databricks

Intelligent Data Engineering

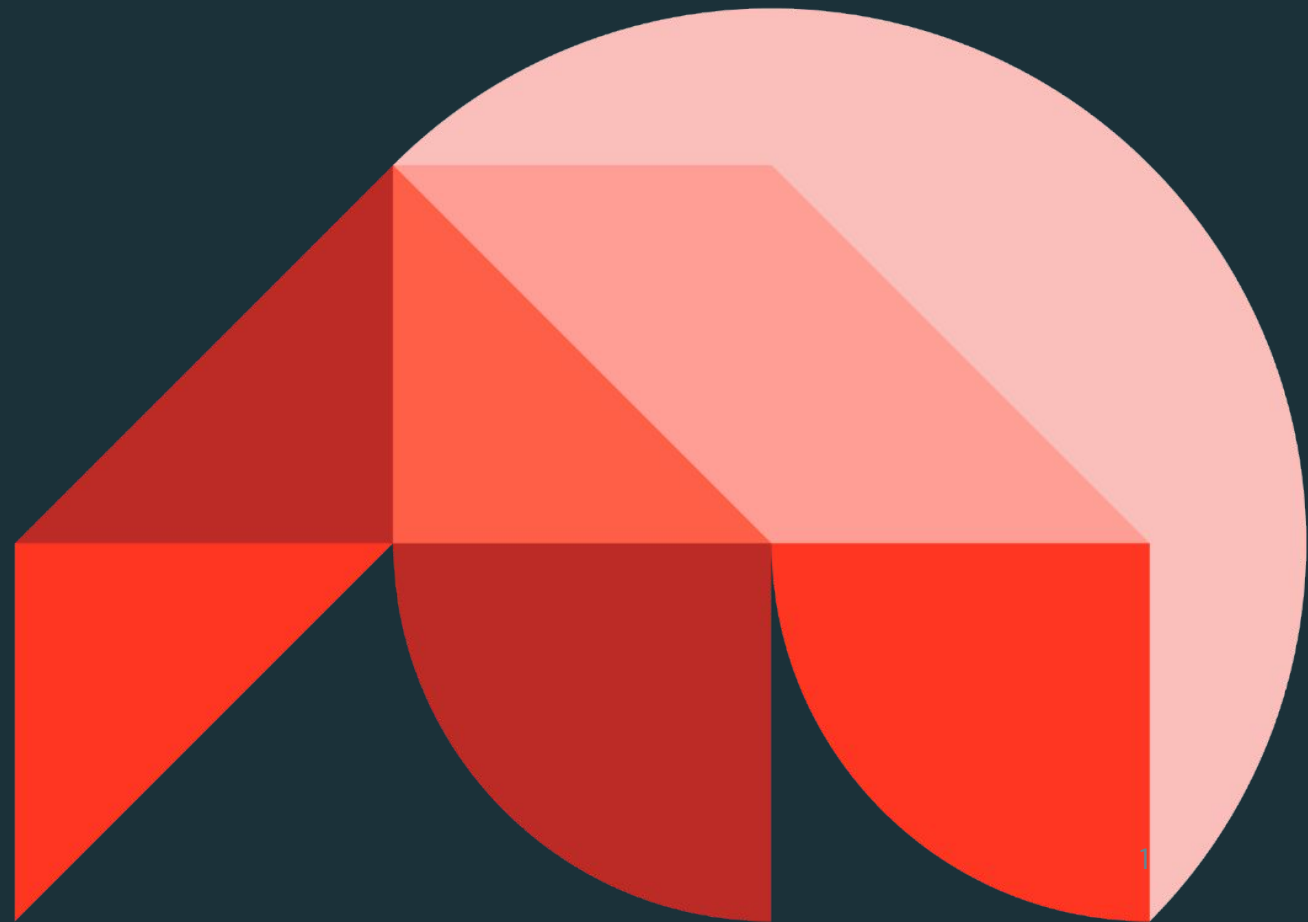
Slide Deck

carahsoft

For more information, contact Carahsoft or our reseller partners:
Databricks@carahsoft.com | 703-581-6693

databricks

Intelligent Data Engineering



Every AI innovation hinges on reliable data

AI initiatives are top of mind...

By 2026, **over 80%** of enterprises will be using GenAI in production environments, up from **less than 5%** in 2023

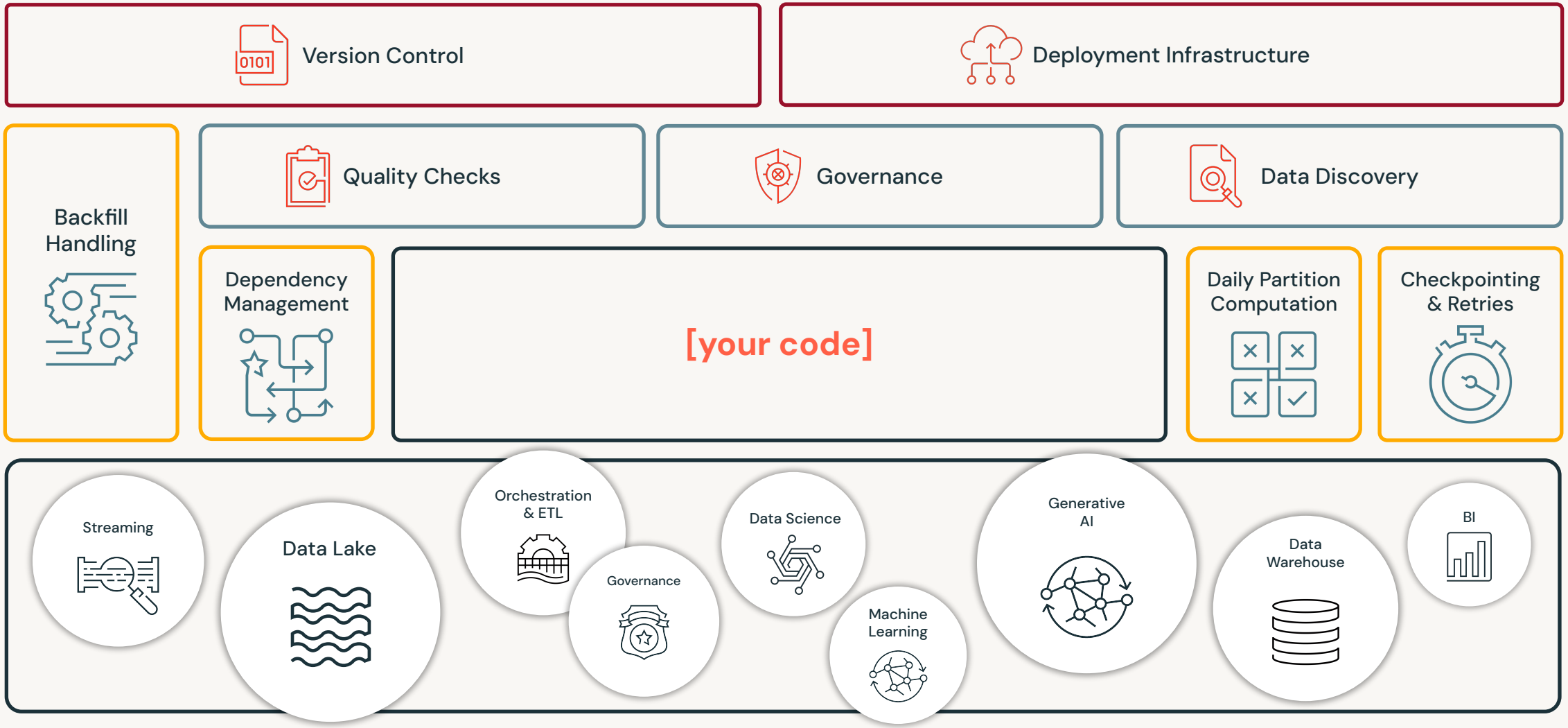
—2023 Gartner Hype Cycle for Generative AI

...but good models can't overcome bad data

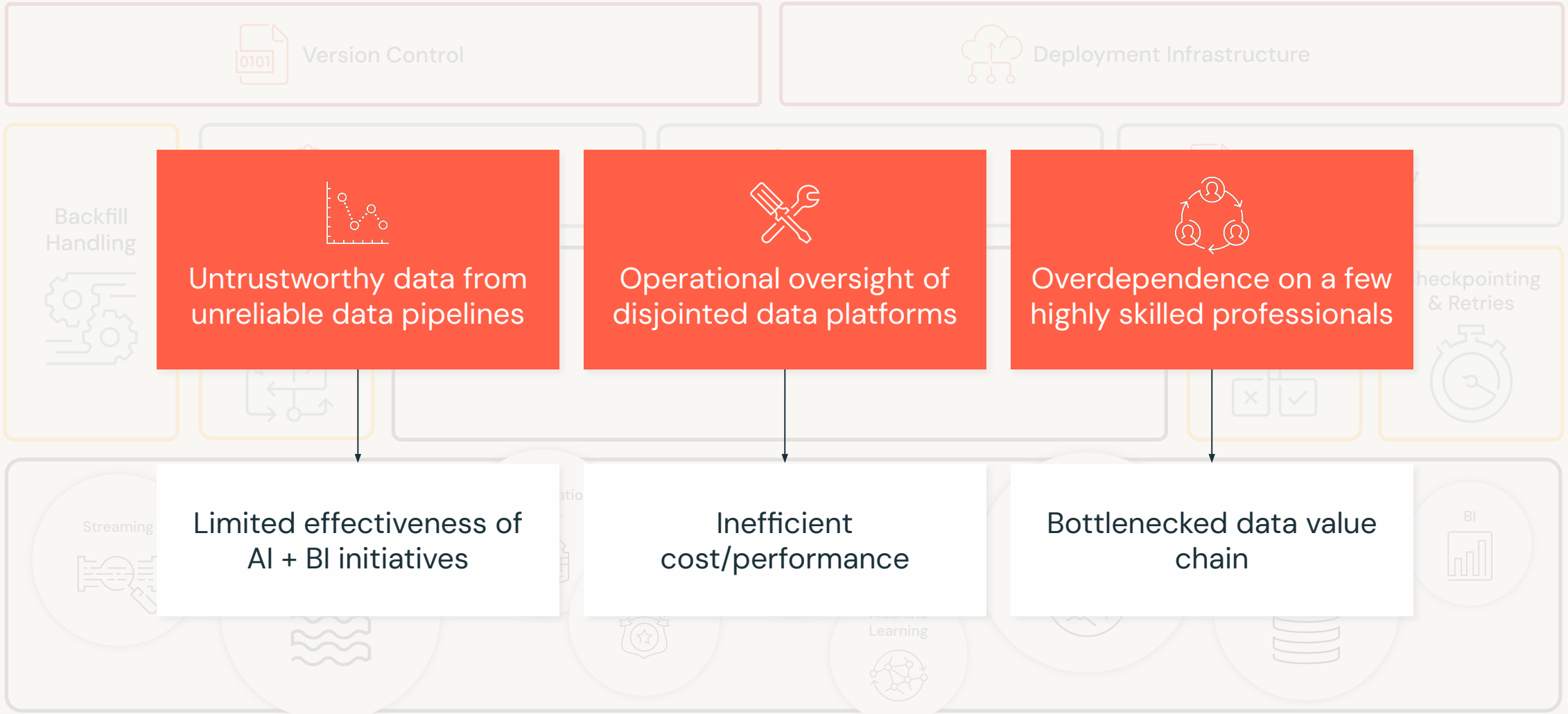
"Data problems are **the most likely factor** to jeopardize our AI/ML goals"

—MIT Technology Review Insights survey, 2022

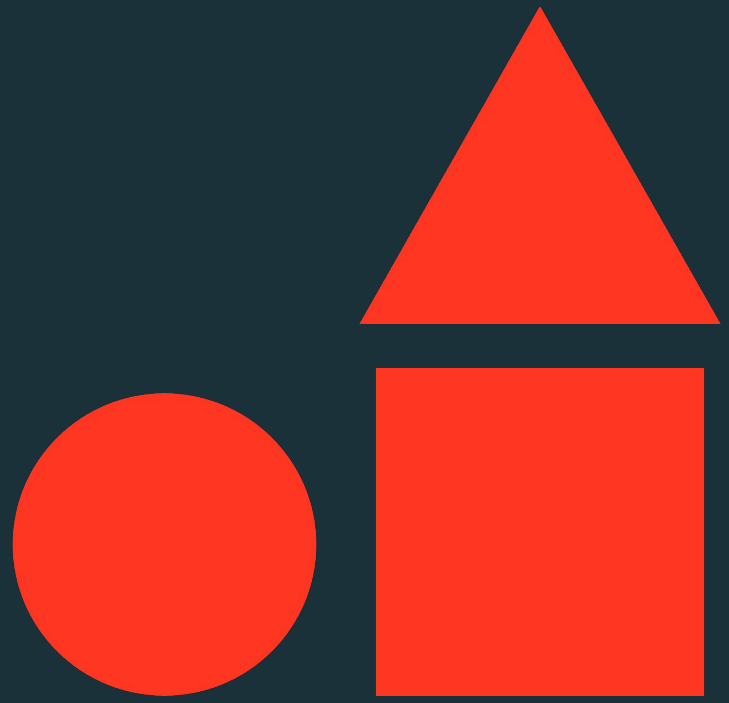
It’s difficult to build and operate reliable data pipelines



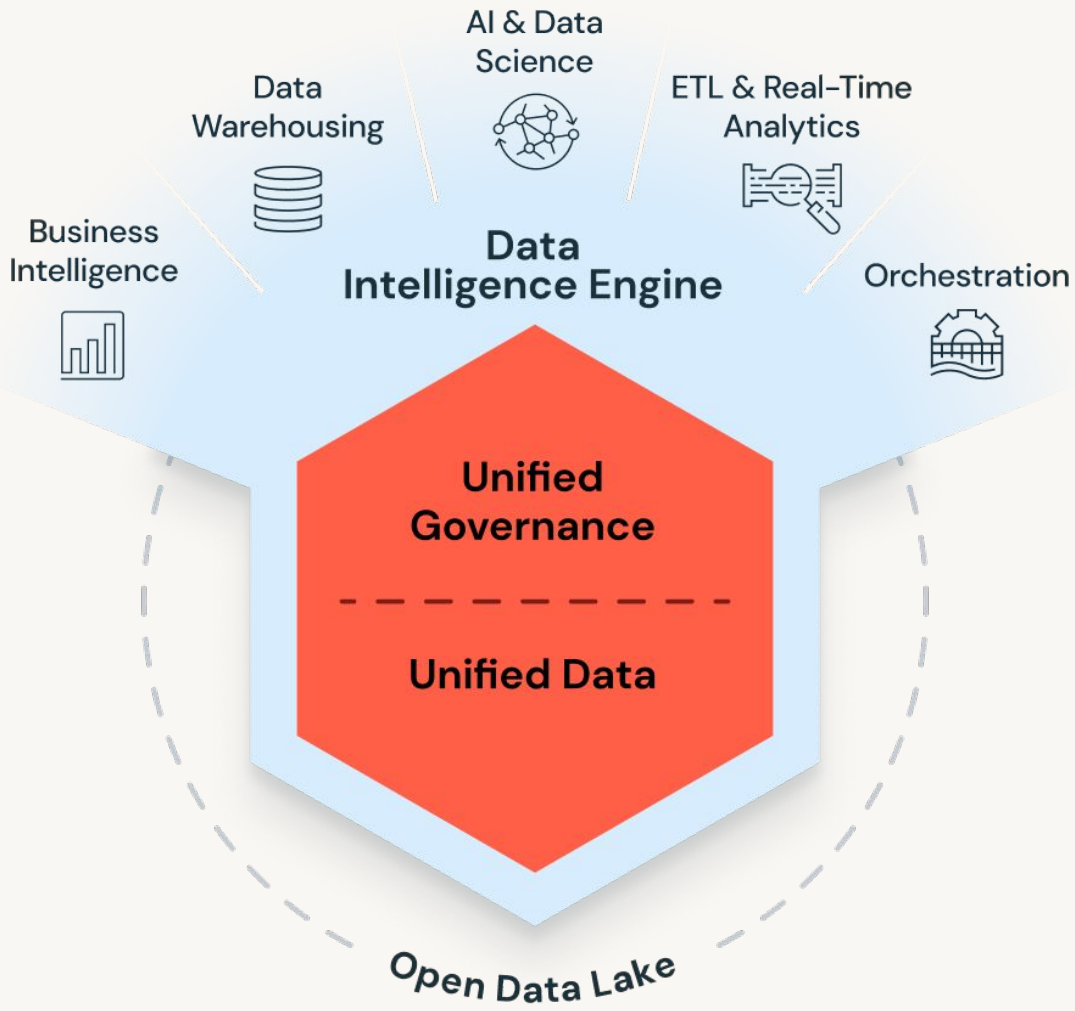
It’s difficult to build and operate reliable data pipelines



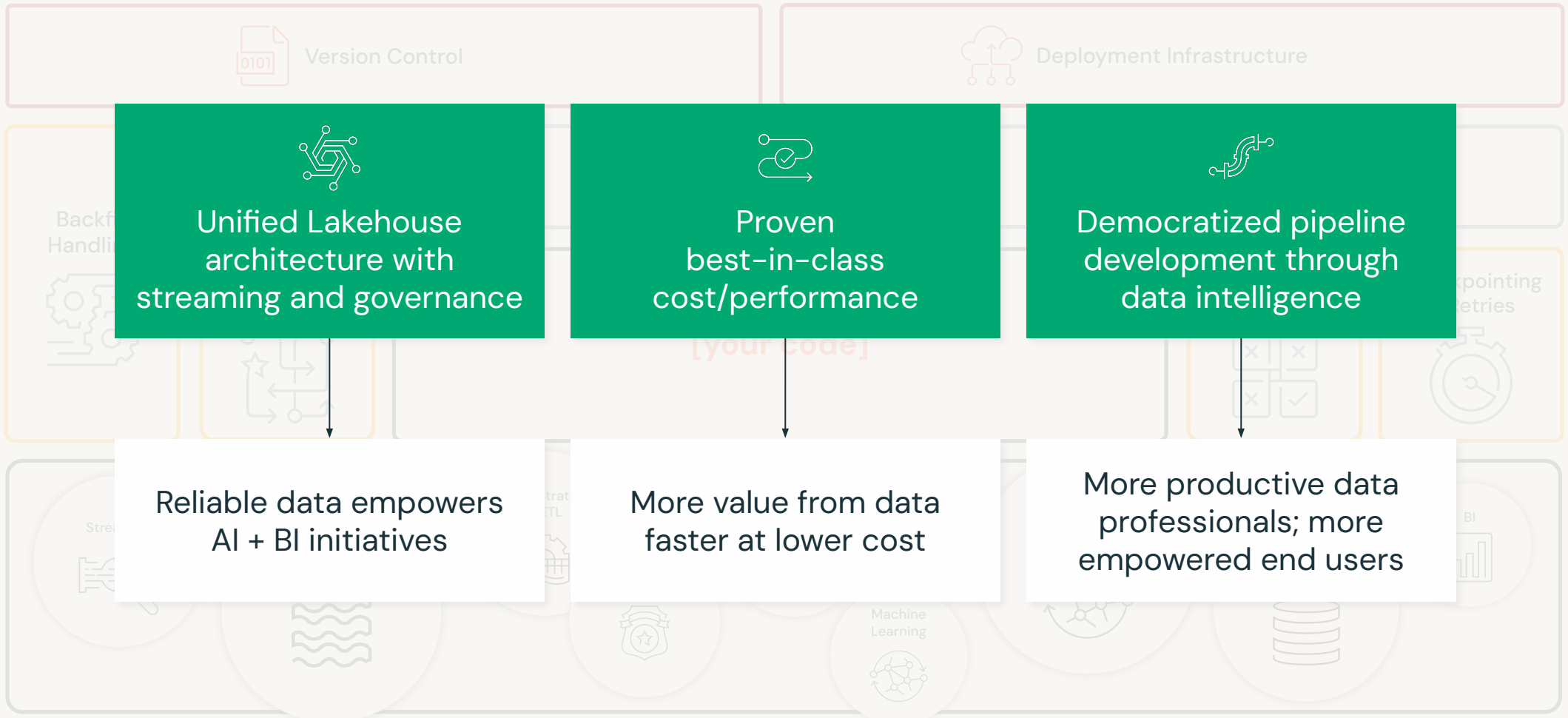
Reliable data pipelines require a unified platform with data intelligence



The Databricks Data Intelligence Platform provides the foundation for Data Engineering in the age of AI

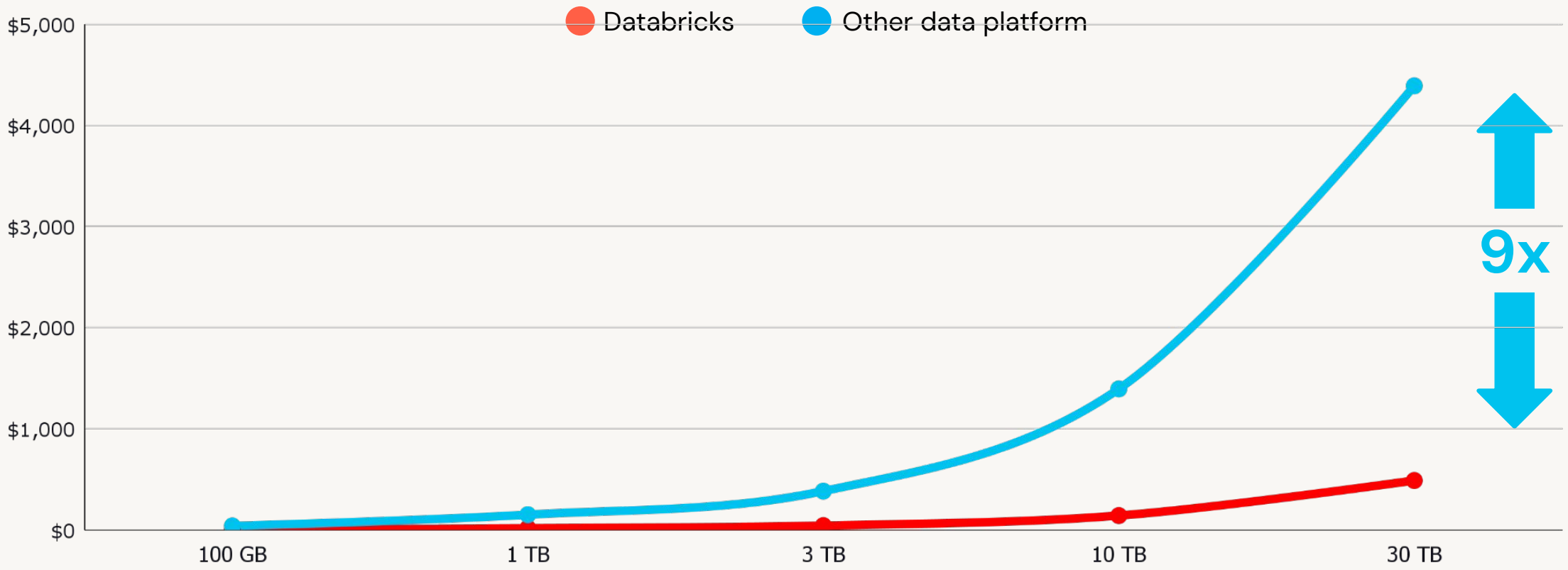


The Databricks Data Intelligence Platform is the best place to do Data Engineering



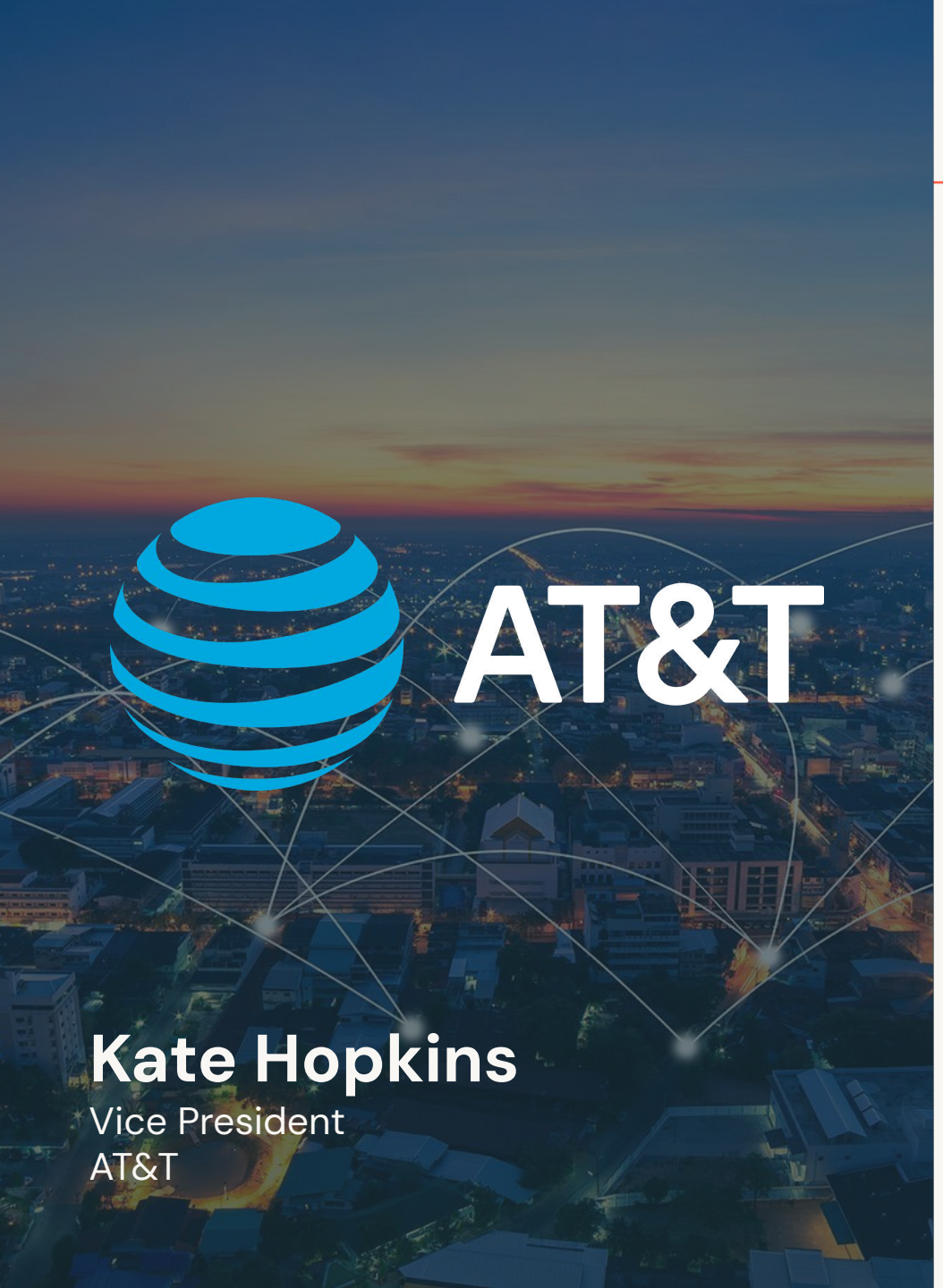
Leading price/performance for ETL

ETL Benchmark shows that Databricks is much more efficient when compared to other data platforms **especially as you scale**



Trusted by organizations of all types

6,500+ Data Engineering customers across industries

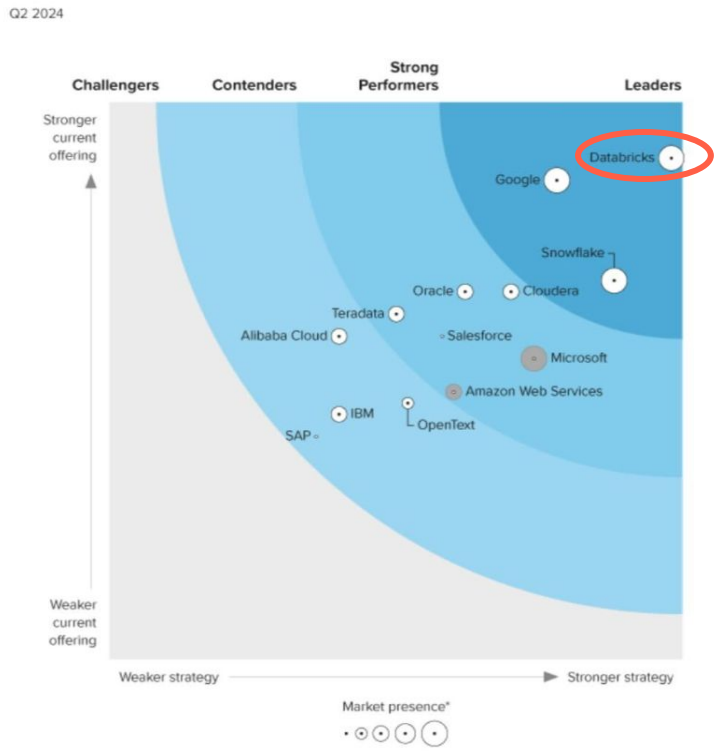


Kate Hopkins
Vice President
AT&T

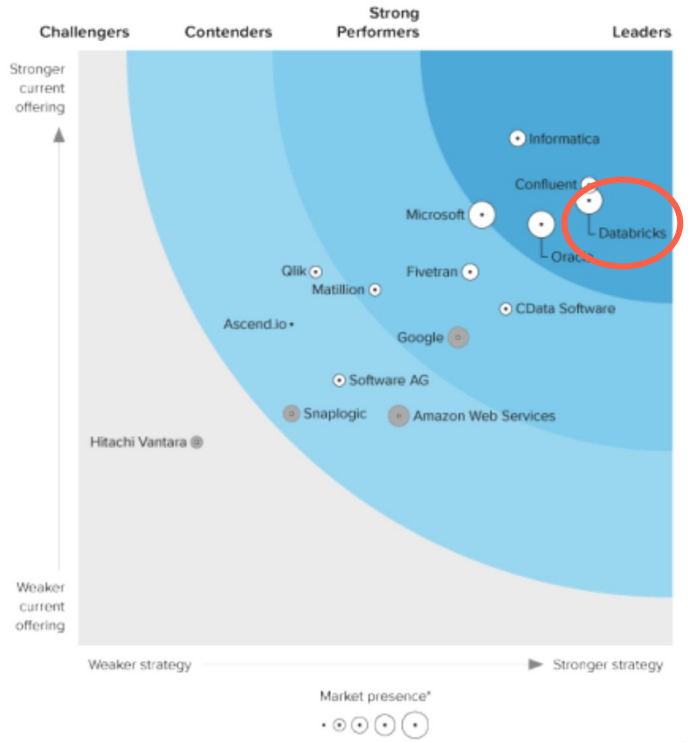
"We're able to ingest **huge amounts of structured and unstructured data** coming from different systems, standardize it, and then build ML models that deliver alerts and recommendations that empower employees in our call centers, stores, and online."

Recognized as a leader in the industry

Forrester Wave: Data Lakehouses

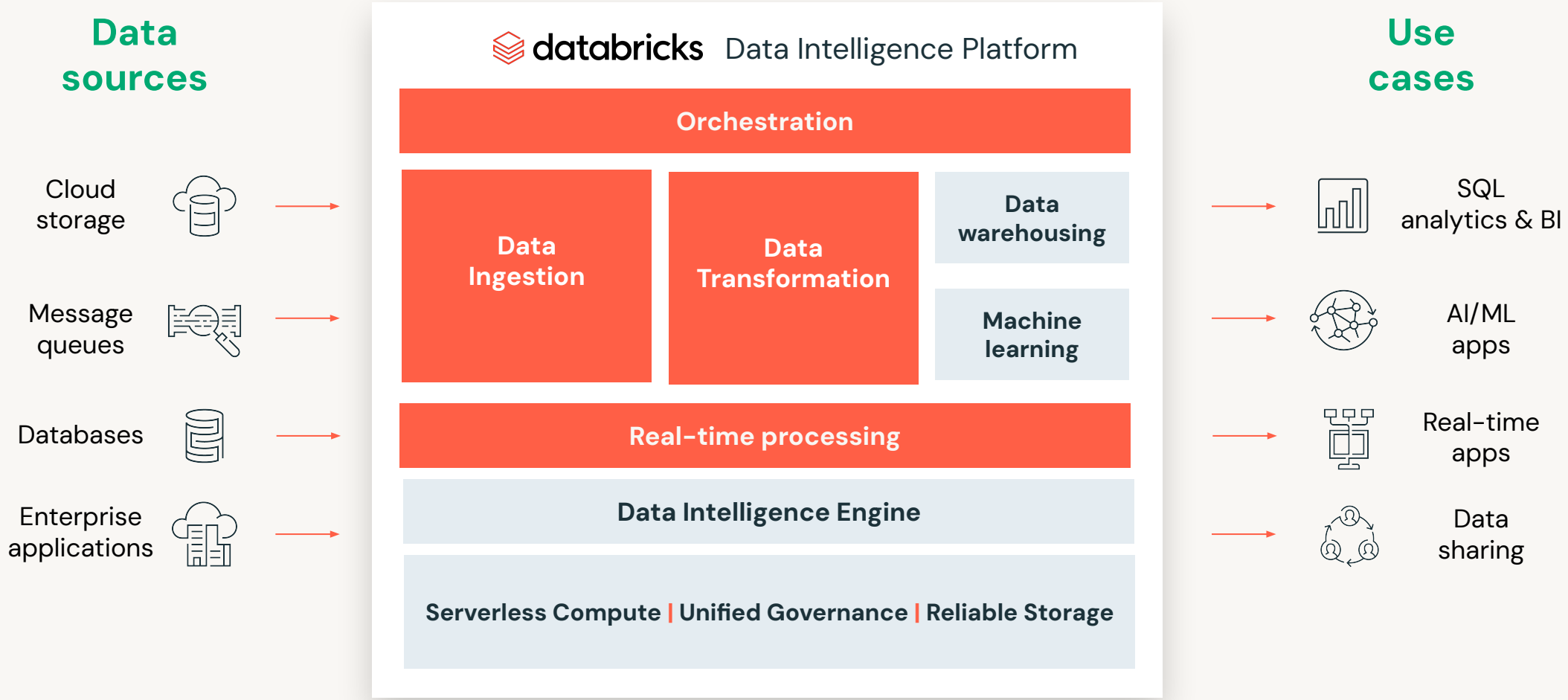


Forrester Wave: Cloud Data Pipelines



IDC MarketScape: Analytic Stream Processing

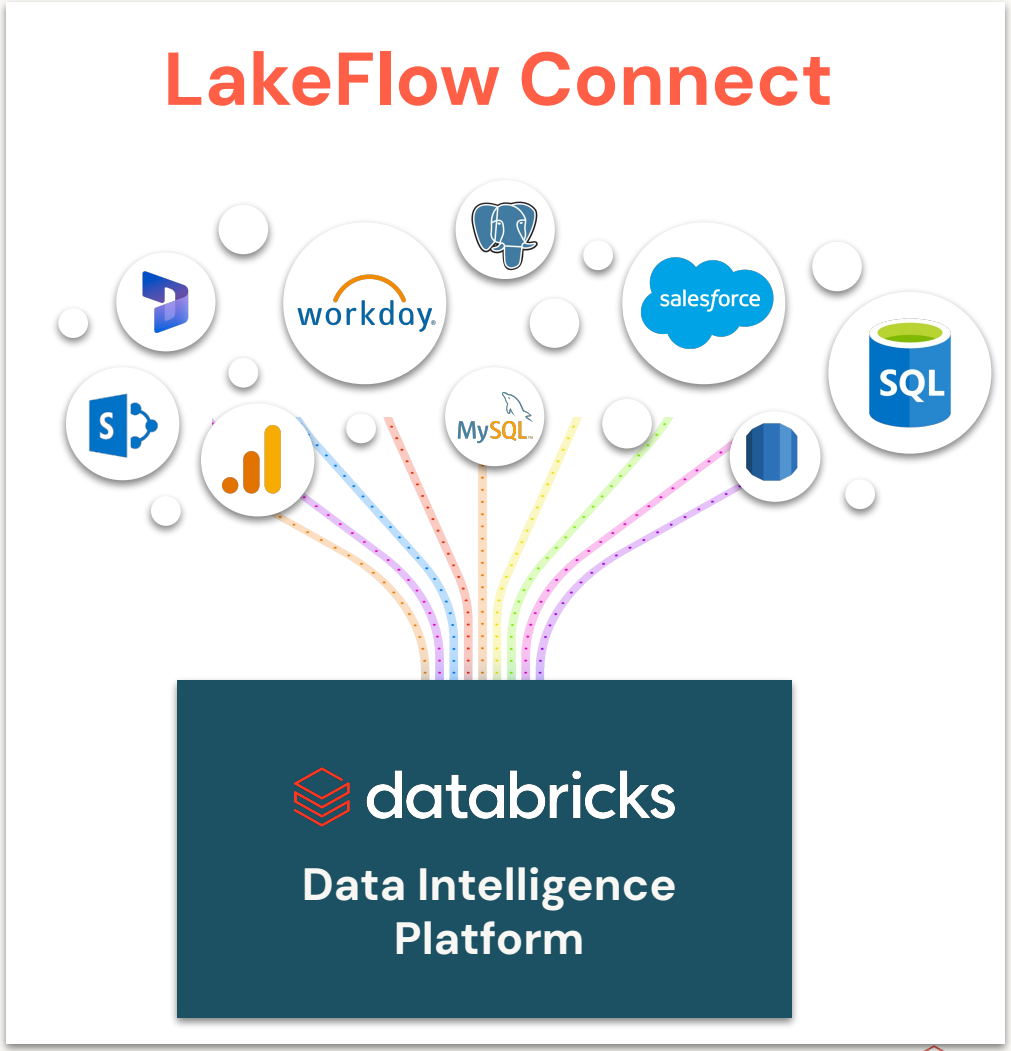
Data Engineering on Databricks



Ingestion

Efficient ingestion for all

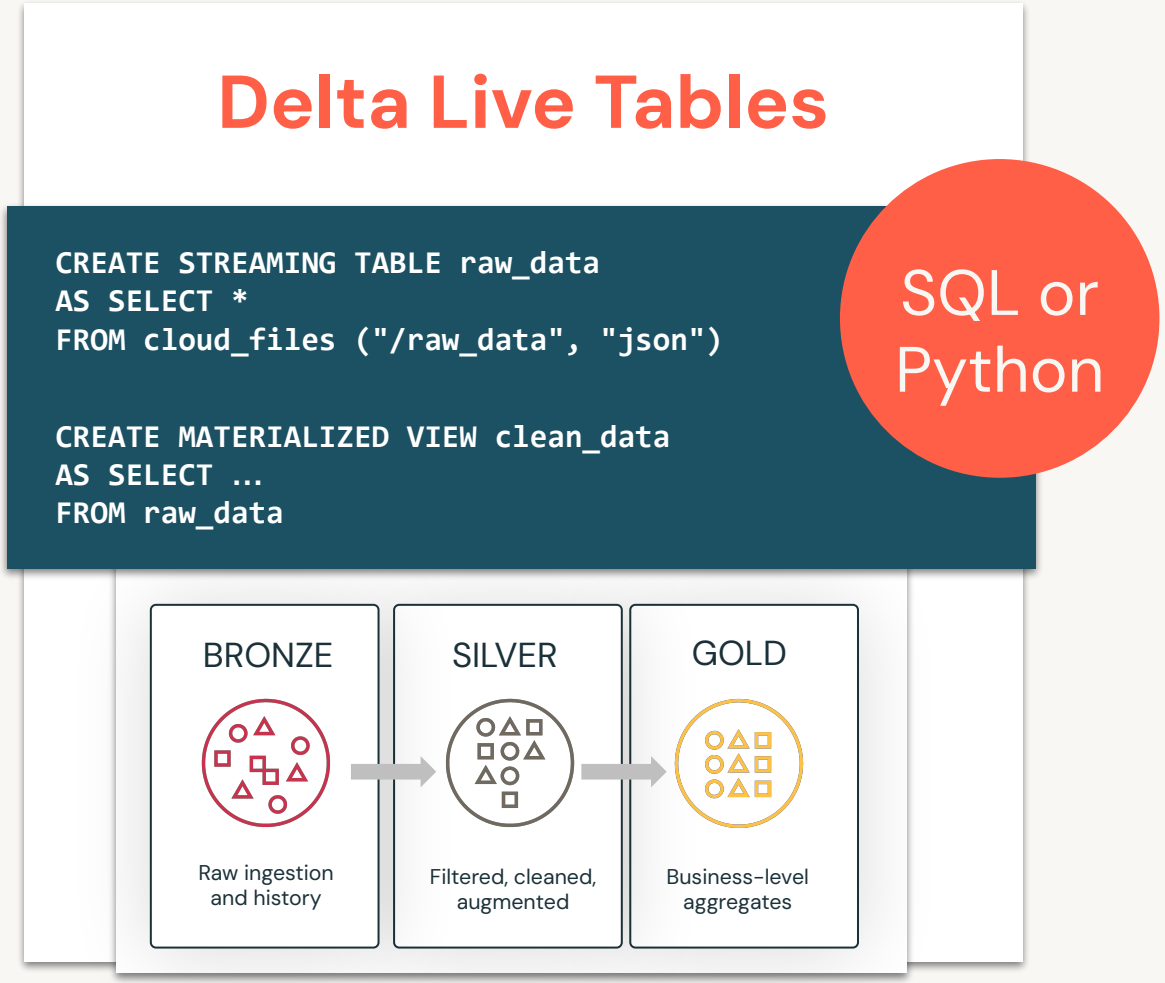
- Cost efficient, highly performant data ingestion with incremental reads
- Native ingestion support for cloud storage, databases, message buses and business applications.
- Large ecosystem of partner solutions



Transformation

Declarative data pipelines with Delta Live Tables

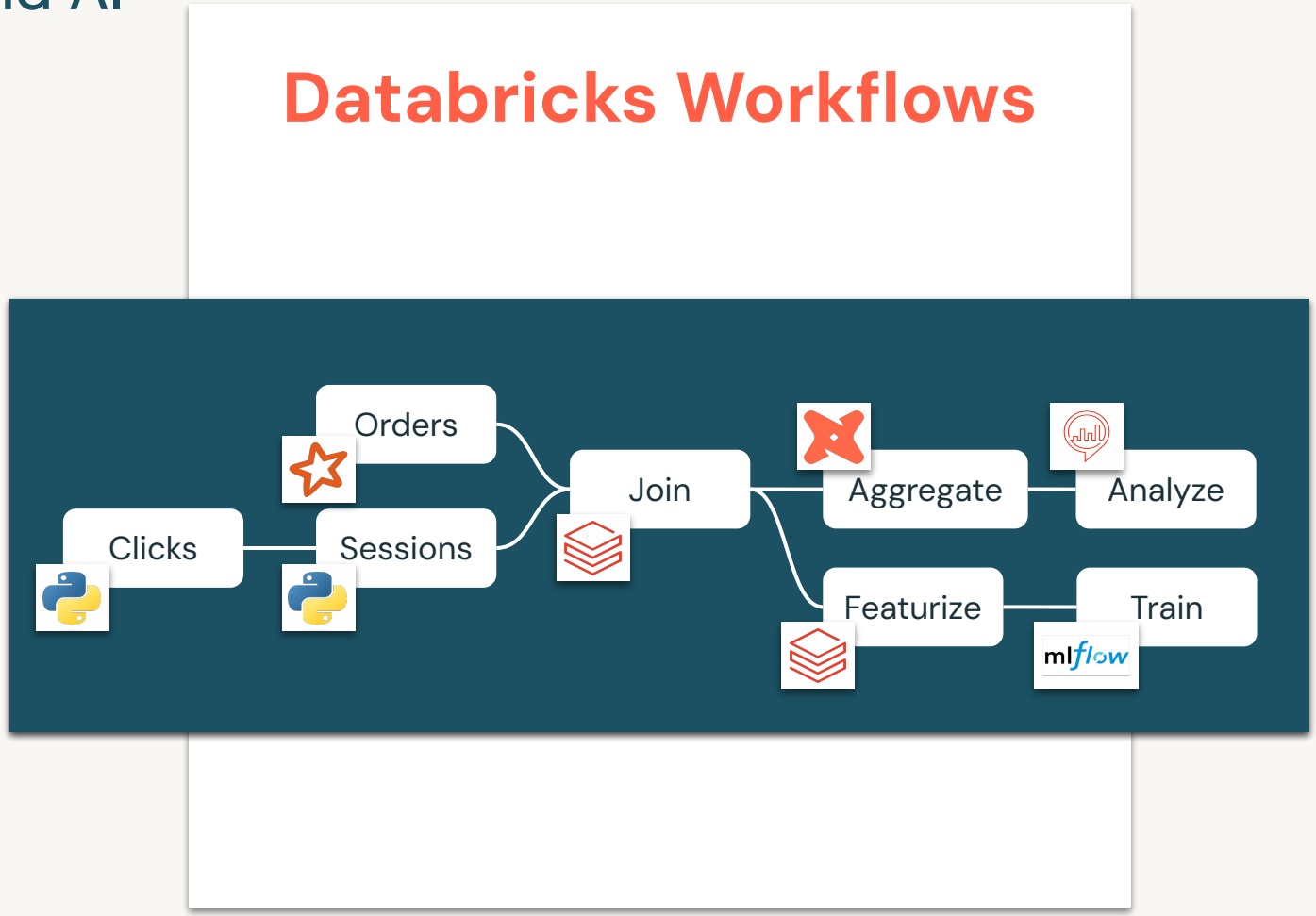
- Faster and cheaper pipelines with automatic incremental processing
- Reliable data through automated quality checks
- Rapid pipeline development through automated infrastructure and operations
- Streamlined data engineering with unified batch and streaming



Orchestration

Unified orchestration for Data and AI with Databricks Workflows

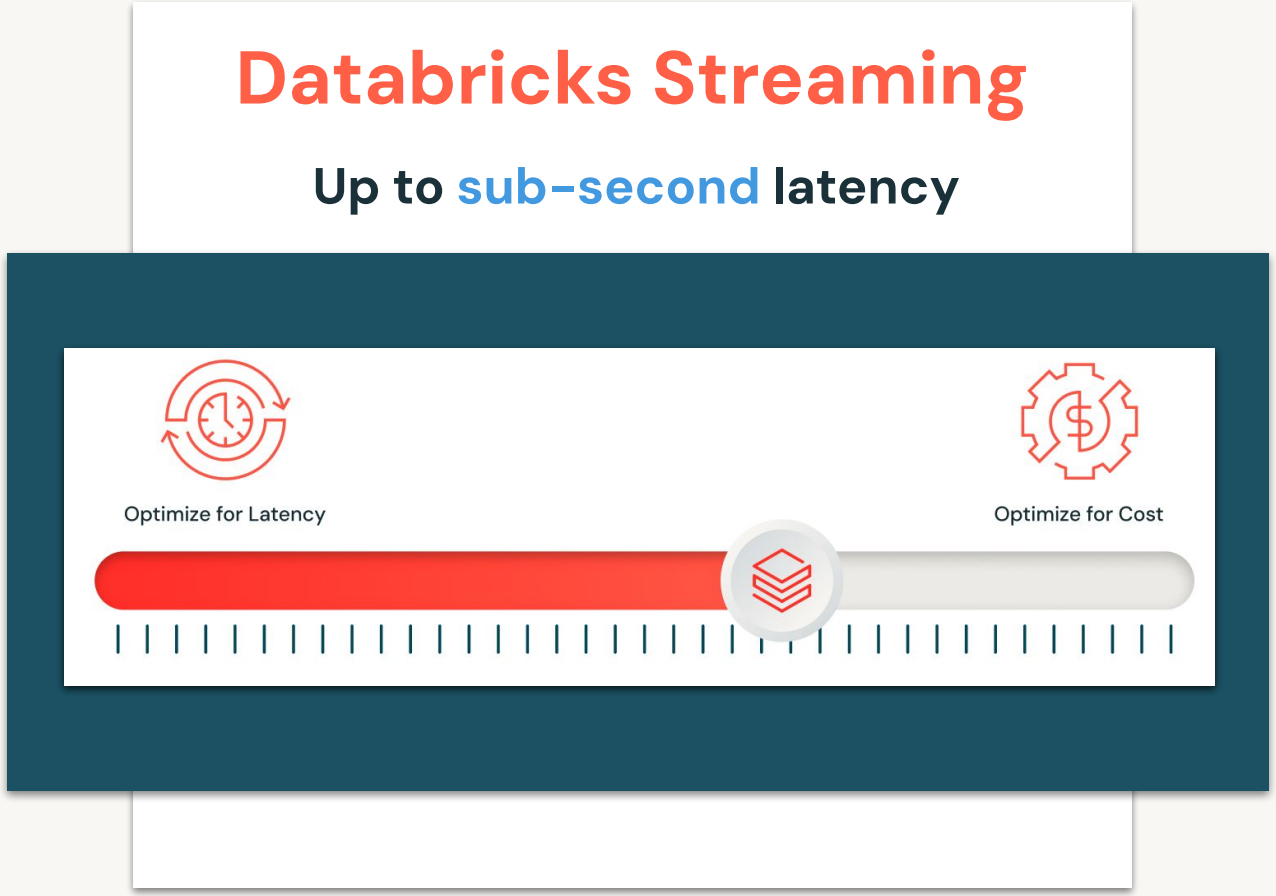
- Simple workflow authoring for any practitioner
- Actionable insights with end-to-end observability
- Reliability at scale in production



Near Real-time processing

Tunable streaming architecture built on the world's most popular streaming engine

- Simplified development with unified batch and streaming APIs
- Reliable operations through automatic checkpointing and failure recovery
- Easily adjustable throughput and latency for enterprise workload requirements



Data Engineering on Databricks

Data sources

- Cloud storage
- Message queues
- Databases
- Enterprise applications

databricks Data Intelligence Platform

Orchestration

Data Ingestion

Data Transformation

Data warehousing

Machine learning

Real-time processing

Data Intelligence Engine

Serverless Compute | Unified Governance | Reliable Storage

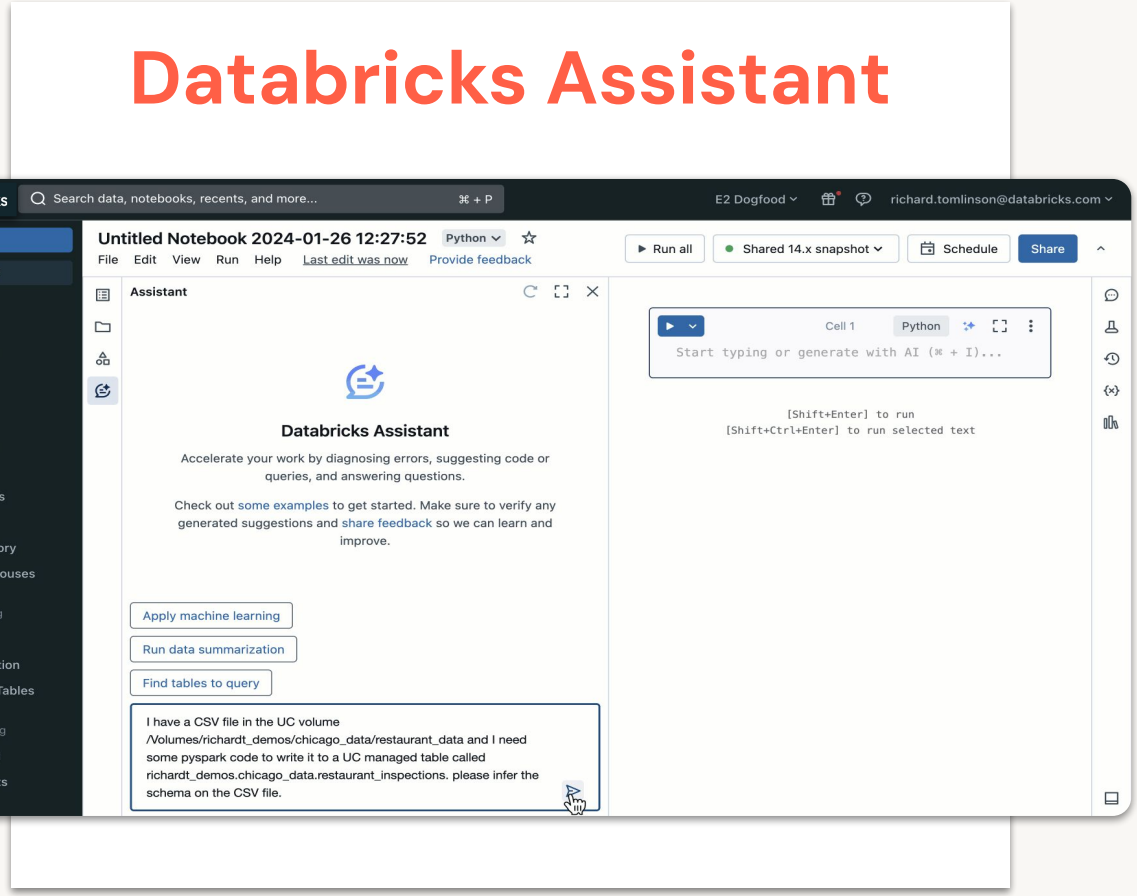
Use cases

- SQL analytics & BI
- AI/ML apps
- Real-time apps
- Data sharing

Empower every data engineer with AI

Boost productivity with your context-aware AI assistant

- Generate, explain, and fix code with natural language
- AI assistance in every user experience
- Powered by the Data Intelligence engine for highly relevant answers



Data Engineering on Databricks

Data sources

- Cloud storage
- Message queues
- Databases
- Enterprise applications

databricks Data Intelligence Platform

Orchestration

Data Ingestion

Data Transformation

Data warehousing

Machine learning

Real-time processing

Data Intelligence Engine

Serverless Compute | Unified Governance | Reliable Storage

Use cases

- SQL analytics & BI
- AI/ML apps
- Real-time apps
- Data sharing

Built on a solid, open foundation

Fully managed serverless compute, unified governance and reliable storage

Serverless Compute

Hands off, fully managed compute

Fast startups

Cost efficient with smart autoscaling

Stable and secured by default

Unity Catalog

Unified governance across all workloads

Single permission model for data and AI

AI powered monitoring, observability and lineage

Secure open data sharing

Delta Lake UniForm

The open format storage layer

High reliability and performance

Automatic instant translation across open formats with UniForm

Open source

Highly productive data engineering teams

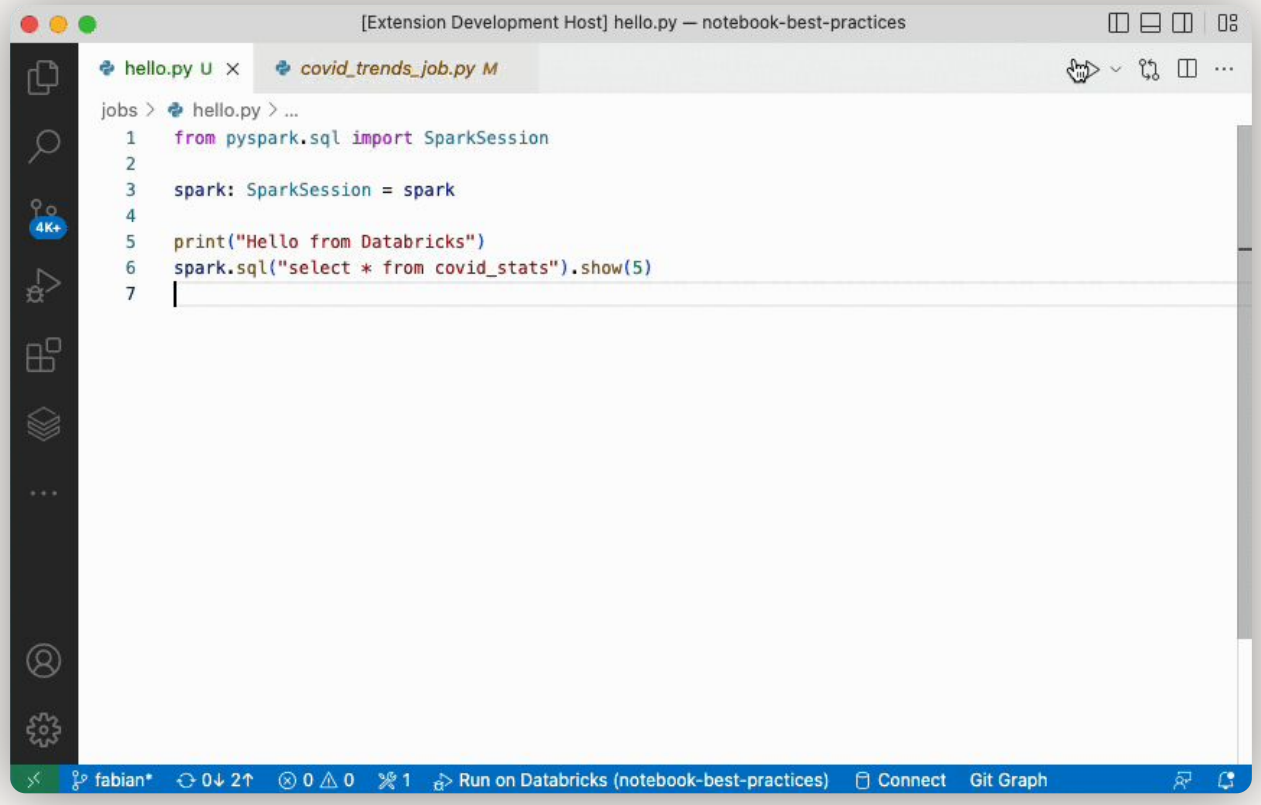
Empower data engineers with developer experience that meets them where they are

Build data pipelines in **Python, SQL, Scala or Java**

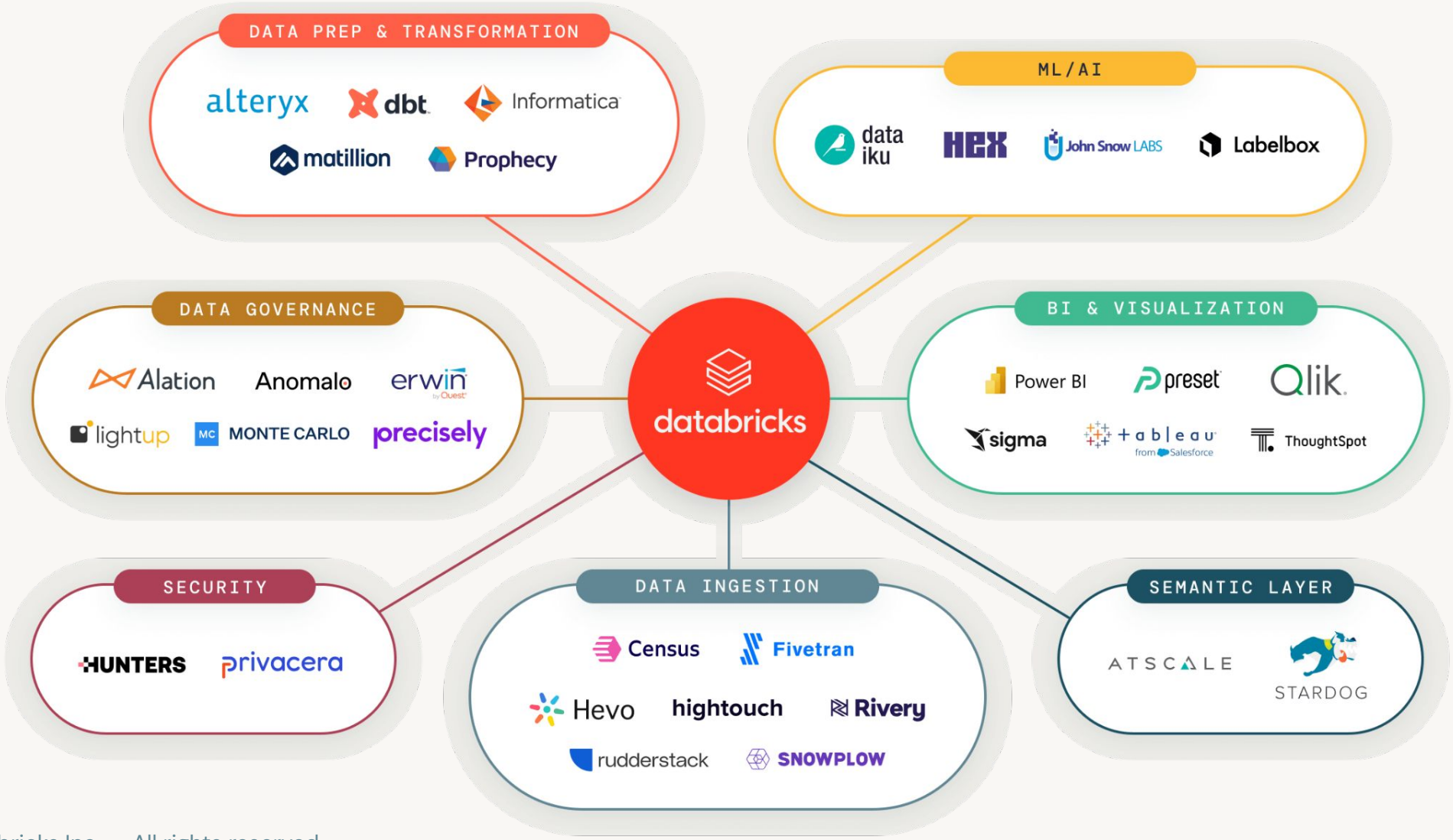
Automate deployments with CI/CD and git integrations

Full integration with **your existing tooling**

Fivetran Informatica dbt Qlik Prophecy hightouch alteryx



Integrated with the tools you know and love



Data engineering: Side by side comparison

Price/performance at scale	Up to 9x higher TCO at high scale	Highly efficient at any scale
Declarative pipelines	Dynamic tables	Delta Live Tables
Serverless with autoscaling	Define warehouse size No autoscaling at job level	No need to select VM size Autoscale at job level
Developer experience	Build pipelines in SQL, Python support limited or in private preview	Build pipelines in SQL, Python, Scala, or Java
DevOps	Git Integration for notebooks only Can execute files saved in git	Git integration, CI/CD, Databricks Asset Bundles
Data Streaming	Latency of 1 min	Sub-second streaming latency
Orchestration	Basic native orchestration complimented with third-party tools	Native orchestrator with advanced features

Tomer Patel
Engineering Manager
Akamai

"Because of our **scale and the demands** of our SLA, Databricks was the right solution for us. If we went with another solution, we couldn't achieve the same level of performance."

Akamai's web security analytics tool handles approximately **10GB of data related to security events per second. That's almost 1PB a day!**

Let’s see the product!

databricks

Thank you for downloading this Databricks presentation! Carahsoft serves as the Master Government Aggregator® and Distributor for Databricks, offering expertise in government procurement processes and practices with purchasing available via GSA, SEWP V, ITES-SW and other contract vehicles.

To learn how to take the next step toward acquiring Databricks' solutions, please check out the following resources and information:

For additional resources, please visit carah.io/DatabricksResources

For additional solutions, visit carah.io/DatabricksSolutions

To speak with our team directly, email Databricks@carahsoft.com or reach out at 703-581-6693.

To view our upcoming Databricks events, visit carah.io/DatabricksEvents

For additional Open Source solutions, visit carah.io/OpenSourceSolutions

To purchase, check out the contract vehicles available for procurement at carah.io/DatabricksContracts

carahsoft

For more information, contact Carahsoft or our reseller partners:

Databricks@carahsoft.com | 703-581-6693