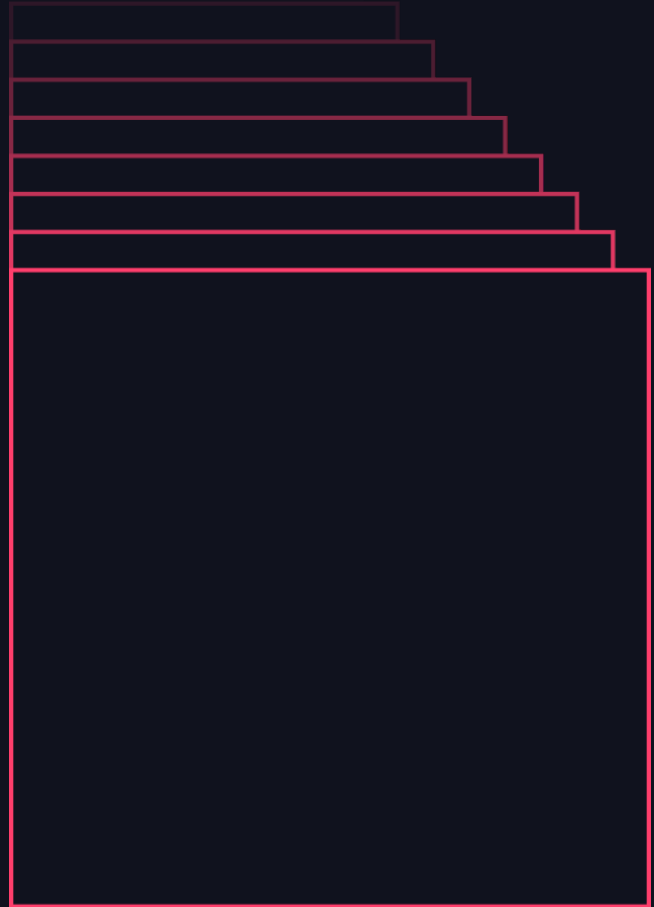


Informatica | Build your Analytics in DB SQL with Trusted High- Quality Data Using IDMC

Pradeep Anandapu & Ajay Gollapalli
Date 12th June 2024



Today's Presenters



Pradeep Anandapu
Sr. Staff Solutions
Architect, Databricks



Ajay Gollapalli
Director, Ecosystems,
Informatica



AGENDA

Build your Analytics in DB SQL with Trusted High-Quality Data Using IDMC

- Informatica & Databricks Partnership
- How Informatica Adds Value On Databricks
- Product integration roadmap
- Demo
- Q&A

Informatica & Databricks Partnership

Informatica extends the value of your Databricks investment



160+
Joint
Customers



Data Integration
Partner of the
year 2024



Top 3 fastest
growing on
Databricks

How Informatica Adds Value On Databricks

Informatica Integrates with Databricks

Data & AI Consumers



ETL Developer



Data Engineer



Citizen Integrator



Data Scientist



Data Analyst



Business Users

Informatica
Intelligent Data Management Cloud

MDM & 360 APPLICATIONS

DATA INTEGRATION

DATA QUALITY & PROFILING

API & APP INTEGRATION

DATA CATALOG

CLOUD DATA MARKETPLACE

GOVERNANCE & PRIVACY

Databricks Data Intelligence Platform

Data Science & AI Mosaic AI	ETL & Real-time Analytics Delta Live Tables	Orchestration Workflows	Data Warehousing Databricks SQL
An AI powered data intelligence engine to understand the semantics of your data			
DatabricksIQ			
Unified security, governance, and cataloging Unity Catalog			
Unified data storage for reliability and sharing Delta Lake UniForm			

Open Data Lake

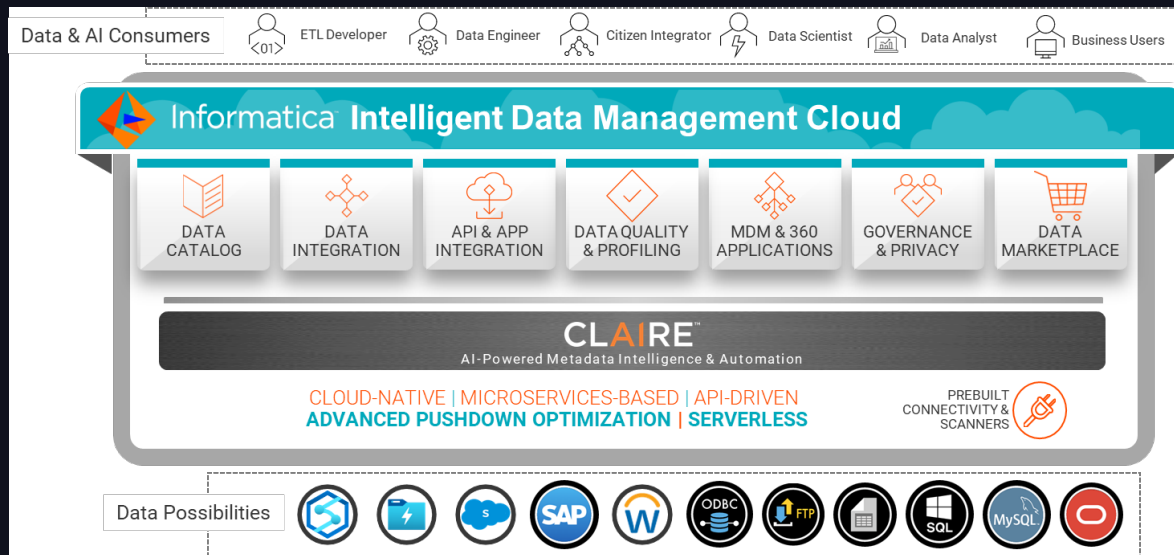
All Raw Data
(Logs, Texts, Audio, Video, Images)



Informatica Data Management Cloud (IDMC)

Processing 95T+ Transactions per month

- Cloud-native, as-a-service microservices based platform
- Optimized for multi-cloud and hybrid data management
- Security-by-design architecture
- Consumption-based pricing
- Available on AWS, Azure and GCP
- Available globally
- The ONLY complete end-to-end data management solution for Databricks



Informatica Supports 300+ connectors

Data Stores: DBs, DWs, Big Data, Cloud

Cloud Apps (SaaS)

Social Apps

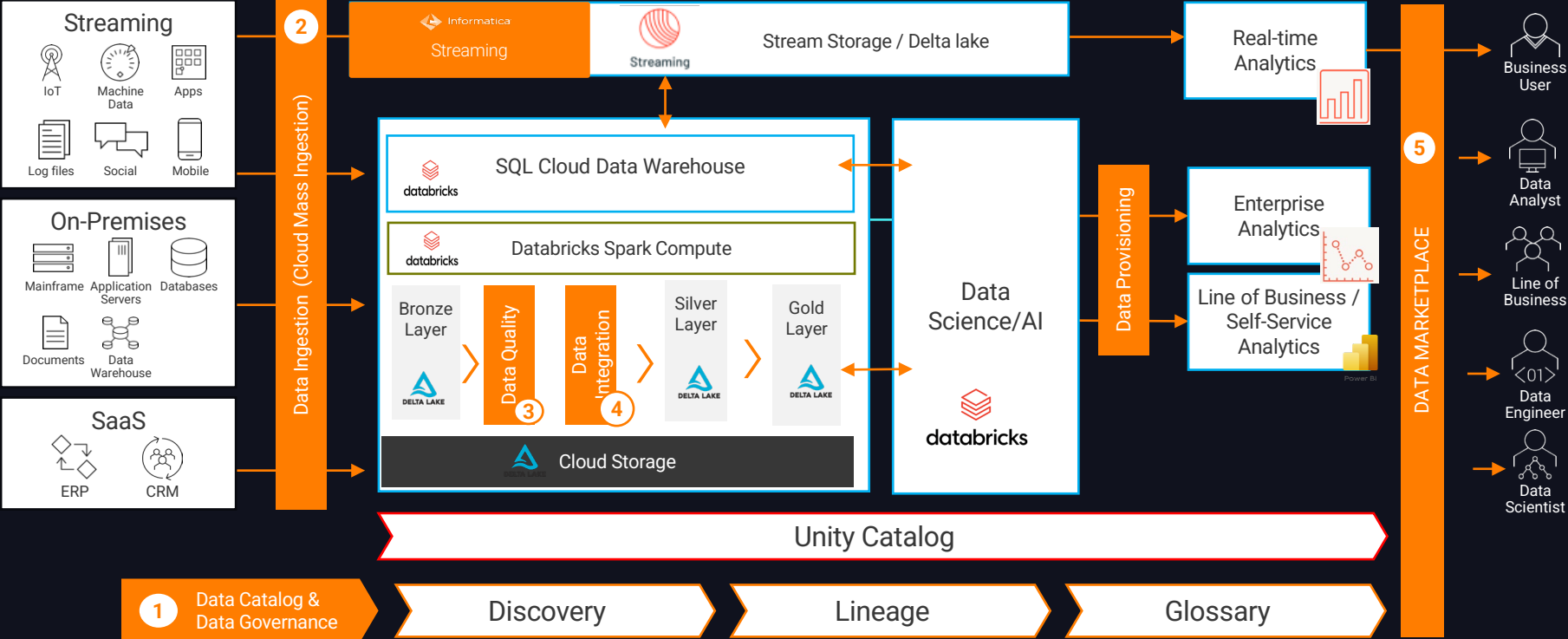
Enterprise Systems

Middleware and Tech

Analytics

B2B

Informatica Empowers Lakehouse Architecture



Product Integration Roadmap

Informatica + Databricks – Integration Roadmap

Now Available on Partner Connect

Generally Available for Customers

Cloud Data Integration

- Support for Unity Catalog with Personal Staging Location
- Support Pushdown Optimization (ELT)
- Integrated with Partner connect (Announcing Today)

Cloud Data Quality

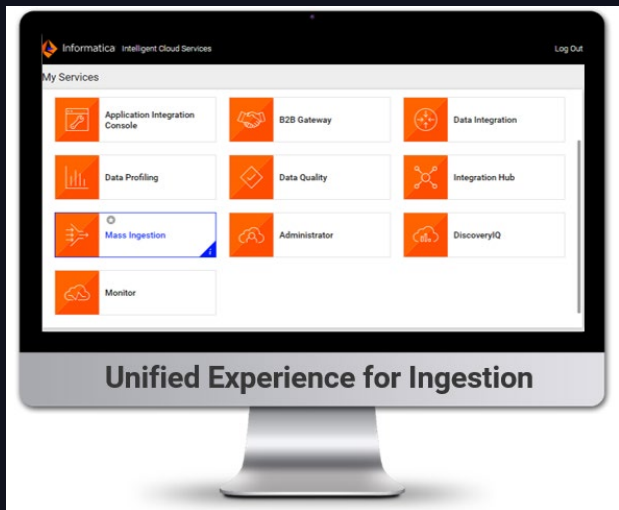
- Support for Data Profiling on Unity Catalog data assets
- Optimizations for shallow profiling
- Support for Unity Catalog and ELT within Databricks SQL

Upcoming Features

- Native SQL ELT leveraging Databricks compute
- Databricks Volumes support
- Databricks DBRX – LLM support
- Databricks Vector Index support

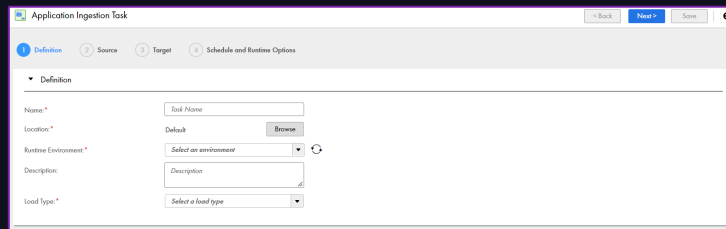
Informatica Data Ingestion & Replication

Cloud Mass Ingestion



✓ Databricks Delta Lake Supported

✓ Step-by-step wizard for designing and creating an ingestion task



✓ Deployment, scheduling, real-time monitoring and lifecycle management



Ingest in Real-Time



Real-Time Monitoring



Automated Schema Drift Handling

✓ Versatile out-of-the-box connectivity to sources and targets



Databases and CDC



Streaming Sources



Files

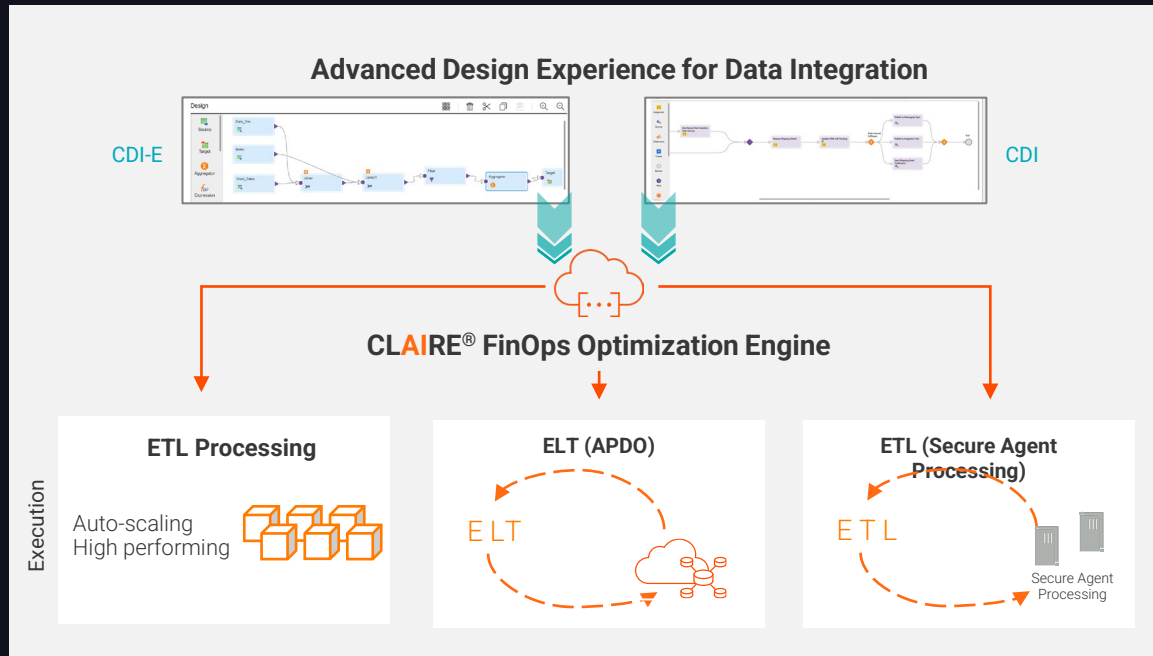


Applications

Informatica Cloud Data Integration

Cost-Performance Optimization

- **Single design time** experience for all your data integration and data engineering needs
- Full Integration with **Unity Catalog**
- **Intelligent (CLAIRE®-driven) optimization** at runtime for **best cost-performance**
 - Driven by job priorities and SLAs
 - **Resources dynamically allocated** for each workload to ensure best price/performance
- **300+ purpose-built, cloud-native connectors**
- **Rich transformation** coverage for any type of workload, at any scale



Informatica Cloud Data Quality

Databricks Customers Are On A Journey That Requires Trusted, Accurate Data

What is Data Quality?

- There are many definitions of data quality – trustworthy & accurate
- Two key characteristics
 - High Quality Data
 - Process by which to achieve high quality data

Measure

- Review Progress
- Threshold Alerts
- Scorecards

Discover

- Profiling
- Identify Data Issues
- Set Data Quality Goals



Apply

- Mapping Generation
- Standardization / Validation
- Matching / Consolidation

Define

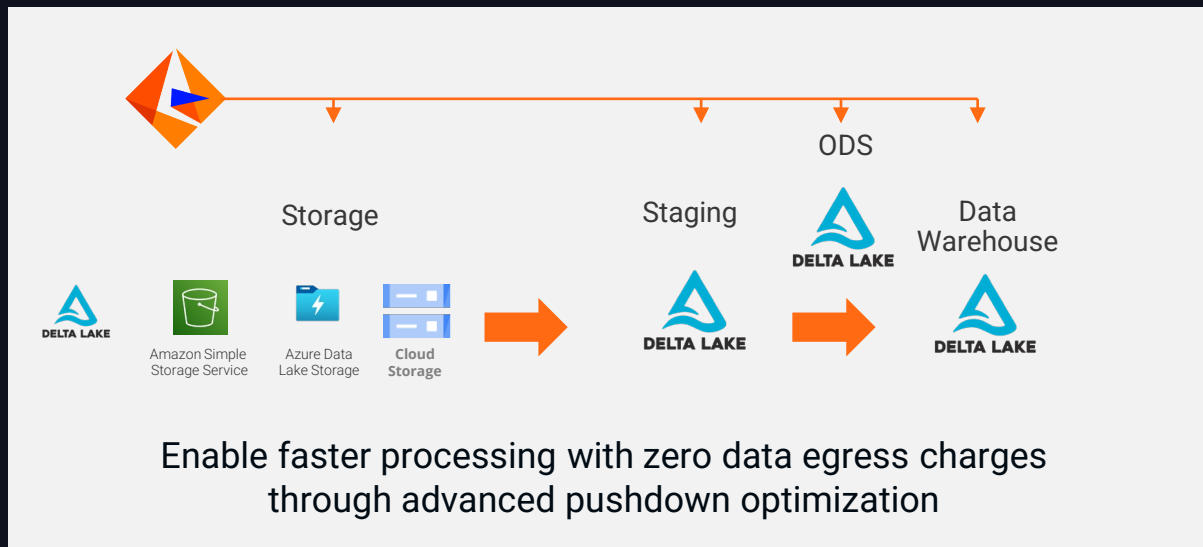
- Dictionaries
- Implement Rules
- Build Cleanse, Parse, Verification, etc. processes

SQL ELT: Advanced Pushdown Optimization

Convert and process data pipelines to native ecosystem commands and SQL queries for faster processing at lower cost while ensuring data stays within the Databricks

Why Advanced PDO?

- Zero data egress charges
- Best-in-class performance
- Best ELT Engine
- Broadest array of connectors
- Single connector for all PDO and advanced features
- Simple drop-down option in GUI—no need to learn proprietary commands



New Ecosystem SQL ELT (Databricks)

Databricks Native experience (July 2024)



Key Highlights

- **Customized** experience for SQL ELT mapping development
- **OOTB support** for 300+ list of Databricks functions



Benefits

- **Guaranteed execution** on Databricks SQL Warehouse Compute.
- **Familiar semantics** for quick onboarding of Databricks CDW users
- **Accelerated development** of SQL ELT mappings

SQL ELT User experience

The screenshot displays two overlapping dialog boxes in the Databricks SQL ELT user interface. The top dialog, titled "New Mapping", asks "What type of mapping do you want to create?". It features two options: "Mapping" (with a description: "Create a mapping with various sources and targets.") and "Mapping - SQL ELT" (with a description: "Create a mapping that is optimized for your company's cloud data warehouse."). The "Mapping - SQL ELT" option is highlighted with a blue border and a red arrow pointing to it from the right, with the word "New" written vertically next to the arrow. Below the options are "Continue" and "Cancel" buttons. The bottom dialog, titled "Databricks Function Grammar", shows a list of functions on the left, including "ABS", "ACOS", "ACOSH", "ADD_MONTHS", "ALL_USER_NAMES", "ASCII", and "ASIN". The "ADD_MONTHS" function is selected, and its SQL syntax is displayed in a text area: `TO_NUMBER(customerid_c, 19, 0)ADD_MONTHS(date_or_time_expr,num_months_expr)`. Below the text area is a description: "Syntax: ADD_MONTHS(date_or_timestamp_expr as TEMPORAL, num_months_expr as NUMERIC) Adds or subtracts a specified number of months to a date or timestamp, preserving the end-of-month information." Below this dialog are "OK" and "Cancel" buttons.

Demo

Empowering all Data Consumers & Data Owners



Cloud Data Integration on Databricks Partner Connect

Simplified Integration



Key Highlights

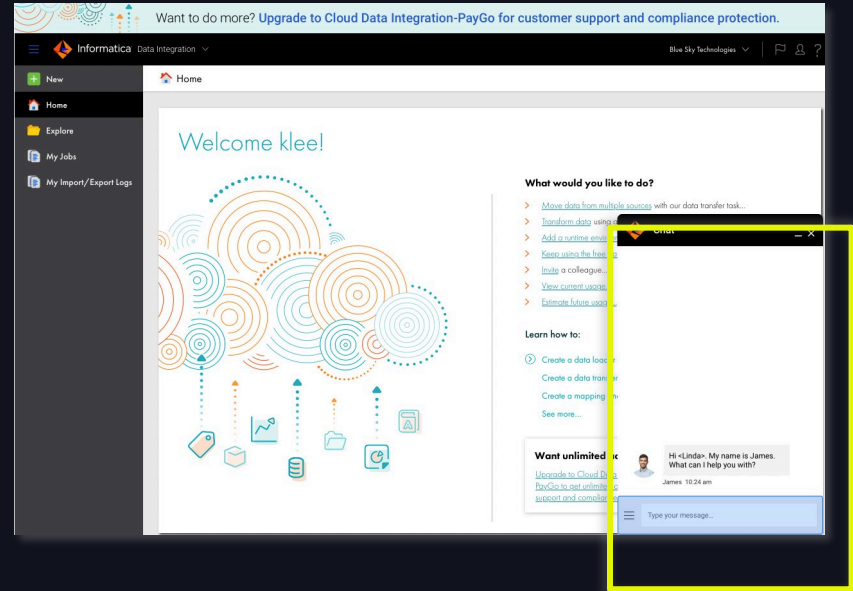
- **Frictionless Registration on Databricks** No OTP, No Password, Immediate Access to CDI Free
- **Free ETL/ELT** with robust transformations and orchestration capability for processing 20Mn rows & 10 hrs. compute
- **Context based video** and chat option for real-time product/feature training and trouble shooting



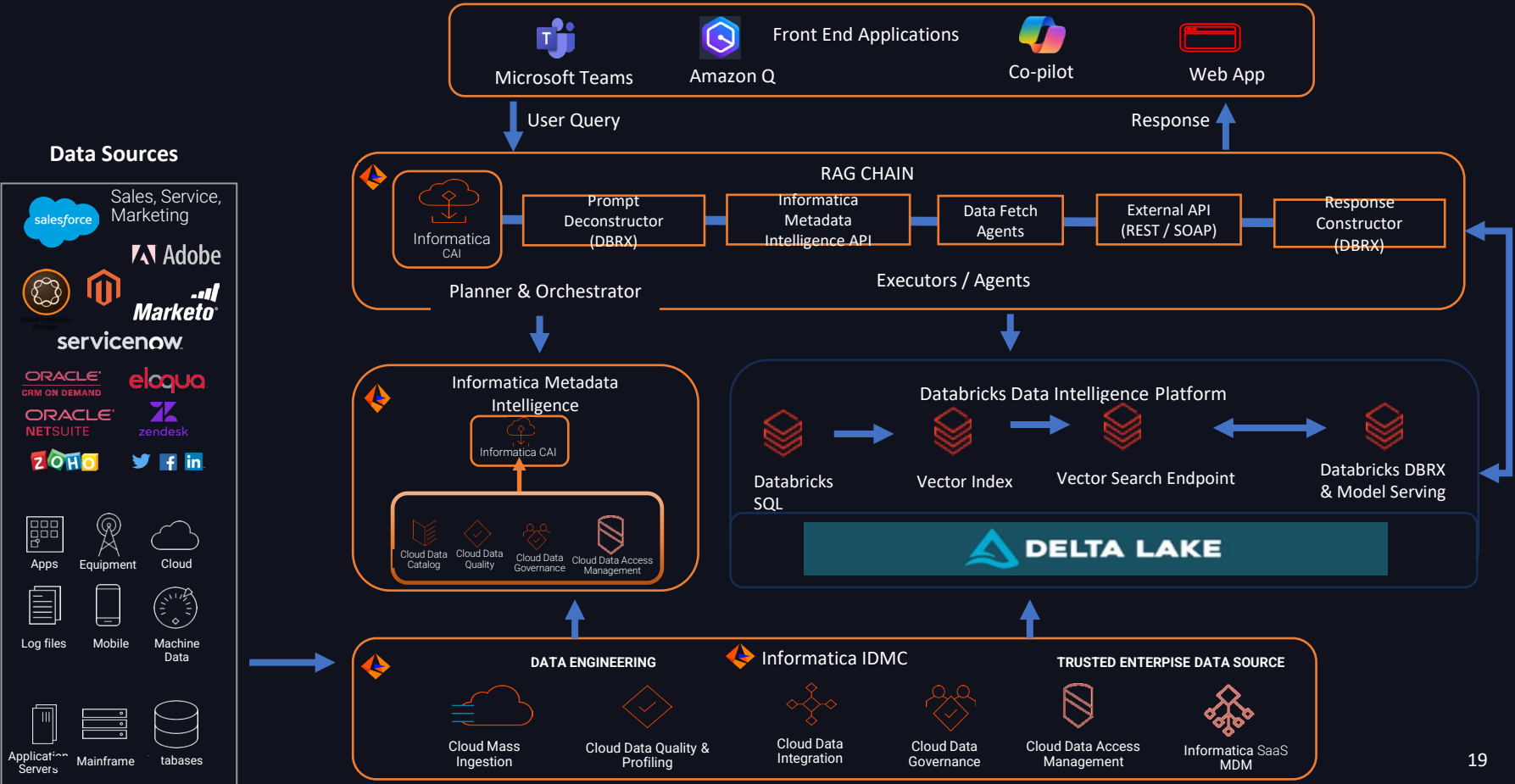
Benefits

- **Easy onboarding and quick value realization** for any Databricks user
- Connectivity to popular data sources for **ingestion & transformation** to get started with AI and analytics projects

Available Now!



Informatica Integrates with Databricks DBRX



DATA+AI SUMMIT

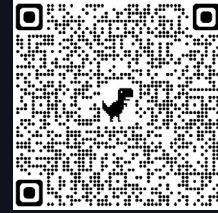
Scan QR codes to learn more



Cloud Data Integration Free
Service via Partner
Connect



Learn more on Informatica
with Databricks



Read about our Latest
Announcements

Thanks

