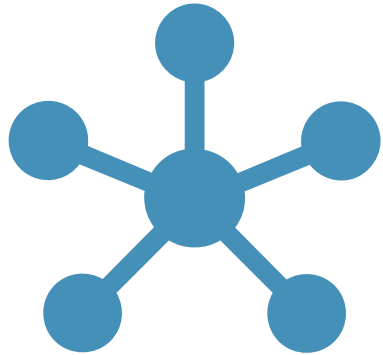


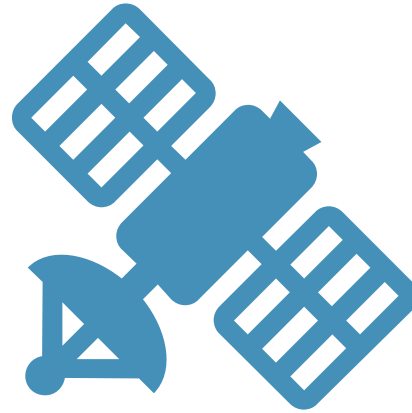


# DATA PIPELINE

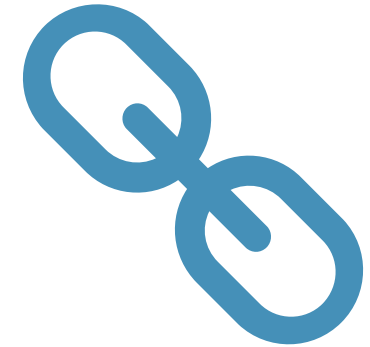
ASTRAFY



Context



Proposition



Sitography

EXIGENCES TECHNOLOGIQUES

# CONTEXT

- Google Cloud as hosting platform
- Open-source technologies as much as possible
- High focus on GitOps and DataOps



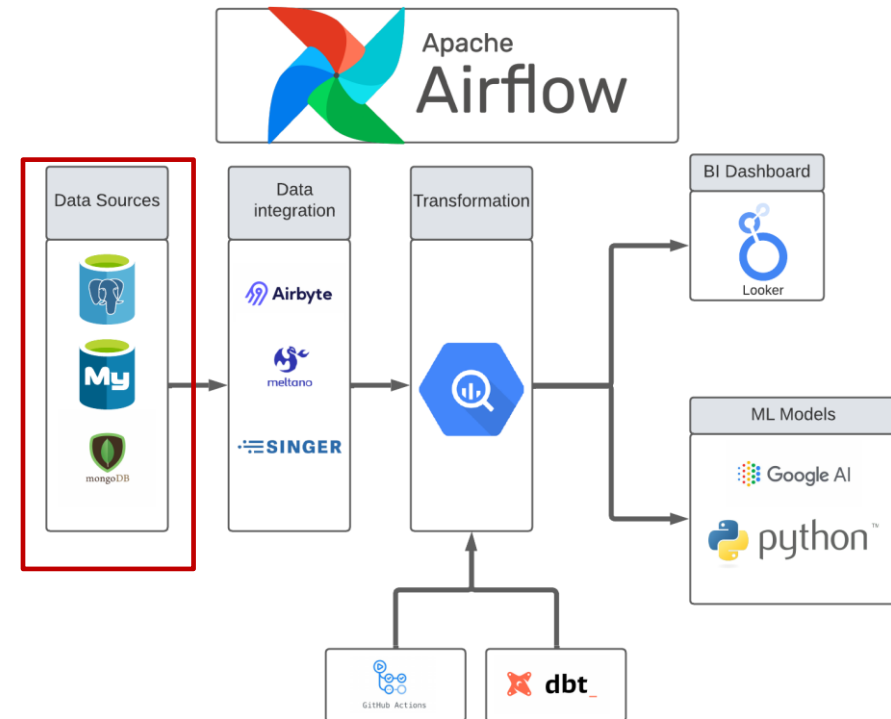
# PROPOSITION

## Data Sources :

- PostgreSQL
- MySQL
- MongoDB

Start of the pipeline

“garbage in, garbage out”



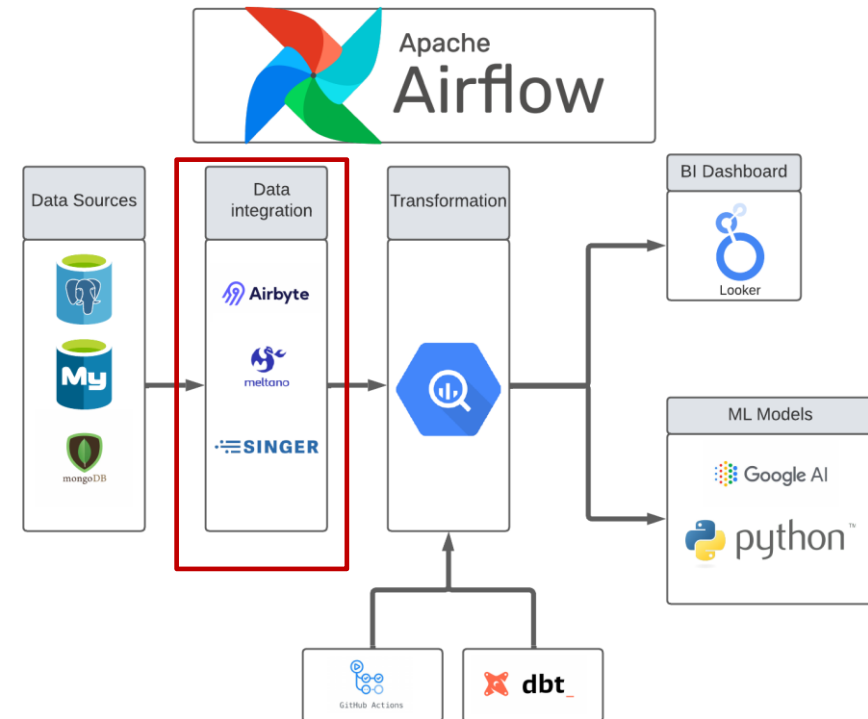
# PROPOSITION

## Data Integration :

- AirByte
- Meltano
- Singer

Ingestion tools open-sources

Exact copy of data sources



# TRANSFORMATION

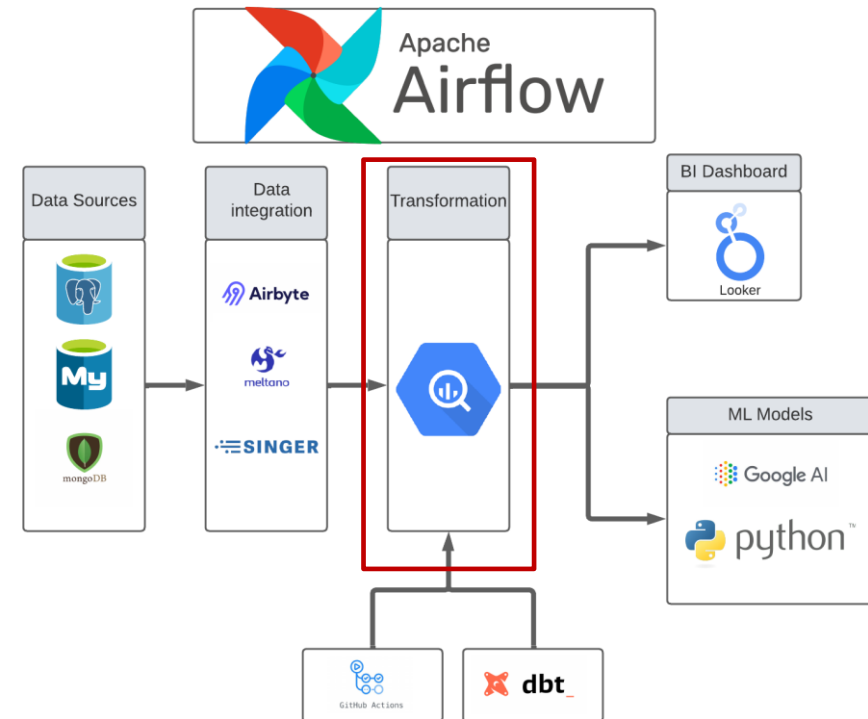
## Transformation :

- BigQuery
- DBT Cloud
- Github Actions

Landing zone → Staging Area

→ Data Warehouse → Data Marts

Use of Github Actions for continuous integration

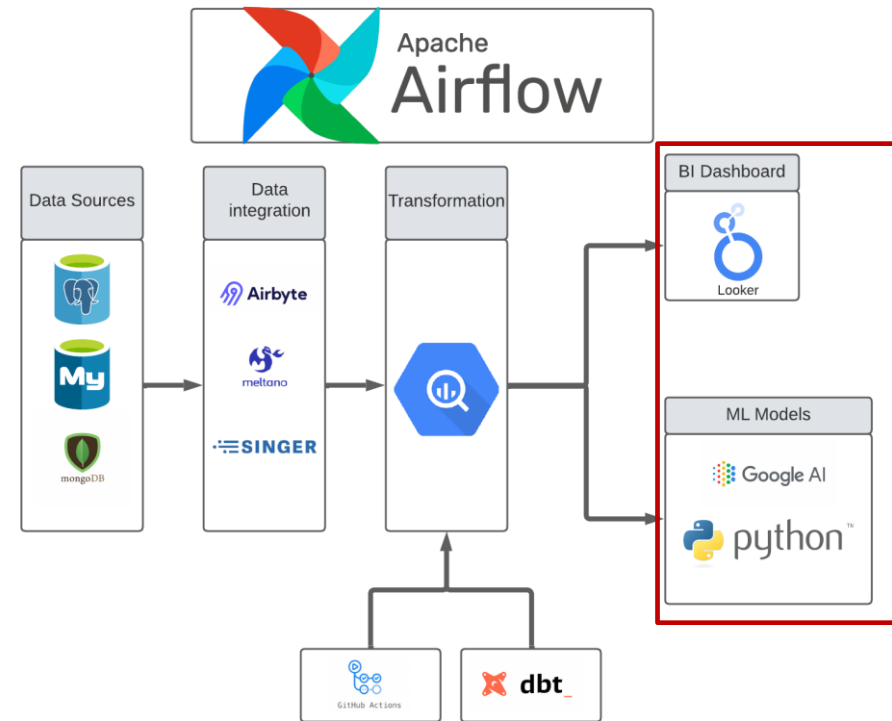


# TRANSFORMATION

## BI Dashboard / ML models :

- Looker
- Google AI
- Python

Visualization Analyze & build AI models



# SITOGRAPHY

- <https://airbyte.com/etl-tools-comparison>
- <https://medium.com/astrafy/>
- <https://www.analytics8.com/blog/best-in-breed-data-stack-platform-bigquery-dbt-and-looker/>
- <https://lucid.app/>





MERCI

ALEXIS.GUILLOTINI@HOTMAIL  
.COM