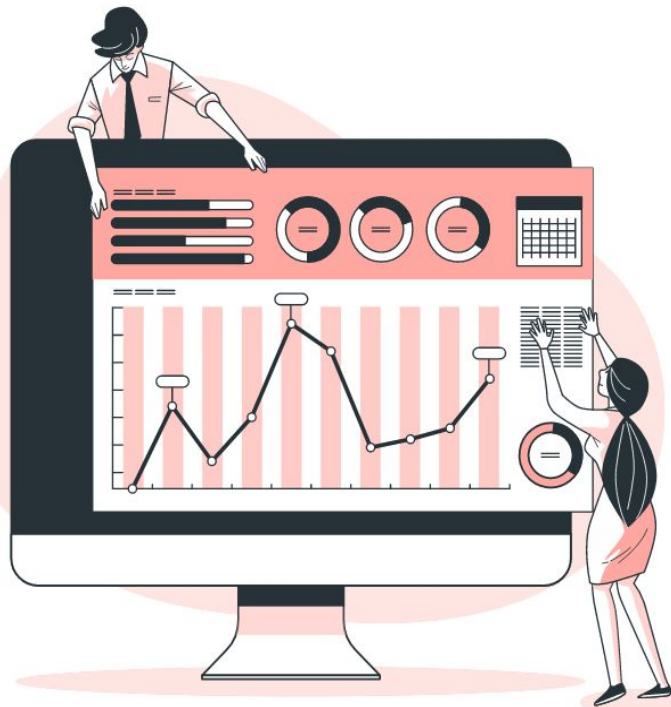


# The Influence of Open Source in the Modern Data Stack

by Victoria Perez Mola



# Hello!

I'm **Victoria Perez Mola**,  
Solutions Architect @  
dbt Labs



# Table of contents

01 About the  
data platform

02 The Modern  
Data Stack

03 Why Open  
Source?

04 How Open  
Source?

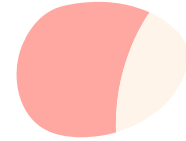


# 01



## About the Data Platform

# The Data Platform



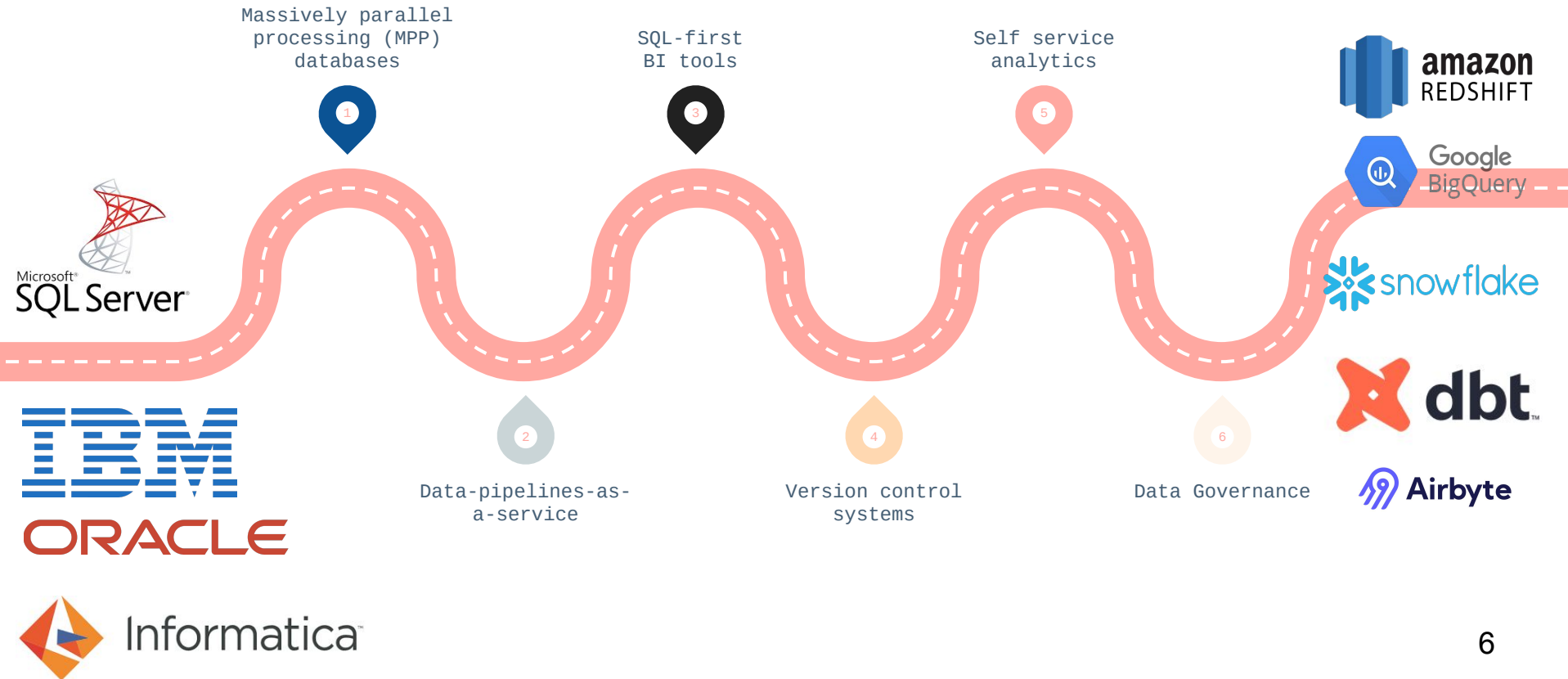
A complete solution that enables:

- Data ingestion
- Data processing
- Data analysis
- Data presentation

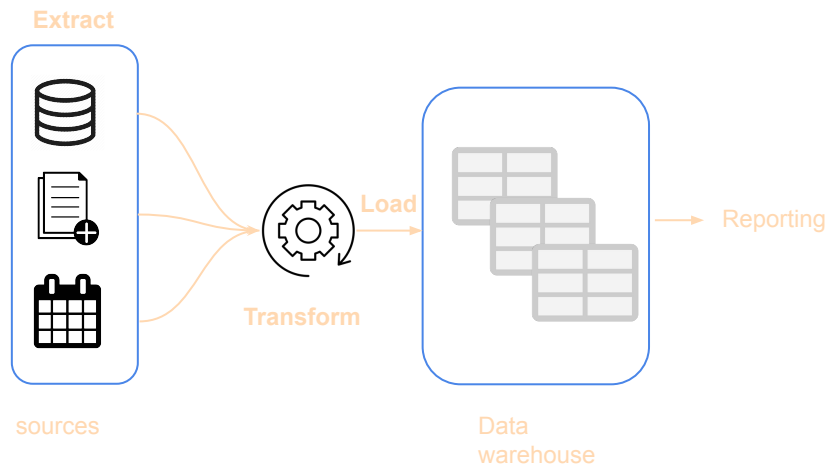
Of the a company's data generated by their own or external systems and processes.



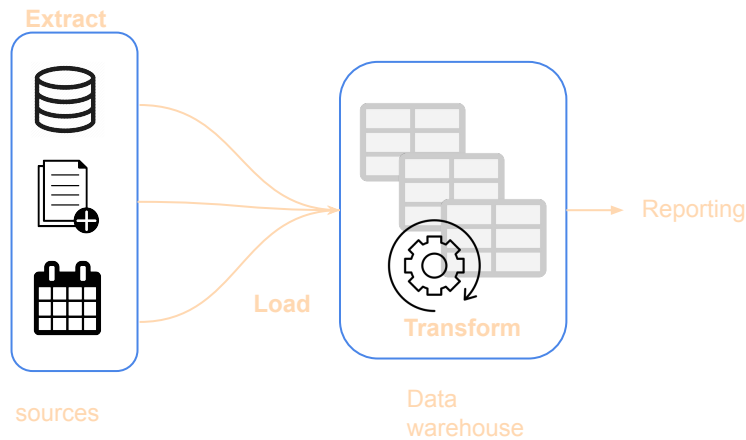
# Data Platform developments



# ETL vs ELT



- Slightly more stable and compliant data analysis
- Higher storage and compute costs
- Engineering usually a bottle neck



- Faster and more flexible data analysis.
- Lower cost and lower maintenance
- Discrepancies in the transformation logic doesn't affect the source data



# 02

## The Modern Data Stack



# The Modern Data Stack

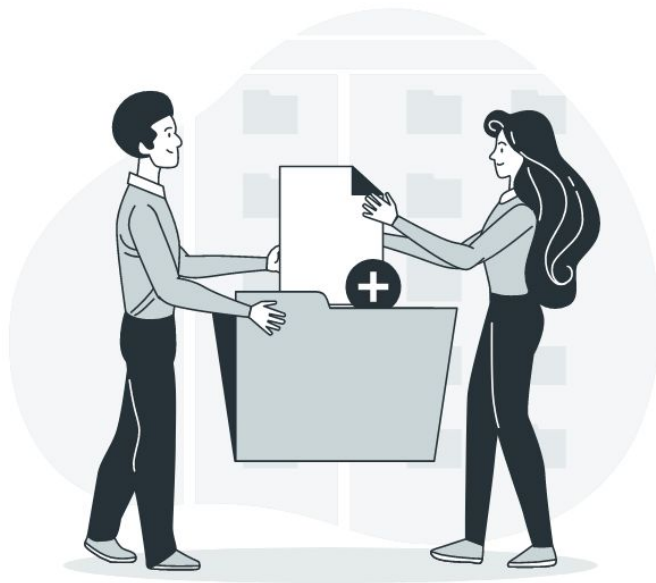
Data  
ingestion



Data  
transformation



Data Analysis



Data  
warehouse

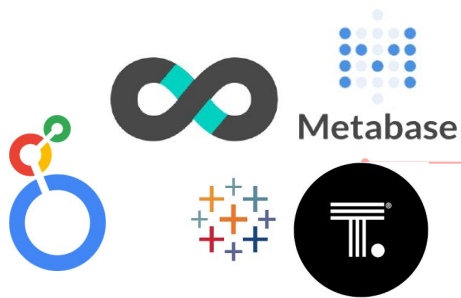


Data Catalog



Data Quality

# TOOLING



Data Storing

Data loading

Data Transformation

Data presentation

Orchestration





# 03

## Why Open Source?

# Open Source Evolution



**Alternative  
Option**



**Innovative  
Option**



**Risky  
option**



**Most secure  
Option**



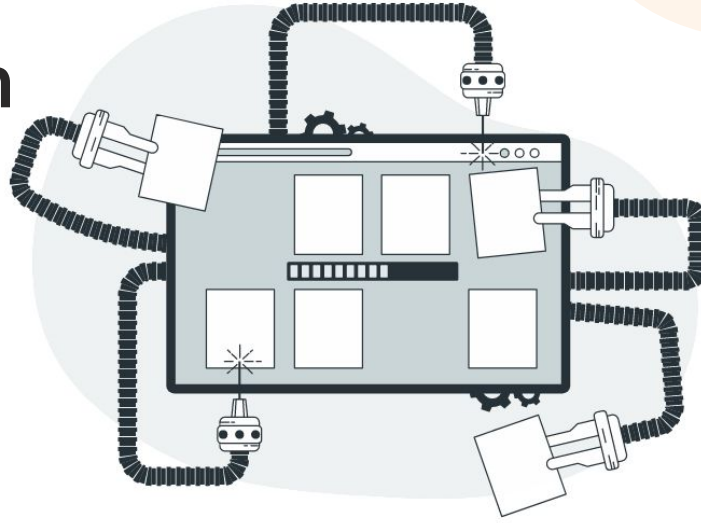
**Limited  
adoption**



**Stronger  
communities**



# Advantages of Open Source



## Free

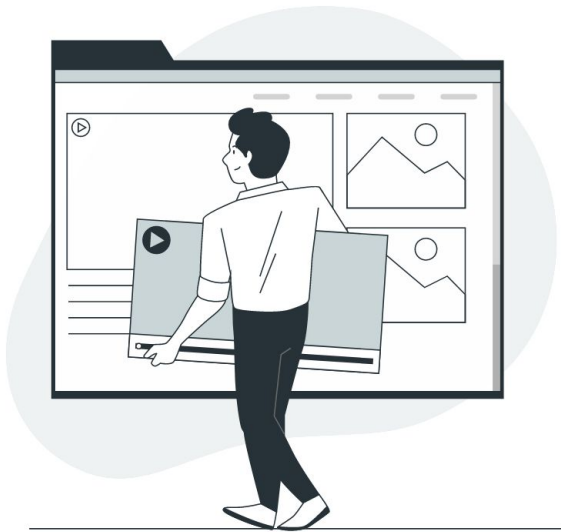
Low to no entry barrier

## Faster development

Features don't depend entirely on the engineering team capacity

## Trust

Code is out in the open, and the tool will outlive the company, if any



# 03

## How Open Source?

# The dbt Story

2016

Started as an  
Open Source  
product

2020

Released the  
first  
commercial  
product on top  
of dbt core

And still going  
strong!



## About us

dbt Labs is on a mission to empower data practitioners to create and disseminate organizational knowledge. Having pioneered the practice of analytics engineering, we're now fortunate to support a community of over 25,000 data practitioners committed to changing how data teams work together.

**9000**

companies using dbt Core

**4500**

trained analytics engineers

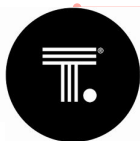
**1800**

dbt Cloud customers

**32000**

dbt Community Slack members

# TOOLING



trino



Data Storing

Data loading

Data Transformation

Data presentation

Orchestration



PREFECT



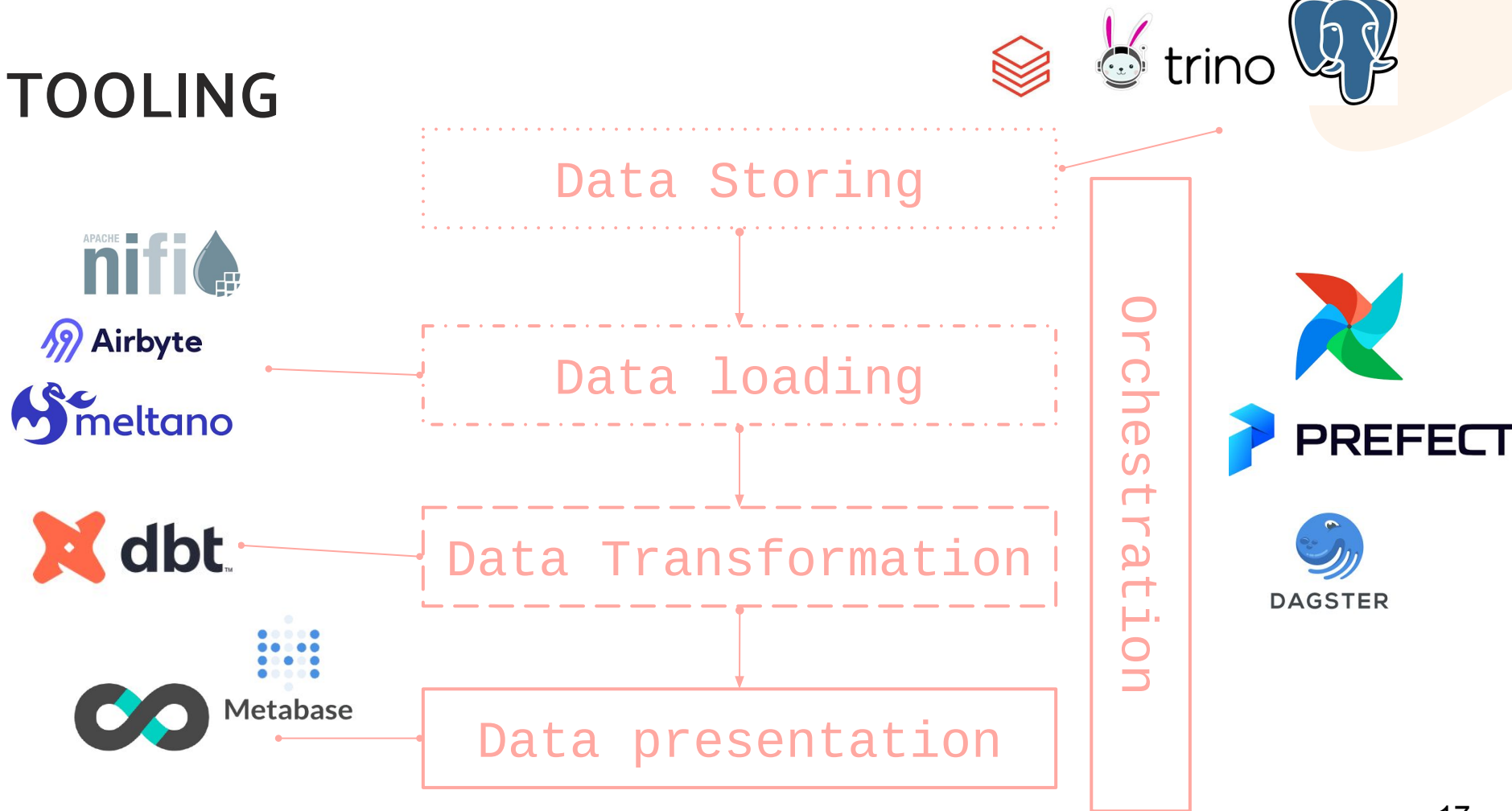
DAGSTER



Azure Data Factory



# TOOLING





# Thanks!

Do you have any questions?



@victoriapm



@victoriaperezmola

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik and illustrations by Storyset

# Resources



- [Building an End-to-End Open-Source](#)
- [Modern Data Platform](#)
- [Open Source Powers the Modern Data Stack](#)
- [Building an Open Source Data Stack](#)
- [Time Till Open Source](#)
- [The asymmetry of open source](#)
- [Understanding the MDS](#)
- [The Modern Data Stack: Open-source Edition](#)
- [Of the Community, By the Community, For the Community](#)
- [Goodbye RJMetrics, Hello Fishtown Analytics](#)
- [2022-05-dbt-a-core-story](#)