

# Adattrendek

## Budapest Data Fórum 2022

Arató Bence

[arato@biconsulting.hu](mailto:arato@biconsulting.hu)



# Elemzői vélemények

Gartner

## Magic Quadrant for Cloud Database Management Systems

Published 14 December 2021 - ID G00740723 - 70 min read

By Henry Cook, Merv Adrian, [and 3 more](#)

Database management systems continue their move to the cloud — a move that is producing an increasingly complex landscape of vendors and offerings. This Magic Quadrant will help data and analytics leaders make the right choices in a complex and fast-evolving market.

### Strategic Planning Assumptions

By 2025, cloud preference for data management will substantially reduce the vendor landscape while the growth in multicloud will increase the complexity for data governance and integration.

By 2022, cloud database management system (DBMS) revenue will account for 50% of the total DBMS market revenue.

# Elemzők - Gartner

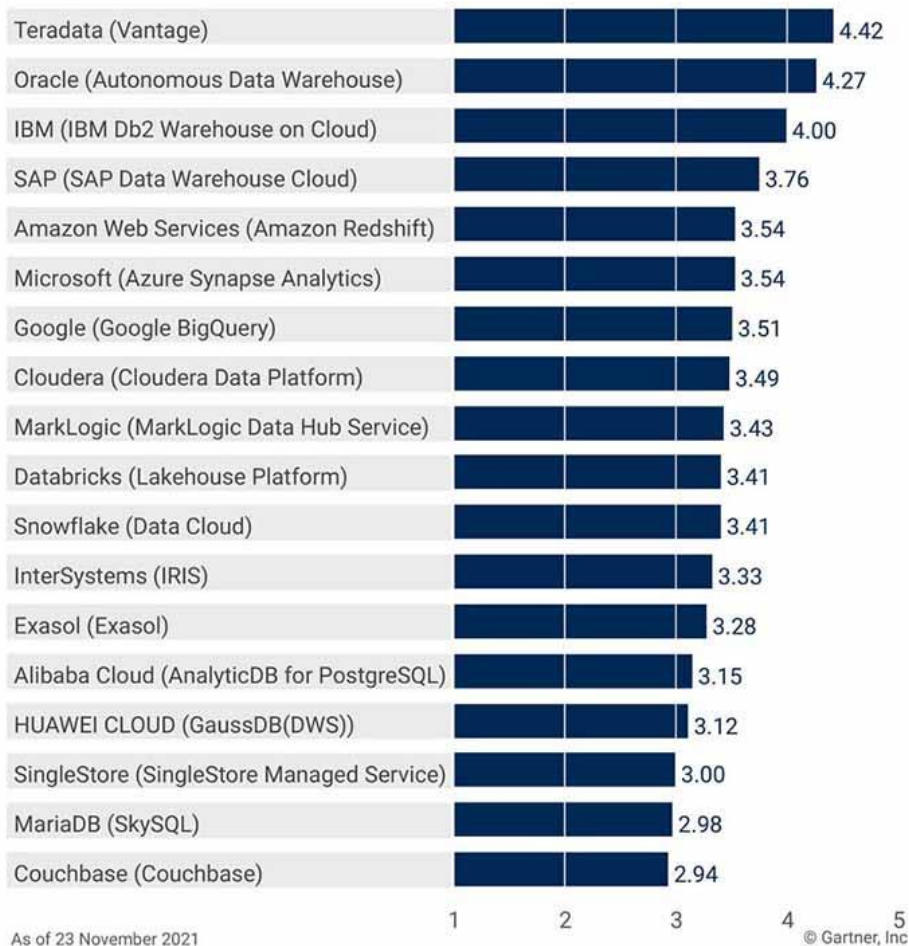


Gartner Magic Quadrant for Cloud Database Management Systems

# Gartner® Critical Capabilities for Cloud DBMSs for Analytical Use Cases

## Vendors' Product Scores for the Data Warehouse Use Case

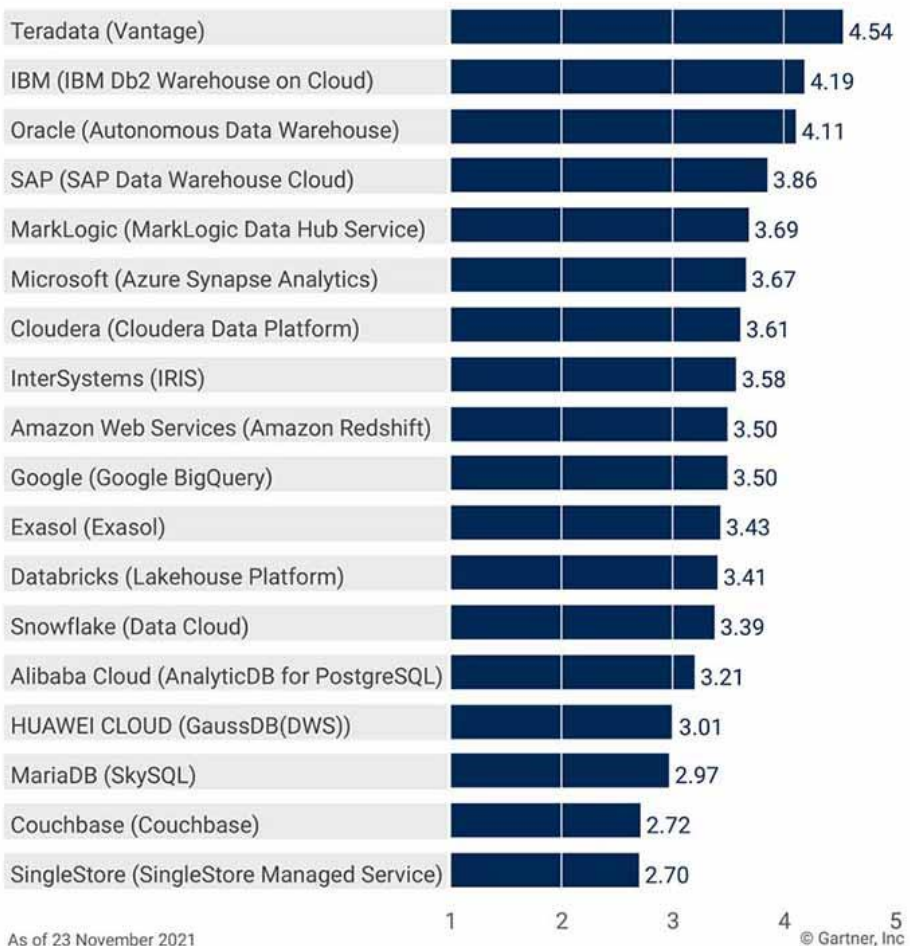
Product or Service Scores for Data Warehouse



Source: Gartner (December 2021)

## Vendors' Product Scores for the Logical Data Warehouse Use Case

Product or Service Scores for Logical Data Warehouse

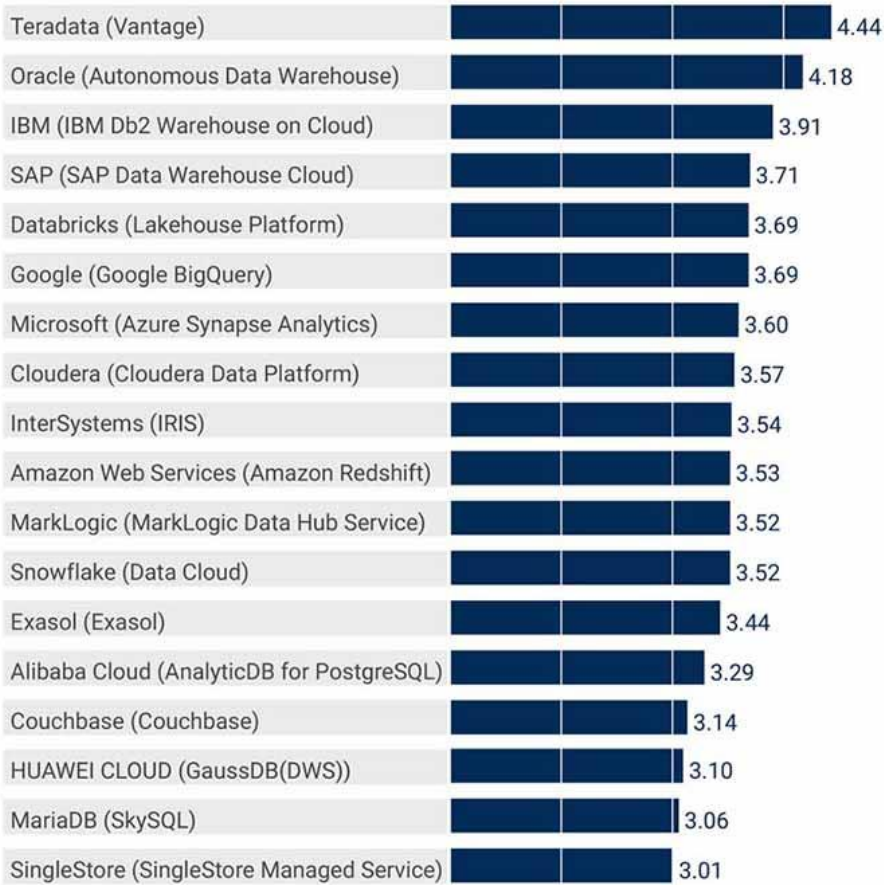




# Gartner® Critical Capabilities for Cloud DBMSs for Analytical Use Cases

## Vendors' Product Scores for the Data Lake Use Case

Product or Service Scores for Data Lake



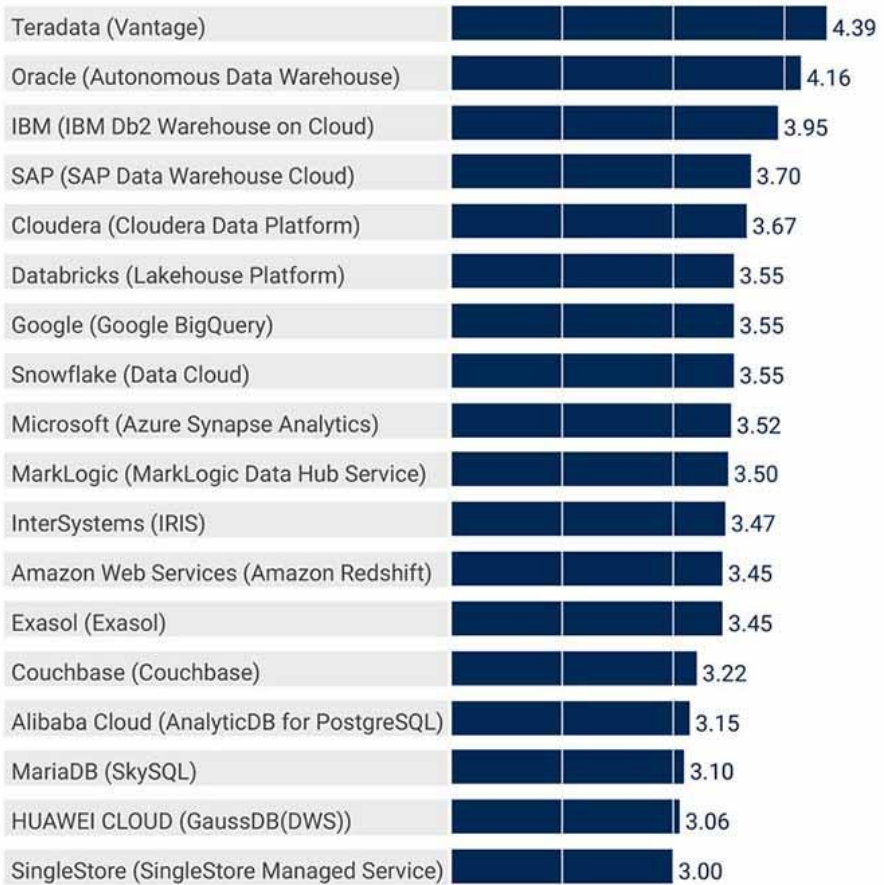
As of 23 November 2021

1 2 3 4 5 © Gartner, Inc

Source: Gartner (December 2021)

## Vendors' Product Scores for the Operational Intelligence Use Case

Product or Service Scores for Operational Intelligence

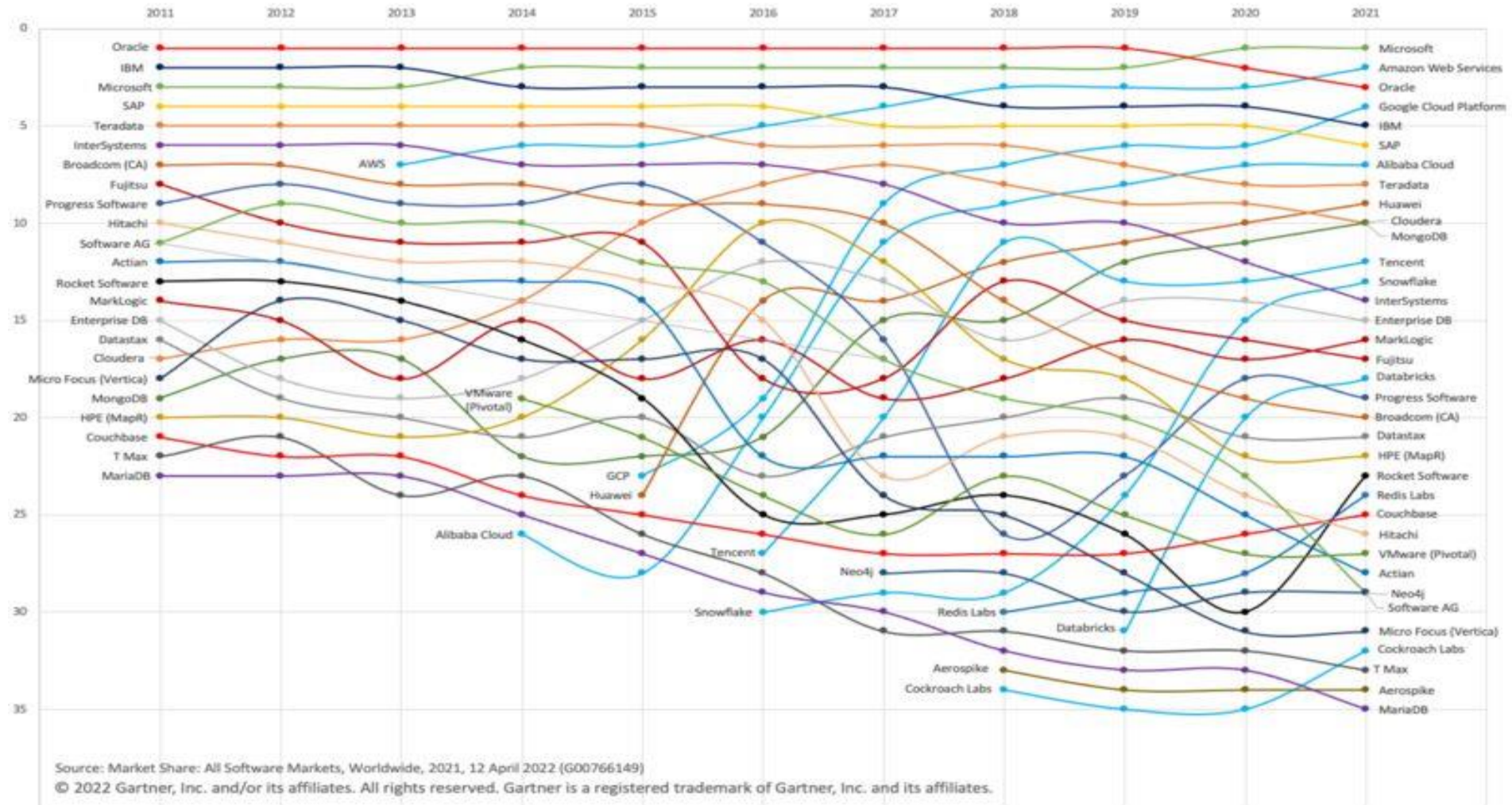


As of 23 November 2021

1 2 3 4 5 © Gartner, Inc

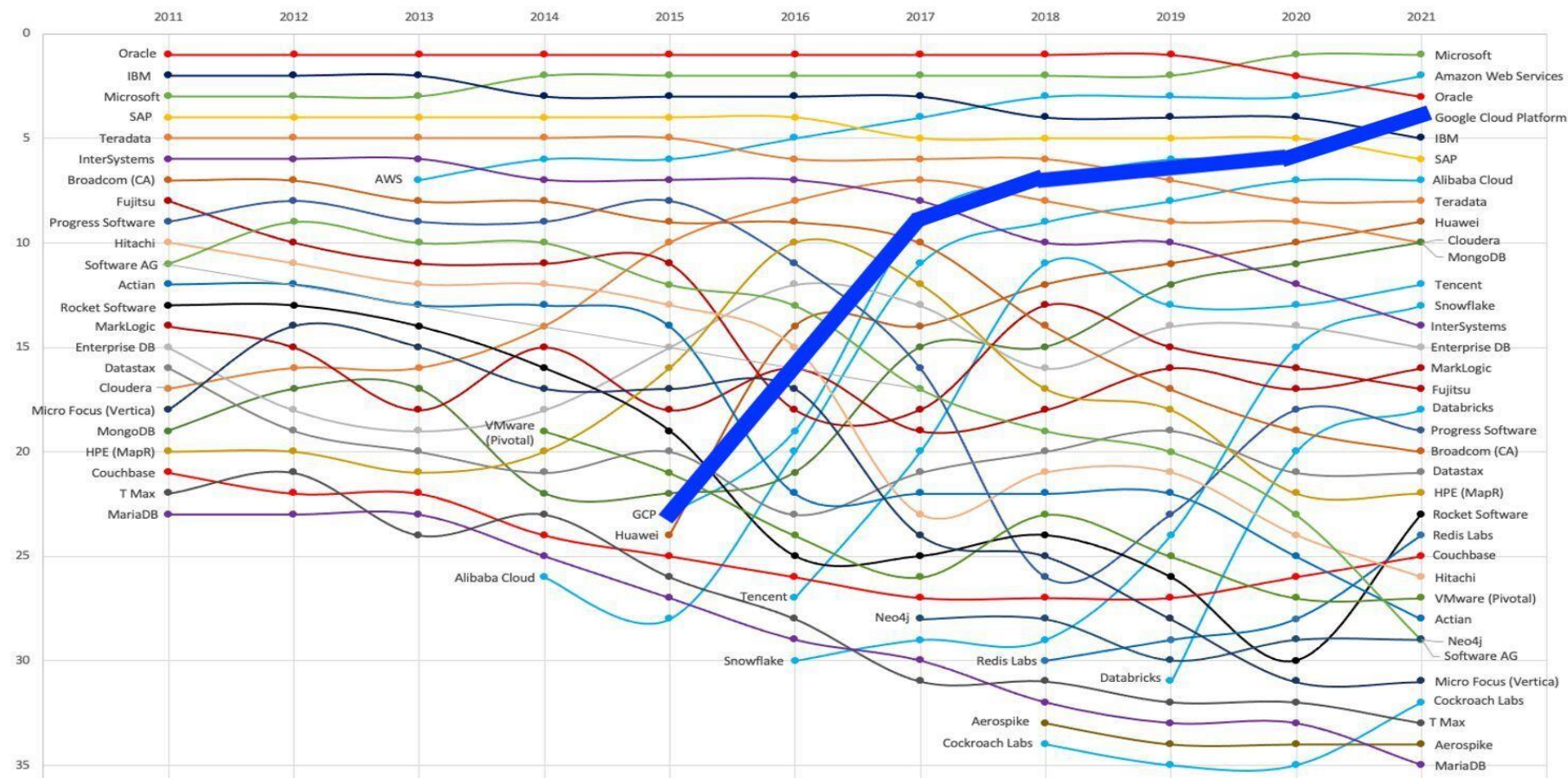
# DBMS Market

## Gartner DBMS Market Share Ranks: 2011-2021

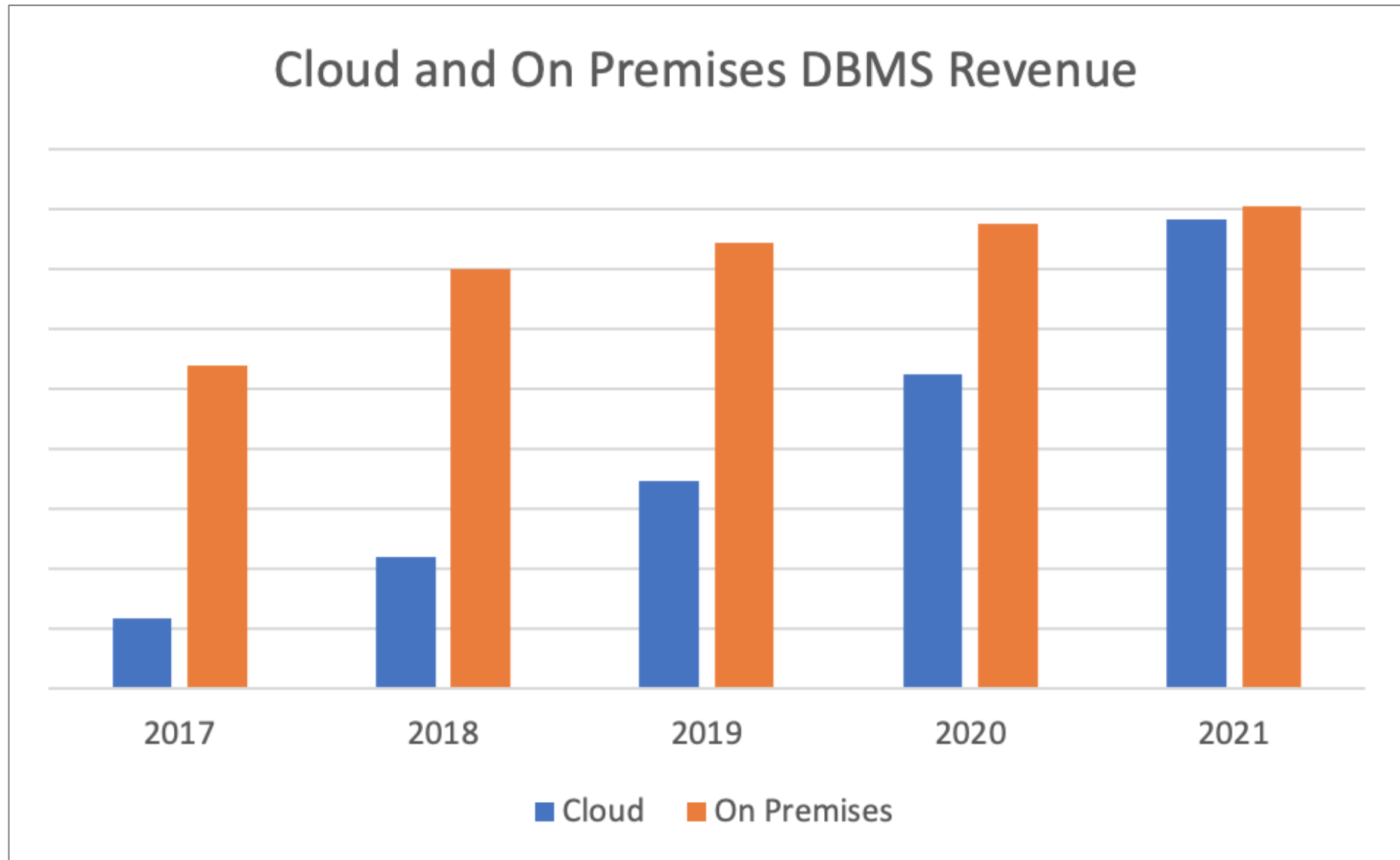




Gartner DBMS Market Share Ranks: 2011-2021



Source: Market Share: All Software Markets, Worldwide, 2021, 12 April 2022 (G00766149)  
© 2022 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.



2017		2018		2019		2020		2021	
Vendor	Share	Vendor	Share	Vendor	Share	Vendor	Share	Vendor	Share
Oracle	36.1%	Oracle	31.1%	Oracle	27.4%	Microsoft	24.3%	Microsoft	24.0%
Microsoft	21.5%	Microsoft	23.6%	Microsoft	24.7%	Oracle	23.8%	AWS	23.9%
IBM	12.7%	AWS	13.5%	AWS	17.1%	AWS	20.6%	Oracle	20.6%
AWS	9.2%	IBM	10.4%	IBM	8.8%	IBM	6.8%	Google	6.5%
SAP	7.4%	SAP	6.9%	SAP	6.5%	SAP	5.6%	IBM	5.6%

# Modern Data Stack

# Technológiai trendek

- **Változatos architektúrák**

- Hagyományos adattárház architektúra
  - Csak on-prem (kihalóban)
  - Egyesített On-prem és felhő támogatás
  - Kizárólag felhőből elérhető DW szolgáltatások
- Data Lake evolúció
  - Klasszikus data lake után Data Lakehouse

- **Cloud centrikusság**

- Különböző megközelítések
  - On-prem + cloud
  - cloud-first és cloud only
  - Single cloud vs multi-cloud
- Dedikált szerver nélküli platformok
  - A „Serverless” előnye a könnyű skálázhatóság
  - Kizárólag használat alapú fizetés

- **Független komponensekből építkezés**

- Storage és Compute szétválasztása
  - Adatok tárolása HDFS vagy object storage rétegben (pl. Amazon S3)
  - Nyílt, optimalizált adatformátumok (Parquet)
- Többféle, alternatív szoftverkomponens
  - SQL, streaming, workflow, metaadat, DQ
  - Általában nagy tech cégek belső fejlesztései

- **Open source népszerűsége**

- Ismét egyre több a jól használható open source
- hagyományos open core + fizetős modell ritkább
- Egyre népszerűbb a teljesen open source szoftver + fizetős felhőszolgáltatás modell



# Emerging Architectures for Modern Data Infrastructure

by Matt Bornstein, Martin Casado, and Jennifer Li

AI, machine & deep learning •  
enterprise & SaaS • big data •  
data infrastructure •  
on the economics of AI/ML & data  
businesses



## Table of contents

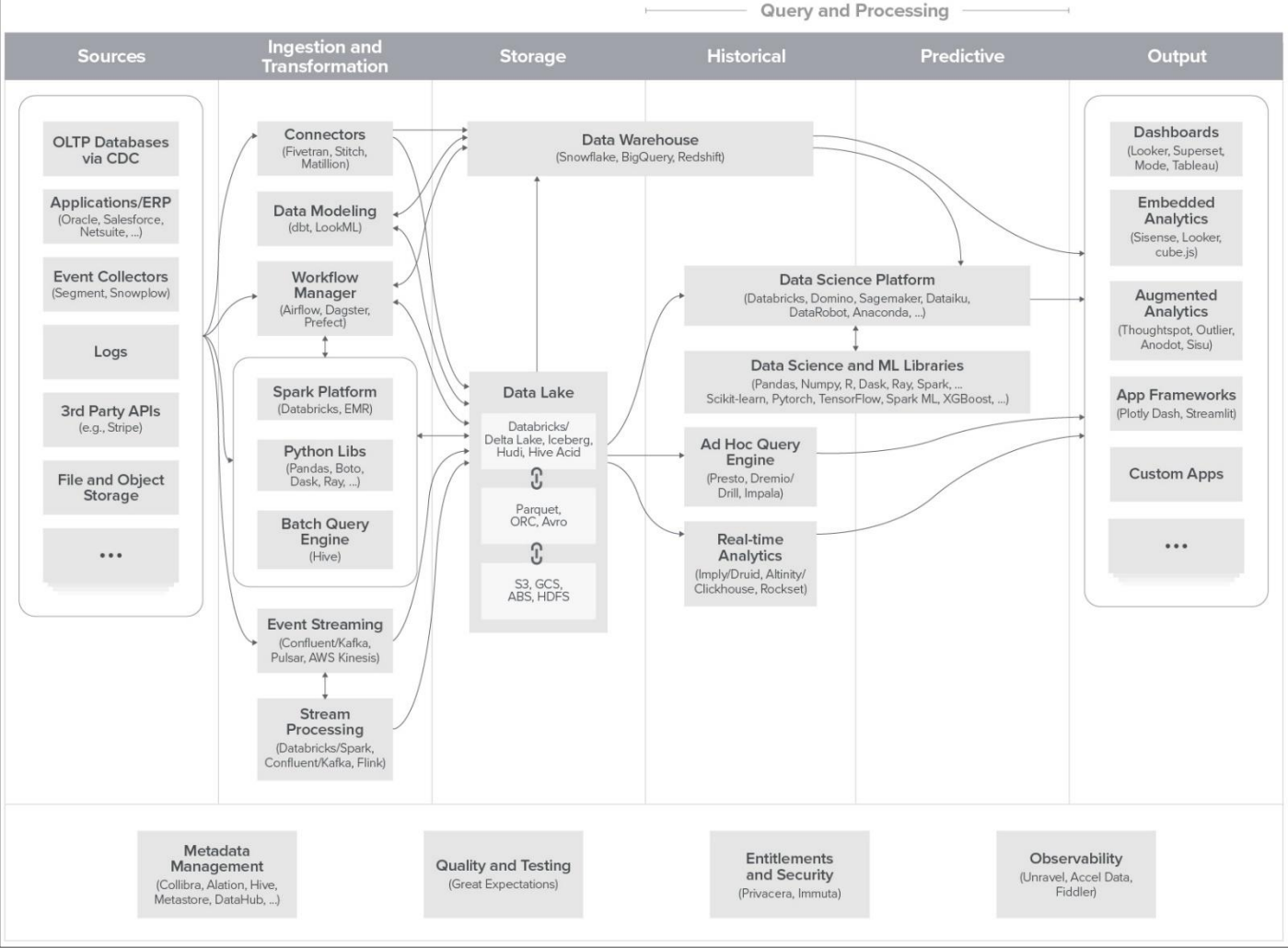
- ▶ 1. The Market
- 2. The Architecture
- 3. Analytics, AI/ML, and the Great Convergence
- 4. Architectural Shifts
- 5. Emerging Capabilities

As an industry, we've gotten exceptionally good at building large, complex software systems. We're now starting to see the rise of massive, complex systems built around data – where the primary business value of the system comes from the analysis of data, rather than the software directly. We're seeing quick-moving impacts of this trend across the industry, including the emergence of new roles, shifts in customer spending, and the emergence of new startups providing infrastructure and tooling around data.


In fact, many of today's fastest growing infrastructure startups build products to manage data. These systems enable data-driven decision making (*analytic* systems) and drive data-powered products, including with machine learning (*operational* systems). They range from the pipes that carry data, to storage solutions that house data, to SQL engines that analyze data, to dashboards that make data easy to understand – from data science and machine learning libraries, to automated data pipelines, to data catalogs, and beyond.

And yet, despite all of this energy and momentum, we've found that there is still a tremendous amount of confusion around what technologies are on the leading end of this trend and how they are used in practice. In the last two years, we talked to hundreds of founders, corporate data leaders, and other experts – including interviewing 20+ practitioners on their current data stacks – in an attempt to codify emerging best practices and draw up a common vocabulary around data infrastructure. This post will begin to share the results of that work and showcase technologists pushing the industry forward.

# A Unified Data Infrastructure Architecture



# Unified Data Infrastructure (2022)

 **Future**

Tech TrendsCompany BuildingSpecial ReportsPodcasts

# Emerging Architectures for Modern Data Infrastructure

Matt Bornstein, Jennifer Li, Martin Casado

Share

**Contents**

**Intro**

Unified data infrastructure (2.0)

Machine learning infrastructure (2.0)

What has changed since 2020

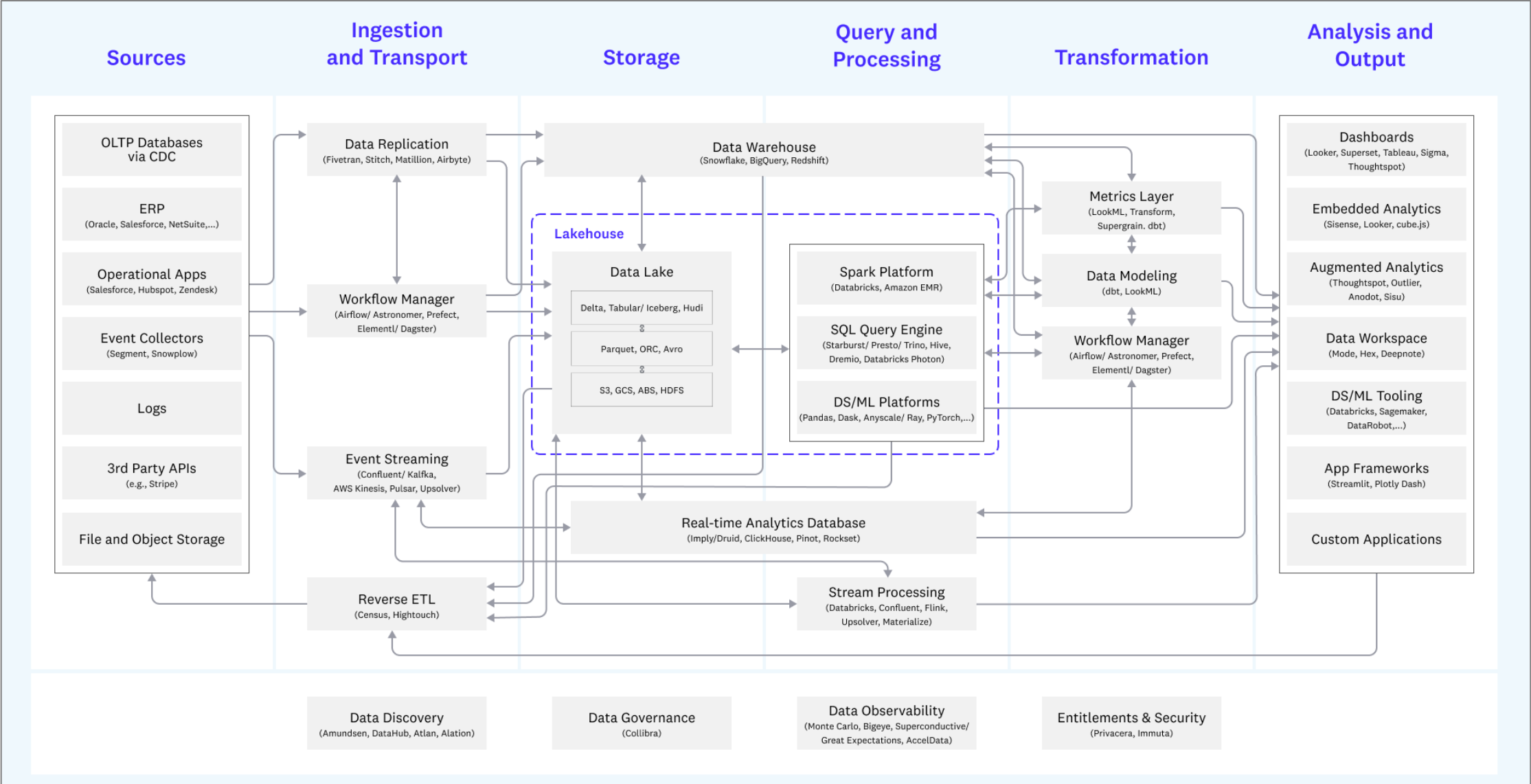
Blueprint 1: Modern business intelligence

*This is an updated version of a post we originally published in 2020. You can read the original version [here](#).*

---

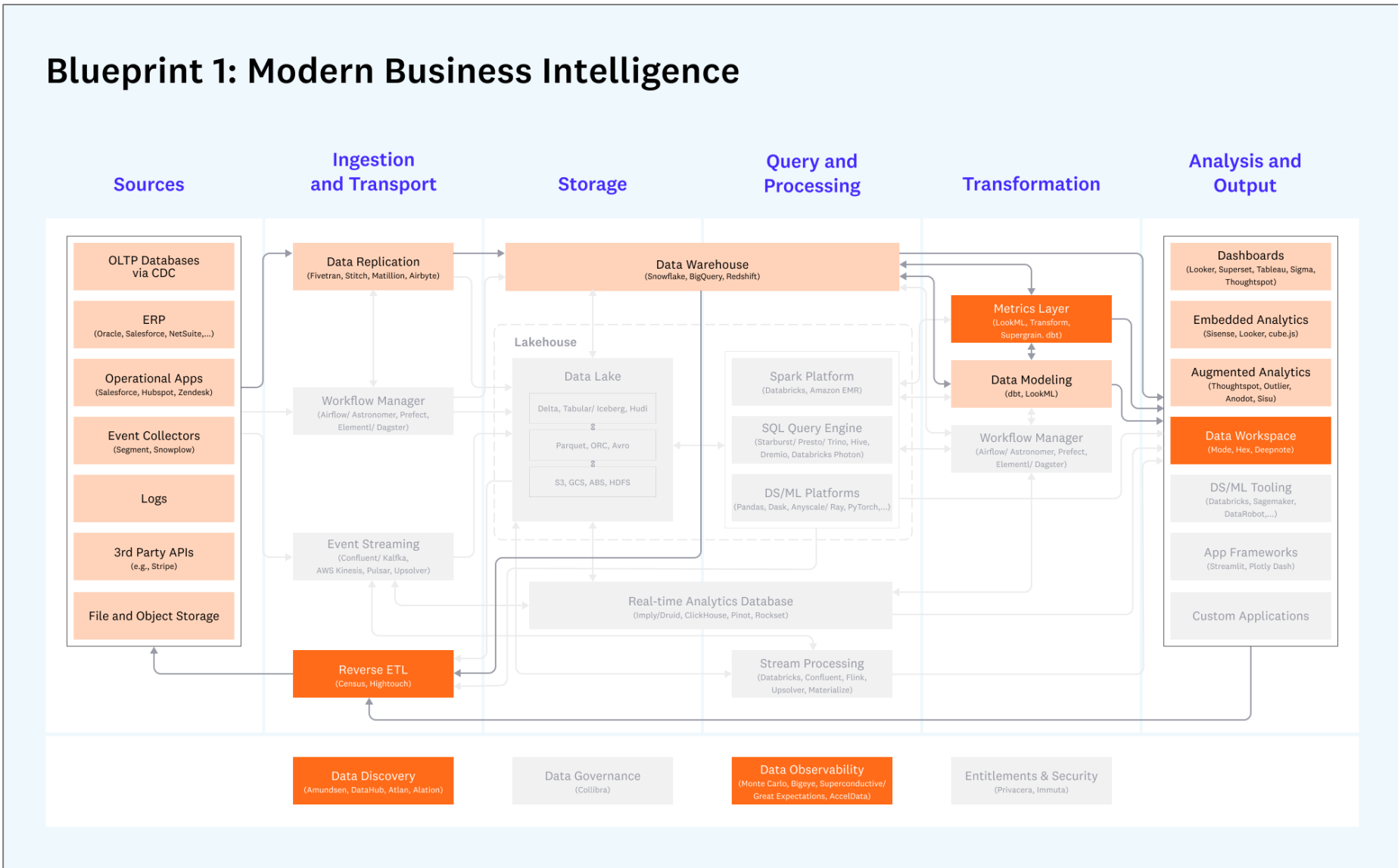
The growth of the data infrastructure industry has continued unabated since we published a set of reference architectures in late 2020. Nearly all key industry metrics hit record highs during the past year, and new product categories appeared faster than most data teams could reasonably keep track. Even the benchmark wars and billboard battles returned.

# Unified Data Infrastructure 2.0



# Unified Data Infrastructure 2.0

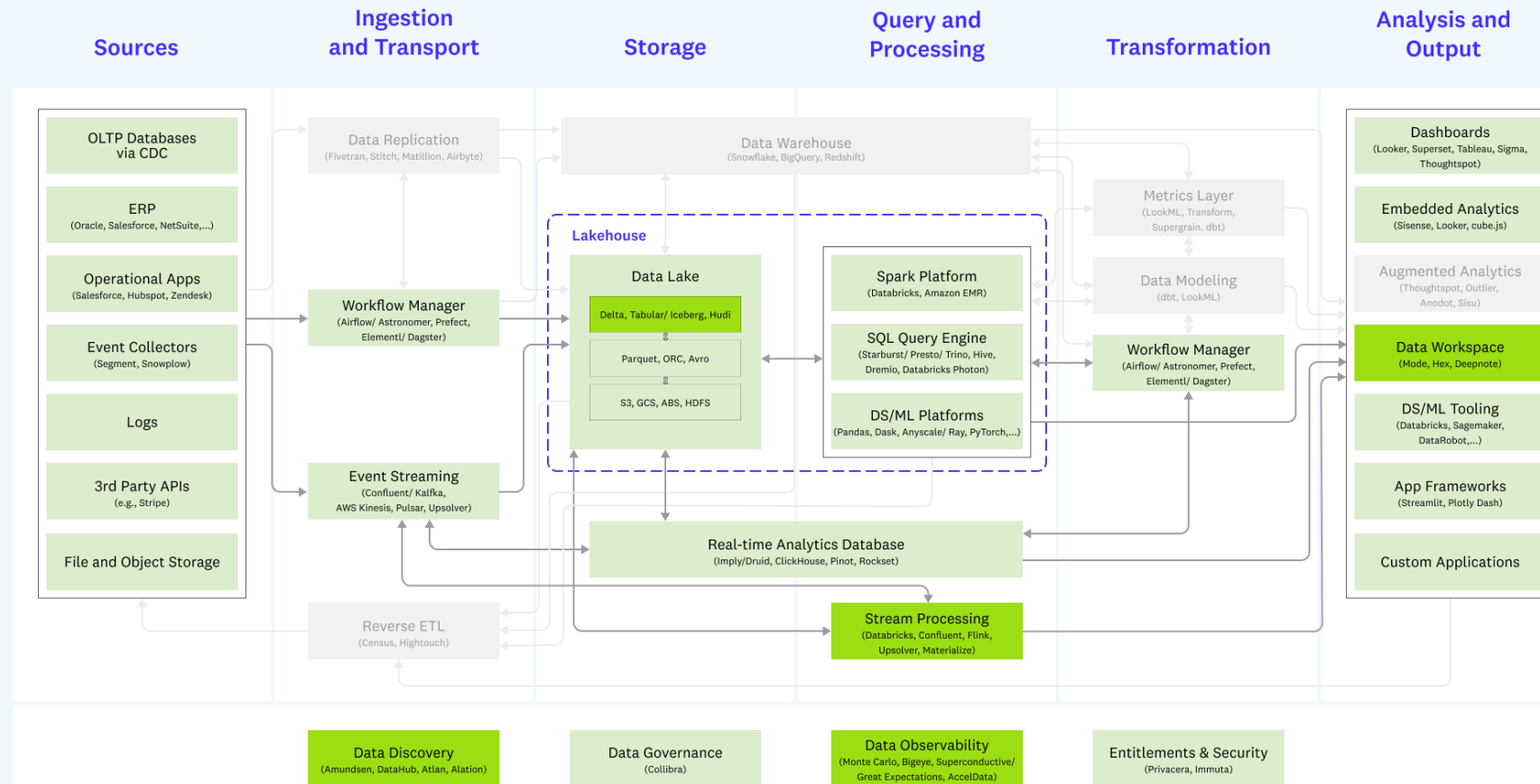
## Blueprint 1: Modern Business Intelligence





# Unified Data Infrastructure 2.0

## Blueprint 2: Multimodal Data Processing



# Az adatplatform receptje

# Az adatplatform receptje - 2020

**Metaadat-kezelés**

**ETL**

**Vezérlés**

**SQL motor**

**Streaming motor**

**Tárolási formátum**

**Tárolási hely**

# Az adatplatform receptje - 2022

**Vezérlés**

**Metaadat-kezelés**

**Adatminőség**

**Streaming motor**

**Adattranszformáció**

**SQL motor**

**ETL / Reverse ETL**

**Tárolási formátum**

**Tranzakciók**

**Adatverziózás**

**Tárolási hely**

# Unbundling vs. Bundling



# Unbundling vs. Bundling

## The Unbundling of Airflow



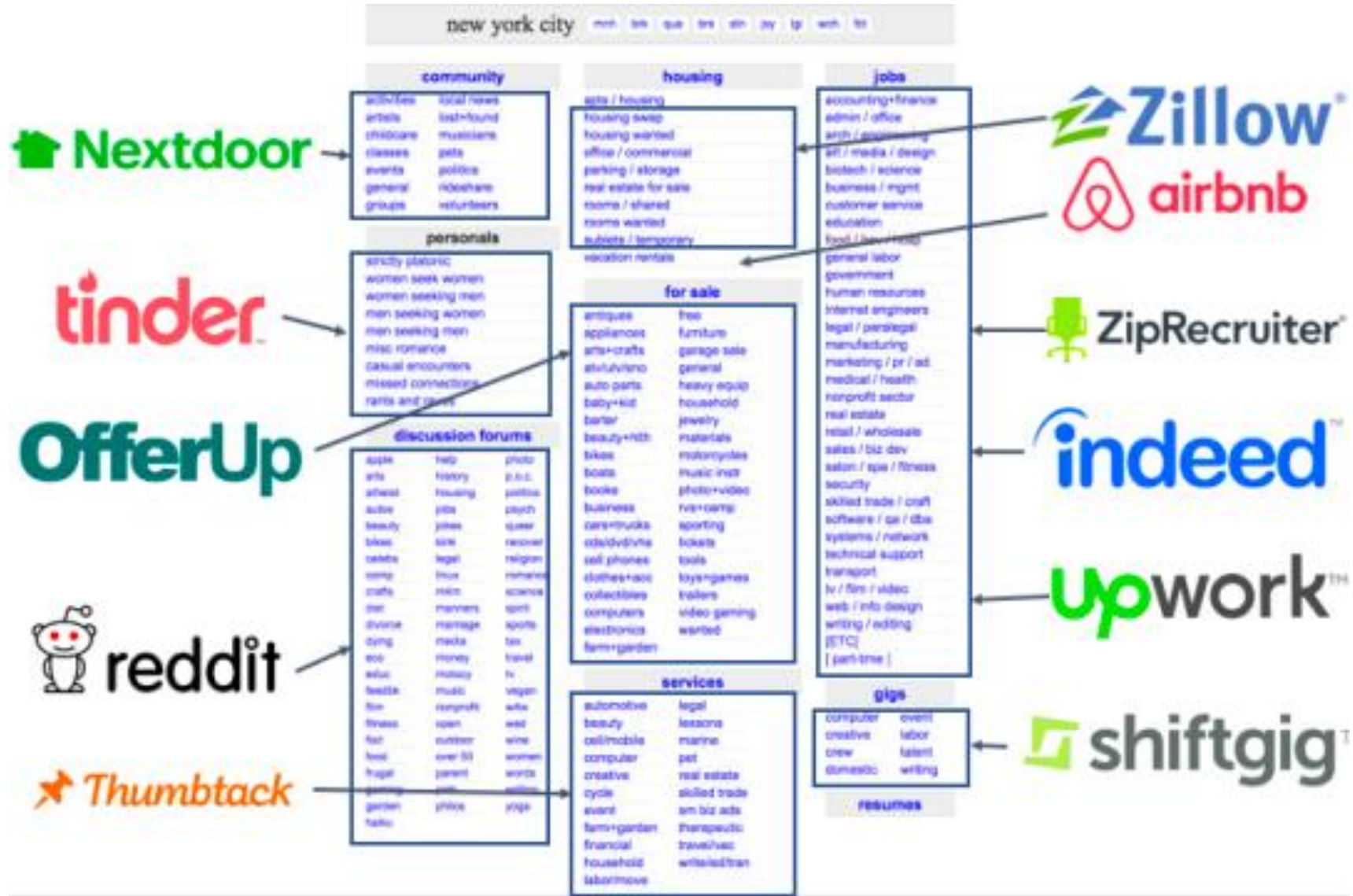
Gorkem Yurtseven

Feb 15, 2022 • 4 min read

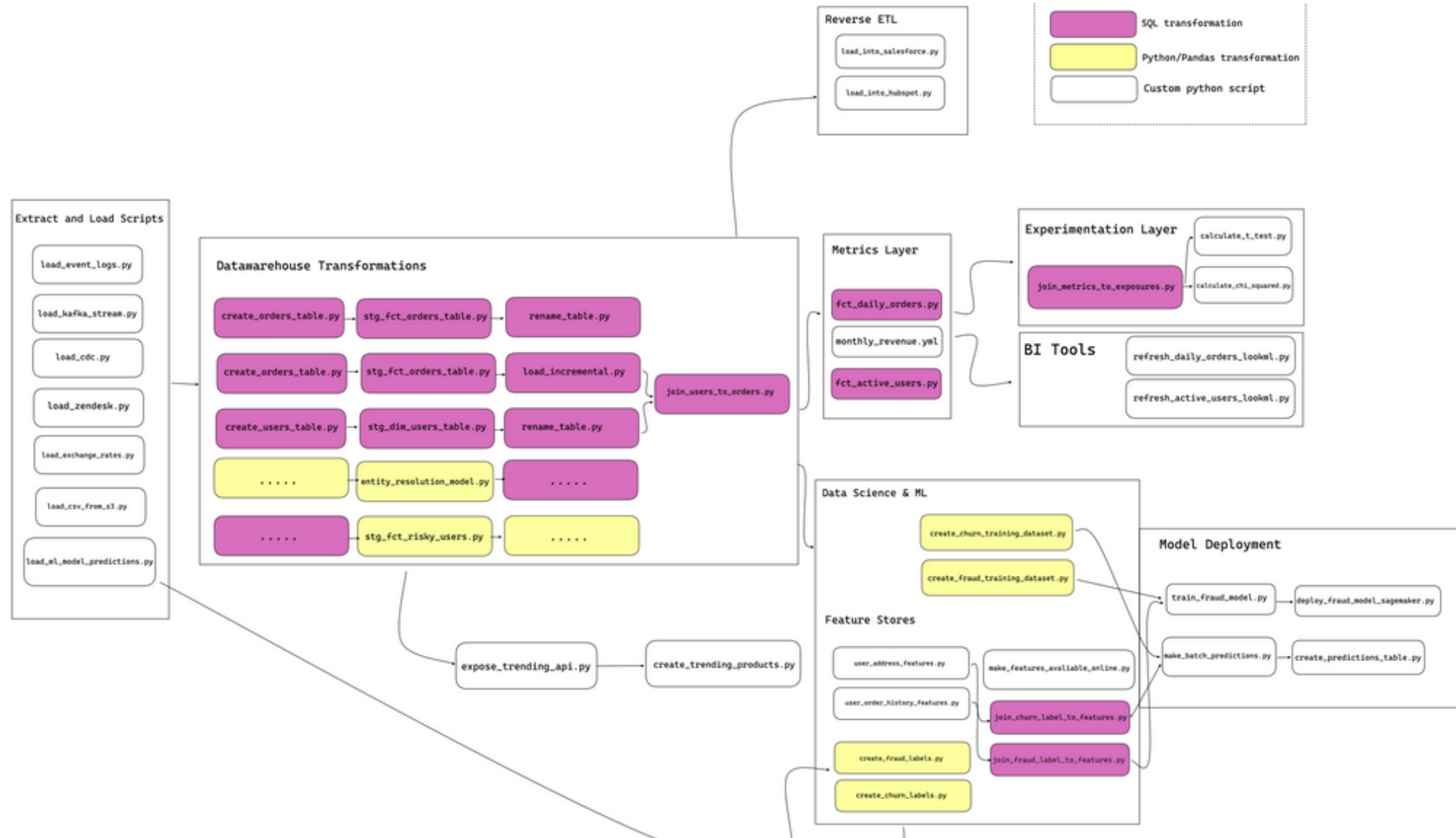
☆ Star us on GitHub 498

In each subdomain of software, we have seen the same story over and over again. Products start small, in time, add adjacent verticals and functionality to their offerings and end up becoming a **platform**. Once these platforms become big enough, people start to figure out how to better serve neglected verticals or abstract out functionality in order to break it down into purpose-built chunks, and *the unbundling starts*. The most classic example of this is the unbundling of Craigslist. You have probably seen a version of this diagram on the internet somewhere:

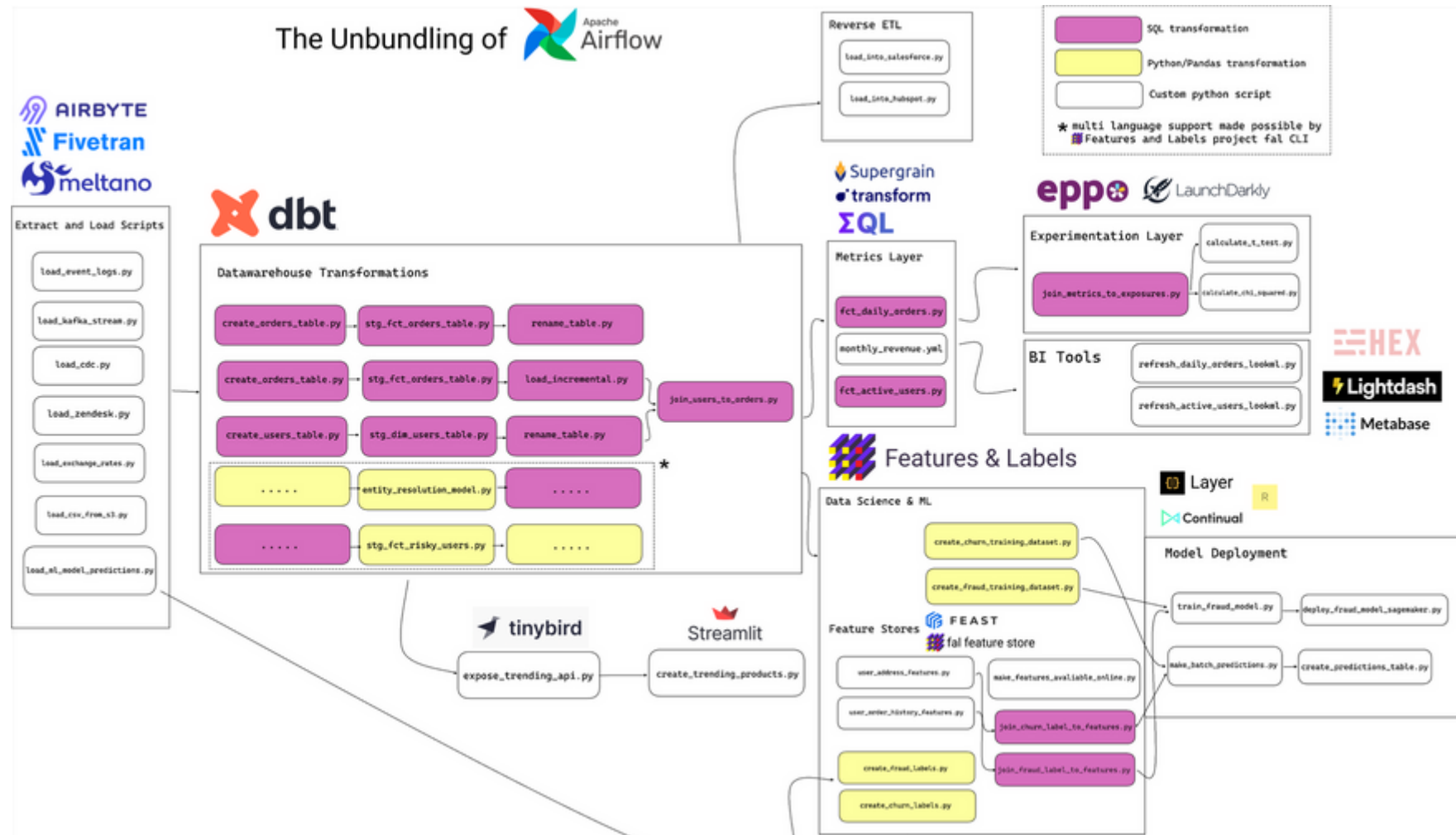
# Unbundling vs. Bundling



# Unbundling vs. Bundling



# Unbundling vs. Bundling



# Unbundling vs. Bundling

February 17, 2022 • 4 minute read

## Rebundling the Data Platform



Nick Schrock  
[@schrockn](#)

---

A fascinating post made the rounds yesterday in the data community entitled [The Unbundling of Airflow](#). It is provocative and deserves the high level of interest it earned.

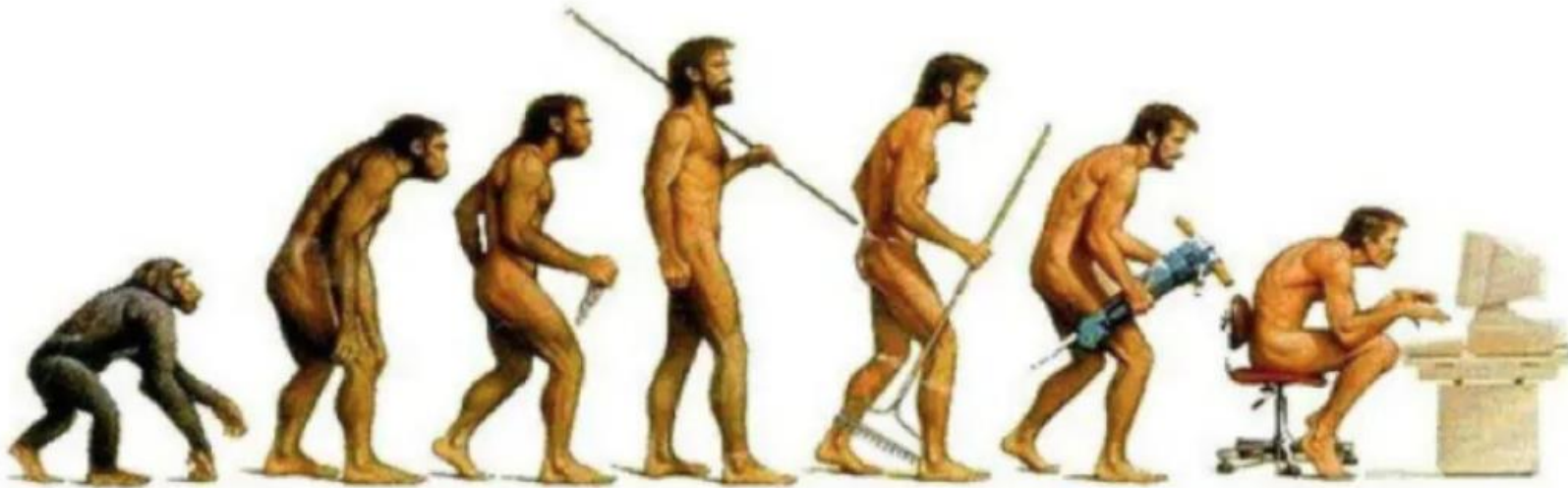
The article describes how the Airflow DAG is being “eaten” from the inside out. Whether it is ingestion tools (Airbyte, Fivetran, Meltano), transformational tools (dbt), reverse ETL tools (Census, Hightouch), metrics layers (Transform), ML-focused systems (Continual) and others: Computations are outsourced to specialized tools that have their own concepts and self-manage some degree of orchestration in their own domain. For example, a dbt project with hundreds of models that would have corresponded to hundreds of manually created Airflow Tasks has been collapsed into a single Airflow task that outsources execution to the dbt runner.

<https://dagster.io/blog/rebundling-the-data-platform>



# Unbundling vs. Bundling

centralized control plane is lunacy, and a terrible endstate for data practitioners and their stakeholders."

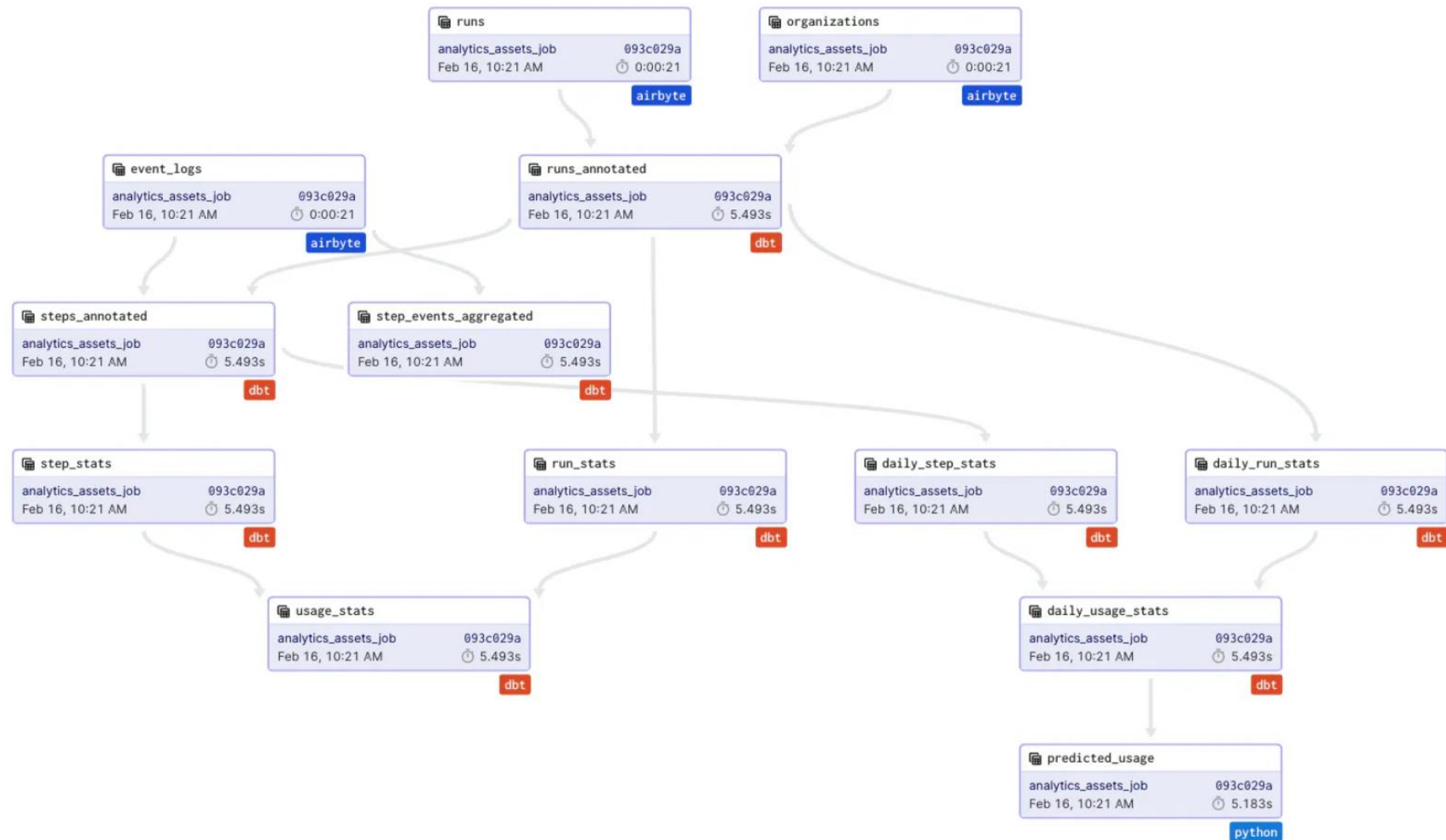


**Overlapping  
Crons**

**Workflow  
Engines**

**Overlapping  
Crons in MDS**

# Unbundling vs. Bundling



# Unbundling vs. Bundling

## **1. There will absolutely be more bundling from our current version of the modern data stack.**

The current version of the modern data stack, with a new company launching every 45 minutes, is unsustainable. We're absolutely in the middle of the golden era of innovation in the MDS, funded quite generously by Venture Capital \$\$ — all in search for the next Snowflake. I've heard stories of perfectly happy (data) product managers in FAANG companies being handed millions of dollars to “try out any idea”.

# Data mesh

# Data Mesh – Principles

## Four principles of Data Mesh

### Domain ownership

Reducing the hops between analytical data consumers and data sources.

[Read more >](#)

### Data as a product

Applying design thinking for data assets. Encapsulating related code, policies and infrastructure in a cohesive product.

[Watch video >](#)

### Self-serve data platforms

Removing friction and technological complexities from the interaction between data producers and consumers.

[Read more >](#)

### Federated computational governance

Automating data governance policies without a centralized authority.

[Read more >](#)

# Data Mesh – Principles

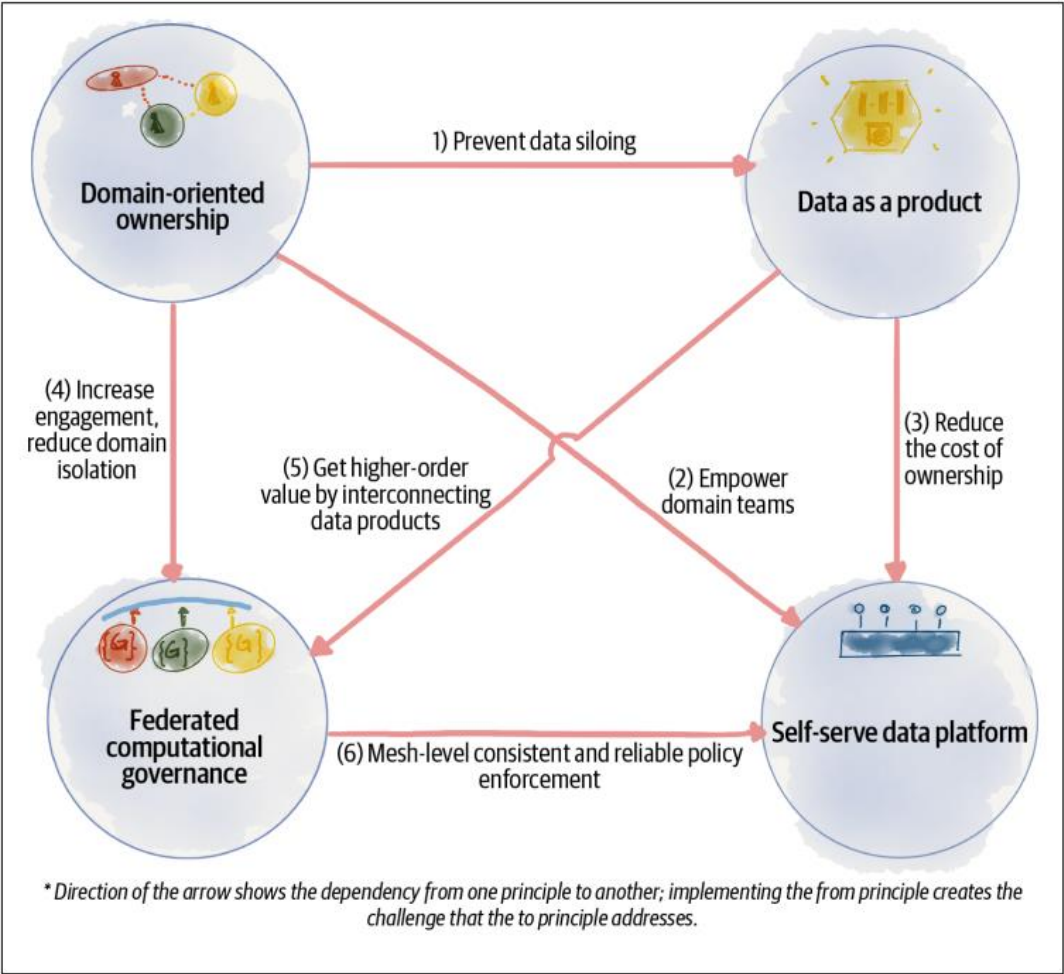


Figure 1-2. Four principles of data mesh and their interplay

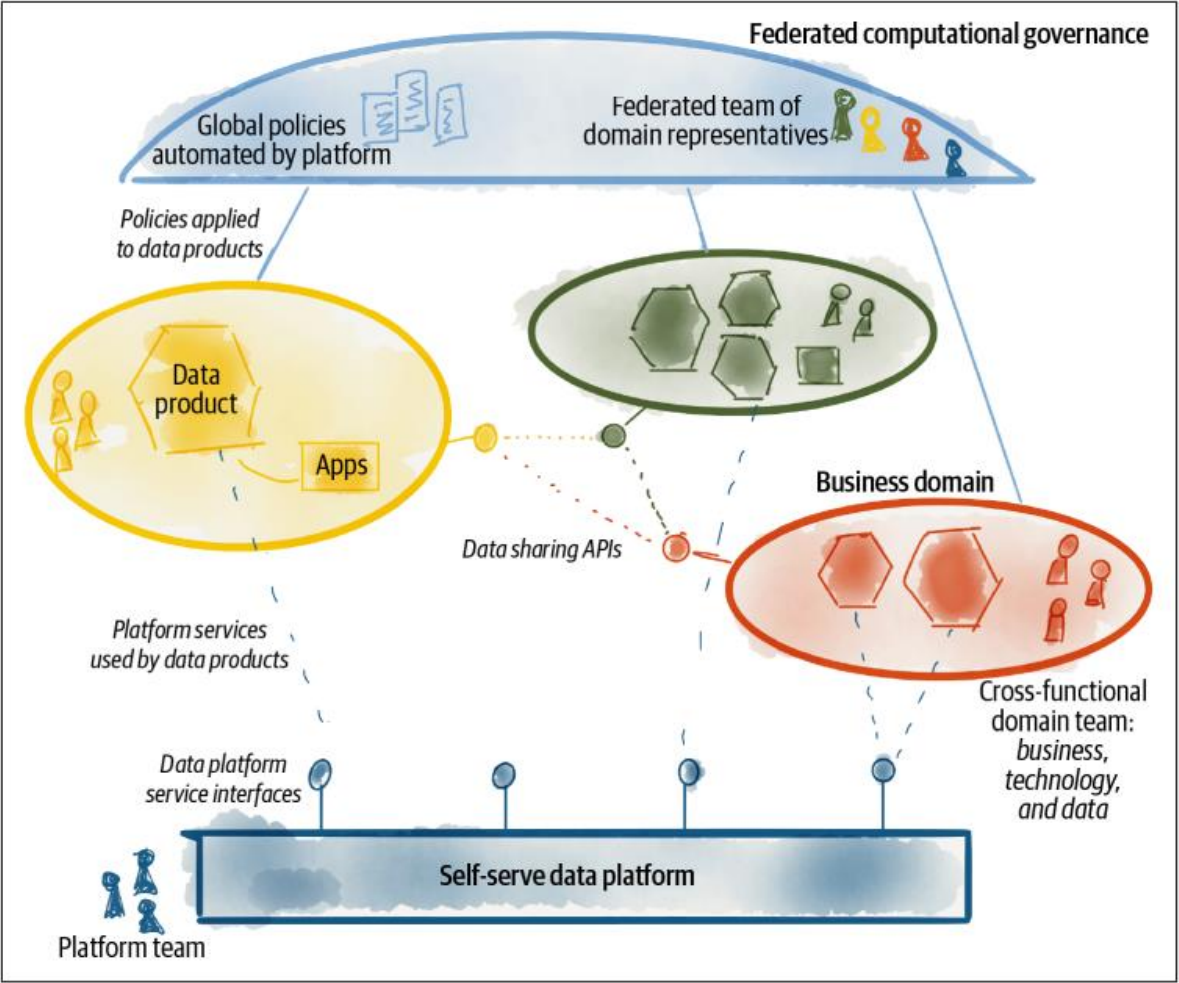
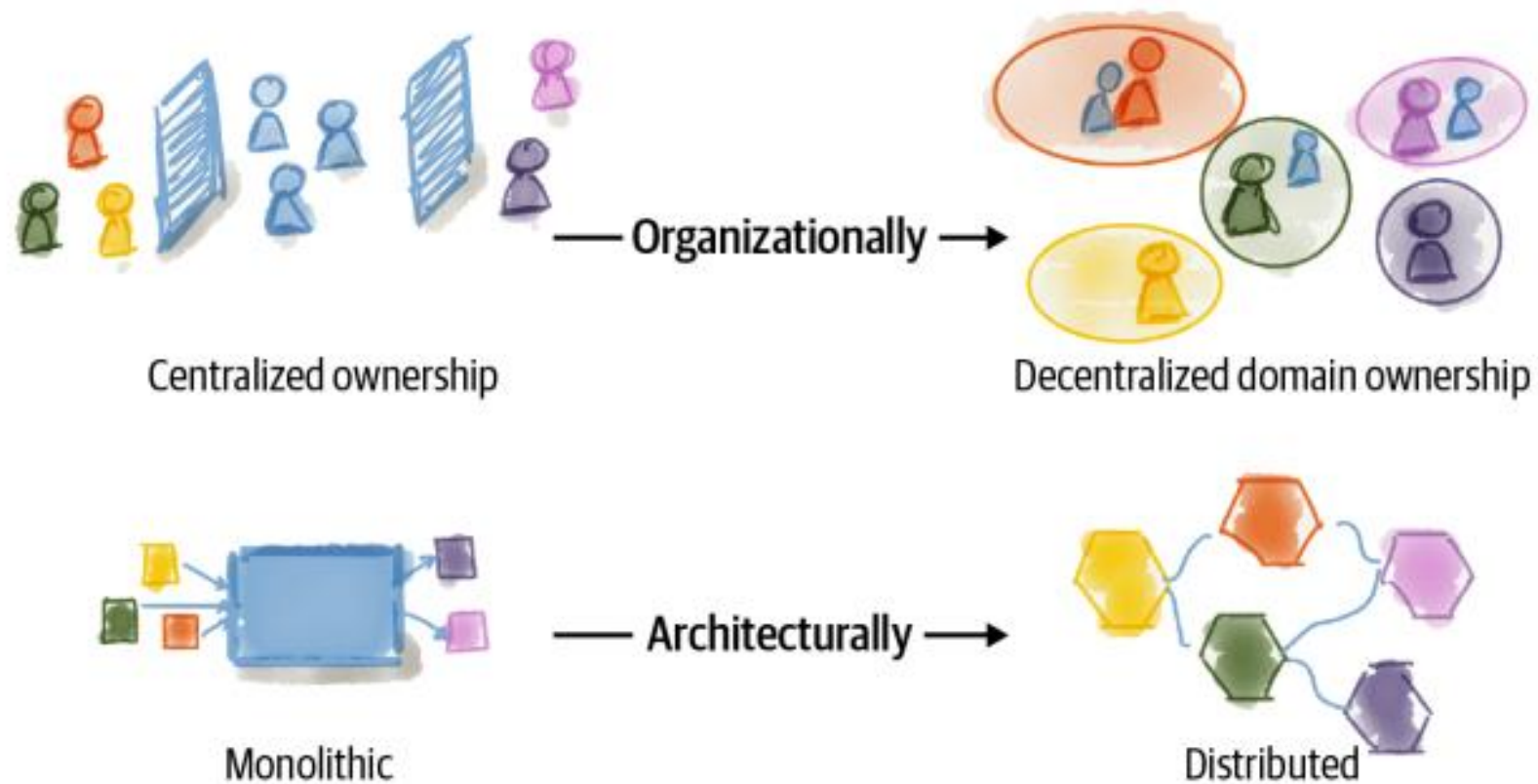


Figure 1-3. Operating model of data mesh principles

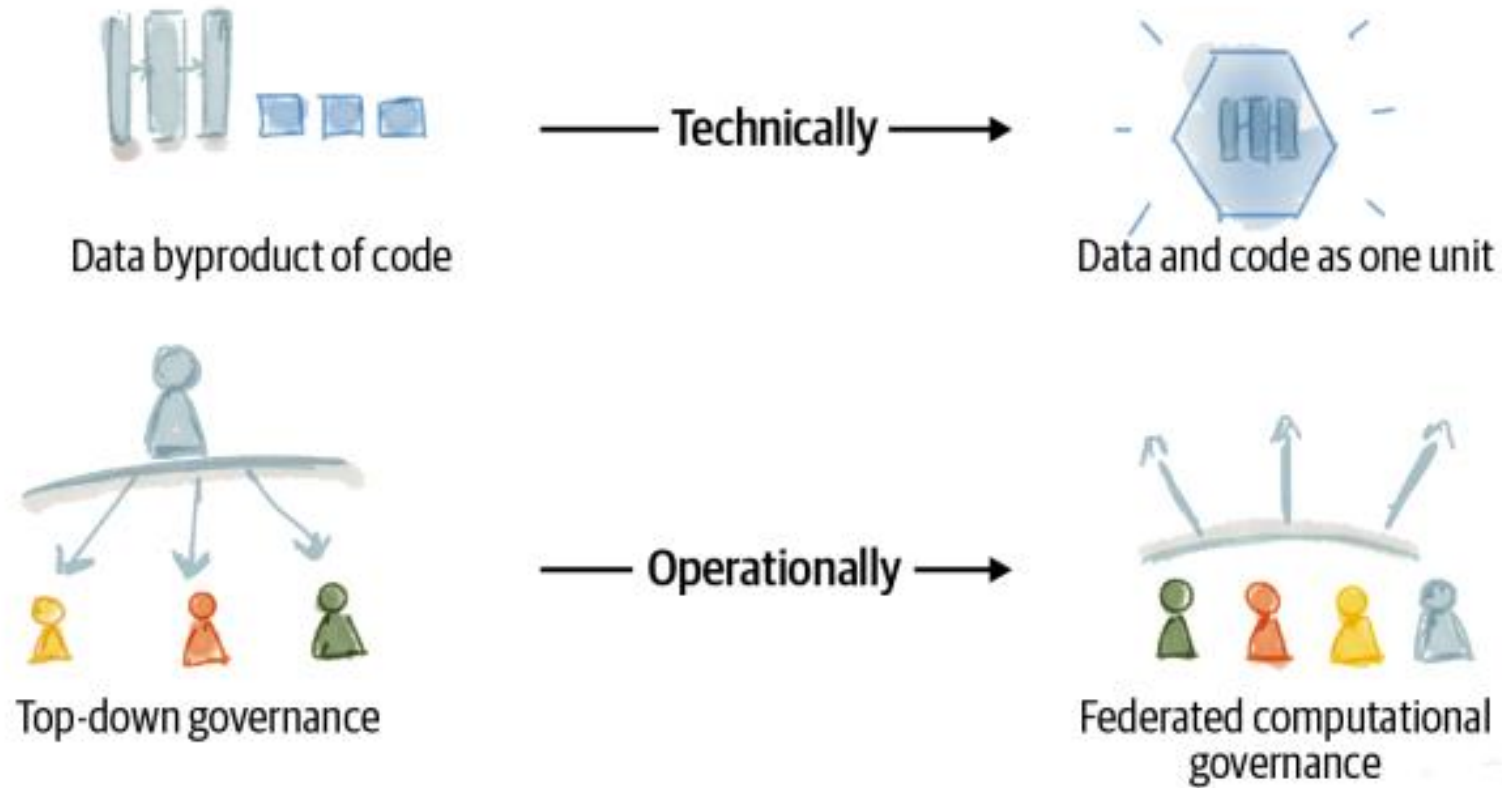


# Data Mesh – Shifts

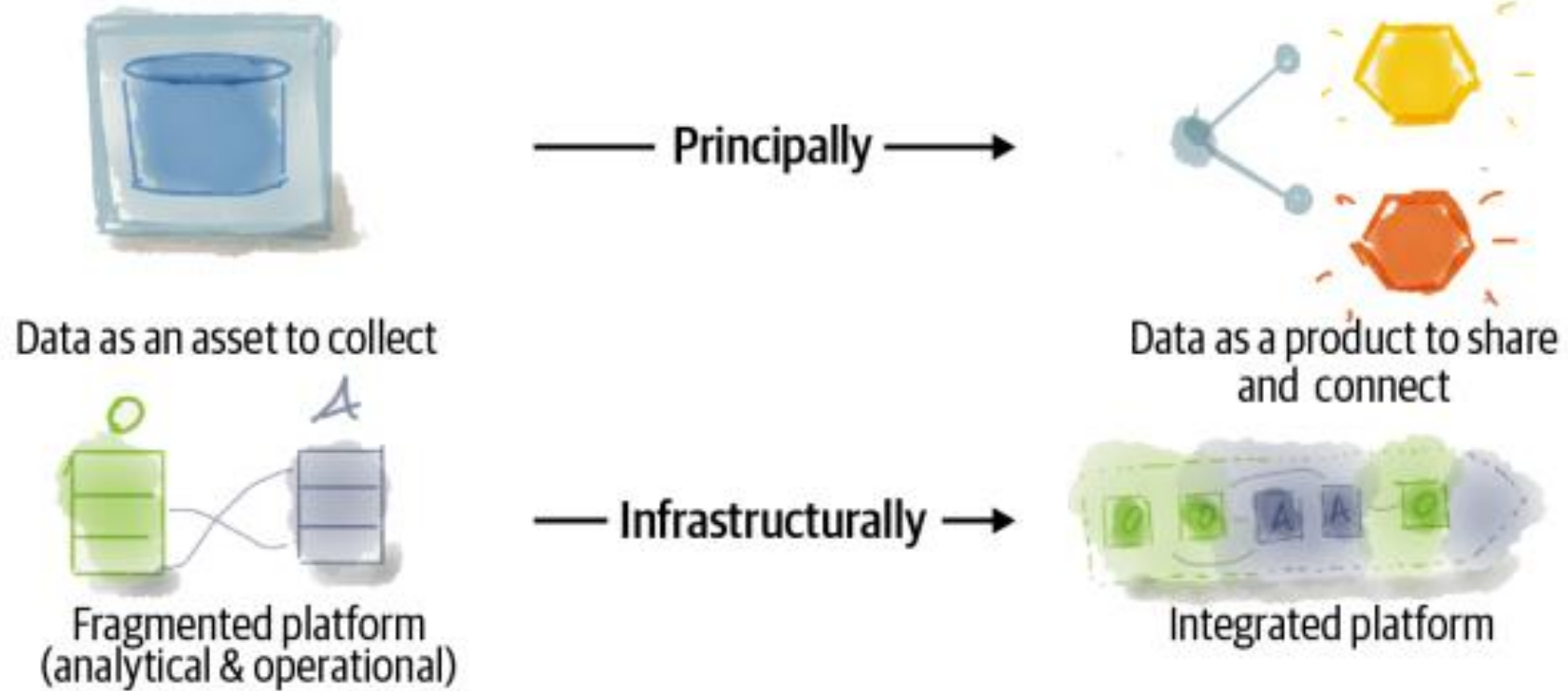




# Data Mesh – Shifts

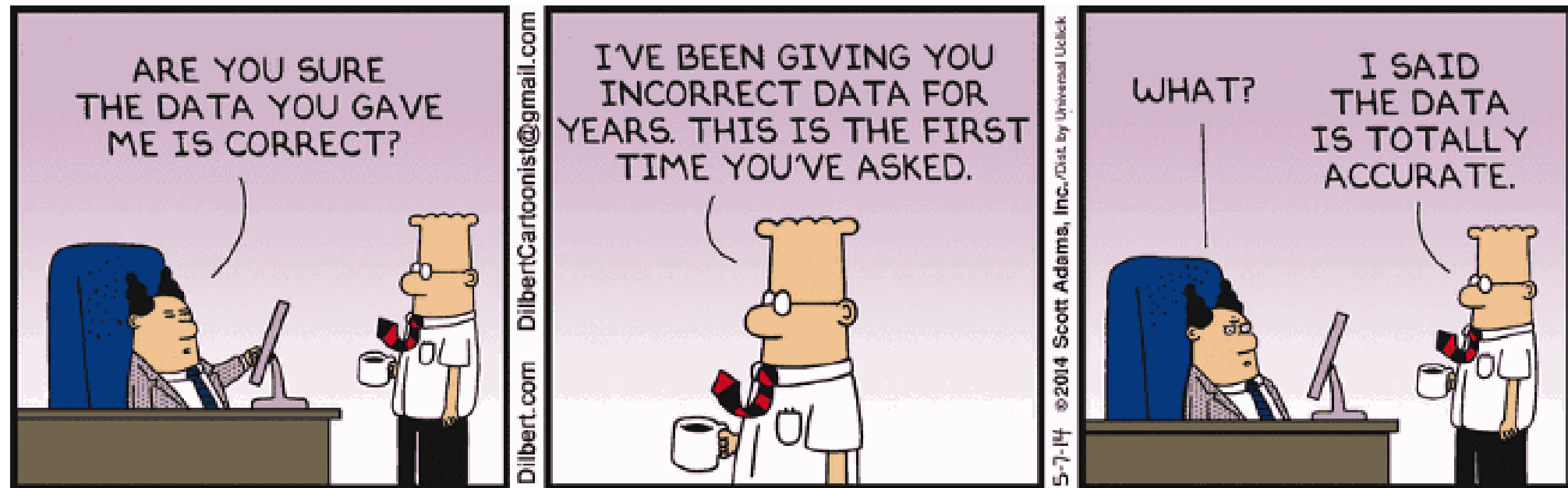


# Data Mesh – Shifts



# Data Quality 2.0

# Data Quality 2.0



## **Data Quality Monitoring is Dead. Say Hello to Full Data Stack Observability**



Post by Salma Bakouk & Benedetta Cittadin

# Data Quality 2.0

- Az egész data stack ellenőrzése és folyamatos monitorozása
- Új terminusok
  - Data Observability
  - Data Reliability
- Nagyon népszerű kategória, számos szereplővel

# Data Quality 2.0


- Részben vagy teljesen open source
  - Superconductive (Great Expectations)
  - Soda
  - Databand
  - Re-data
- Csak fizetős / cloud
  - Monte Carlo Data
  - BigEye
  - Anomalo
  - AccelData
  - Datafold



## Monte Carlo raises \$25M for its data observability service

Alex Wilhelm @alex / 3:01 PM GMT+1 • February 9, 2021

 Comment

 **Image Credits:** baranozdemir / Getty Images

This morning [Monte Carlo](#), a startup focused on helping other companies better monitor their data inflows, announced that it has closed a \$25 million Series B.

The round, which was co-led by [GGV](#) and [Redpoint](#), comes mere months after its [September Series A that was worth \\$15 million](#). [Accel](#) led the company's Series A and seed deals, participating in its Series B as well.

The round caught our attention not only for the speed at which it was raised following Monte Carlo's preceding investment, but also because your humble servant had no idea what data observability, the startup's niche, really was.

So we got Monte Carlo co-founder and CEO [Barr Moses](#) on a call to explain both her company's space, and how it managed to attract so much more capital so quickly.

## Monte Carlo closes \$60M Series C on the back of rapid ARR growth

Alex Wilhelm @alex / 2:00 PM GMT+2 • August 17, 2021

 Comment

 Image Credits: Vithun Khamson / Getty Images

Monte Carlo has returned to the capital well for the second time this year, adding \$60 million to its accounts thanks to a new, ICONIQ Growth-led Series C. The company helps customers monitor their data inflows, finding issues and errors that could foul downstream data analysis.

The company raised a Series B earlier this year, a \$25 million round led by GGV and Redpoint. Both Series B lead investors participated in Monte Carlo's Series C, joined by Salesforce Ventures, a new investor in the company.

### Monte Carlo raises \$25M for its data observability service



This morning Monte Carlo, a startup focused on helping other companies better monitor their data inflows, announced that it has closed a \$25 million Series B. The round, which was co-led by GGV and Redpoint, comes mere months after its September Series A that was worth \$15 million. Accel led the company's Series A and ... [Continue reading](#)

 TechCrunch

 0 


## Monte Carlo Raises \$135M Series D to Accelerate the Rapid Growth of the Data Observability Category

Share article



Today, I'm excited to announce that Monte Carlo, the data reliability company, has raised **\$135M in Series D funding** from IVP, with participation from Accel, GGV Capital, Redpoint Ventures, ICONIQ Growth, Salesforce Ventures, and GIC Singapore. With this round, we've raised a total of \$236M in a 20-month period, most recently announcing their Series C in August 2021 and a suite of new product functionalities to help data teams achieve more reliable data.

# Monte Carlo

 Monte Carlo announces \$135M Series D funding to accelerate the growth of data observability → [Learn More](#)

**MC** MONTE CARLO

[Product](#)

[Docs](#)

[Customers](#)

[About us](#)

[Resources](#)

[Request a demo](#)

[Sign in](#)

## Data Reliability Delivered.

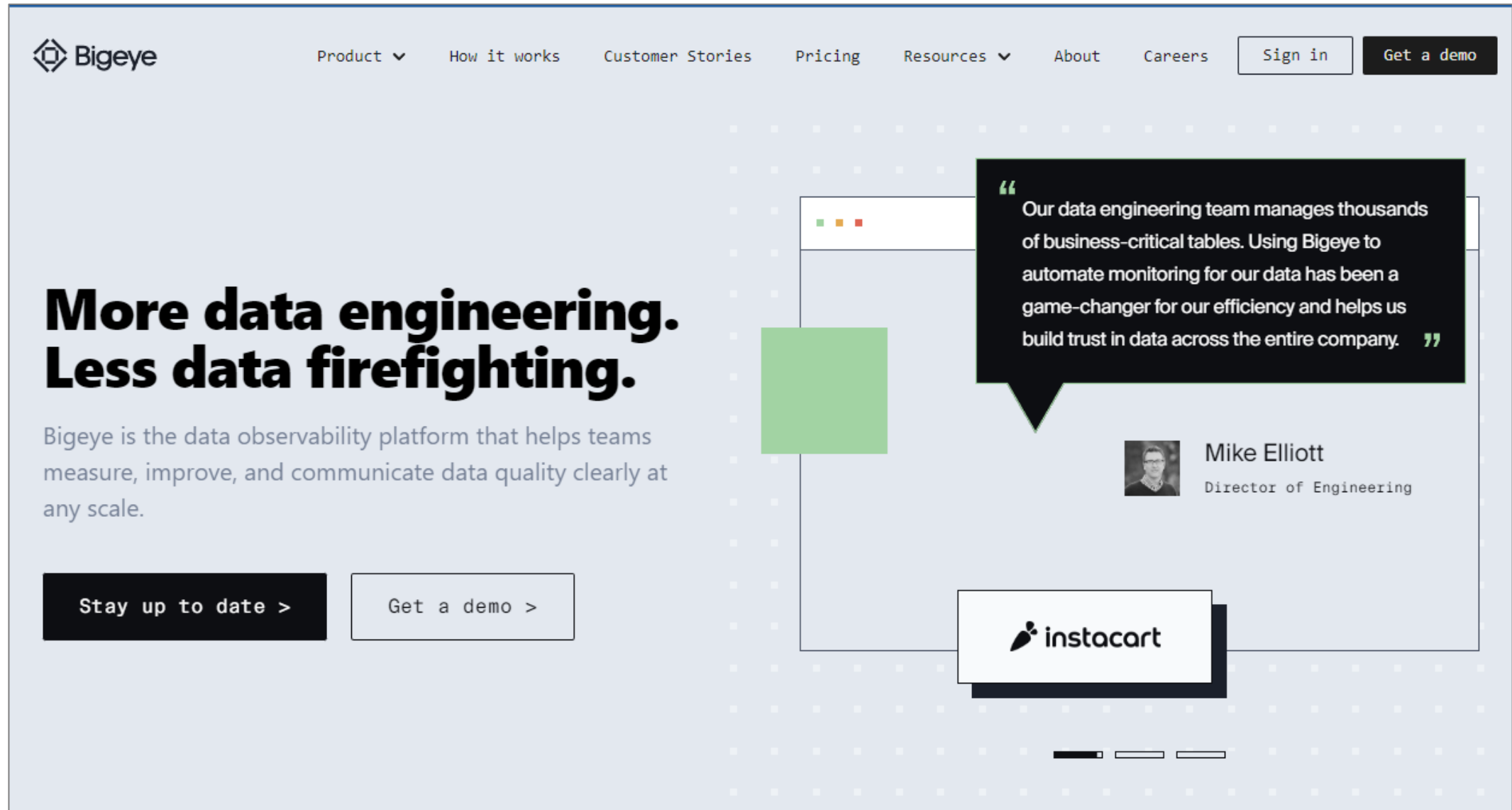
Data breaks. We ensure your team is the first to know and the first to solve with end-to-end data observability.


[Product tour](#)

[▶ What we do](#)



<https://www.montecarlodata.com/>

The image shows the BigEye website landing page. The background is a light blue-grey color with a subtle grid of small white squares. On the left side, there is a large heading "More data engineering. Less data firefighting." in bold black text. Below it, a paragraph describes BigEye as a data observability platform. At the bottom left, there are two buttons: "Stay up to date >" and "Get a demo >". On the right side, there is a testimonial from Mike Elliott, Director of Engineering at Instacart. The testimonial is enclosed in a dark grey speech bubble. Above the speech bubble, there is a small window icon with three colored dots (green, orange, red) and a green square. Below the speech bubble, there is the Instacart logo. At the bottom right, there are three small horizontal bars of different lengths (black, white, white).

 **Bigeye**


[Product](#) [How it works](#) [Customer Stories](#) [Pricing](#) [Resources](#) [About](#) [Careers](#) [Sign in](#) [Get a demo](#)


## More data engineering. Less data firefighting.

Bigeye is the data observability platform that helps teams measure, improve, and communicate data quality clearly at any scale.

[Stay up to date >](#) [Get a demo >](#)

**“** Our data engineering team manages thousands of business-critical tables. Using Bigeye to automate monitoring for our data has been a game-changer for our efficiency and helps us build trust in data across the entire company. **”**

 **Mike Elliott**  
Director of Engineering



# Tőkepiac

# Tőkepiac – 2021 első félév

*Data engines and platforms*  
Starburst · Dremio · Databricks · StarTree · Firebolt · Imply

*Databases*  
Cockroach Labs · TigerGraph · Yugabyte · Timescale · Neo4j

*ETL and workflow*  
Prefect · Matillion · Airbyte

*Data Quality and observability*  
Monte Carlo · BigEye

*Data catalog and metadata*  
Alation · Acryl Data · Atlan · Stemma

*Other*  
Snowplow · Jedox · SafeGraph · Streamlit





# Tőkepiac – 2021 teljes év

## Fundraising by data companies in 2021

Data platform and infra	Databricks \$1B \$1.6B	Dremio \$135M	AnyScale \$100M	StreamNative \$23M	Treeverse \$23M	Coiled \$21M
Database & SQL	Neo4j \$325M	Clickhouse \$50M \$250M	YugaByte \$48M \$188M	Cockroach Labs \$160M	Firebolt \$127M	Redis Labs \$110M
	TigerGraph \$105M	SingleStore \$80M	SingleStore \$80M	Imply \$70M	TimeScaleDB \$40M	ArangoDB \$27,8M
	StarTree \$24M	Ahama \$20M	QuestDB \$12M	MemGraph \$9,3M		
ETL & reverse ETL	Fivetran \$656M	AirByte \$5.2M \$26M \$150M	Dbt Labs \$150M	Matillion \$100M	Hightouch \$12M \$40M	Prefect \$11.5M \$32M
	Hevo Data \$30M	Census \$16M	Elementl \$14M	Meltano \$4.2M		
Data quality & observability	Monte Carlo \$25M \$60M	BigEye \$17M \$45M	AccelData €35M	Anomalo €33M	Datafold €20M	Soda €11.5M
Metadata & data governance	Collibra \$250M	Alation \$100M	Atlan \$16M	Acryl Data \$9M	Stemma \$4.8M	
BI & Dataviz	Grafana \$220M	ToughtSpot \$100M	Jedox \$100M	Preset \$35.9M	Metabase \$30M	Noteable \$21M
AI & ML	Dataiku \$400M	Scale.ai \$325M	DataRobot \$300M	Weights & Biases \$135M	Snorkel.ai \$35M \$85M	H2O.AI \$100M
	Sama \$70M	Comet \$50M+ \$13M	Gretel.ai \$50M	Streamlit \$35M	Tonic.ai \$35M	Hex \$5.5M \$21M

adat.blog/funding · Created by @BenceArato

# Tőkepiac – 2021 teljes év

## Data platform and infra

Databricks  
\$1B \$1.6B

Dremio  
\$135M

AnyScale  
\$100M

StreamNative  
\$23M

Treeverse  
\$23M

Coiled  
\$21M

## Database & SQL

Neo4j  
\$325M

Clickhouse  
\$50M \$250M

YugaByte  
\$48M \$188M

Cockroach Labs  
\$160M

Firebolt  
\$127M

Redis Labs  
\$110M

TigerGraph  
\$105M

SingleStore  
\$80M

SingleStore  
\$80M

Imply  
\$70M

TimeScaleDB  
\$40M

ArangoDB  
\$27,8M

StarTree  
\$24M

Ahama  
\$20M

QuestDB  
\$12M

MemGraph  
\$9,3M

## ETL & reverse ETL

Fivetran  
\$656M

AirByte  
\$5.2M \$26M \$150M

Dbt Labs  
\$150M

Matillion  
\$100M

Hightouch  
\$12M \$40M

Prefect  
\$11.5M \$32M

Hevo Data  
\$30M

Census  
\$16M

Elementl  
\$14M

Meltano  
\$4.2M

# Tőkepiac – 2021 teljes év

## Data quality & observability

Monte Carlo  
\$25M \$60M

BigEye  
\$17M \$45M

AccelData  
€35M

Anomalo  
€33M

Datafold  
€20M

Soda  
€11.5M

## Metadata & data governance

Collibra  
\$250M

Alation  
\$100M

Atlan  
\$16M

Acryl Data  
\$9M

Stemma  
\$4.8M

## BI & Dataviz

Grafana  
\$220M

ToughtSpot  
\$100M

Jedox  
\$100M

Preset  
\$35.9M

Metabase  
\$30M

Noteable  
\$21M

## AI & ML

Dataiku  
\$400M

Scale.ai  
\$325M

DataRobot  
\$300M

Weights & Biases  
\$135M

Snorkel.ai  
\$35M \$85M

H2O.AI  
\$100M

Sama  
\$70M

Comet  
\$50M+ \$13M

Gretel.ai  
\$50M

Streamlit  
\$35M

Tonic.ai  
\$35M

Hex  
\$5.5M \$21M

*adat.blog/funding · Created by @BenceArato*

## 2022. első félév (részleges)

Hónap, év	Cég	Szektor	Mennyiség \$M
February 2022	Starburst	Database/SQL	250
April 2022	Grafana Labs	BI/Dataviz	240
February 2022	dbt labs	Data tranformation	222
March 2022	Astronomer	Data Integration	213
January 2022	Dremio	Data platform	160
May 2022	Monte Carlo	Data Quality	135
January 2022	Labelbox	Data platform	110
February 2022	Voltron Data	General	110
January 2022	Firebolt	Data platform	100
May 2022	Imply	Database/SQL	100

# Meddig mehet ez így?

The current version of the modern data stack, with a new company launching every 45 minutes, is unsustainable. We're absolutely in the middle of the golden era of innovation in the MDS, funded quite generously by Venture Capital \$\$ — all in search for the next Snowflake. I've heard stories of perfectly happy (data) product managers in FAANG companies being handed millions of dollars to “try out any idea”.

# Meddig mehet ez így?



# Meddig mehet ez így?

Piaci összegzés > Microsoft

**244,49** USD

-13,87 (-5,37%) ↓ az elmúlt évben

Zárva: jún. 14. 17:18 GMT-4 • Nyilatkozat

Zárás után 244,49 0,00 (0,00%)

1 N

5 N

1 H

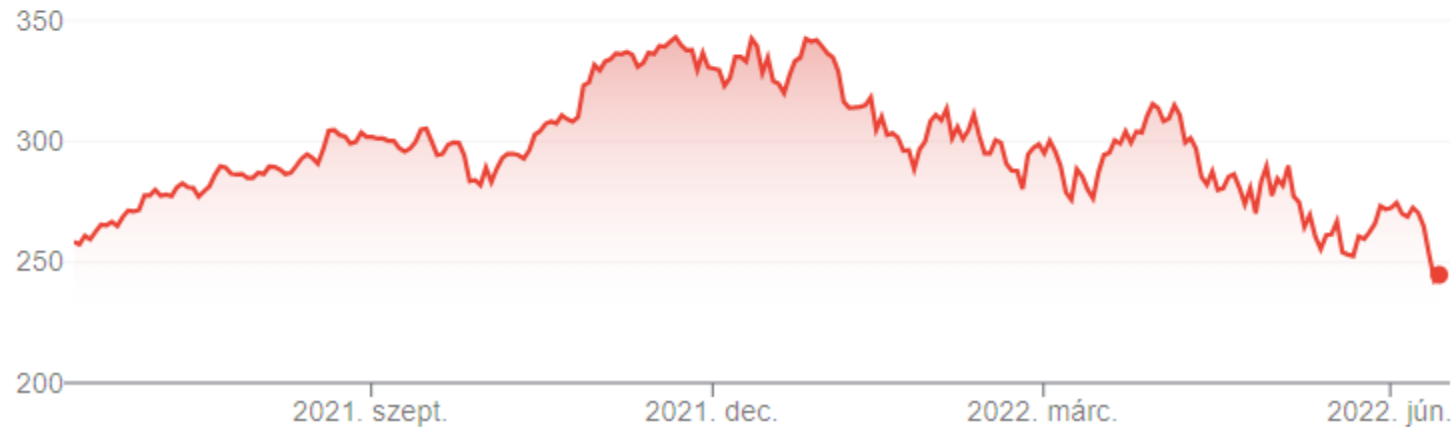
6 H

YTD

1 É

5 É

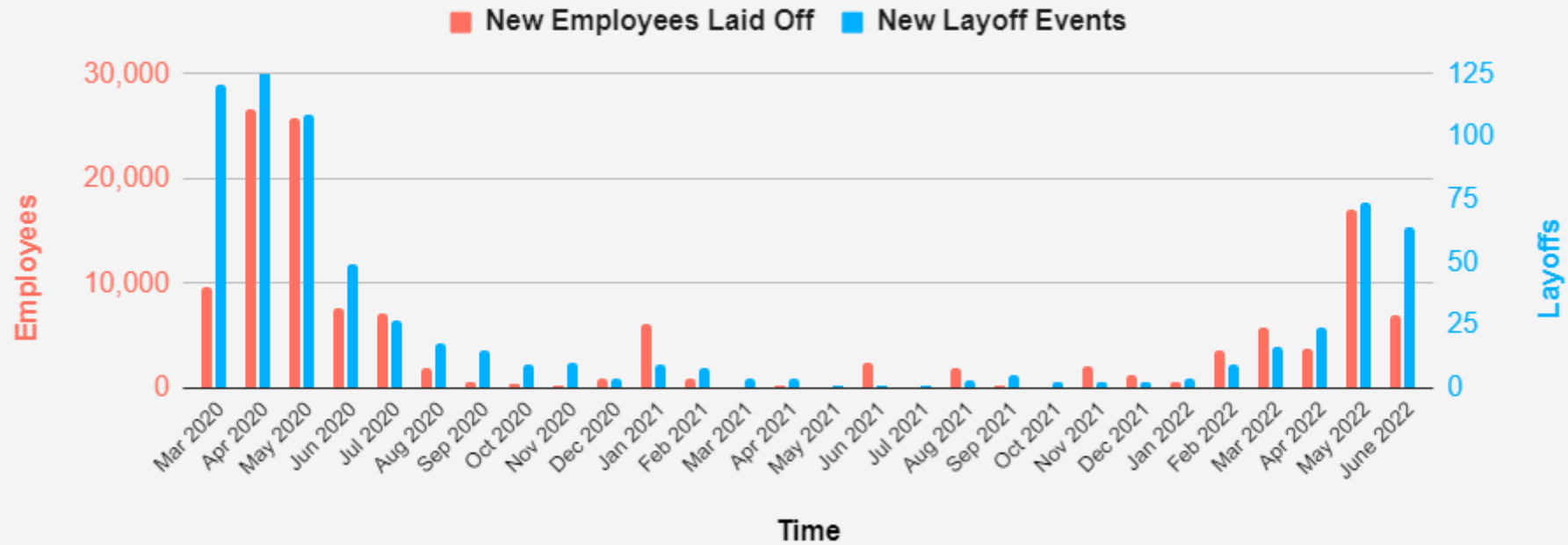
Max.





## New startup layoffs since COVID-19 (by month)

Source: <https://layoffs.fyi>



# Leépítések

<input type="checkbox"/>	Company	Location	# Laid Off	Date	%	Industry	Source
1	Coinbase	SF Bay Area	1100	6/14/2022	18%	Finance	<a href="https://www.cnbc.com/...">https://www.cnbc.com/...</a>
2	Redfin	Seattle	470	6/14/2022	8%	Real Estate	<a href="https://www.streetinsid...">https://www.streetinsid...</a>
3	Compass	New York ...	450	6/14/2022	10%	Real Estate	<a href="https://www.bloomberg...">https://www.bloomberg...</a>
4	Shopee	Singapore		6/14/2022		Food	<a href="https://asia.nikkei.com/...">https://asia.nikkei.com/...</a>
5	BlockFi	New York ...	250	6/13/2022	20%	Finance	<a href="https://blockfi.com/a-m...">https://blockfi.com/a-m...</a>
6	Studio	SF Bay Area	33	6/13/2022	40%	Education	
7	Automox	Boulder		6/13/2022		Infrastruct...	<a href="https://www.linkedin.co...">https://www.linkedin.co...</a>
8	Desktop Metal	Boston		6/13/2022	12%	Other	<a href="https://www.marketwat...">https://www.marketwat...</a>
9	Crypto.com	Singapore	260	6/10/2022	5%	Finance	<a href="https://blockworks.co/c...">https://blockworks.co/c...</a>
10	FarEye	New Delhi	250	6/10/2022		Logistics	<a href="https://inc42.com/buzz/...">https://inc42.com/buzz/...</a>
11	Sanar	Sao Paulo	60	6/10/2022	20%	Healthcare	<a href="https://www.terra.com,...">https://www.terra.com,...</a>
12	Freetrade	London	45	6/10/2022	15%	Finance	<a href="https://www.altfi.com/a...">https://www.altfi.com/a...</a>
13	Albert	Los Angeles	20	6/10/2022	8%	Finance	<a href="https://dot.la/albert-fint...">https://dot.la/albert-fint...</a>
14	Keepe	Seattle		6/10/2022		Real Estate	<a href="https://www.geekwire.c...">https://www.geekwire.c...</a>
15	Liongard	Houston		6/10/2022		Infrastruct...	<a href="https://www.crn.com/n...">https://www.crn.com/n...</a>
16	Ziroom	Beijing		6/10/2022	20%	Real Estate	<a href="https://pandaily.com/zir...">https://pandaily.com/zir...</a>

<https://layoffs.fyi/>

# Mire lehet számítani?

- Ki fog derülni, hogy kinek sikerült valóban jó üzleti modellt találni
- Konszolidációs hullám
  - Nem kell feltétlenül X darab cég / eszköz minden egyes részterületre
  - Nagyok megveszik a kicsiket
- Első jelek már láthatóak
  - Airflow-GROUPAROO (Reverse ETL)
  - Astronomer – Datakin (Data Lineage)
- Az óriáscégek is lehetnek felvásárlók

# Thank You

