Sources: Superset supports integration with a variety of data sources, including relational databases, data warehouses, and big data platforms. This allows you to visualize data from multiple sources in a single interface, making it easier to derive insights and correlations.
4. Security and Access Control: Superset offers robust security features, including role-based access control and integration with authentication providers like LDAP and OAuth. This ensures that sensitive data is protected and that access to data and visualizations is controlled and audited.

Result

```
2024-03-13 22:58:10 [info     ] Df length: 10000
2024-03-13 22:58:11 [info     ] Ingest to raw_customers successfully.
2024-03-13 22:58:11 [info     ] Start ingesting to table: raw_items
2024-03-13 22:58:11 [info     ] shape: (5, 13)
```

ORDER_ID	ID	GIFT_CARD	GRAMS	…	TITLE	TOTAL_DISCOUNT	PRE_TAX_PRICE	HAS_MESSAGE
i64	i64	bool	i64		str	f64	f64	i64
196682711066	407559864346	false	34	…	Japanese Maple	0.0	13.0	0
5345444941	9575806221	false	34	…	Heart Bench	0.0	null	0
165021876250	339999293466	false	40	…	Angel 2.0	0.0	13.0	0
5002792013	8930786957	false	54	…	Love Turtle	0.0	null	0
5756484429	10319954061	false	40	…	The Beatles Abbey Road	0.0	null	0

Fig. Example ingest a raw table

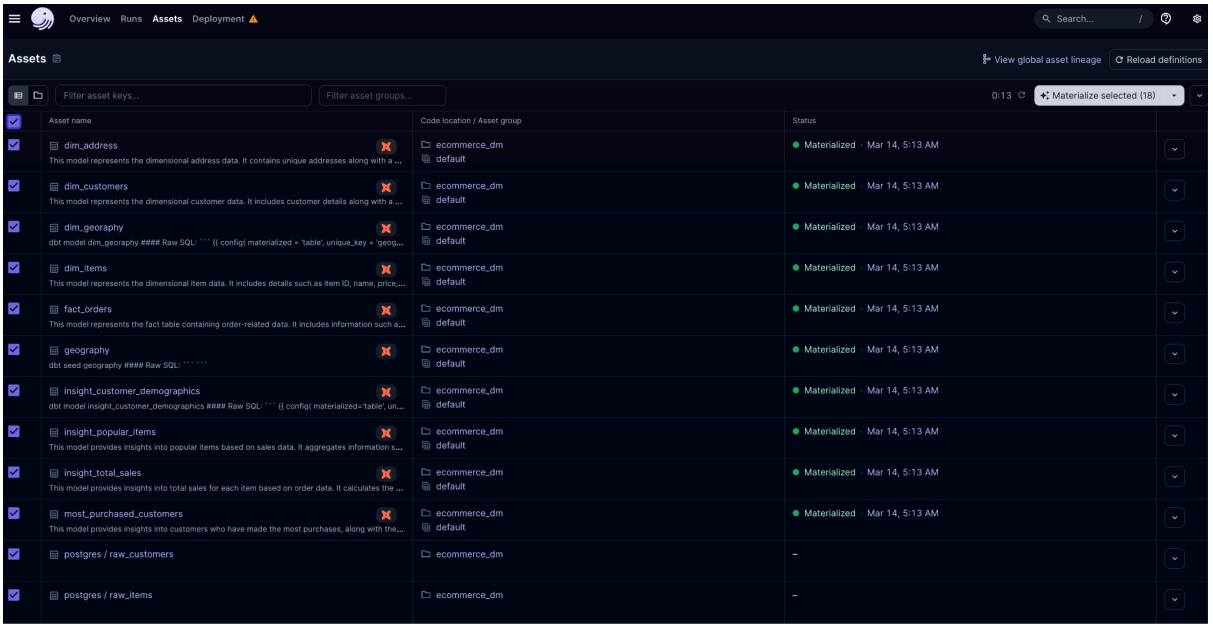


Fig. Dagster assets

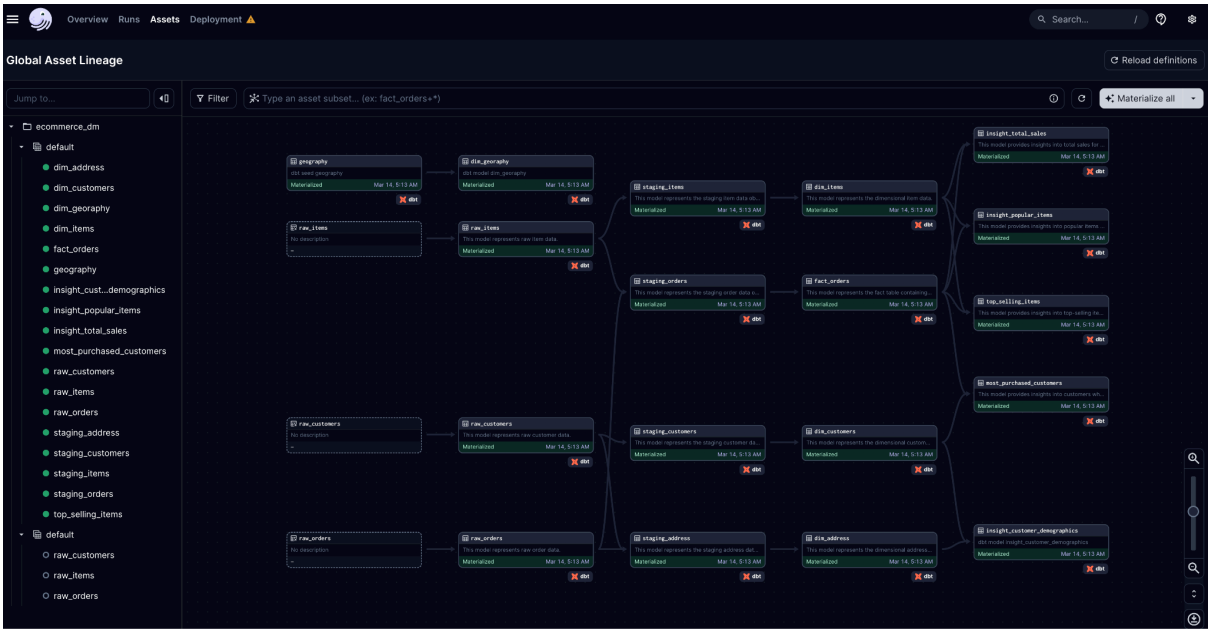
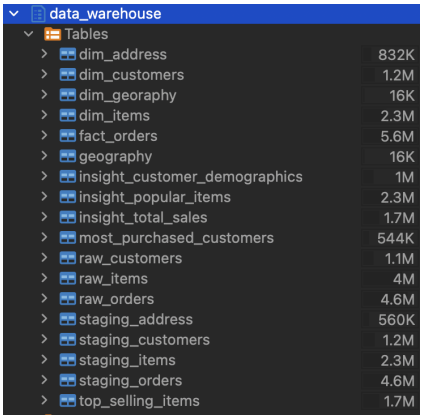


Fig. Dagster orchestration dbt

Ater run materialize all dbt dags by Dagster:



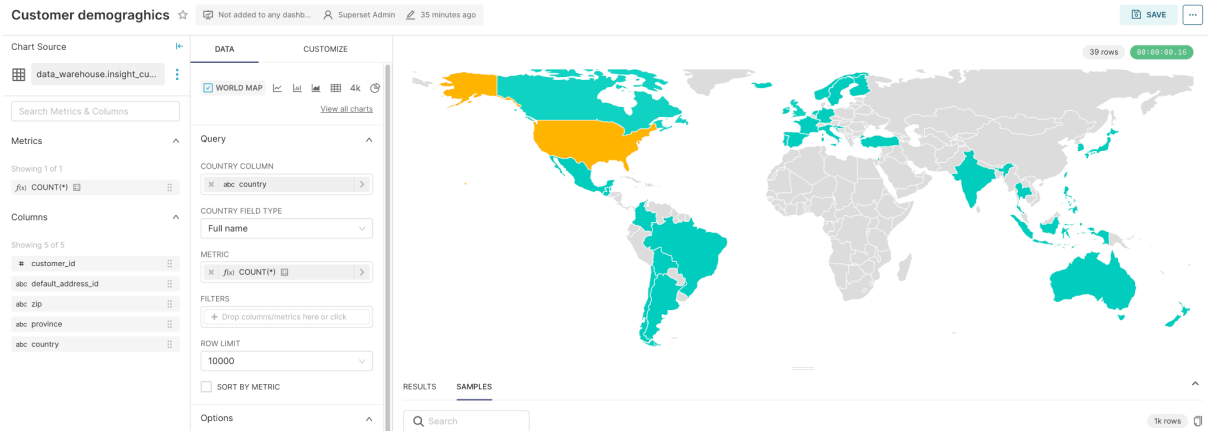


Fig. Customer demographics

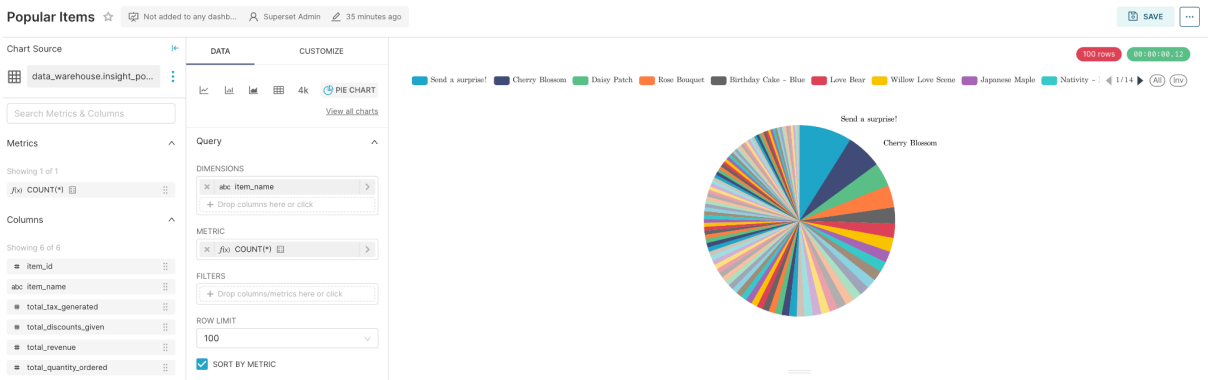


Fig. Popular Items

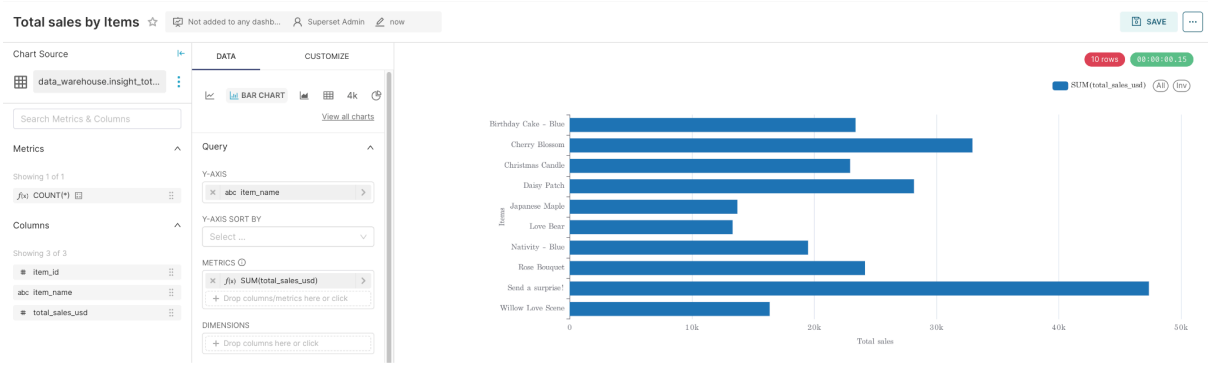


Fig. Top 10 total sales by Items