EDM Webinar

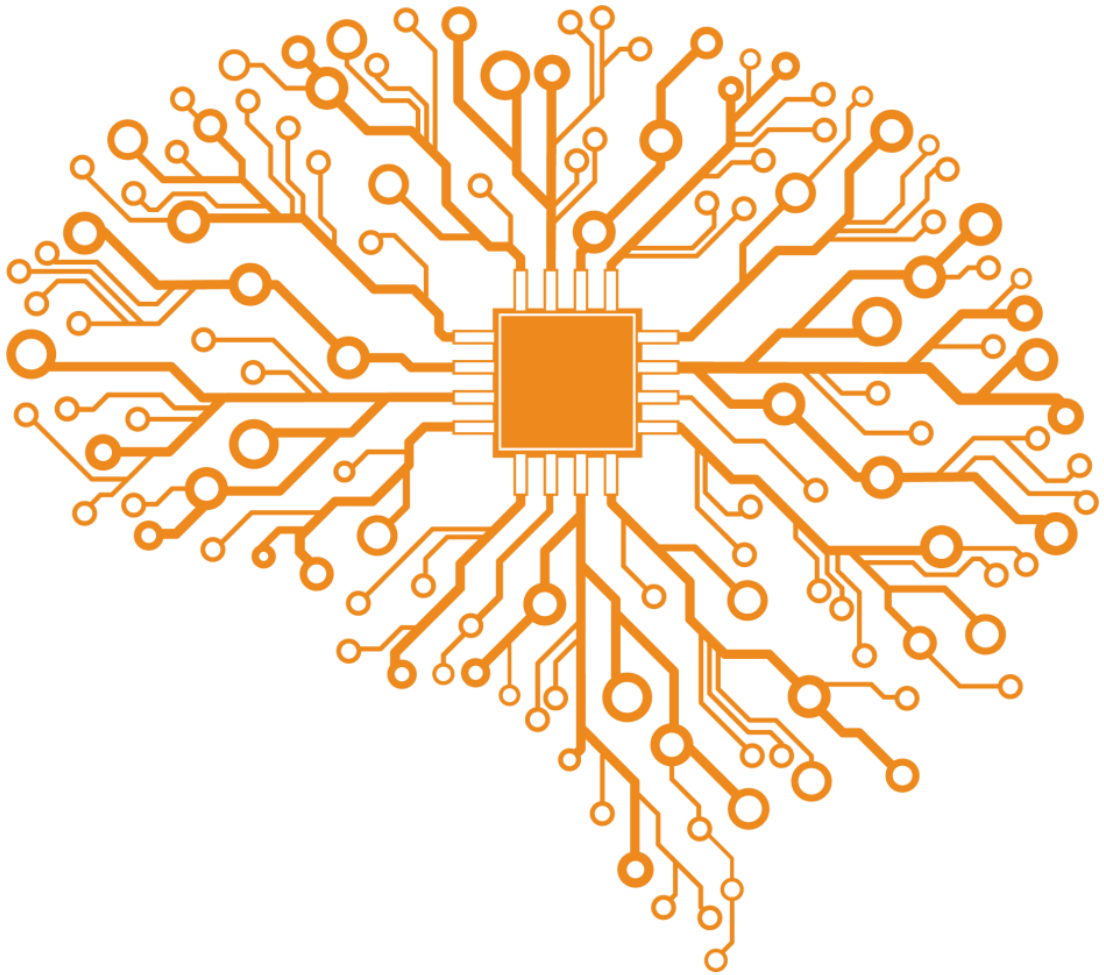
AI Needs Data Products (And Data Products Need AI)

A conversation with

Stephen Brobst
Chief Technology Officer
Ab Initio Software

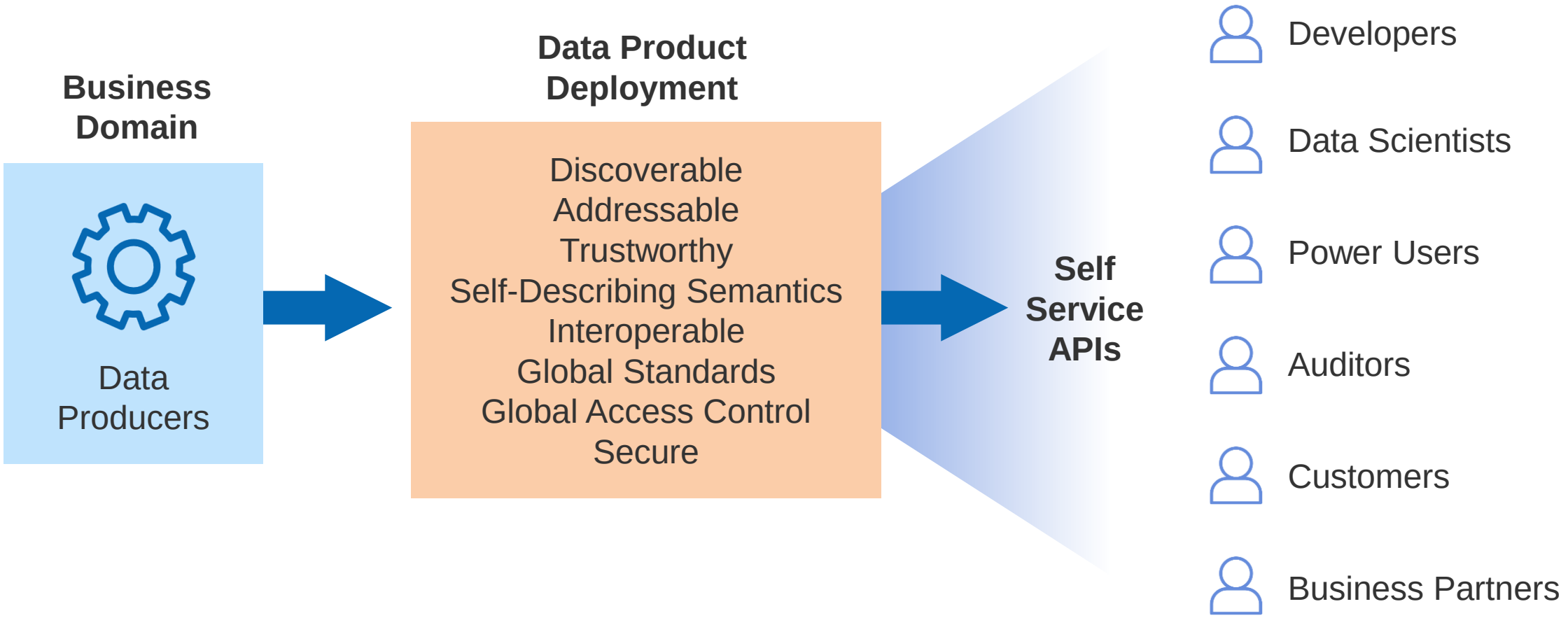
Dr. Debarag Banerjee
Chief AI & Data Officer
L&T Finance

The Resurgence of AI



> All companies are data and AI companies in the 21st century.

Anatomy of a Data Product



The Future of Data

“If you want to be data-driven…

…you need to be metadata-driven.”

–Gartner Data and Analytics Summit, London, 2024

Current Data Ecosystem

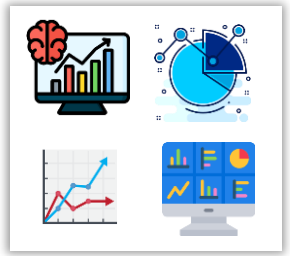
Data Governance

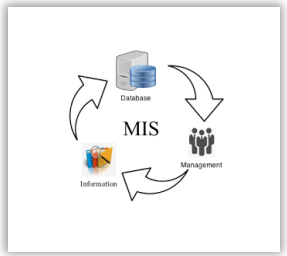
Qlik & Looker

MIS & Reporting

AI/ML

Democratized Views





Data Consumer Layer

Data Mart Layer

Sourcing | Collections | Finance/ Regulatory | Cyclops | Nostradamus | Gold | HRMS

Transformation

Data Lake Layer

Google Big Query & Bucket

Extract & Load

LOS | Collections | Bureau | Cyclops

LMS | Planet | BRE | HRMS

Enterprise Application Layer

Ab Initio at LTF

L&T Finance

Core Source Systems

- Apps: Planet, Brake, Clutch, Omni, Spoors
- LOS: MERC, Kuliza, Salesforce
- LMS: ONE, MERC, TCS
- Underwriting: CYCLOPS
- Accnt Aggregator: Perfios, Pirimid
- BRE: Experian
- Banking, Credit: Credentek, CIBIL, CRIF, Multi Bureau, Posidex, RMC
- Customer Connect: Axiom
- Internal Ops: Onex, Website, Workline

Source Types

- File Data
- DB Data
- ETL Extracts
- API Data

Storage: Data Lake Google Big Query

- Data Lake
- Data Mart

Ab Initio Data Governance Suite (GCP HOSTED)

- Data Lineage
- Data Quality
- Data Catalogue
- Business Glossary

Ab Initio Observability

- Management Dashboard
- Operational Dashboards
- Alerts Dashboards
- Quick Insights

Security

- Authentication
- Authorization
- Validation
- Encryption

Business Insights

- BI Reports
- AI/ML Insights
- Exploratory Analytics
- Gen AI Apps
- Data Exchange

Return On Data Investments - RODI

Objective

Measures the return of additional data investments

RODI Calculated for Existing DATA for Two Wheeler Cyclops

- Banking (Account Aggregator)
- Geo-Location Profiling
- Payment Aggregator
- Premium Bureau
- Alternate Bureau

Calculation Expected Return with Investment

$$RODI = \frac{\textit{Change in Interest Income \& Credit Loss due to use of Data}}{\textit{Cost of Acquisition of Data}}$$

Methodology is Submitted for Peer Reviewed Publication

A framework for measuring return on data investment (RODI)*

Debarag Banerjee† Kamlesh Kumar‡ Anik Paul§

The return of data investment on data asset Z is

$$RODI_Z = \frac{\Delta R_Z}{I_Z}$$

Where ΔR_Z is the total return, which is the sum of the change in expected credit loss and the expected return on loans.

$$\Delta R_Z = (ER_l(X,Z) - ER_l(X)) + (ECL_l(X) - ECL_l(X,Z))$$

The expected interest return is calculated as follows:

$$ER_l(X) = \sum_{i=1}^{N} M^i \left(1 - P_d(D(X^i, M^i)|\widehat{P}_d(D(X^i)) < \Delta)\right) \left[\frac{rT}{(1+r)^T - 1} + (rT - 1)\right]$$

The expected credit loss is calculated as follows:

$$ECL_l(X) = \sum_{i=1}^{N} MP_d\left(D(X^i, M^i)|\widehat{P}_d(D(X^i)) < \Delta\right)$$

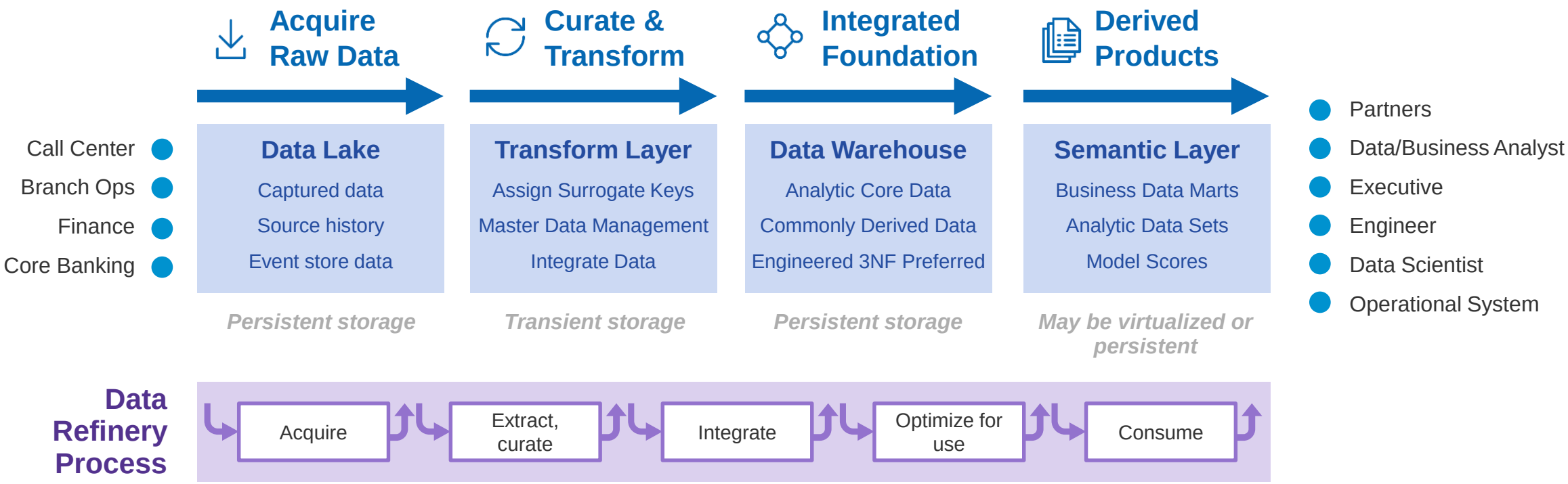
Last, I_Z is the investment in procuring a data asset.

Generative AI is Revolutionizing Data Products

Faster and easier to create data products.

Faster and easier to consume data products.

Data Product Construction



Data Product Construction Using Agentic AI

- Automate the mapping from raw data in the data lake to curated data in the data warehouse.
- Automate the creation of data quality controls.
- Automate the creation of business metadata for describing data products.
- Automate the creation of semantic models for accessing data products.

Critical Success Factor for Agentic AI: Metadata Context

Example input to Agentic AI execution for data product construction:

- Technical metadata and profiles for raw data (capture via automated scanners).
- Knowledge graph to describe raw data relationships.
- Internal documentation describing rules and operating procedures related to target data product.
- External and regulatory documentation describing rules and operating procedures related to target data product.
- Target (foundation) data model.

Data Product Consumption using Agentic AI

- Marketplace for finding the “right” data for the task at hand.
- Natural language interface for conversations with metadata.
- Natural language interface for conversations with data.

- Knowledge worker should *not* need to understand location of data.
- Knowledge worker should *not* need to understand underlying technology for storing data.

Critical Success Factor for Agentic AI: Metadata Context

Example input to Generative AI execution for data product consumption:

- Business metadata describing data product.
- Technical metadata describing data product location, technology, and structure.
- Semantic metadata describing mapping from knowledge worker view of data into foundation data products and/or derived data products..
- Data product lineage, data quality metrics, data product contracts.

The Future of Data

“If you want to be AI-driven…

…you need to be metadata-driven.”

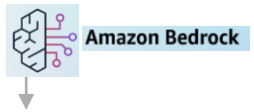
–Gartner Data and Analytics Summit, London, 2025

Future Proofing Deployment of Generative AI

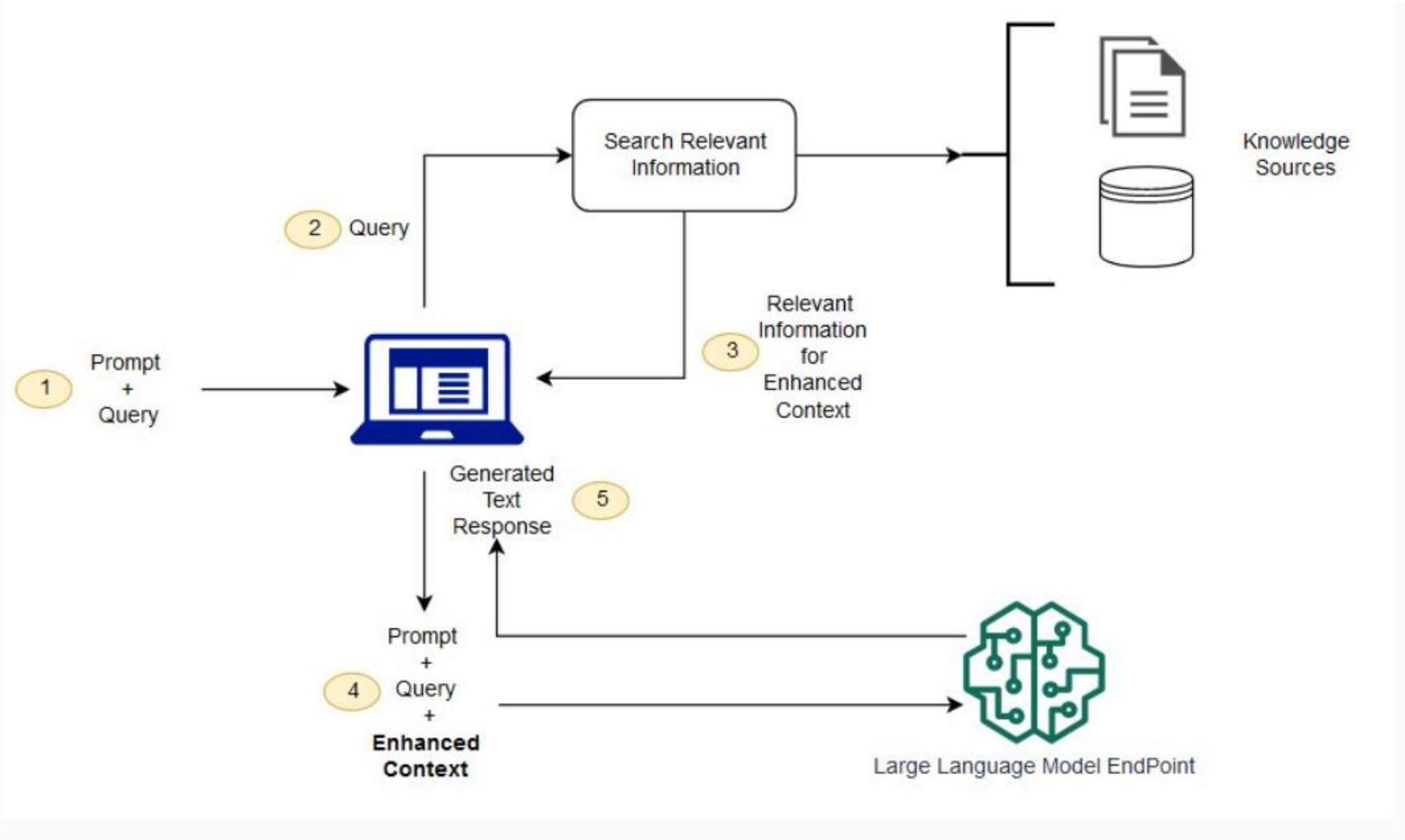
Whatever LLM framework is the “best” for the job today is unlikely to be the “best” answer in 6-12 months.

Whatever cloud is the “best” for the job today is unlikely to be the “best” (or only) answer in 24-36 months.

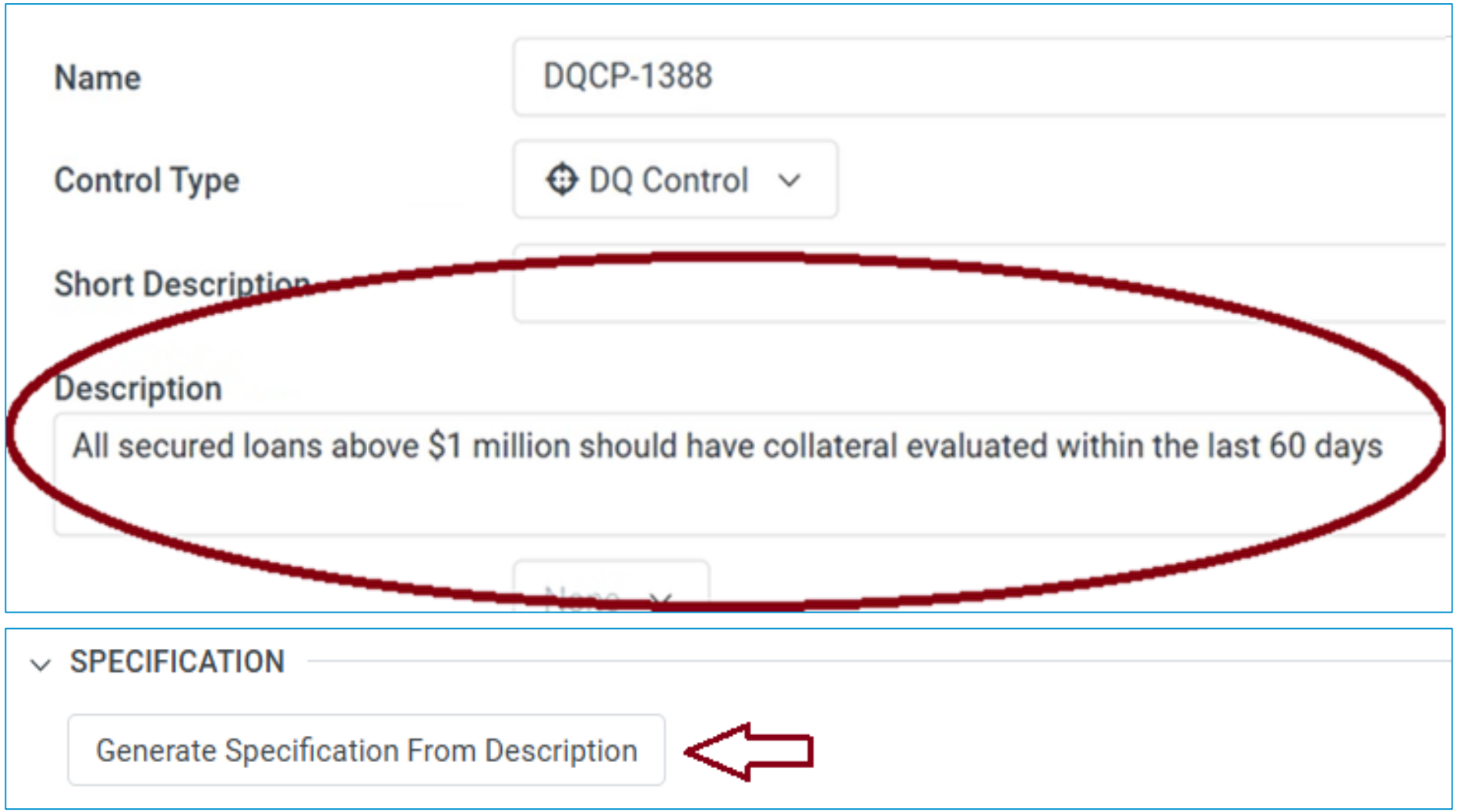
Bottom Line: Design for portability.



Retrieval Augmented Generation (RAG)

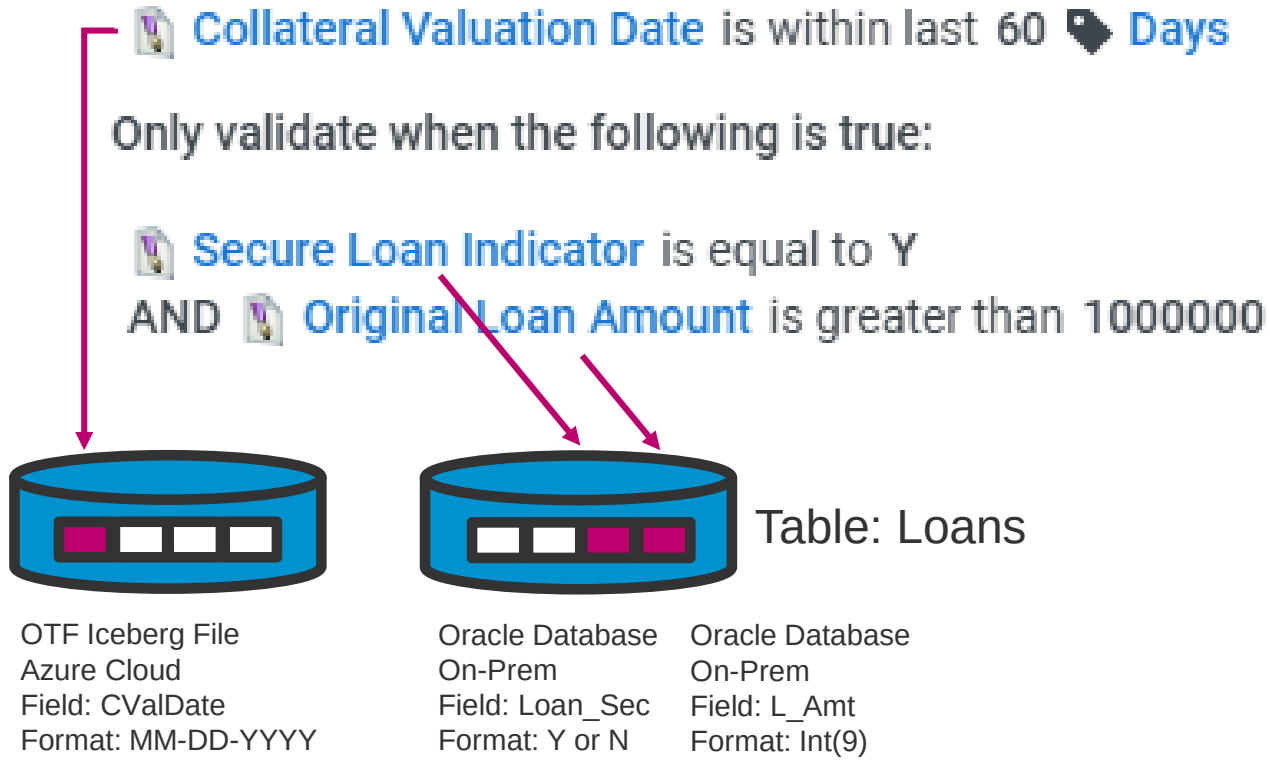


Knowledge Fabric + AI to Generate DQ Controls



Architectural Principle:
Low / No Code with Self Service Computing

- Co-pilots using Generative AI for constructing data products
- Co-pilots using Generative AI for conversations with data as well as metadata
- Create and manage metadata
- Deploy integration patterns
- Empower business users



Fastest Path to Data Products

- Use of Generative AI powered by an LLM and underlying metadata to automate creation of data products.
- Comprehensive data governance to create trust in the data product content with lineage, data quality rules, access contracts, semantic layer aligned to business knowledge worker communities, etc.
- API framework for programmatic access to data as well as business self-service access to data via no/low code and Generative AI.
- Data marketplace allowing business knowledge workers to shop for data using rich metadata enabled for discovery (by humans or AI algorithms).
- Data products can by virtualized or physicalized (or hybrid) using best practices Data Fabric capabilities.

Call to Action

Embrace Data as a Product

- Package data to be reused across multiple knowledge worker communities
- Align semantic models to knowledge worker communities
- Establish service levels for data (time, durability, quality)
- Deliver radical self-service
- Move with agility from discovery to production
- Determine a balance among conflicting goals or different perspectives

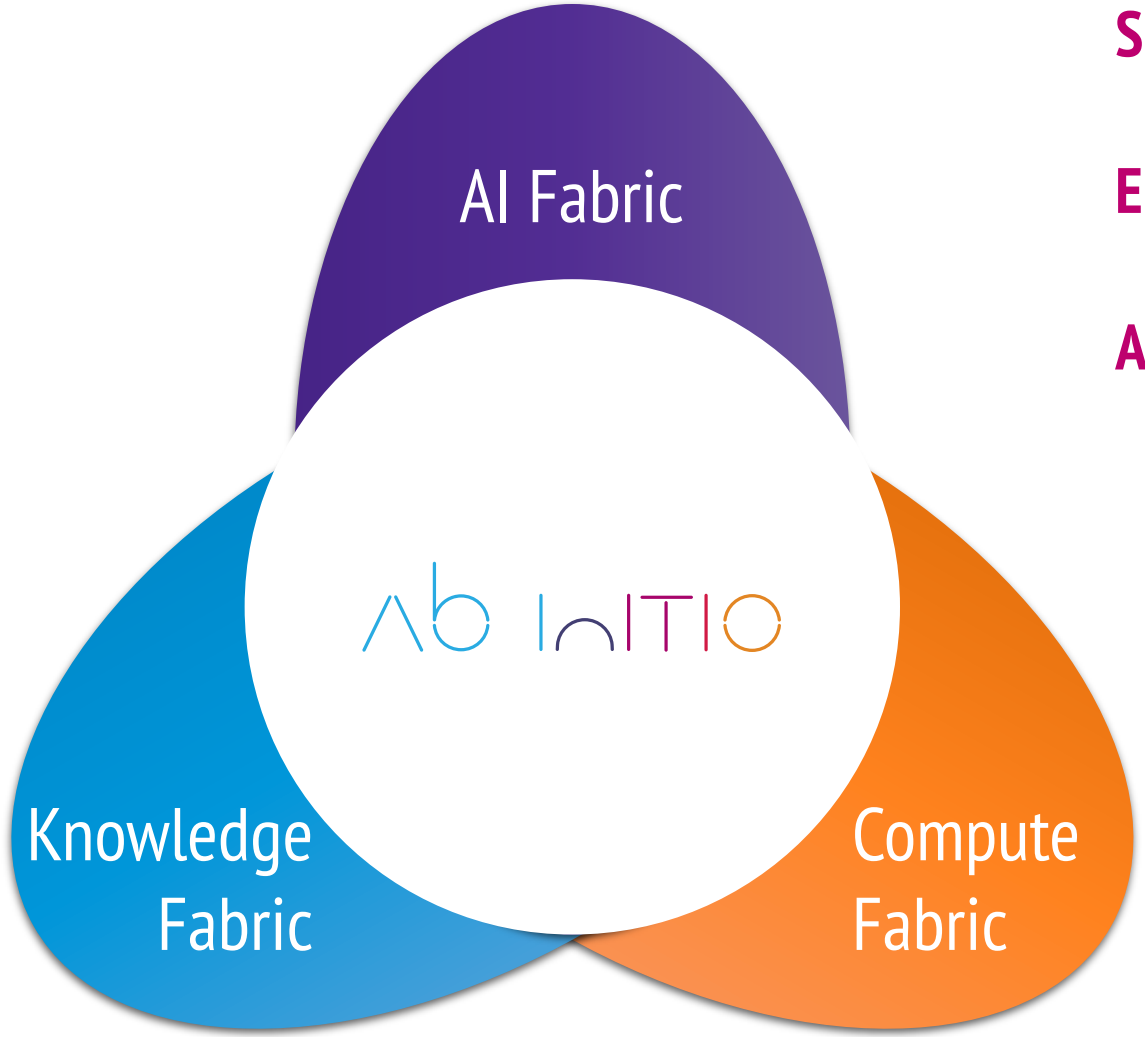
Embrace Generative AI with Metadata Context

- To accelerate construction of data products
- To drive natural language conversations with data
- To drive natural language conversations with metadata

Embrace the Ecosystem

- Leverage MCP to allow AI Agents to interoperate across multiple platforms.
- Support for heterogeneous platforms, as well as an evolving LLM landscape
- Leverage best of breed services and LLMs from both public and private clouds, including open source and ISVs

Ab Initio Converged Data Management Platform

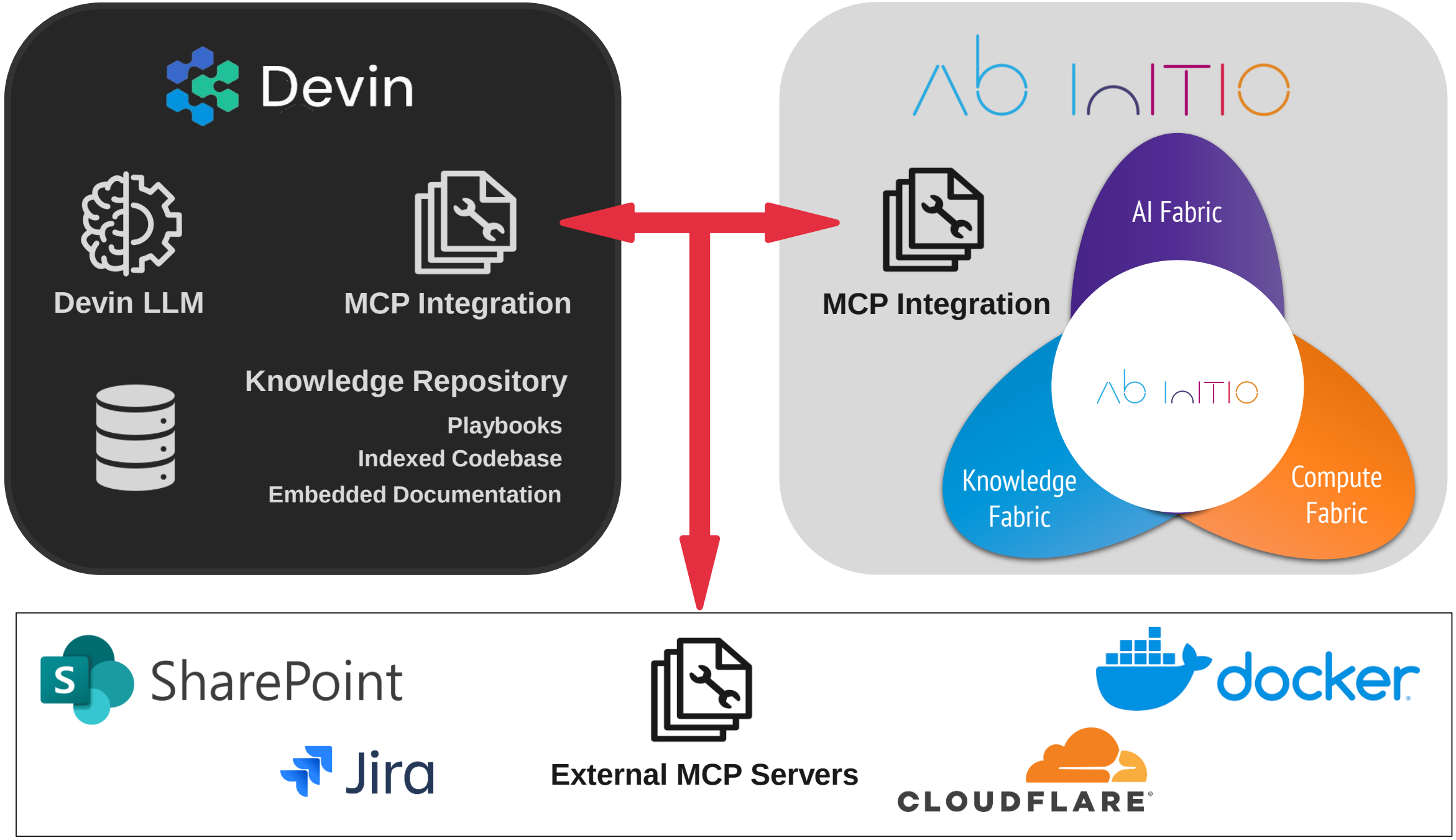


SIMPLIFY the data landscape

ELEVATE the business user

ACCELERATE data management

- data products
- data governance
- data integration
- data acquisition
- data compliance


Devin
Devin LLM
MCP Integration
Knowledge Repository
Playbooks
Indexed Codebase
Embedded Documentation
MCP Integration
AI Fabric
Knowledge Fabric
Compute Fabric
SharePoint
Jira
External MCP Servers
docker
CLOUDFLARE

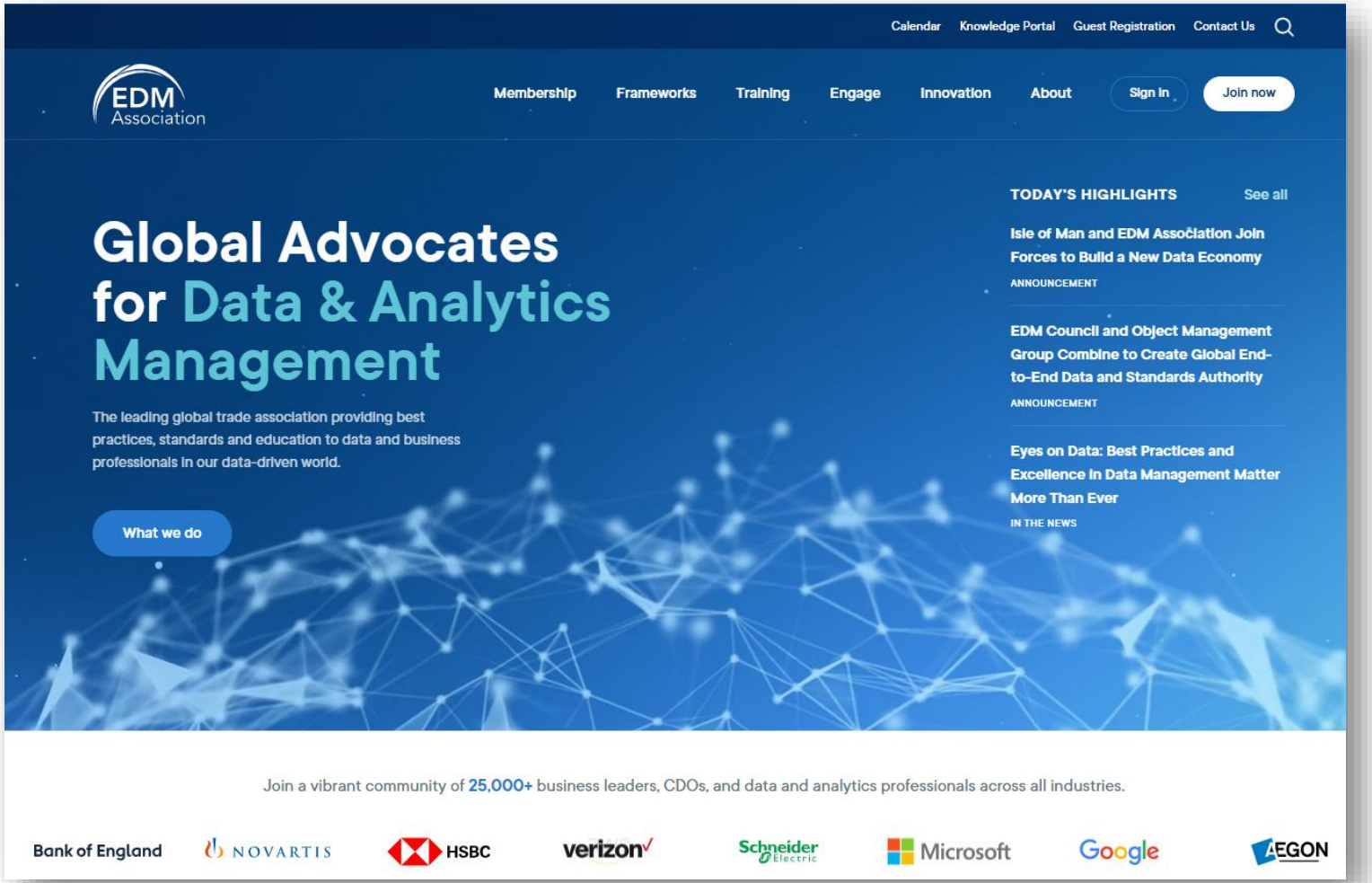
EDM Webinar

Questions?

EDM Association

Learn more about EDM Association

(formerly EDM Council)



EDM Association

350+ Member Firms
Cross-industry,
including Regulators

25,000+
Professionals

Worldwide
Americas, Europe,
Africa, Asia, Australia

edmcouncil.org

DCAM

CDMC EDMAssociation
Cloud Data Management Capabilities

Data ROI EDMAssociation

elc eLearningCurve

EDM Association