Unifying Enterprise Data

How a Data Fabric Converges Application and Data Integration

BY JAY PISCIONERI
MARCH 2024

About the Author

Jay Piscioneri has over 25 years of experience in data technologies, including data warehousing, business intelligence, data quality, and data governance. He's worked with organizations in various industries, helping them plan and build their next-generation data platforms. As a leader of many initiatives, Jay brings extensive experience in the nuts and bolts of implementing data solutions and the organizational challenges of adopting new priorities, processes, and tools.

About Eckerson Group

Eckerson Group is a global research, consulting, and advisory firm that helps organizations get more value from data. Our experts think critically, write clearly, and present persuasively about data analytics. They specialize in data strategy, data architecture, self-service analytics, master data management, data governance, and data science. Organizations rely on us to demystify data and analytics and develop business-driven strategies that harness the power of data. **Learn what Eckerson Group can do for you!**

Eckerson Group

RESEARCH   **CONSULTING**   ADVISORY

G E T · M O R E · V A L U E · F R O M · Y O U R · D A T A

About This Report

This report is sponsored by Promethium, which has exclusive permission to syndicate its content.

Table of Contents

Executive Summary

Segregated Enterprise Data

Companies today face ongoing pressure from fast-changing customer expectations, innovative rivals, and new regulations. They understand that data is the key to quickly identifying and responding to a fluid business environment. Thus, many organizations have a growing ecosystem of applications and data sources that help them run their business and make decisions. With data in different locations and diverse formats, they need to share data between applications or prepare it for analytics through integration.

Integration tools and methods fall into two distinct categories: application integration and data integration. Application integration distributes data between different operational systems to improve efficiency. Data integration aggregates and harmonizes diverse data for analytical purposes. Unfortunately, this separation complicates things, making it hard to use data from one sphere in the other. For developers, this creates redundant work and data inconsistencies. For data consumers, it limits their self-serve options to those assets managed with data integration. They can't access valuable operational data because it's ensconced behind technical barriers they don't know how to surmount.

Solution Approach

Organizations can bridge the gap between the operational and analytical data realms by using a data fabric that connects to enterprise application integration (EAI) tools. Data fabric is an architectural approach for analytics that integrates and simplifies data, making it more easily accessible for business users. On the other hand, EAI tools are designed to enable software engineers to synchronize data among operational applications in real-time or in small batches. Integrating EAI tools with a data fabric provides a comprehensive view of operational and analytical data and facilitates the bidirectional flow of data between them. This approach also provides a common user experience that serves both business and technical users.

Recommendations

To bridge the worlds of operational applications and analytics, consider the following steps:

> *Identify and prioritize analytical use cases that require operational data. Evaluate analytics demand for operational data to determine the extent to which your organization can benefit from implementing EAI as a source for a data fabric. Interview data analysts, data scientists, and business leaders. Review backlogged requests for application data.*

> *Identify and prioritize use cases that require analytical data embedded in operational applications. Ask business leaders and operational workers how having analytical data points in their operational systems can streamline their work and improve their decisions.*

> **Evaluate current integration capabilities.** *Assess your existing data and application integration landscape to determine how ready your organization is to implement a combined EAI-and-data fabric solution.*

> **Prioritize data quality.** *Implement or enhance data quality monitoring to prevent misguided analytical conclusions.*

> **Educate and empower users.** *Provide training and resources to both technical and business users on the benefits of a unified data and application integration framework.*

> **Monitor and optimize costs.** *Keep a close watch on compute and storage costs associated with increased data access. Evaluate and prioritize analytical initiatives based on the cost of resource consumption versus the benefit to the organization.*

Two Worlds

In today's rapidly evolving digital landscape, organizations rely on a myriad of application and data sources to run their business and inform decisions. To make data accessible when and where it's needed, they use integration approaches that fall into two distinct categories: application integration and data integration. Application integration enables organizations to share operational data between applications, while data integration allows them to collect and prepare data for analytics.

However, these two approaches leave enterprise data divided, making it harder for organizations to make full use of their data. In this report, we'll explore how this separation complicates the enterprise data landscape, and we'll look at a novel solution using data fabric to break down the artificial barriers between data residing in both operational and analytical realms.

Barriers to Data. Application integration and data integration have evolved as distinct disciplines, each with its own dedicated platforms, tools, and methodologies. The divergence of these two approaches makes sense, given their different objectives. However, data is the common denominator.

The separation between application and data integration creates a barrier preventing businesses from gaining a holistic view of their data landscape.

The separation between application integration and data integration creates a barrier preventing businesses from gaining a holistic view of their data landscape. It also makes it more difficult for them to leverage data from one sphere that's valuable in the other. The result is redundant work, inconsistencies in data handling, and a fragmented user experience.

The Role of a Data Fabric. Organizations can bridge the two worlds to address these challenges with a data fabric that manages connections to disparate sources for discovery and other purposes. This unifying layer provides a cohesive user experience for building federated queries and data pipelines and accessing source documentation such as lineage, acceptable usage, data classifications, and quality.

Extending a data fabric to include operational data via EAI tools is an ideal solution to unify the two worlds.

Extending a data fabric to include operational data via EAI tools is an ideal solution to unify the two worlds. It empowers organizations to fully leverage their data wherever it is and in any format. It also provides a standardized user experience for finding, understanding, and working with data. (See the **Appendix** for more details about data fabric.)

Current State Challenges

Challenges for Analytics Users. The separate worlds of application and data integration present significant challenges, particularly for analytics data users. They often can't access or even find operational data in SaaS applications, such as **Salesforce** or **Shopify**, or data from ERP applications, such as **SAP**.

Analytics users often can't access or even find operational data in SaaS applications.

These enterprise applications don't store their data in simple tables or lists. Instead, they employ an abstraction layer over their physical data storage. This presents several challenges for analytics users:

> **An unfamiliar environment.** Analytics users have come to expect a self-serve environment for data discovery, data access, and pipeline development. The lack of these familiar features in the application sphere puts operational data out of reach.

> **Insufficient technical knowledge.** Analytics users lack sufficient technical knowledge to find and use the application data they need on their own.

> **Complex Integration tools.** The enterprise application integration (EAI) tools, such as **Boomi** or **Workato** that help software engineers navigate application data are too complex for non-technical users.

> **Help from software engineers.** Even if an analytics user knows what operational data they need from an application, they don't have the means to access it without help from overburdened software engineers.

Challenges for Software Engineers. Software engineers face other challenges in a segregated data landscape, such as:

> **Barriers to cross-sphere data flows.** Software engineers need to embed analytical data in applications. Both are hampered by a lack of familiarity with the corresponding environment, which slows down solution development.

> **Insufficient application knowledge.** Even with the help of EAI tools, software engineers need to know an application well. For example, to configure an integration in an EAI tool, they need to know the specific data object names in each application they want to integrate.

> **The curse of "the expert".** The software engineer who knows an application well is often inundated with requests for help from colleagues who can't access or navigate its data themselves. So, the expert has less time for their own work.
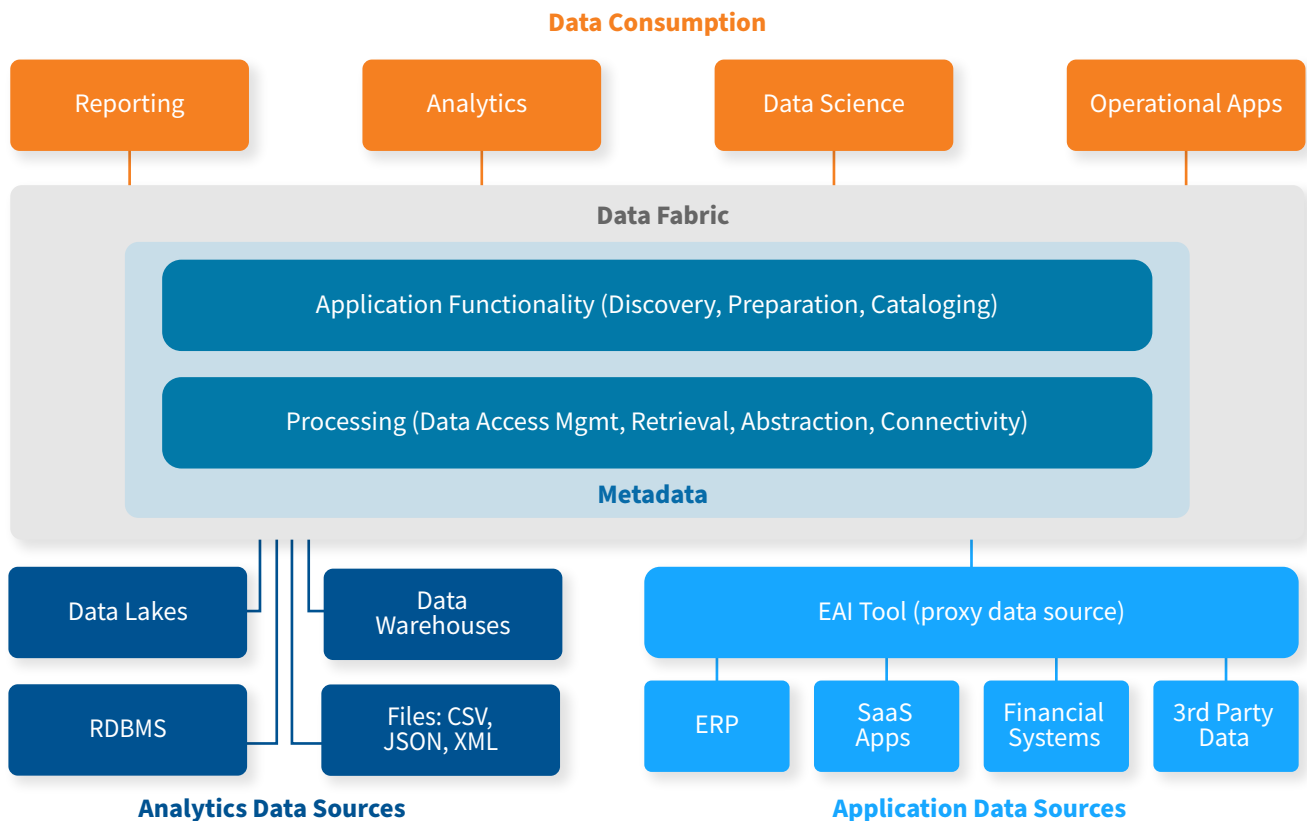
Data Fabric Solution

A data fabric integrated with an EAI tool bridges the gap between operational and analytical data. Here, an EAI tool acts as a single proxy data source for all the operational data it manages. This enables a user to find and work with complex application data sources in the same way they're accustomed to working with analytical data.

An EAI tool acts as a proxy data source for all operational data it manages.

For instance, an analyst can self-serve data, such as customer touch-point records from **Salesforce**, transaction data from **Shopify**, and product inventory data from **SAP**, without the help of developers

Figure 1. Data fabric brings enterprise data together with a unified user experience

Data Consumption

| Reporting | Analytics | Data Science | Operational Apps |

Data Fabric

Application Functionality (Discovery, Preparation, Cataloging)

Processing (Data Access Mgmt, Retrieval, Abstraction, Connectivity)

Metadata

| Data Lakes | Data Warehouses |

| RDBMS | Files: CSV, JSON, XML |

Analytics Data Sources

EAI Tool (proxy data source)

| ERP | SaaS Apps | Financial Systems | 3rd Party Data |

Application Data Sources

who have deep knowledge of those applications. Thus, users can add operational data to their analytical toolbox.

By extending data fabric with an EAI tool, companies can enable their business teams to leverage the entire data ecosystem for informed decision-making and innovation. Figure 1 illustrates how an EAI tool fits into a data fabric.

Software engineers also benefit from an EAI-enhanced data fabric. They may have the technical skills to determine how to find and extract operational data from three Saas applications. But it will take them much more time to do so without the fabric's abstraction layer and AI-augmented application functions.

Benefits

An EAI-enabled data fabric offers several benefits:

> **Efficiency and Productivity for Analytics Data Consumers**. Data fabric's unified user experience significantly lowers barriers for data analysts and data scientists to access operational data. With simplified data discovery and integration, it makes operational data readily accessible to users who don't have extensive technical knowledge.

> **Efficiency and Productivity for Data Engineers**. An EAI tool acts as an abstraction layer within the fabric—an abstraction layer for the abstraction layer. So, data engineers don't need to know the many technical details about connections and APIs for each application. Instead, they use the EAI tool as a single data source that obscures those details.

> **Improved Data Governance and Compliance**. Managing application and analytical data assets with data fabric ensures a uniform data access and quality management approach. It enables organizations to enforce policies consistently across diverse data sources and types.

Considerations

Integrating an EAI tool into a data fabric offers promising advantages, but it also comes with a set of limitations and considerations:

> **Data Fabric and EAI Prerequisite**. Organizations must have a data fabric and an EAI tool in place to integrate them.

> **Data Security and Access Controls**. Protecting data while enabling appropriate access for authorized users is an ongoing challenge, even for an individual application. However, organizations face additional challenges when unifying data from multiple sources. For example, diverse

applications often have different role structures and RBAC (Role-Based Access Control) privileges that must be reconciled.

> **Risks Associated with Data Democratization**. Enabling users to access more data exposes them to assets they may not fully understand how to use, which can lead to inaccurate conclusions and unsound decisions. Also, increased data access can lead to greater processing loads and latency in operational systems. The elastic scalability of cloud data platforms can address these challenges. But simply scaling up can result in unpleasant surprises in compute and storage costs.

These considerations show that to integrate EAI and data fabric successfully, organizations must devise a strategic approach that takes both the technical challenges and organizational impacts into account.

Integration Use Cases

Now let's explore two use cases that illustrate the advantages of extending data fabric to include data managed by an EAI tool. The first use case embeds analytic data in a SaaS application. The second use case involves an ad hoc study to find new customer segments based on use of the company's eCommerce gift registry.

Note: These use cases are greatly simplified so we can focus on the advantages of including an EAI tool in a data fabric

Use Case 1: Bringing Analytics Data into Applications

The objective of our first use case is to retrieve personalized offers from an analytics system using a data fabric. On the analytics side, we have cleansed shopping data in a Postgres database and customer demographic data in Snowflake. Those applications feed data to an ML model that generates customized offers for a customer based on their shopping behavior(see Figure 2).

Thanks to data fabric, the data engineer responsible for this integration doesn't have to know how to connect to **Postgres** or **Snowflake** or what **Shopify** APIs to call to store the personalized offers. The data fabric and EAI tool abstraction layers both manage those complexities.

Figure 2. Data fabric simplifies the flow of data from analytics to applications
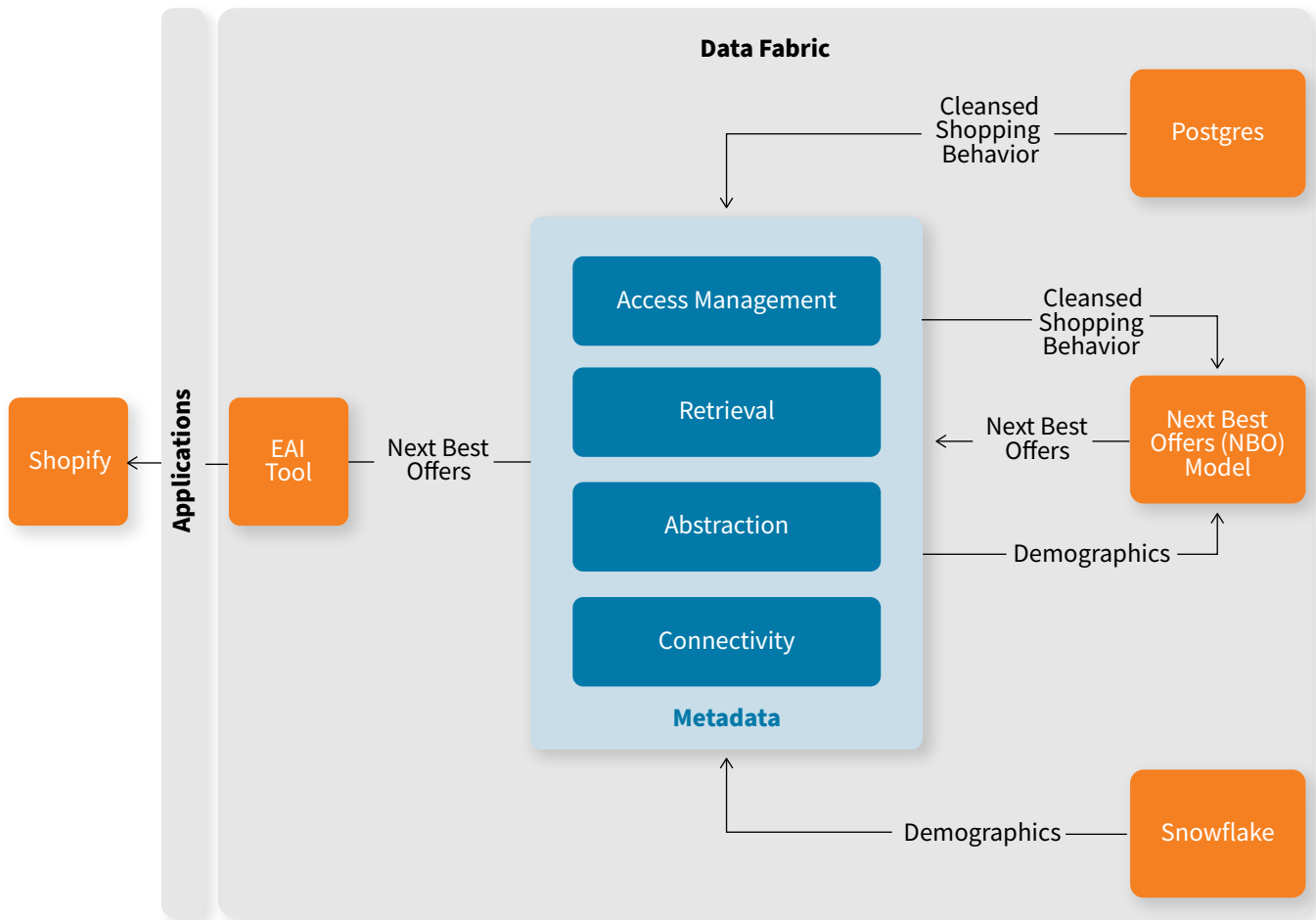

Use Case 2: Making Application Data Accessible for Analytics

The scenario for this use case is a data analyst hypothesizes that customers who use the gift registry make more frequent purchases. He wants to apply statistical models to data from the eCommerce website and bring in customer data from **Salesforce** to associate customer characteristics with their eCommerce behavior. If the data supports his theory about purchasing patterns, then he can use the CRM data to propose customer segmentation for marketing campaigns to promote the gift registry.

In this case, data fabric simplifies the step of gathering data and lets the analyst focus on analysis. He doesn't have to know where the eCommerce or CRM data comes from. The data engineers who manage the data fabric don't have to create separate custom connections to the SaaS applications. Instead, they use the EAI tool as a single proxy data source, enabling them to onboard that data for analytics in less time. The analyst finds the application data he needs using data fabric's unified discovery functions. And the fabric's GenAI-augmented data preparation process helps him write his analytical programs.

Figure 3. Data fabric uses the EAI tool as a proxy data source to SaaS applications
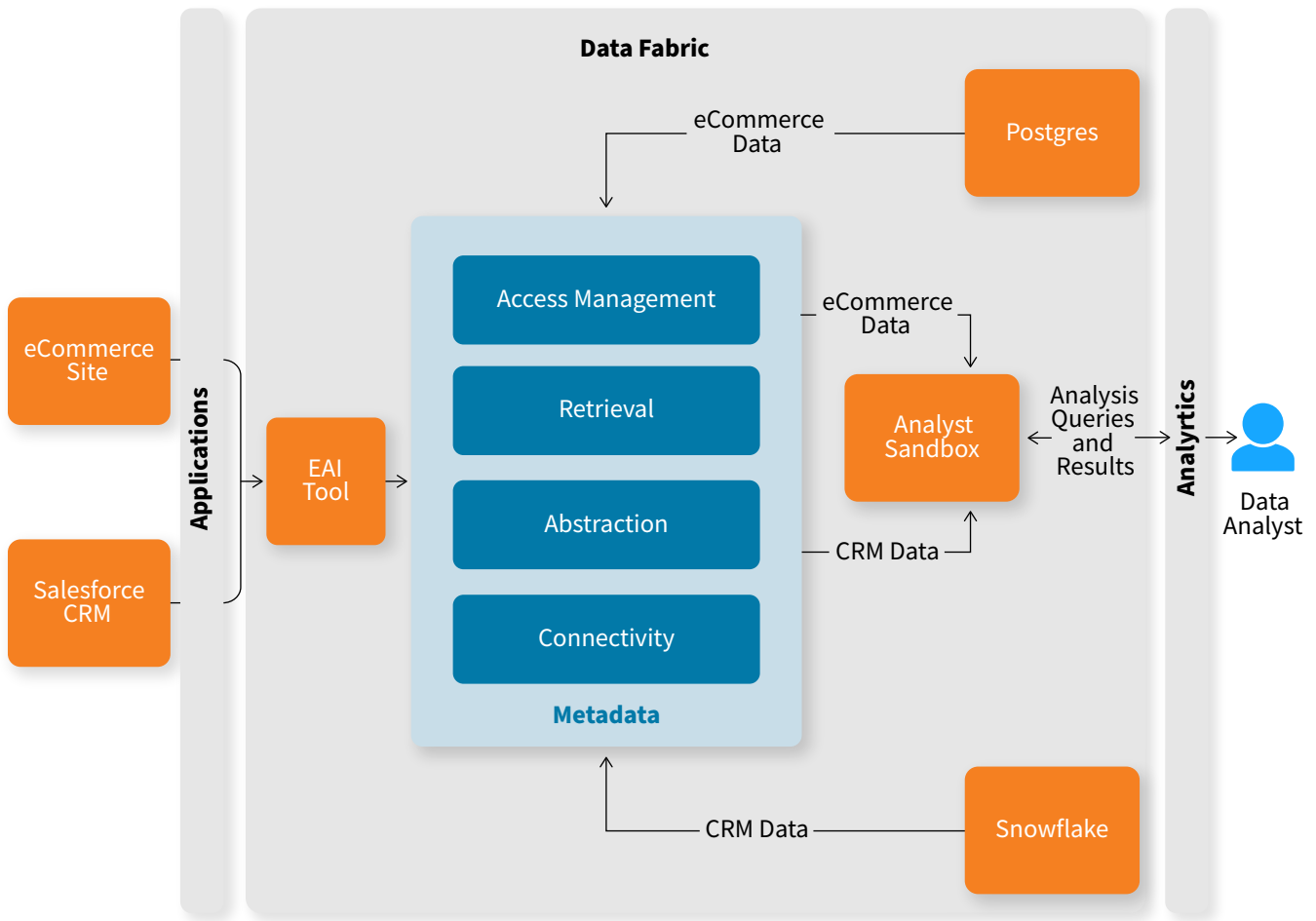
Conclusion and Recommendations

A Unified Enterprise Fabric

EAI tools and data fabric serve different users with different needs and objectives. Integrating them to create a unified enterprise fabric gives analytics consumers a comprehensive view of their organization's data landscape in a simplified and familiar environment. However, it leaves in place the features that software engineers require for application development and integration.

As organizations increasingly adopt a data fabric, integrating it with EAI tools will become a standard.

We anticipate that as organizations increasingly adopt data fabric, integrating it with EAI tools will become a standard. This evolution will further blur the lines between operational and analytical data. Bringing all enterprise data under a common democratizing framework will lead to more agile, responsive, and data-driven businesses.

Recommendations

To get started with bridging the worlds of applications and analytics through EAI and data fabric, consider the following:
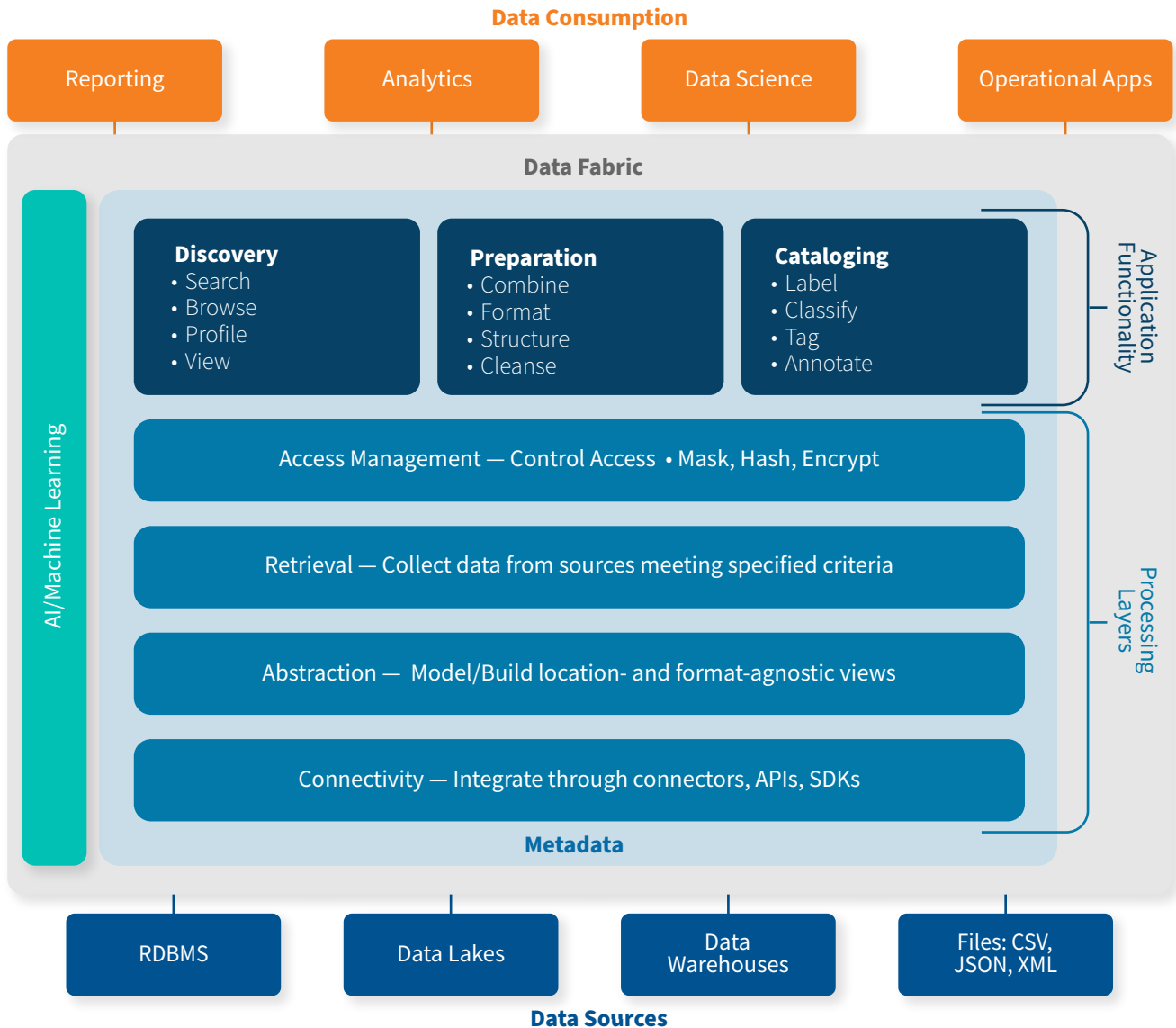
> **Identify and prioritize analytical use cases that require operational data.** Evaluate analytics demand for operational data to determine the extent to which your organization can benefit from implementing EAI as a source for a data fabric. Interview data analysts, data scientists, and business leaders. Review backlogged requests for application data.

> **Identify and prioritize use cases that require analytical data embedded in operational applications.** Ask business leaders and operational workers how having analytical data points in their operational systems can streamline their work and improve their decisions.

> **Evaluate current integration capabilities**. Assess your existing data and application integration landscape to determine how ready your organization is to implement a combined EAI-and-data fabric solution.

> **Prioritize data quality**. Implement or enhance data quality monitoring to prevent misguided analytical conclusions.

> **Educate and empower users**. Provide training and resources to both technical and business users on the benefits of a unified data and application integration framework.

> **Monitor and optimize costs.** Keep a close watch on compute and storage costs associated with increased data access. Evaluate and prioritize analytical initiatives based on the cost of resource consumption versus the benefit to the organization.

Appendix

Overview of Data Fabric

Data fabric is an integration approach for analytics that uses an abstraction layer to hide the complexity of working with distributed data and disparate formats. It unifies the separate functions of data management—integration, preparation, cataloging, security, and discovery—into a cohesive process through intelligent automation. Figure 4 below illustrates data fabric's core components. Let's review these components to discover their purpose.

Figure 4. The Core Components of Data Fabric

Data Consumption

| Reporting | Analytics | Data Science | Operational Apps |

Data Fabric

AI/Machine Learning

Application Functionality:

Discovery
• Search
• Browse
• Profile
• View

Preparation
• Combine
• Format
• Structure
• Cleanse

Cataloging
• Label
• Classify
• Tag
• Annotate

Processing Layers:

Access Management — Control Access • Mask, Hash, Encrypt

Retrieval — Collect data from sources meeting specified criteria

Abstraction — Model/Build location- and format-agnostic views

Connectivity — Integrate through connectors, APIs, SDKs

Metadata

Data Sources

| RDBMS | Data Lakes | Data Warehouses | Files: CSV, JSON, XML |

Metadata

Metadata serves as the foundation of a data fabric. It uses metadata to create an abstraction layer that hides the complexity of working with distributed data and diverse formats, whether for operational or analytical use cases.

Application Functionality

A data fabric's application functions provide a common interactive experience that enables business users, power users, and data teams to find and work with analytical data. Adding EAI-managed data assets to a data fabric extends the productivity benefits of its optimized discovery, preparation, and cataloging capabilities to operational data.

> **Discovery:** Makes data assets easy to find. It enables users to browse through or search for available data sets and preview data from different sources to better understand whether it helps them in their analysis. Search functions vary from basic keyword search to AI-assisted search with natural language processing (NLP).

> **Preparation:** Helps data consumers, using AI, to gather and organize data for ad hoc analysis. For example, data scientists use preparation functions to gather and organize the data they need for their machine-learning models, while data teams use preparation functions to conform and aggregate data for reporting.

> **Cataloging:** Enables users to create, view, and manage metadata. It's the interface that displays information about data resources during discovery and allows users to enrich the information with tags, comments, ratings, and certifications.

Processing Layers

Data fabric's processing layers do the heavy lifting of fulfilling analytical data requests by providing connections to data sources, managing workflows, abstracting data to create a unified view, and ensuring access to data is appropriately controlled.

> **Connectivity:** Provides a communication channel between data sources and the data fabric.

> **Abstraction:** It hides the disparate formats and distributed locations of data sources from consumers, making it easier to use enterprise data. Since they interact with abstracted versions of source data, they don't have to worry about where the data is or what platform it's coming from.

> **Retrieval:** A query engine that collects data from source systems that meet specified criteria for a given purpose.

> **Access management:** Determines who can use certain data based on the consumer's role and the classification of data requested.

About Eckerson Group

Wayne Eckerson, a globally known author, speaker, and consultant, formed **Eckerson Group** to help organizations get more value from data and analytics. His goal is to provide organizations with expert guidance during every step of their data and analytics journey.

Eckerson Group helps organizations in three ways:

> **Our thought leaders** publish practical, compelling content that keeps data analytics leaders abreast of the latest trends, techniques, and tools in the field.

> **Our consultants** listen carefully, think deeply, and craft tailored solutions that translate business requirements into compelling strategies and solutions.

> **Our advisors** provide competitive intelligence and market positioning guidance to software vendors to improve their go-to-market strategies.

Eckerson Group is a global research, consulting, and advisory firm that focuses solely on data and analytics. Our experts specialize in data governance, self-service analytics, data architecture, data science, data management, and business intelligence.

Our clients say we are hardworking, insightful, and humble. It all stems from our love of data and our desire to help organizations turn insights into action. We are a family of continuous learners, interpreting the world of data and analytics for you.

Get more value from your data. Put an expert on your side. **Learn what Eckerson Group can do for you!**

Eckerson Group

RESEARCH   CONSULTING   ADVISORY

G E T · M O R E · V A L U E · F R O M · Y O U R · D A T A

About the Sponsor

Be Ready to Answer Tomorrow's Questions Today

Answering important business questions with enterprise-wide data has become an unsustainable and slow cycle of repeatedly moving and transforming data. This can't happen without data products. Good data products can only be built on a solid data fabric foundation.

By presenting a unified and consistent view of all of an enterprise's data and providing the ability of fast & secure access to the data, Promethium can enable trusted data products to be assembled with far greater agility and ease.

At Promethium, we are relentlessly focused on helping organizations break free from the inefficient data relay holding them back. We have invented the future of data analytics with a solution that prioritizes fast access to trusted answers in minutes instead of months so you will never be surprised by unexpected outcomes.

With Promethium, our customers find new insights and answer pressing business questions while reducing project backlogs, decreasing the total cost of ownership, while simplifying the quest to a single and consistent view + access of their data.

For more information visit https://www.promethium.ai/