What is

Analytics Engineering?

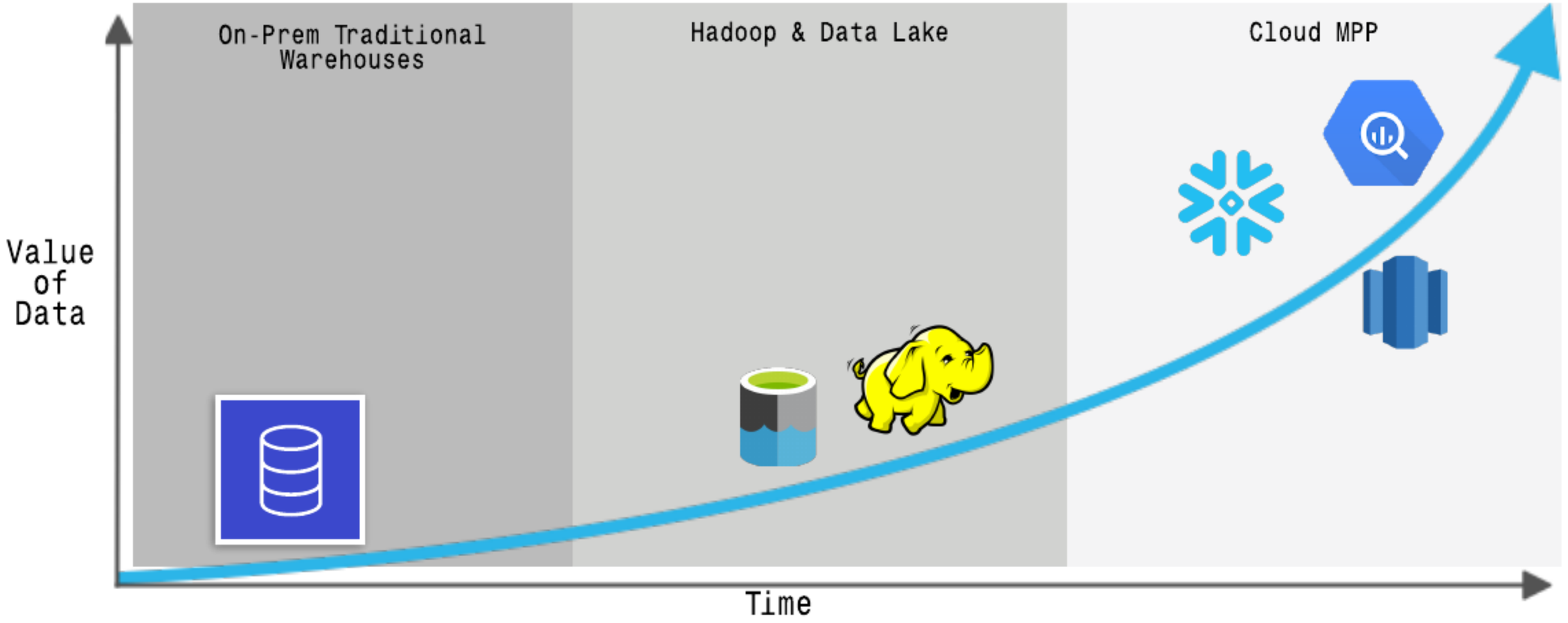
About me

— Lead data @ SnapTravel

— Stitch + Airflow + Snowflake + DBT stack

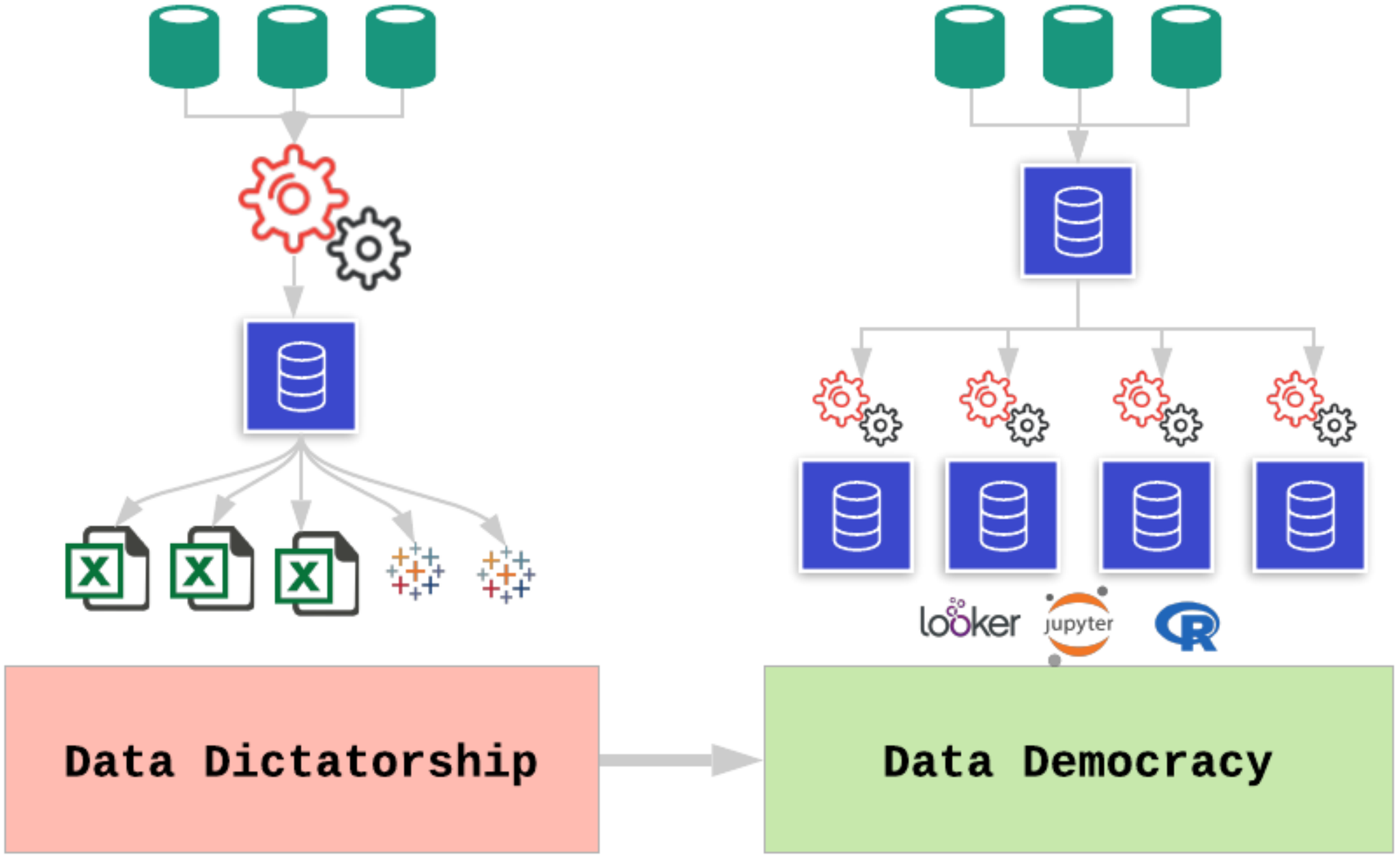
— Love helping people build their data stacks

Journey so far - Data Storage

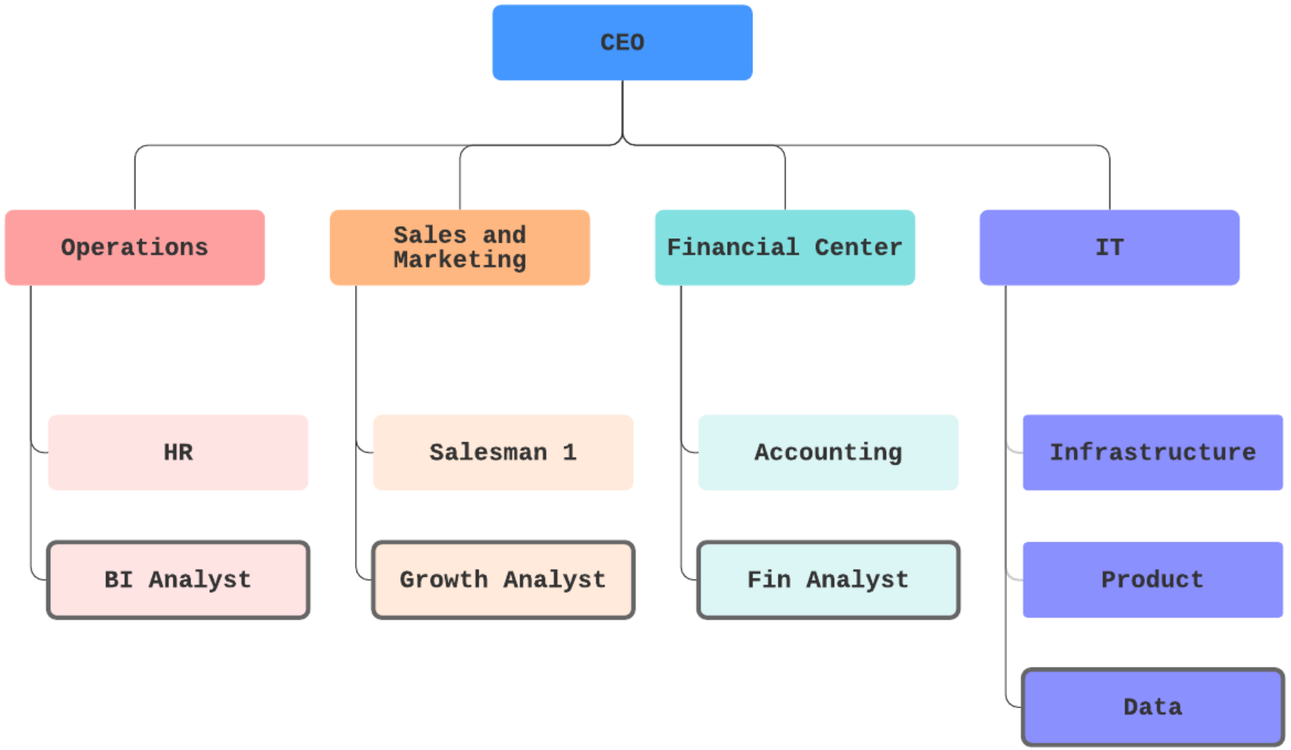

On-Prem Traditional Warehouses
Hadoop & Data Lake
Cloud MPP
Value of Data
Time

- MPP in the cloud
- Benefits
  - Costs
  - Low Infra complexity
  - SQL
  - Variety, Velocity

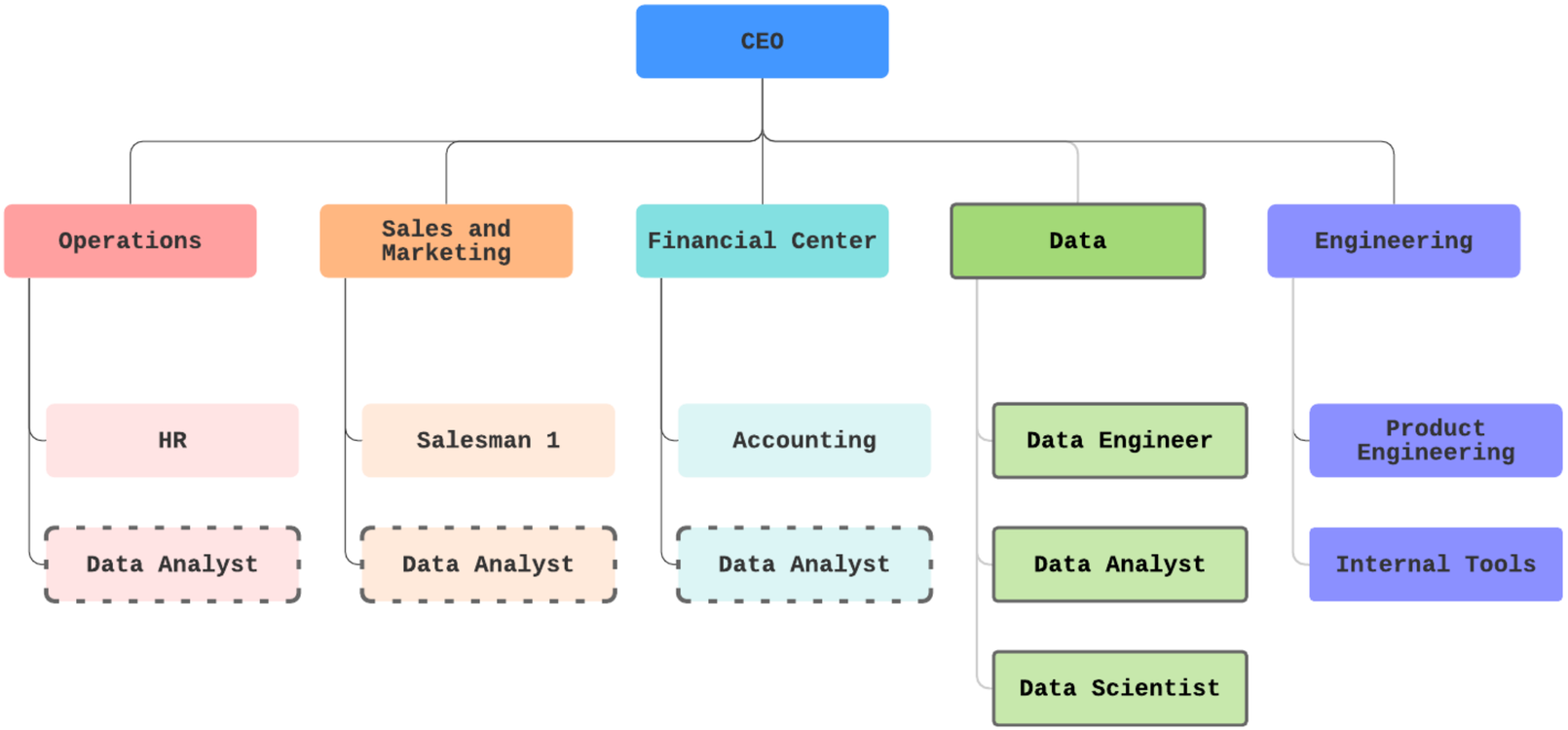
Journey so far - Analytics


looker
jupyter
Data Dictatorship
Data Democracy

Journey so far - Data Teams

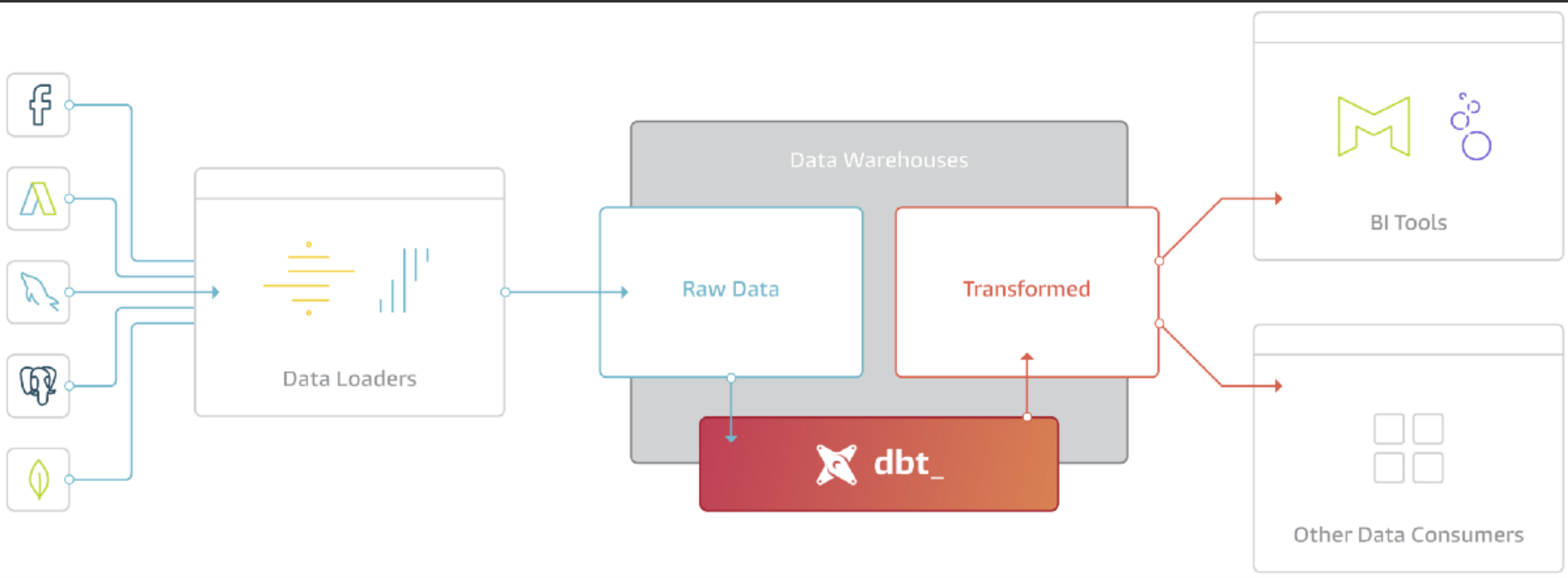


— Compute was expensive or complex
— Few Data team(s)
— Slow and bottleneck
— Away from domain experts



— Domain experts to work on data from source
— Collaborate via metadata
— Nimble
— Reuseable reports

A modern analytics stack


Data Warehouses
Raw Data
Transformed
dbt_
Data Loaders
BI Tools
Other Data Consumers

A modern analytics stack

Salient Features
- Consolidated data ingestion: Stitch Fivetran
- Data warehouse as service
- ETL -> ELT
- DBT to transform
- Mode, Looker, Sisense

A modern data team

Data Engineer	Analytics Engineer	Data Scientist
Enablers		
Custom Data Integration		
Data Platform		
DW Optimization		
ML endpoints		
Data Security		

A modern data team

Data Engineer	Analytics Engineer	Data Scientist
Enablers		Domain expertise with data
Custom Data Integration		Deep insight work
Data Platform		A/B Experiments
DW Optimization		Statistical/ML algos to solve business problems
ML endpoints		
Data Security		

A modern data team

Data Engineer	Analytics Engineer	Data Scientist
Enablers	**Clean and Modeled data**	Domain expertise with data
Custom Data Integration	**SDLC Analytics**	Deep insight work
Data Platform	**Data catalog**	Predictions
DW Optimization	**Self-serve BI platform**	Staticical/ML algos to solve business problems
ML endpoints	**Educate**	A/B Experiments
Data Security		

Things I propose we tease out

- Best practices to structure analytics repos/projects
- How to structure data teams?
- Data as a service vs Data as a product
- Data democratization strategies
- Observability/Trust in Data

Future

Where do we go from here?