The Modern Geospatial Data Stack

Layer by layer

Ramiro Aznar

bit.ly/siglibre23-mgds

Girona, June 14th 2023

🌿 Environmental Biologist

⭕ Solutions & Support Engineer at **CARTO**

🛰️ Geospatial Data Engineer, Manager at **Planet**

Ramiro Aznar

How to navigate the data mesh 🧭

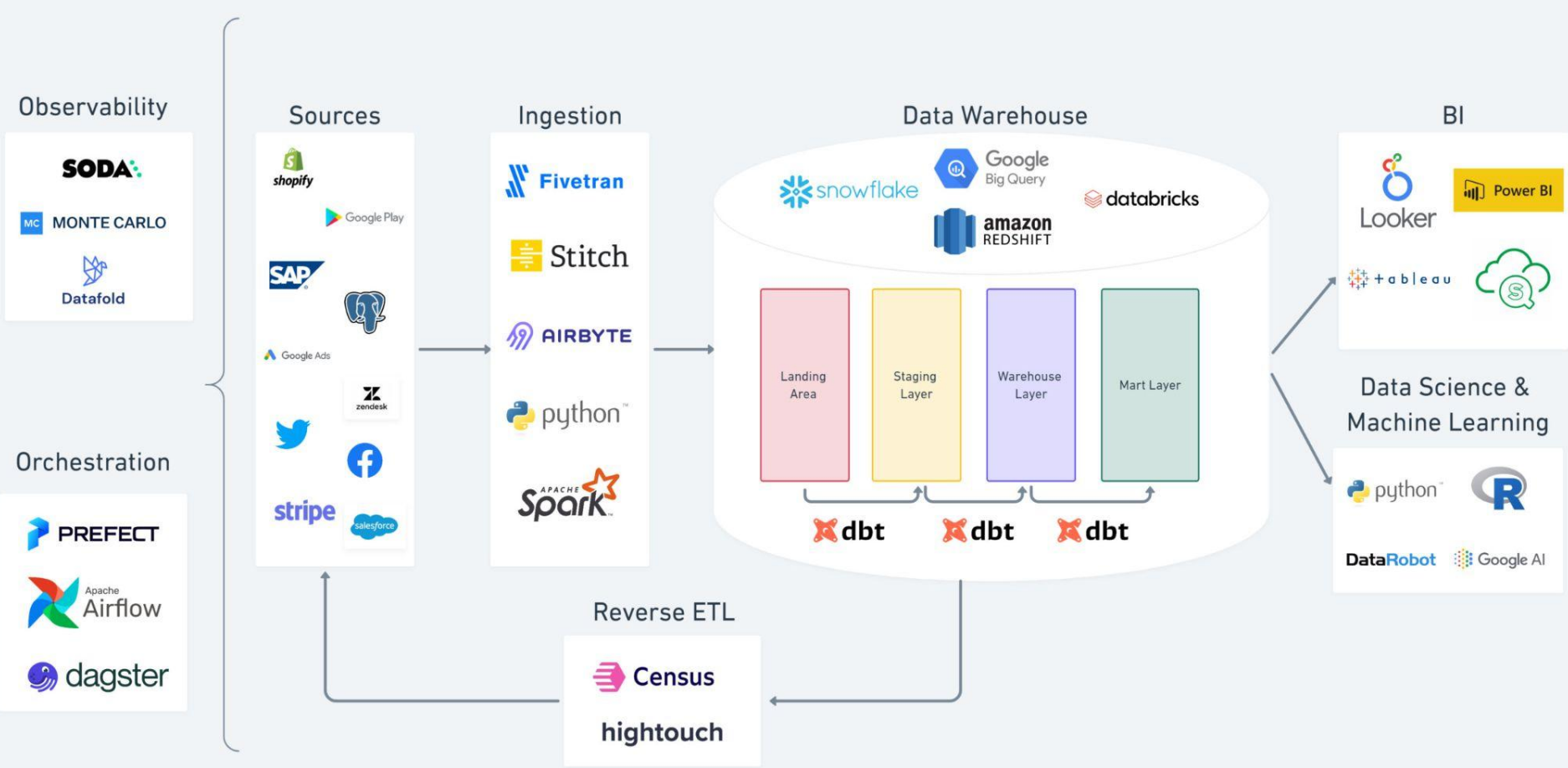
101 in geospatial data engineering

The MGDS Laws 🔭

☁️ Extract the data and load it into the cloud

⚙️ Transform raw data into data models

🌌 Expose the data models as metrics in a viz tool


Observability
SODA
MC MONTE CARLO
Datafold
Orchestration
PREFECT
Apache Airflow
dagster
Sources
shopify
Google Play
SAP
Google Ads
zendesk
stripe
salesforce
Ingestion
Fivetran
Stitch
AIRBYTE
python
APACHE Spark
Data Warehouse
snowflake
Google Big Query
databricks
amazon REDSHIFT
Landing Area
Staging Layer
Warehouse Layer
Mart Layer
dbt
dbt
dbt
BI
Looker
Power BI
tableau
Data Science & Machine Learning
python
R
DataRobot
Google AI
Reverse ETL
Census
hightouch

Extract And Load ☁️

ELT is the new ETL

The data sources could vary from APIs, remote databases, cloud buckets…

Depending on your data source, you could do batch loads or real time ingestions, but they should be orchestrated

All big tech providers have analytical databases to store and compute your geospatial data

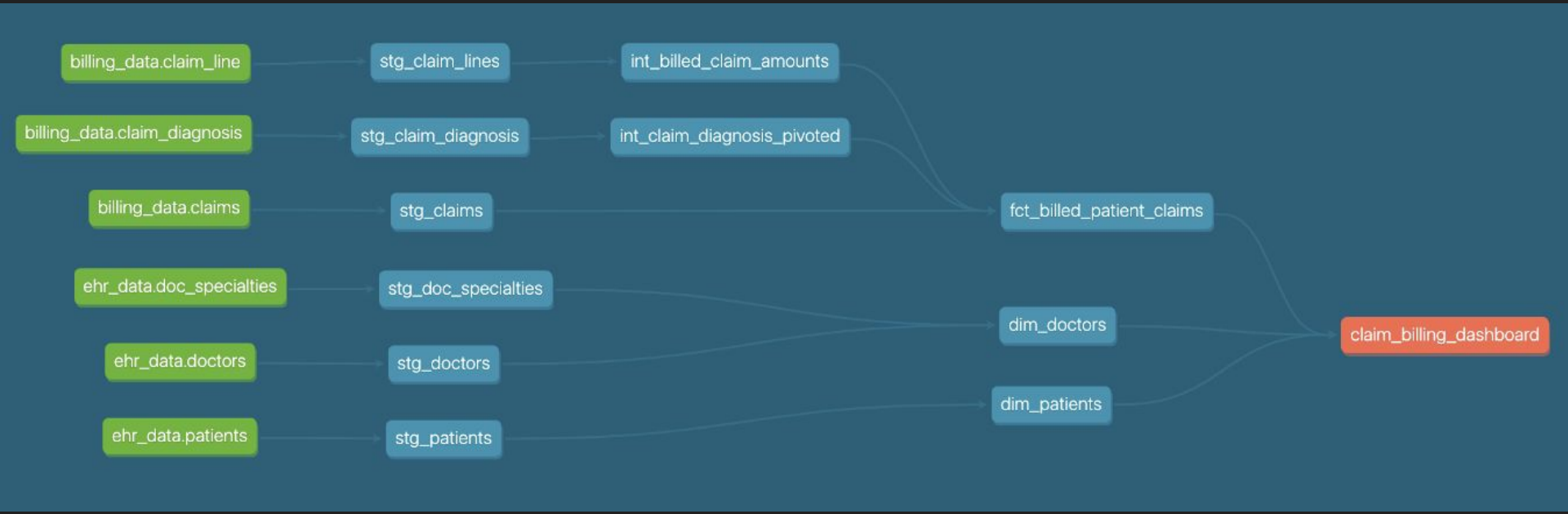
Transform ⚙️

dbt, dbt and dbt

SQL is the lingua franca

Data modeling is becoming a very demanding skill

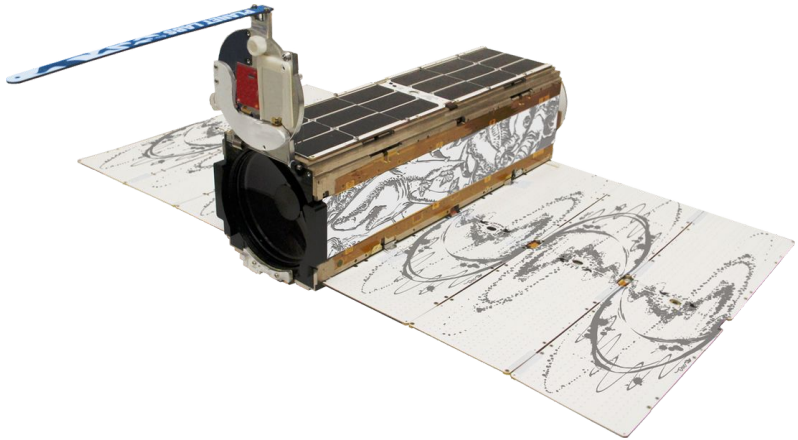
Geospatial transformations usually mean aggregating by spatial indexes and build tilesets


billing_data.claim_line
stg_claim_lines
int_billed_claim_amounts
billing_data.claim_diagnosis
stg_claim_diagnosis
int_claim_diagnosis_pivoted
billing_data.claims
stg_claims
fct_billed_patient_claims
ehr_data.doc_specialties
stg_doc_specialties
dim_doctors
claim_billing_dashboard
ehr_data.doctors
stg_doctors
dim_patients
ehr_data.patients
stg_patients

Visualize 🌌

BI tools include very basic web mapping capabilities

CARTO integrates pretty smoothly into the MDS



Geospatial Data Roles 🦄

Looking for an unicorn

The Data Family 👩‍👧

👷 Data Ops

🚇 Data Engineers

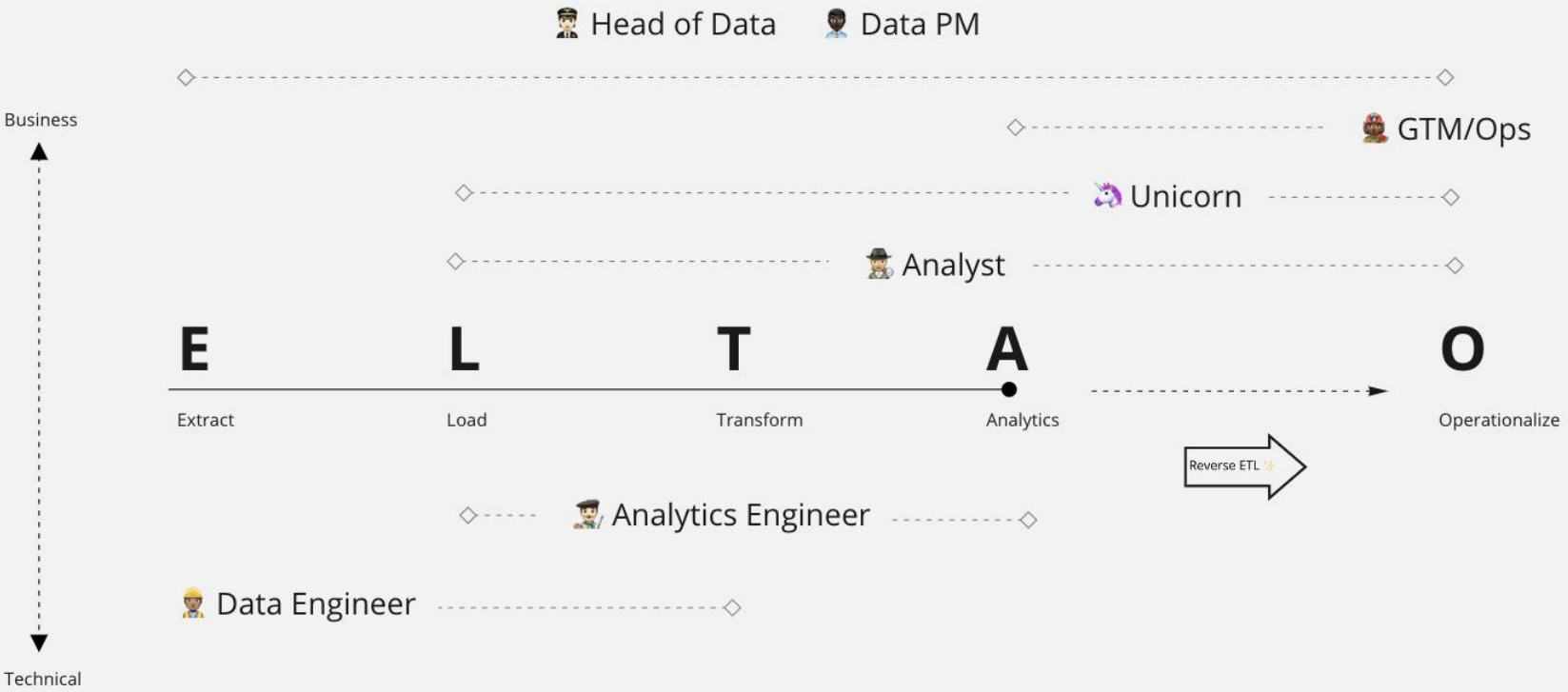
🗒️ Analytics Engineers

📊 Data Analysts

👩‍🔬 Data Scientists

🤖 ML Engineers

💡 Prompt Engineers


Head of Data
Data PM
Business
GTM/Ops
Unicorn
Analyst
E
L
T
A
O
Extract
Load
Transform
Analytics
Operationalize
Reverse ETL
Analytics Engineer
Data Engineer
Technical


© IHarryH

Thanks! 🚀

@ramiroaznar
bit.ly/siglibre23-mgds
elgrancirculo.substack.com

References 📚

Connors, D. 2021. *How to build a mature dbt project from scratch.* Coalesce 2021 Conference. [link]

Rehana, S. 2022. *How dbt Can Help Solve 4 Common Data Engineering Pain Points.* dbt blog. [link]

Handy, T. 2021. *The Magic of the Analyst. Fictional (but all-too-real) Data Teams. Technology is Not the Hard Part. [#256].* The Analytics Engineering Roundup substack. [link]

Resources 🔑

dbt. *Docs* [link], slack [link], newsletter [link] and github [link]

Deziel, Z. 2022. *From ETL to ELT: Transforming Geospatial Data Processing with PostGIS and dbt.* PostGIS Day 2022. The author explains basically the same discovery and processes I came up when implementing dbt at Planet [link]

Kahan Data Solutions. 2021. *dbt - All videos*. A long Youtube list of short videos to get you onboard in dbt [link]

Schott, M. 2021. *The ABCs of Analytics Engineering*. A book that goes letter by letter through all key concepts and tools that Analytics Engineers use, being dbt one of the most important ones [link]

Schott, M. 2021-. *Learn Analytics Engineering.* A substack newsletter that covers in each post an analytics engineering topic, from data modeling to data testing [link]