

# Unifying Data Worlds

Python, SQL, and Data Governance  
in BigQuery

Stanisław Magierski, Customer Engineer, Google Cloud



# Data World speak in two languages



# The use cases are converging



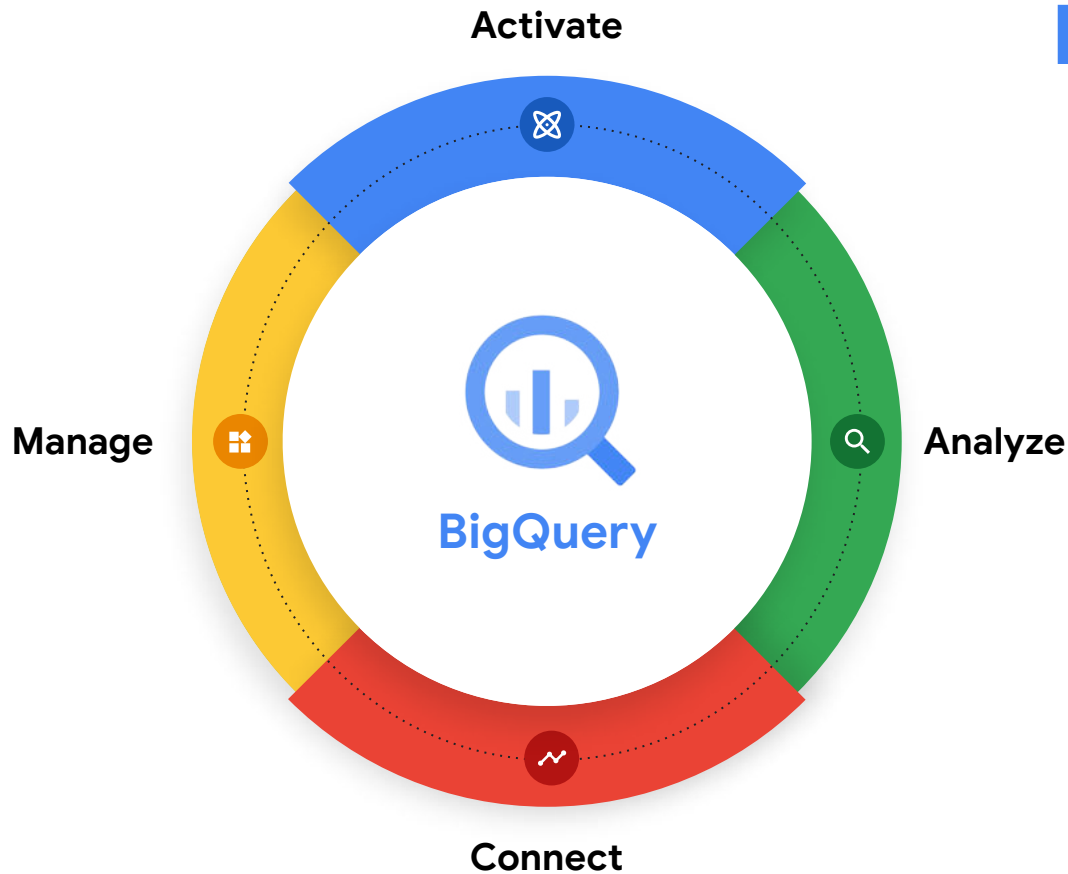
**Data scientist**  
Models that work



**Data engineer**  
Get clean, useful data



**Data analyst**  
Query and analyze



# BigQuery constant evolution:

Cloud Data Warehouse



Data LakeHouse



Data & AI Platform

Unified

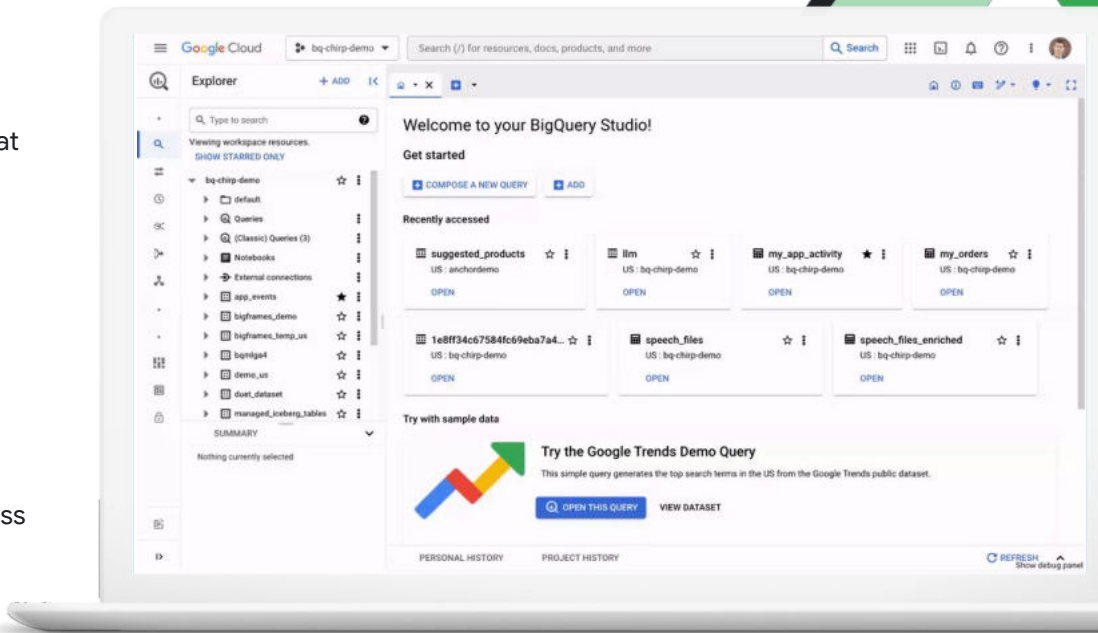
Preview

New!

# Introducing BigQuery Studio

A collaborative data workspace for all data practitioners to accelerate data to AI workflows

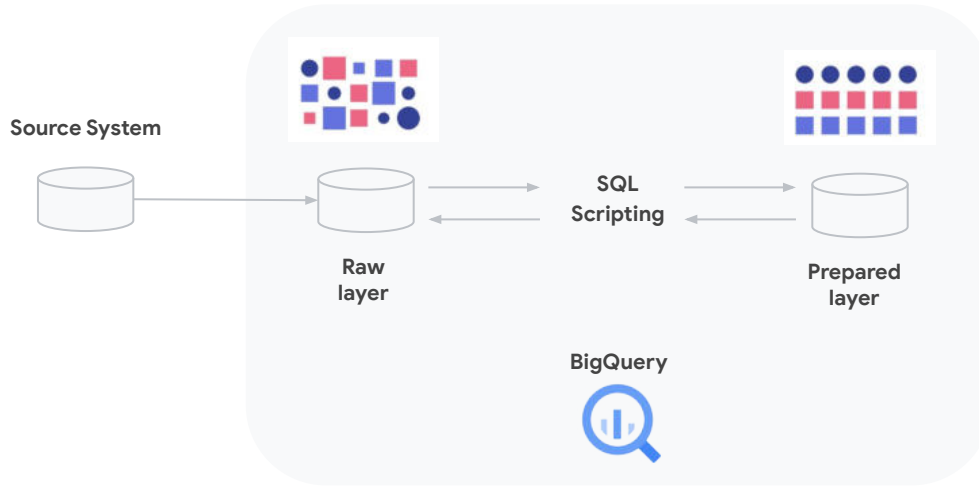
- **Single unified workspace** including a SQL and notebook interface for all data users
- Use **SQL, Python, Spark, Javascript and natural language** for analytics regardless of data scale, format and location
- **BigQuery DataFrames:** Familiar Pandas interface to prepare, train on and score data at PB scale
- **Centralized source control and revision history:** improved collaboration by extending software development best practices to data and code with Dataform
- **Automated data profiling, quality and lineage** across all data assets
- **Maximize productivity** by collaborating with an AI-powered chat and code assistant



So how does it affect

# Modern Data Pipelines?

# The good: “T” in ELT pipeline



## Pros

- Orchestrate
- Test & Data Quality
- Documentation
- Centrally version controlled

## Dataform

Data  
transformation

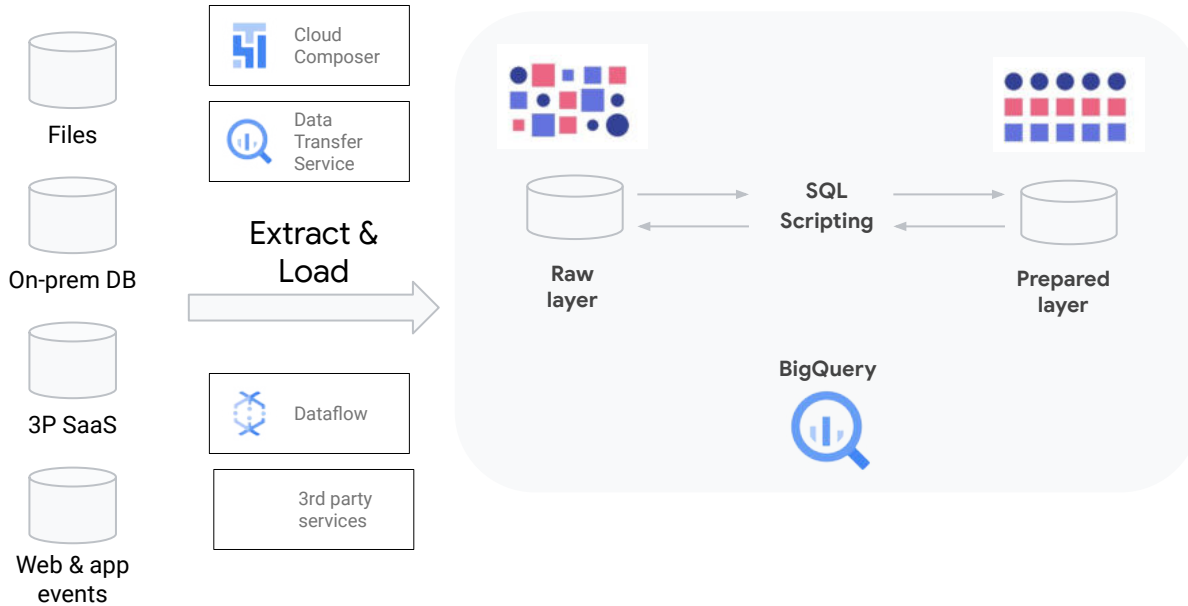
Data quality  
enforcement

Metadata  
generation

Pipeline  
orchestration

Pipeline  
monitoring

# Not sooo good: “EL” in ELT pipeline

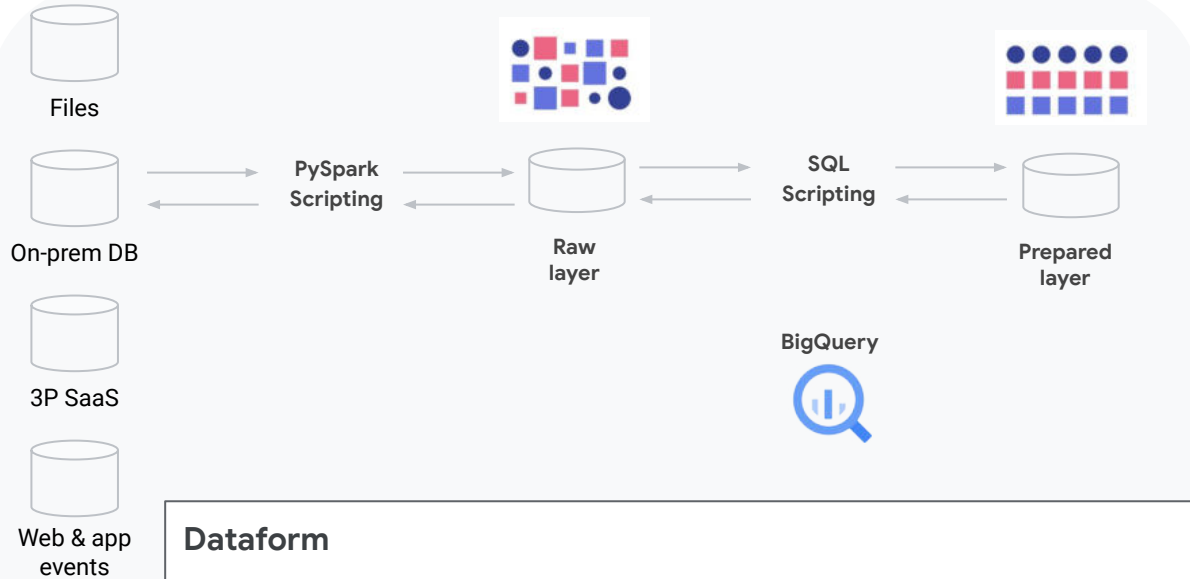


## Cons

- No central versioning
- Harder to orchestrate
- Sparse Data Lineage
- Sparse Documentation



# Solution: ELT pipeline in BQ



# Entire data journey not leaving BigQuery!

Data exploration  
and prototyping

Data enrichment  
with LLM

Automation into a  
full ELT Pipeline

Presentation

05.10.2023 - CONFERENCE DAY

**FEBRUARY SESSION**

**8:30 - 9:00** Registration of participants and meeting coffee

**9:00 - 9:30** Opening

**9:30 - 9:50** Advancing education with responsible AI **MultiCloud**

From the 2023 presentation positions of around 200,000 to the world to gather historical data that can be used for analysis and insights. The result of large historical data set resulting 50 billion rows processing this is done using Spark in DataLake environment

Public viewing

**9:50 - 10:00** Tracking all ships in the world - sea of opportunities **Amazon**

From the 2023 presentation positions of around 200,000 to the world to gather historical data that can be used for analysis and insights. The result of large historical data set resulting 50 billion rows processing this is done using Spark in DataLake environment

Public viewing

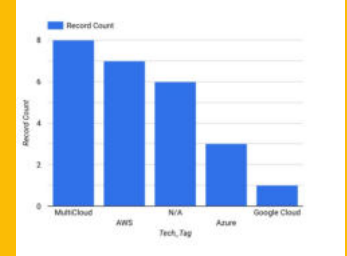
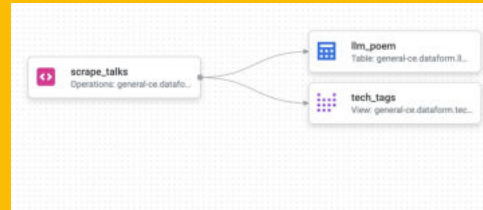
**10:00 - 10:20** Adoption of Generative AI

2023 is becoming the pivotal year with most of the organizations already adopting Generative AI in their products and services. However, this might also mean that there will still be a long way for integrating Generative AI into data and analytics. In this panel we will talk with our experts about an adoption of Generative AI projects at their companies.

Public viewing

Speakers:

- Adam Kawa  
CEO and Co-Founder  
DataScribe
- Karper Łodziejewski  
Vice President of Analytics  
Polaris
- Lukasz Deruski  
Head of the Artificial Intelligence Team  
Allego Intelligence
- Dariusz Krukucki  
CEO of DataScribe  
Polaris



# Walkthrough:

<https://youtu.be/KN9ZluKeZSQ>

Thank you

