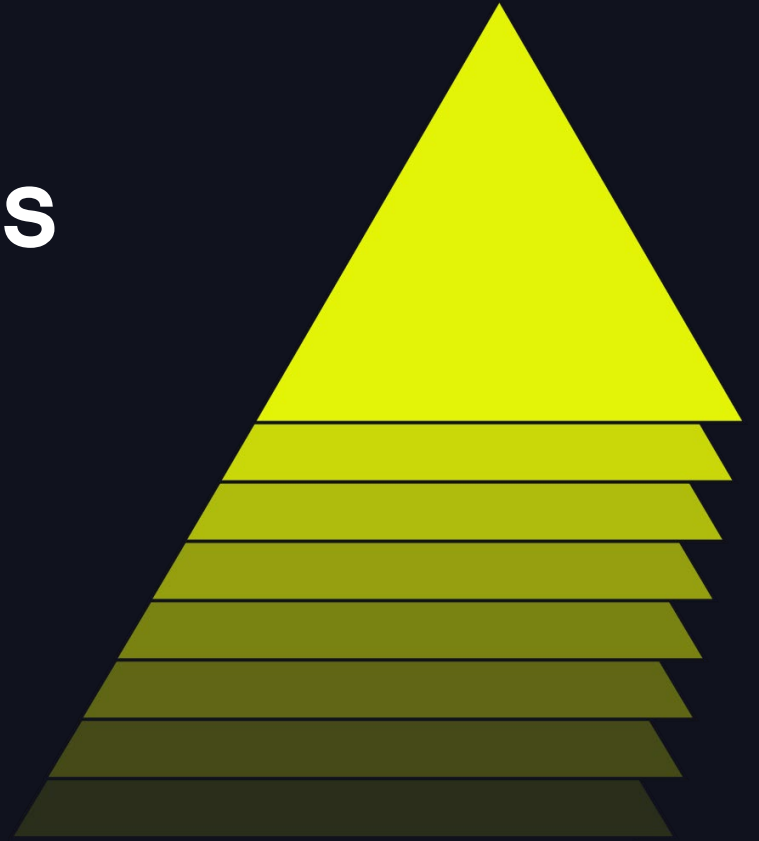


How KPMG Enables Trusted AI with Databricks and Informatica



Rik Tamm-Daniels & Dennis Tally
June 2024

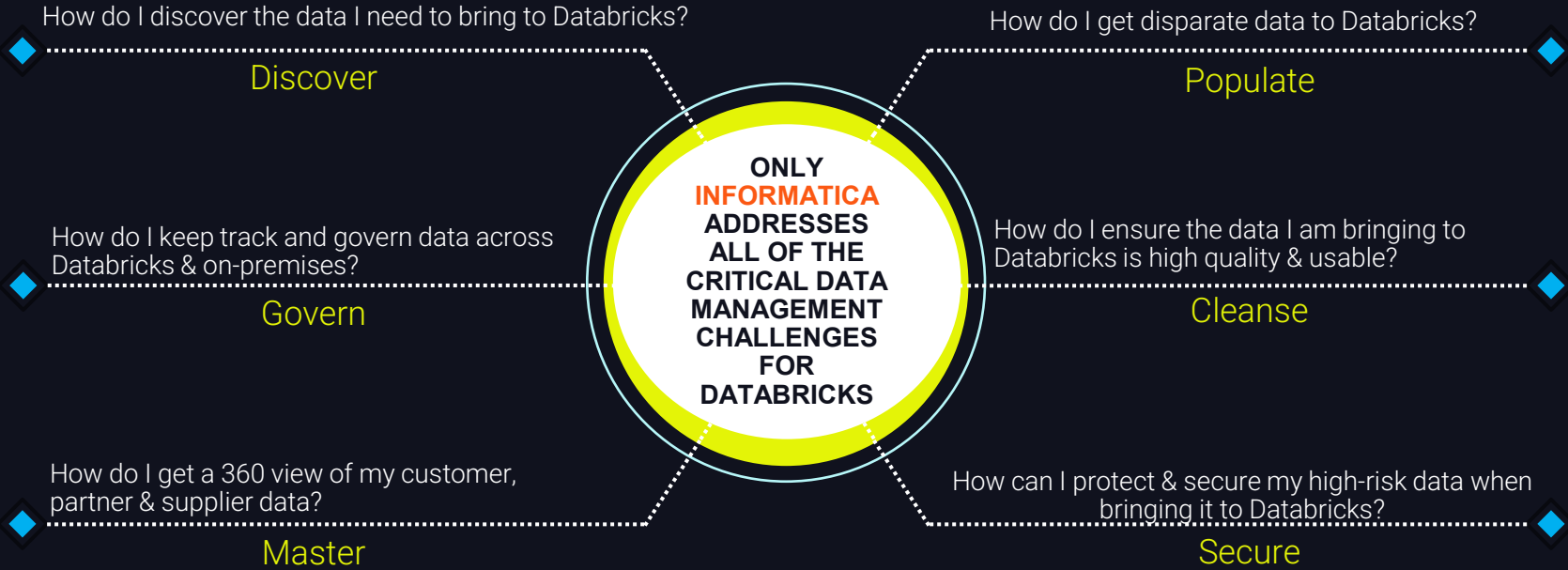
Winner

2024 Databricks
Partner Awards

Data Integration
Partner of the Year



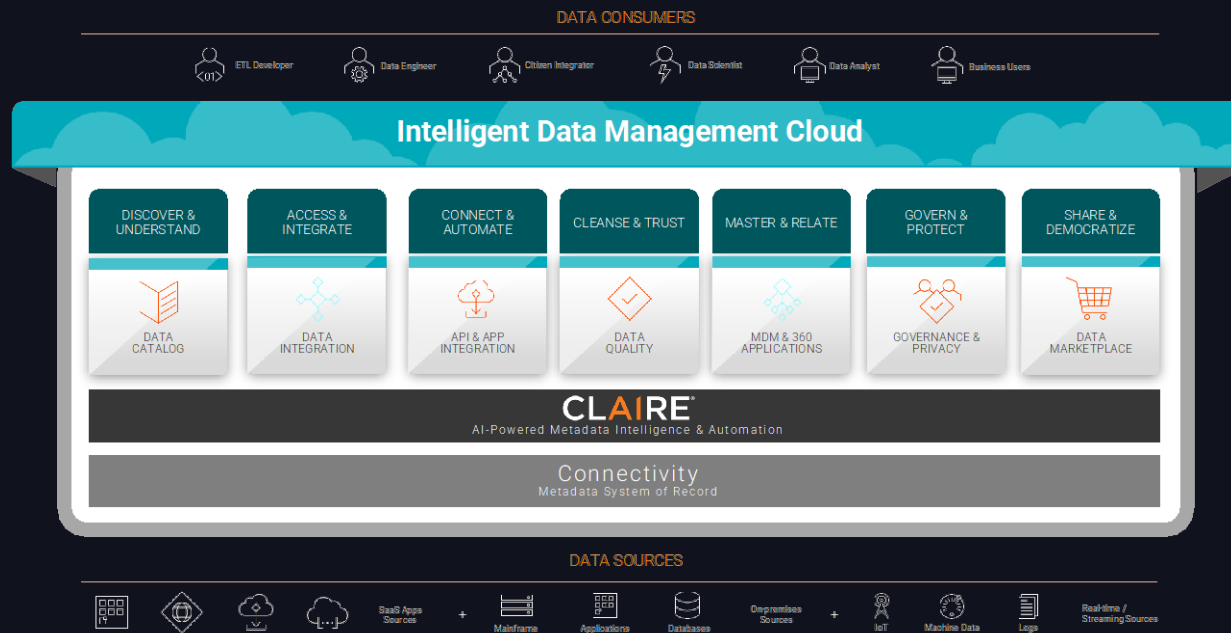
Informatica Solves Critical Data Challenges for Databricks Customers



Informatica Cloud Data Management (IDMC)

Processing 92T+ Transactions per month

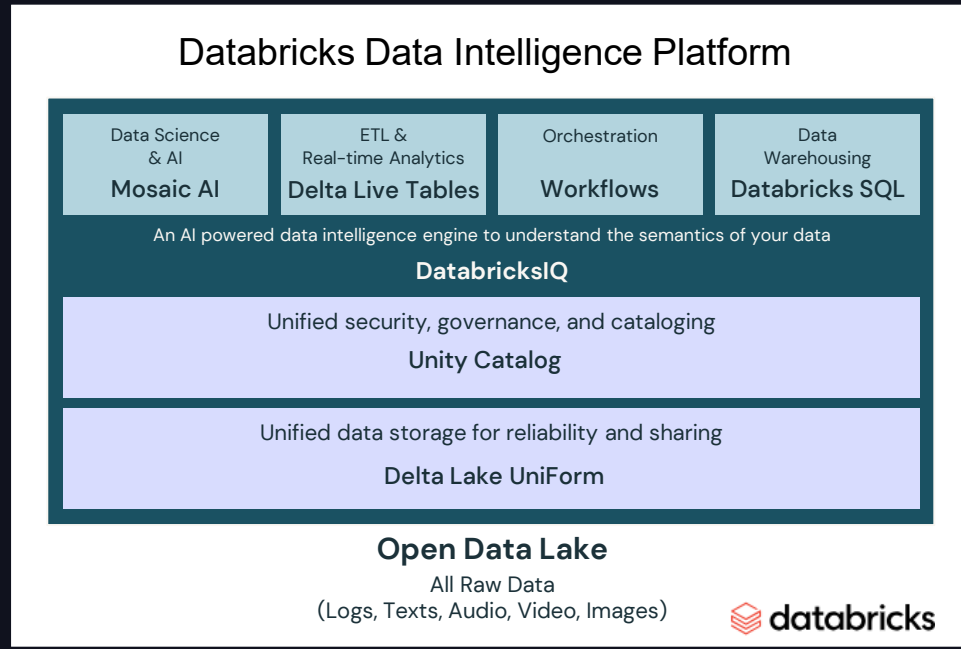
- Cloud-native, as-a-service microservices based platform
- Optimized for multi-cloud and hybrid data management
- Security-by-design architecture
- Consumption-based pricing
- Available on AWS, Azure and GCP
- Available globally
- **The ONLY complete End to End Data Management solution for Databricks**



Complete Data Management for Databricks



Informatica
Intelligent Data Management Cloud





Delivering Gen AI Dreams as Enterprise Reality

Requirements of Enterprise Generative AI Apps

Create foundational success with your enterprise context & metadata intelligence



Grounded

- ✓ Models hallucinate
- ✓ Fine-tuning is hard to manage and ensure consistent quality



Contextualized

- ✓ Models don't understand your enterprise terminology and semantics
- ✓ Ensure prompts are enriched with your business context
- ✓ Rich summarizations applicable to your business



High Quality

- ✓ Not all data is created equal
- ✓ Responses need to be accurate



Easy to Develop and Deploy

- ✓ Develop and deploy without lots of hand-coding
- ✓ Rapidly adopt new & evolving Generative AI capabilities



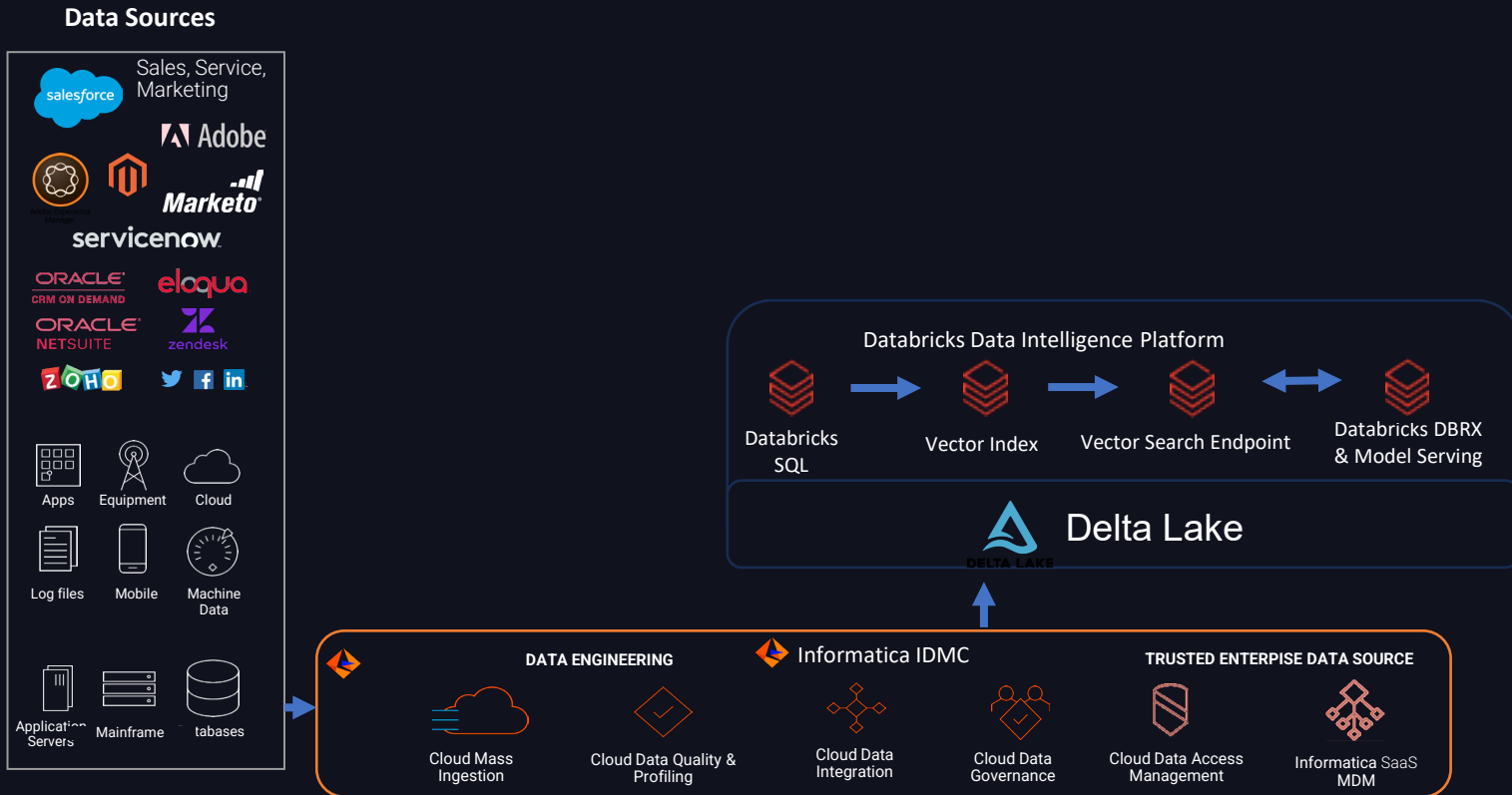
Governed & Secure

- ✓ Transparent and traceable
- ✓ Enforce data access policies
- ✓ Cost and usage control

Informatica Blueprint for Databricks DBRX

Prepare Enterprise Data Sources

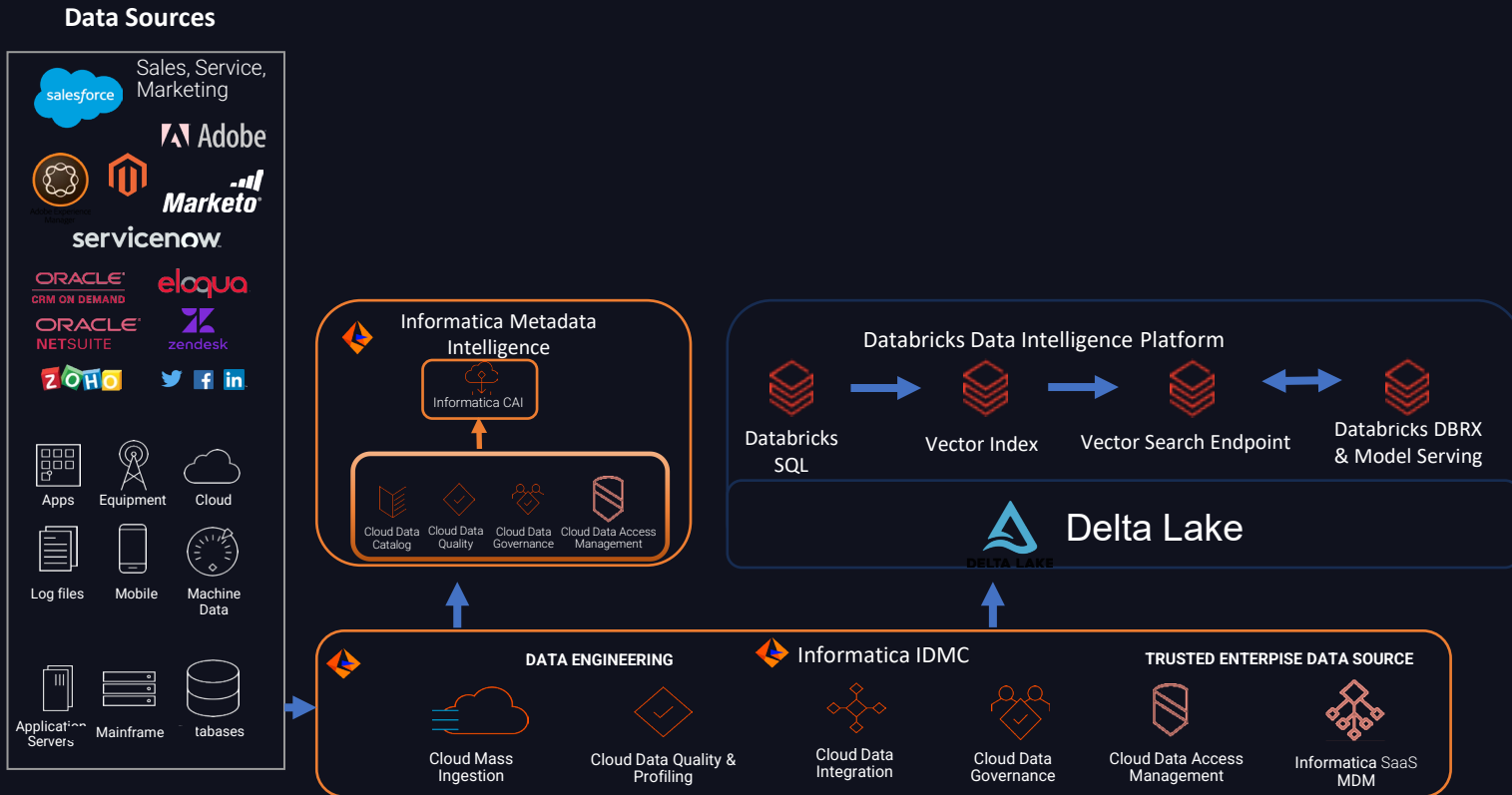
- Ingest from internal and external data sources
- High-quality Master Data Management Data is a key data source for Gen AI apps
- Vectorize input and store in a Vector database



Informatica Blueprint for Databricks DBRX

Prepare Enterprise Metadata

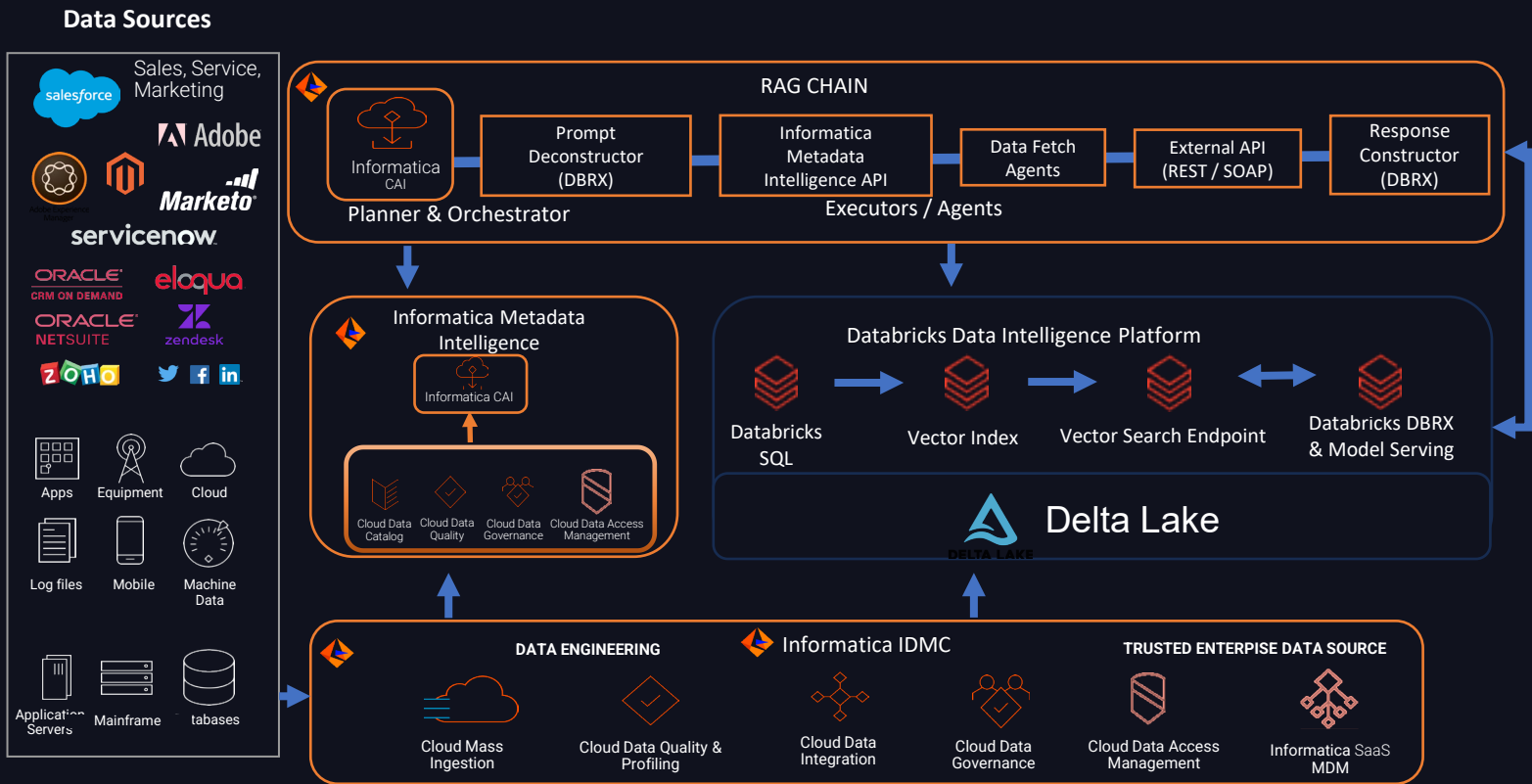
- Vectorize semantic context, data quality metrics, data access policies and data lineage input and store in a Vector database



Informatica Blueprint for Databricks DBRX

Define Agentic RAG Executors

- Annotate prompts with Semantic metadata
- Prioritize and select RAG datasets based on quality of the data
- Enhance and standardize response summaries
- Enforce access controls and role-based usage restrictions



Informatica Blueprint for Databricks DBRX

Frontend

Frameworks for building GenAI apps (e.g., Teams, Copilot, Native App etc.)

API Management

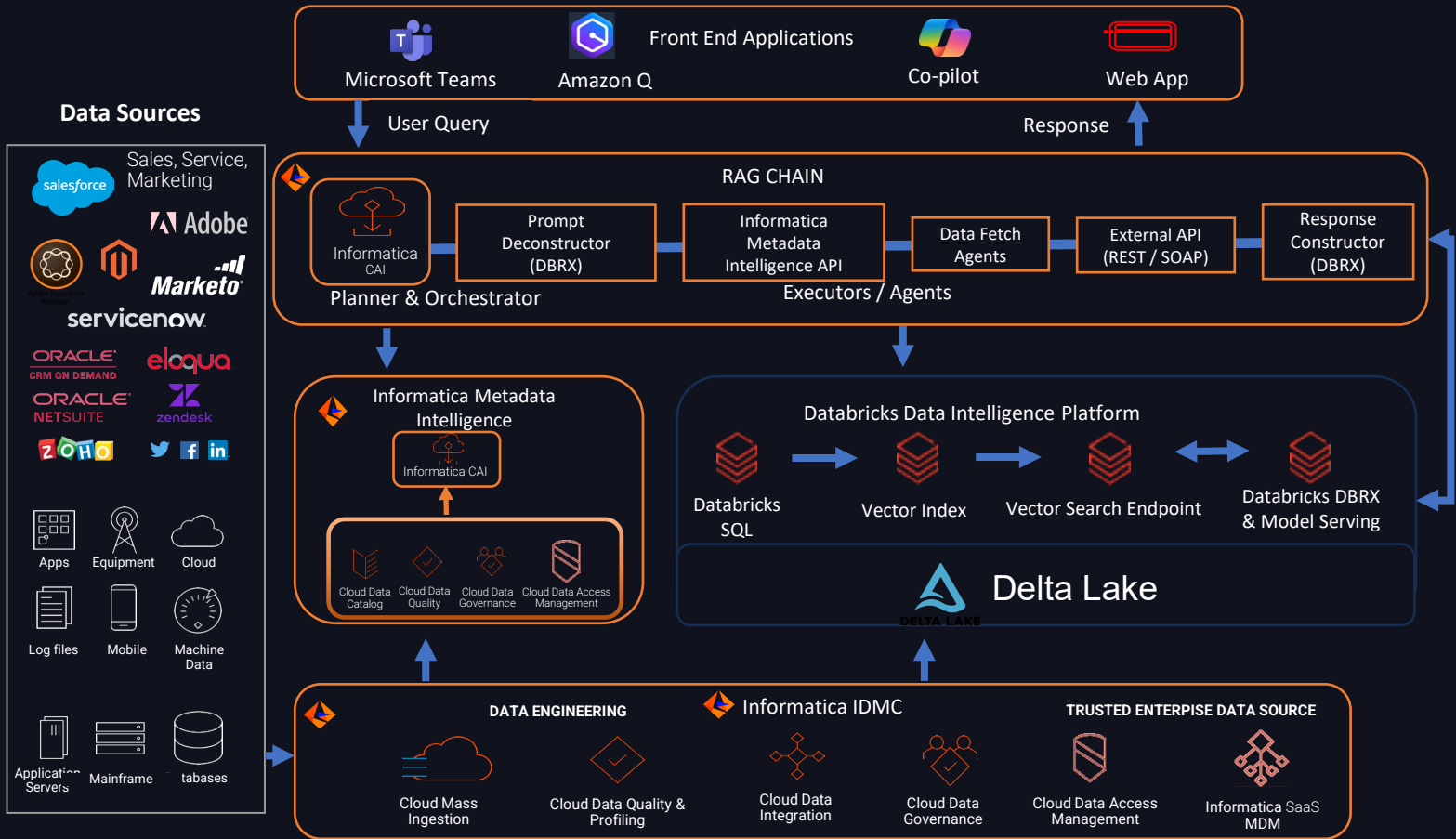
- Rate limiting
- Advanced authentication

Planning Agent

- Orchestrate API calls to
- Retrieve relevant context
- Bring metadata intelligence
- Enterprise apps incl. Informatica MDM

Databricks DBRX

- Provides LLM Inference



How KPMG puts this into practice

Introduction



Share how KPMG US transformed the data estate from constrained, on-premises infrastructure to a robust, cloud-based data estate leveraging the integrated power of the Azure ecosystem including Azure Databricks, and Informatica Intelligent Data Management Cloud (IDMC).



Highlight the decisions and best practices that we followed to achieve this transformation and, the benefits that we gained by modernizing our enterprise data capability.



Inform other enterprises looking to optimize data management and analytics capabilities using market leading technologies.



Background

KPMG US is a leading professional services firm that provides audit, tax, and advisory services to a wide range of clients across various industries.

As a data-driven organization, KPMG US relies on the people, process, and technology supporting the data estate to deliver high-quality, timely, and relevant insights to its clients and stakeholders.

However, the legacy data estate that KPMG US had in place in 2021 was not able to meet the growing and changing demands of the business.



Legacy limitations



The legacy data estate was based on an on-premises architecture that had several limitations.

KPMG initiated a cloud transformation enabling a more agile, efficient, and innovative data management and analytics environment capable of supporting Trusted AI objectives.

- High maintenance and operational costs
- Limited scalability and sub-optimal performance
- Data silos and access limitations
- Disparate data governance and security
- Inconsistent data quality and reliability
- Difficulty integrating and analyzing varieties of data



Transforming the data estate

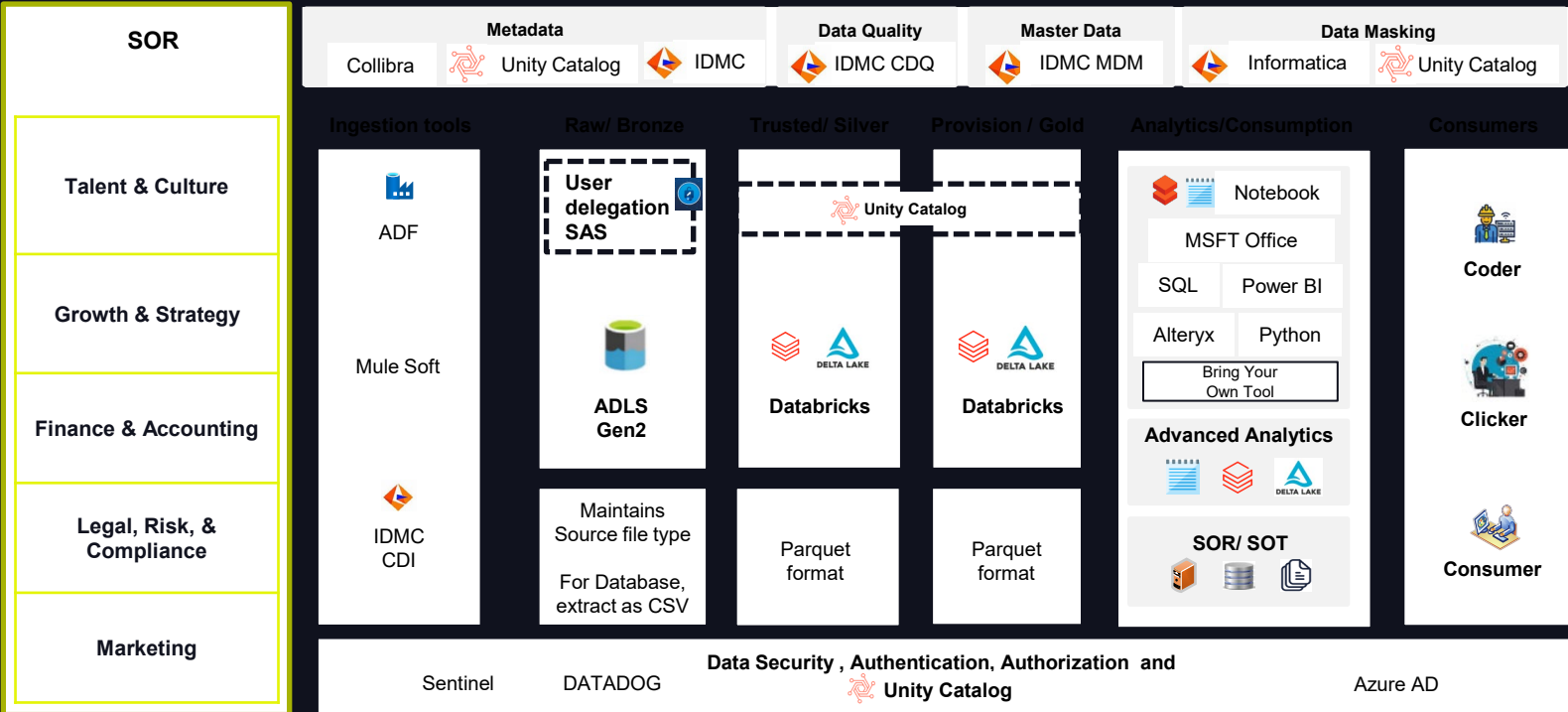
KPMG chose Microsoft Azure as its cloud platform, as it offered a comprehensive and integrated suite of services and tools for data storage, processing, and analytics.

- Understanding the future data and artificial intelligence (AI) requirements of our business
- Assessing the current state of the data estate and identifying the data sources, data types, data volumes, and data dependencies
- Designing the target state of the data estate and defining the data architecture, data models, data schemas, and data standards
- Planning the migration strategy and determining the migration tools, methods, and phases
- Executing the migration process and validating the data quality, accuracy, and completeness
- Optimizing the data estate and implementing data governance and security policies and practices



Current state architectural view

(2Q - FY24)



Integrated Data + AI Capabilities

Leveraging the capabilities of key partners to further optimize the data management and analytics processes by integrating these services with the Azure data estate.

Azure Databricks

- Accelerated data innovation and insights supporting scalable, performant, advanced data processing, analytics, and machine learning workflows
- Enabled data democratization and collaboration providing interactive, self-service for data exploration, visualization, and sharing
- Empowered data users and applications to access, analyze, and act on data using an enterprise-grade, AI-powered platform
- Secured data using metadata-driven approach and robust classification techniques

Informatica IDMC

- Enhanced integration of data quality dashboarding
- Improved reliability profiling, cleansing, enriching, and validating data from various formats and sources
- Improved data consistency and accuracy by centrally synthesizing master data and reference data across different systems and domains
- Increased data discoverability and understanding through cataloging and documenting data assets, metadata, lineage, and relationships



Benefits of cloud transformation

As a result of the transformation, KPMG US is realizing important benefits





Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

Learn about us:



kpmg.com

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without appropriate professional advice after a thorough examination of the particular situation.

© 2024 KPMG LLP, a Delaware limited liability partnership and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.

Latest Informatica Innovations for Databricks



Gen AI Solution Blueprint for Databricks DBRX

- Provides roadmap for development of RAG-based GenAI applications
- Grounds GenAI application inferencing, leading to high-fidelity, trusted responses
- Enables rapid development, maintenance, and scaling for GenAI apps by leveraging no-code/low-code orchestration



Informatica Cloud Data Integration-Free on Partner Connect

- CDI-Free now available via Databricks Partner Connect
- Quick, easy, and free access to data ingestions and transformation capabilities
- Up-to 20M rows or 10 ELT compute hours per month free



Native SQL ELT

- Enables in-database transformations with full pushdown execution
- Leverages Databricks native functions and computing power
- Allows customers to develop ETL pipelines with enhanced performance, plus scalable data transformations and integrations



IDMC Validated for Unity Catalog

- IDMC platform is now validated with Databricks Unity Catalog
- Includes all critical IDMC services
- Enhances Informatica's ability to manage data lineage and governance on the IDMC platform



Informatica Changes the Game



Ingest and Prepare Data Quickly and Easily

- Zero-cost Data Loading and Data Pipelines running natively on Databricks
(Now on Partner Connect!)



Scale Adoption with Trusted Data

- Enterprise-wide Data Governance, Data Quality and Catalog certified
Interoperability with Unity Catalog to accelerate onboarding and consumption of trusted data

Realize the Potential of Gen AI for the Enterprise

- Comprehensive Blueprint for Enterprise-Grade Gen AI applications with Databricks and Informatica IDMC



Next Steps

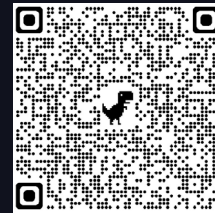
Scan QR codes to learn more



Cloud Data Integration
Free Service via
Partner Connect



Learn more on Informatica
with Databricks



Read about our Latest
Announcements