MCKNIGHT CONSULTING GROUP

Choosing Your Provider for Implementing a Data Fabric

Presented by: William McKnight

"#1 Global Influencer in Big Data" Thinkers360

President, McKnight Consulting Group

3 X Inc 5000

/in/wmcknight

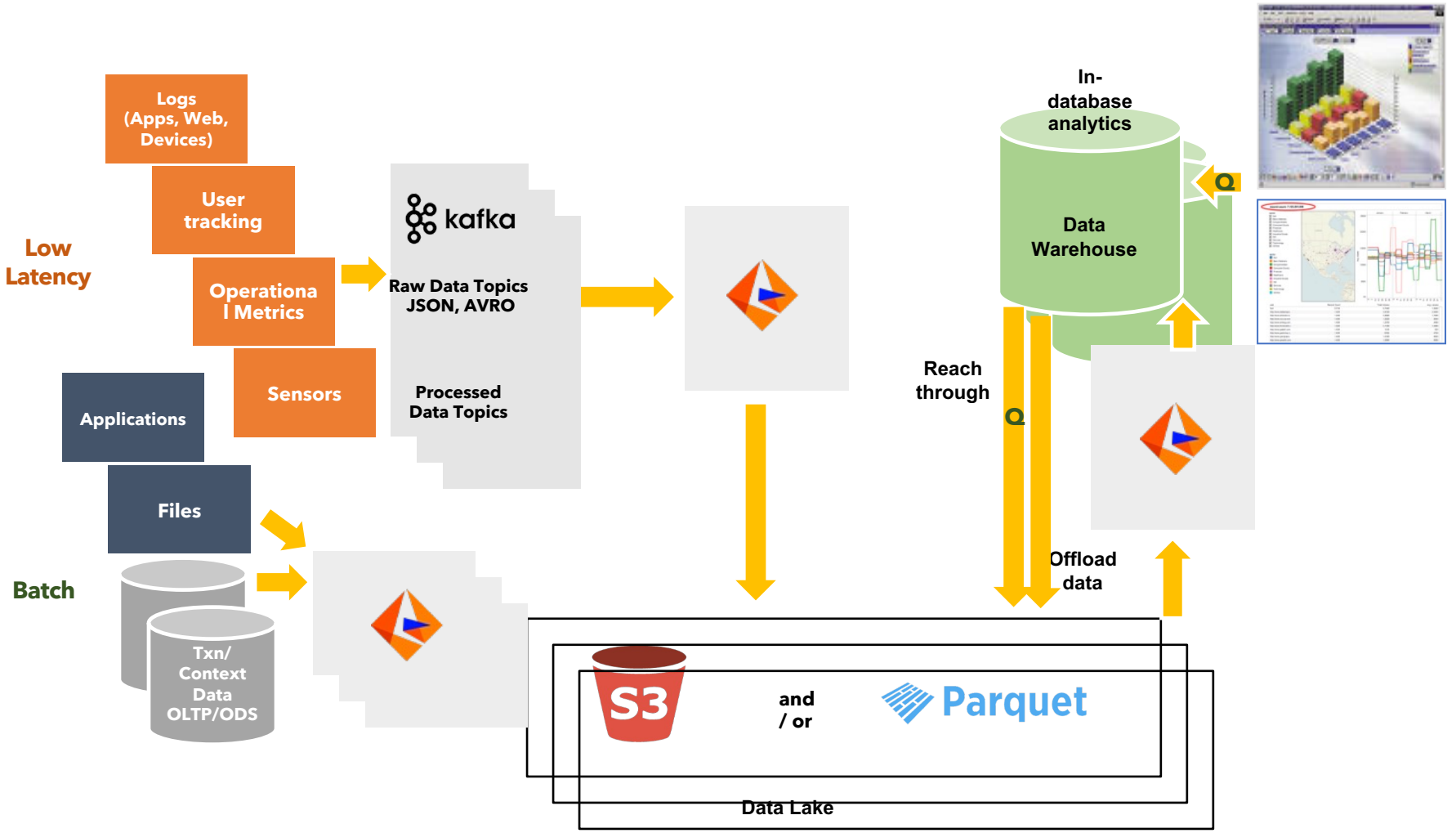
www.mcknightcg.com
(214) 514-1444

https://tinyurl.com/McKnightDI

The 10 critical capabilities are:
Native Connectivity and Multi-Latency Data Ingestion
Data Transformation
Data Security and Access Control
Data Quality and Data Governance
Workflow Orchestration
Analytics, Automation and AI
Data Cataloging and Metadata Management
Enterprise Scaling with Performance
Ecosystem and Platform Versatility
Financial Operations, Compliance and Data Auditing

In alphabetical order, the 10 vendors chosen were:			
1	Airbyte	6	Matillion
2	AWS: Glue	7	Microsoft: Azure Data Factory
3	Fivetran	8	Oracle
4	Google: Alooma	9	Qlik: Talend
5	Informatica	10	SAP

Data Mesh


Logs (Apps, Web, Devices)
User tracking
Low Latency
Operational Metrics
Sensors
Applications
Files
Batch
Txn/ Context Data OLTP/ODS
kafka
Raw Data Topics JSON, AVRO
Processed Data Topics
In-database analytics
Data Warehouse
Reach through
Offload data
S3
and / or
Parquet
Data Lake

Data Mesh Principles and Benefits

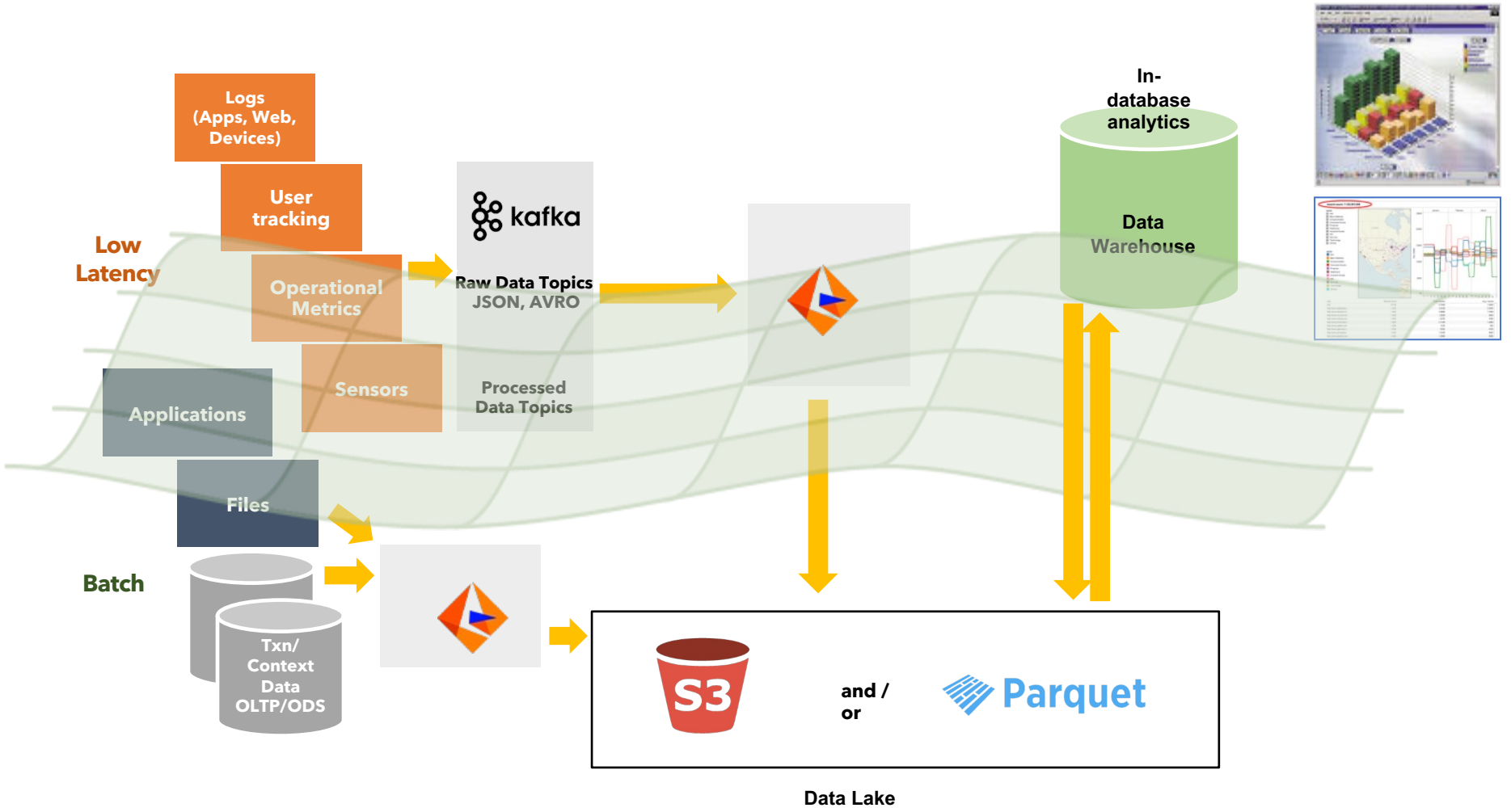
- Domain Ownership
- Data as a Product
- Self-Service Data
- Federated Governance

- Democratization of Data
- Cost-saving Measures
- Reduced Technical Debt
- Collaboration
- Safety and Adherence

What is a Data Fabric?

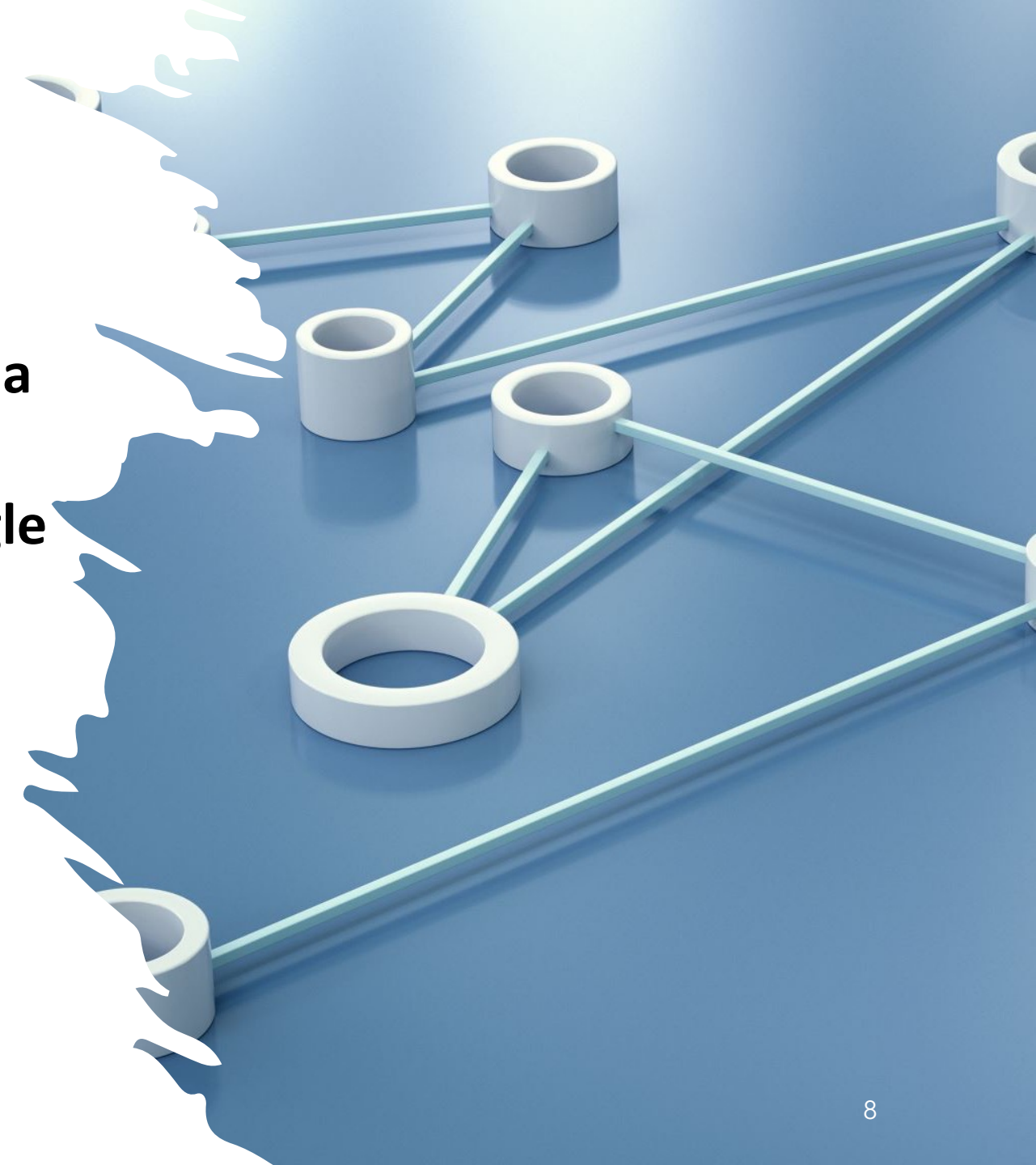
MCKNIGHT CONSULTING GROUP

Data Fabric


Logs (Apps, Web, Devices)
User tracking
Low Latency
Operational Metrics
Sensors
Applications
Files
Batch
Txn/ Context Data OLTP/ODS
kafka
Raw Data Topics JSON, AVRO
Processed Data Topics
In-database analytics
Data Warehouse
S3
and / or
Parquet
Data Lake

Relationship to Data Cloud

- Data fabric can operate on top of a data cloud
- Data fabric is not limited to a single data cloud



Factors Driving the Need for a Data Fabric

- **Big Data:**
  - Processing data at high scale, low latency, and in various formats
- **Cloud Computing:**
  - Data storage, access, and processing outside the enterprise (multi-cloud, hybrid cloud)
- **Edge Computing:**
  - Computing and acting on data from gateway and edge devices

Data Fabric Principles

- Intelligent and Automated
- Unification of Disparate Data Systems
- Access to Integrated Enterprise Data
- Scale Efficiently
- Multi-Cloud Awareness

Benefits of a Data Fabric

- Integrated Intelligence
- Data Democratization
- Improved Data Security
- Universal access to all data sources
- Standardized data format
- Reduced complexity for end users
- Improved data integration and standardization
- Enables access to data from various sources

Key Features of a Data Fabric

- Connects all data sources and processing
- Unified layer with single interface
- Accessible regardless of location (hybrid, multi-cloud, on-premises)
- Standardizes data format (schema, models, nomenclature, file types)

Components of a Data Fabric

- Data Fabric is composed of several key components that work together to provide a unified and integrated data management solution. These components include:
  - Data Integration
  - Data Governance: Ensures the security, compliance, and quality of data by implementing policies, procedures, and controls.
  - Data Catalog: Provides a centralized repository for metadata, allowing users to easily discover, understand, and access data assets.
  - Data Virtualization: Enables real-time access to data regardless of its physical location, eliminating the need for data duplication and improving data agility.
  - Data Analytics: Empowers organizations to derive insights from data through advanced analytics techniques, such as data mining, machine learning, and predictive modeling.
  - Data Security: Protects data from unauthorized access, ensuring data privacy and compliance with regulatory requirements.

Use Cases for Data Fabric

- Fraud Detection
- Preventative Maintenance
- Data Discovery
- Customer Profiling
- Risk Modeling

How Data Fabric Benefits AI Applications

- **Unified data access:** Provides a single point of entry for all AI applications to access data.
- **Improved data quality:** Ensures consistency and accuracy of data used by AI models.
- **Simplified data governance:** Streamlines data security, access control, and lineage tracking.
- **Enhanced data preparation:** Enables efficient feature engineering and data transformation for AI workflows.

AI Use Cases:

- **Machine learning model training:** Faster access to high-quality training data improves model performance.
- **Real-time AI applications:** Data fabric facilitates data flow for real-time AI decision making.
- **Automated analytics:** Seamless data access empowers AI-powered analysis and insights generation.

Data Fabric in Action

Integrates data from various sources (e.g., Snowflake, Kafka, Salesforce)

Enables data reuse for different use cases (e.g., identifying investment opportunities, microsegmentation trends, churn detection)

Reduces time and cost for data re-engineering

Solving Data Distribution Challenges

- Addresses scattered data landscape
- Access data from everywhere uniformly
- In-house data experience for all sources
- Easy-to-understand, business-friendly format

Data Fabric Implementation Types

Physical Data Fabric:

- Data is moved to a common ecosystem (e.g., hyperscaler cloud)

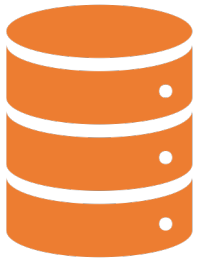
Logical Data Fabric:

- Leverages data virtualization, materialization, and query federation
- Data remains at source or moves only at last moment

Implementation Considerations

- **Time investment:**
  - Effort for establishing semantic models, mapping data, and fulfilling technical requirements
- **Value realization:**
  - Accelerated deployments for use cases not requiring extensive upfront data management

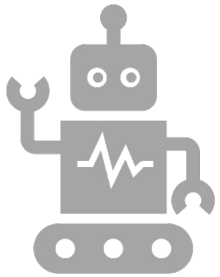
Data Fabric with AI and Metadata



Automates Data Integrations:

Artificial intelligence

Metadata



Semantic Inferencing and Machine Learning:

Scalable automation for data fabric deployments

Key Considerations for Choosing a Provider

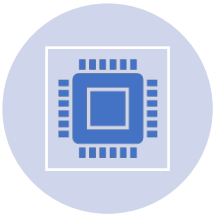
MCKNIGHT CONSULTING GROUP

The DIY Approach to Data Fabric

- The do-it-yourself approach to a data fabric involves leveraging tools and platforms to build a customized solution that meets specific data integration and management needs
- Organizations can easily adapt and incorporate new technologies or features into their data fabric as they become available
- You sculpt it to your unique needs, scaling it with your ambitions
- This self-assembly route demands broad expertise
- Integrating diverse tools and weaving in new technologies comes at a cost to keep it from becoming a patchwork architecture

Provider Approaches > DIY Approach

We compared the tasks of creating, integrating, distributing, and managing a data fabric with a common DIY set to a leading fabric provider.



Looking across all organization sizes, we found that the cost of constructing a data fabric infrastructure with a leading fabric provider was about one-third that of developing a fabric on your own, providing a savings of 69%.

Our analysis revealed three additional, significant cost reductions with a leading fabric provider compared to a DIY approach.

We found that a leading fabric provider costs were just 37% that of DIY for initial build costs (a 63% savings) and 47% that of DIY for CI/CD costs (53% savings).

Administration costs with a leading fabric provider were 59% that of DIY, a savings of 41%

In our analysis, we found the cost of a leading fabric provider deployment to be 42% that of the DIY technique for creating, integrating, distributing, and maintaining a data fabric—a 58% savings.

Ease of Use Differences Versus DIY

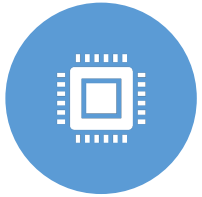
- Large:
  - System Administration
  - Data Modeling
  - Data Security
- Medium:
  - Data Acquisition
  - Data Integration
  - Data Consumption
- Small:
  - Data Preparation
- Negligible:
  - Data Catalog

Choosing a Provider

A product can weave a unified path across all your information, whether nestled in applications, clouds, or on-premises servers

It allows you to see the wider picture, identify hidden trends, and make more informed judgments thanks to a networked environment and a corporate data fabric

A product can dissolve barriers, opens data governance, and maintains security



Integrating these capabilities under one roof is beneficial because it saves time and effort and also ensures seamless collaboration between different teams working on data-related tasks

A product can facilitate this, performing normally distinct tasks like defining, modeling, and integrating data to enable governance, modeling, and data engineering all at once

With a product, organizations can streamline data management processes and achieve greater efficiency in handling complex data projects, as we saw in our test

Complete connectivity to data sources

- **Universal data source connectivity**
- **Abstraction of technical complexities**
- **Equal treatment of diverse datasets**
- **Focus on unification and simplification**

Centralized Access

- Universal data access
- Centralized data understanding
- Actionable insights
- Single point of access
- Just-in-time access
- "Single pane of glass" view

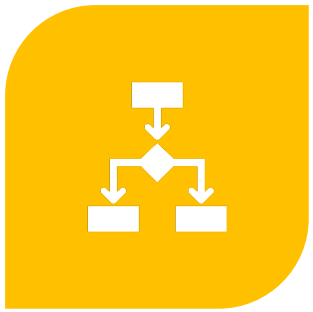
Preparation of Data

- Predefined transformations
- Versioning mechanisms for data tracking
- Automatic schema drift detection and resolution
- Workflow automation for repetitive tasks (e.g., preparing Salesforce data)

Integration of Data

STANDARDIZATION IS KEY

UNIFIED DATA FOR DIVERSE APPLICATIONS

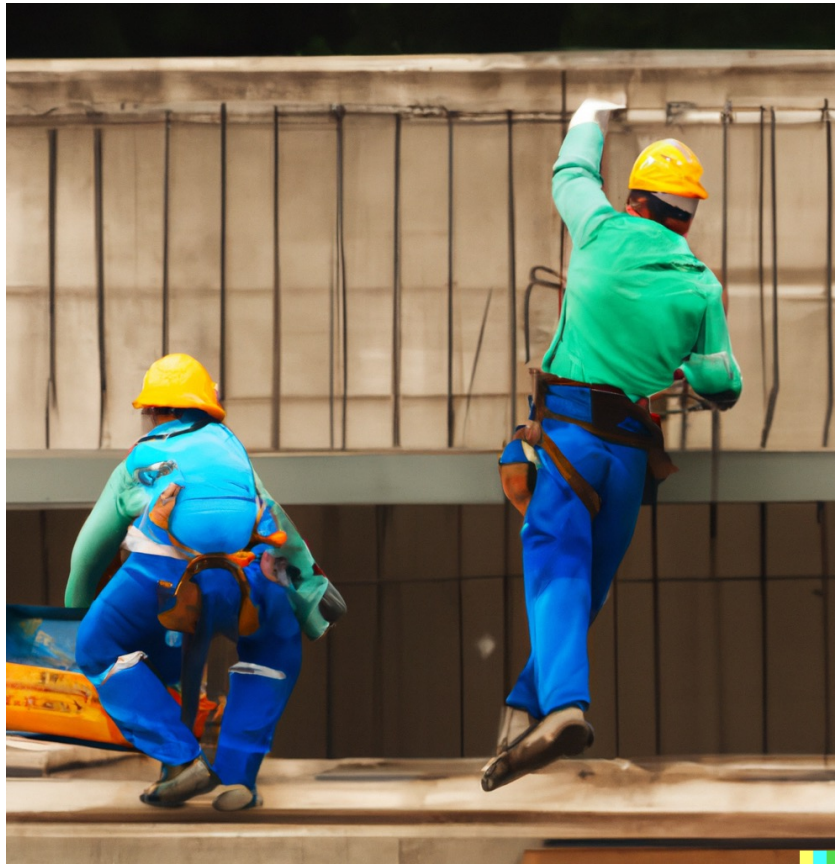


MULTIPLE INTEGRATION METHODS

AUTOMATION WITH METADATA

Semantic Models

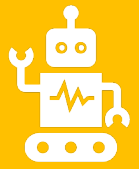
- A semantic model helps business-users understand the meaning of data and its metadata for achieving business goals.
- Data fabrics require a semantic layer to make sense of data in nontechnical terms, making it meaningful for business purposes.
- Methods for facilitating these models include taxonomies or glossaries of business concepts, which can be stored in knowledge graphs or data catalogs.
- The semantic layer maps these descriptions to business concepts, making it useful for downstream purposes like data integration.
- Data fabric semantic models differ from dedicated semantic layer/metrics store platforms, which function as translation layers between storage and BI tools.



Data Management

Data fabrics must support data governance, including security and access control, for sustainable data reuse.

Vendors should integrate data profiling, quality, modeling, master data management, lifecycle management, and other traditional aspects of data governance.



User behavior monitoring, access control methods, and detailed data provenance.

Data Security

- **Data fabrics implement various security measures to protect sensitive information:**
  - **Access control:** Granular access controls restrict who can access and modify data based on predefined roles and permissions.
  - **Data encryption:** Data is encrypted at rest and in transit, protecting it from unauthorized access even if intercepted.
  - **Data masking and anonymization:** Sensitive data can be masked or anonymized, reducing the risk of exposure while still enabling data analysis.
  - **Auditing and logging:** All data access and modifications are logged for audit purposes and potential forensic investigation.
- **Data fabrics leverage existing security tools and infrastructure for a layered approach:**
  - Integration with identity and access management (IAM) systems.
  - Compliance with relevant security regulations and standards (e.g., GDPR, HIPAA).

Market for Data Fabric Providers

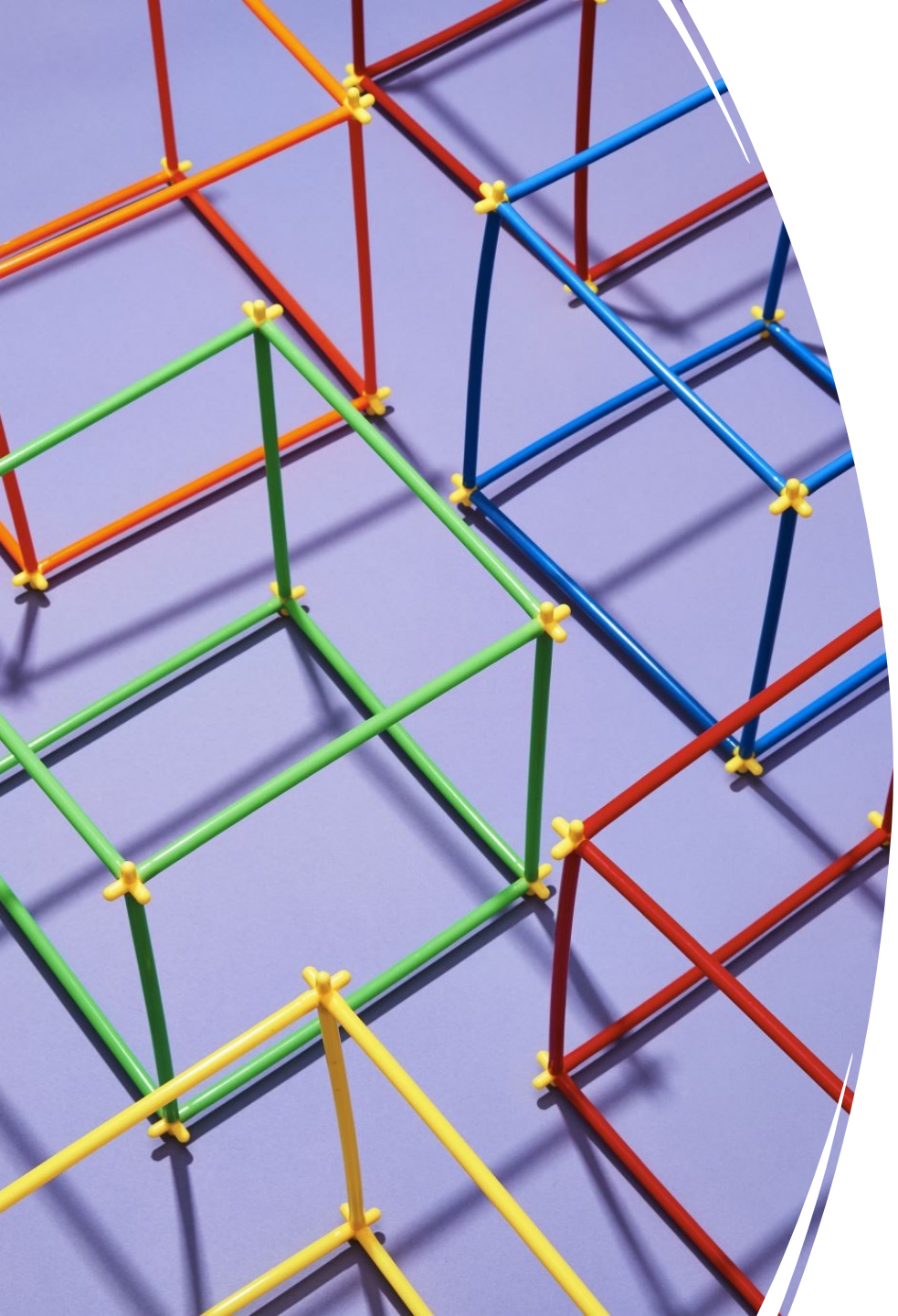
- Data Virtualization/Data Movement/Data Federation
- Low-Code/No-Code
- 10s-100s of connectors
- Data Harmonization
- Knowledge Graphs
- Data Catalog
- Data Profiling
- Data Marketplace
- Natural Language Query

Evaluating Potential Providers

MCKNIGHT CONSULTING GROUP

Pricing

- Data Warehouse Capacity Units
- Data Lake Compute Terabytes
- Data Integration jobs per hour per month
- Optional Data Catalog
- Catalog Crawling Hours per Month



The Future of Data Fabric

- **Integration with artificial intelligence (AI) and machine learning (ML):** Data fabric will play a crucial role in providing data for AI/ML models, enabling real-time decision-making and automated workflows.
- **Decentralized data management:** Data mesh architectures will be integrated with data fabric to create a hybrid approach, enabling better data ownership and governance at the business unit level.
- **Focus on interoperability and open standards:** Increased adoption of standardized APIs and protocols will facilitate seamless data exchange and collaboration across diverse platforms and tools.
- **Enhanced security and privacy features:** Data fabric will continue to evolve with robust security measures to address evolving threats and ensure responsible data governance.
- **Self-service analytics and citizen data science:** Data fabric will empower business users with simplified data access and visualization tools, fostering data-driven decision-making across the organization.

Best Practices for Data Fabric

- Define clear goals and objectives
- Assess your data landscape
- Consider user needs and adoption
- Standardize data formats and metadata
- Implement a robust data governance framework
- Start small and iterate
- Focus on data quality and cleansing
- Prioritize security and compliance
- Embrace a collaborative approach

Summary

- A data fabric is a unified and integrated architecture that provides a consistent and scalable approach to managing and accessing data across multiple sources and formats.
- Benefits include Integrated Intelligence, Data Democratization, Improved Data Security, Universal access to all data sources, Standardized data format, Reduced complexity for end users, Improved data integration and standardization and access to data from various sources.
- It includes Data Integration, Data Governance, Data Catalog, Data Virtualization, Data Analytics and Data Security.
- Purchase considerations are complete connectivity to data sources, centralized entry, preparation of data, integration of data, support for semantic models, data management and data security.
- In our analysis, we found the cost of a leading fabric provider deployment to be 42% that of the DIY technique for creating, integrating, distributing, and maintaining a data fabric—a 58% savings.

MCKNIGHT CONSULTING GROUP

Choosing Your Provider for Implementing a Data Fabric

Presented by: William McKnight

"#1 Global Influencer in Big Data" Thinkers360

President, McKnight Consulting Group

3 X Inc 5000

/in/wmcknight

www.mcknightcg.com
(214) 514-1444