# Why Data Unification and Fabrics?

For Organizations to Become Truly Data Driven, They Have to Commit to a Unifying Data Integration Strategy

Introduction

Typically, organizations treat each system integration as a separate application. Developers have done their best to find or build the right technology for the job but over time, without an overall data integration strategy, an enterprise's integration topology can begin to resemble the following:
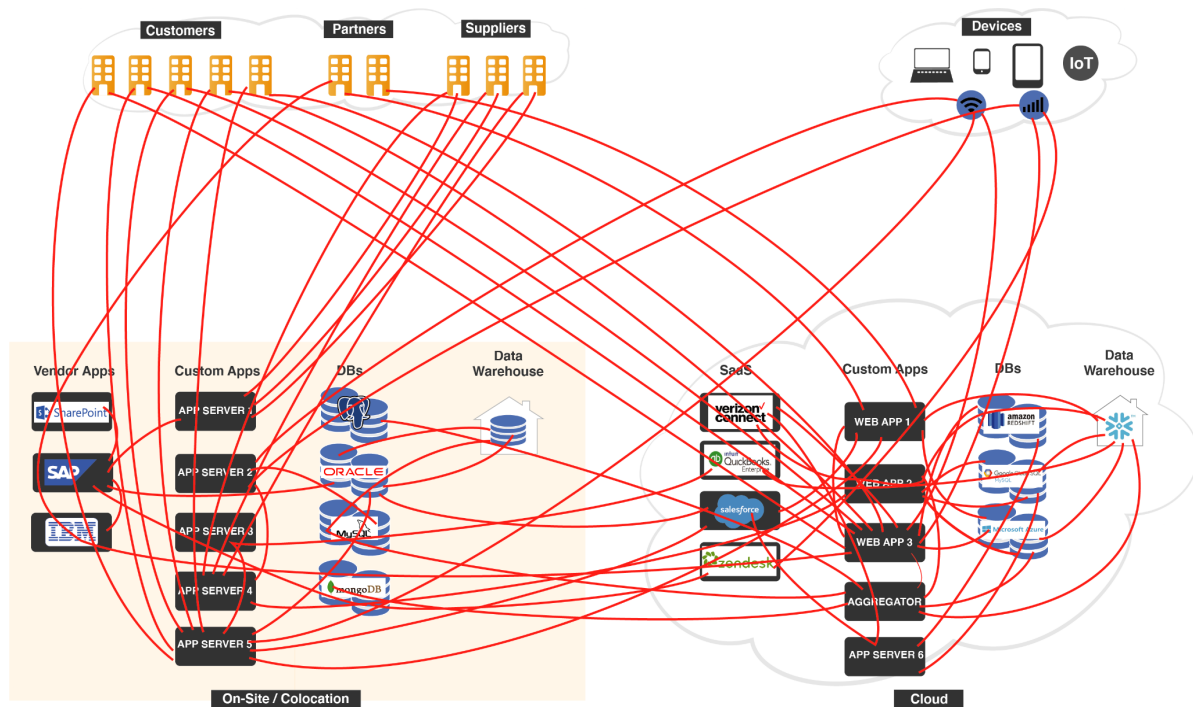
An [IDG poll](#) of more than 200 companies with at least 1,000 employees saw companies drawing from over 400 different data sources.[1]

---

[1] King, Timothy. "Companies Are Drawing from over 400 Different Data Sources on Average." Solutions Review, 1 Nov. 2019, https://solutionsreview.com/data-integration/companies-are-drawing-from-over-400-different-data-sources-on-average/

To stay competitive and relevant, many organizations recognize that they must be more data driven but much of their data is inaccessible because it is owned by different business units and siloed across the enterprise. Because of this, a large share of enterprise data remains unleveraged (42% according to an [IDC study](#)[2]).

Most senior IT professionals have taken on data unifying initiatives to see them fall by the wayside or cancelled in favor of more urgent business objectives. Common reasons are:

- Senior management didn't prioritize the initiative and anticipate the requisite resources to complete and maintain it - this includes not giving the proper incentives to developers, and IT staff to get onboard with the initiative
- Locking into proprietary and over-priced proprietary data integration tools and platforms that did not deliver on their promises

Beyond Analytics

Not only do decision makers and data scientists need frictionless access across the data tier for better analytics, but they need it to better power:

- Machine Learning (ML)
- AI for IoT and other machine intelligence initiatives
- Search - Users want to search enterprise data as they search the web with Google

Enter the Data Fabric, Data Unification and Knowledge Graphs

Gartner describes a [data fabric architecture](#) as:

> "... a robust solution to ever-present data management challenges, such as the high-cost and low-value data integration cycles, frequent maintenance of earlier integrations, the rising demand for real-time and event-driven data sharing and more[3]."

---

[2] Condon, Stephanie. "Enterprises Are Collecting More Data, but Do They Know What to Do with It?" ZDNet, ZDNet, 15 July 2020, www.zdnet.com/article/enterprises-are-collecting-more-data-but-do-they-know-what-to-do-with-it/.

[3] Gupta, Ashutosh. "Data Fabric Architecture Is Key to Modernizing Data Management and Integration." Smarter With Gartner, Gartner, 11 May 2021, www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration/.

Central to data fabrics is an event-driven federated data unification platform coupled with graph technology. Event-driven, federated data unification manages the integration, governance and access of data across the data tier while being sensitive to the needs of business units to manage their own data. Graph databases are designed to treat the relationships between data as equally important to the data itself. They are intended to hold data without constricting it to a predefined model. Instead, the data is stored as it drawn out - showing how each individual entity connects with or is related to others[4]. This semantic data is used to build knowledge graphs that help build search engines (like Google's) that can leverage structured data across an enterprise data tier about topics vis-avis to the organization.

By placing data graph technology atop an event-driven, federated data unification platform, organizations create a data fabric architecture that enables data-driven initiatives while accommodating the realities of their decentralized computing ecosystems.

Data Integration vs Data Unification

Traditional data integrations typically work in batches (ETL, most MDM) and often have data quality and business process flows attached to them. The data integration level across the data tier is quite low when compared to data unification.

Data unification provides deeper integration across the data tier when compared to traditional integration methods and gives additional value to your data that allows your organization to be more data driven.

Data Unification Attributes

Event-driven, Federated Data Routing

Federated data unification is the event-driven interface between an organization's disjointed systems. Adaptors detect changes in the source systems (data events) and publish them to the Federated Unification Server to perform the task of distributing (routing) the data between the subscribing systems in the data tier.

---

[4] "DBMS A Brief Look into Database Models, Data Warehousing, Mining, Trends and an Overview on Data Flow Diagram", Course Hero, Acharya Institute of Technology

Requests can be made to retrieve data records from various source systems and assemble them into one federated data record for the user's unique requirements based on which systems they consider the best source of truth while considering data governance rules that need to be applied to the request.

Governance

Governance is the process of managing data access to and between source systems in the data tier (i.e., who accesses certain data sets based on their role, application, etc.). More specifically, governance is used to describe if a consumer (API consumer or other source system) has access to data. Governance includes what data movement between applications is allowed. Governance must provide granular access control for a given data record by allowing/disallowing specific data record properties from:

- Leaving one system for another
- Responding to federated data access requests through APIs

Enterprise-Wide Data Access

Developers and data wranglers need to access the entire data tier through a single interface to drive their application, ML, AI, analytics and data graph initiatives using REST API endpoints using modern interfaces such as Lucene and GraphQL.

Data analysts need enterprise-wide access through a single JDBC connector to power tools such as business intelligence (BI) .

Lineage

Data lineage (tracing) includes a data record's origin, what happens to it and where it moves through the data tier over time. Data lineage gives visibility for data compliance while greatly simplifying the ability to trace errors back to the root cause in a data analytics process.

Deduplication

De-Duplication refers to the responsibility of a federated data unification platform to identify and then either report and/or link duplicate data records across the enterprise. As something

of a centralized clearing house for shared data across the organization, the federated system is in a unique position to help clean up duplicate data and report cases where ambiguous data must be resolved manually by the appropriate data stewards..

Data Quality

Prior to mapping data records into the federated unification platform, data quality tools should be employed to cleanse, enhance and enrich the data. Most organizations have invested heavily in data cleansing with their existing data integrations so it is imperative that a federated data unification platform can integrate with other custom and vendor data quality tools so enterprises don't lose their existing data quality investments.

Contrasting Integration vs Unification

ENTERPRISE WIDE FEATURES	DATA INTEGRATION	FEDERATED DATA UNIFICATION
Event-driven, Federated Data Routing	No	Yes
Governance	No	Yes
Data Access	No	Yes
Data Lineage	No	Yes
Deduplication	No	Yes
Data Quality	Possibly	If desired

Giving CIOs What They Want

To put it in a different perspective, CIOs want single pane views across the entire multi-cloud data tier.
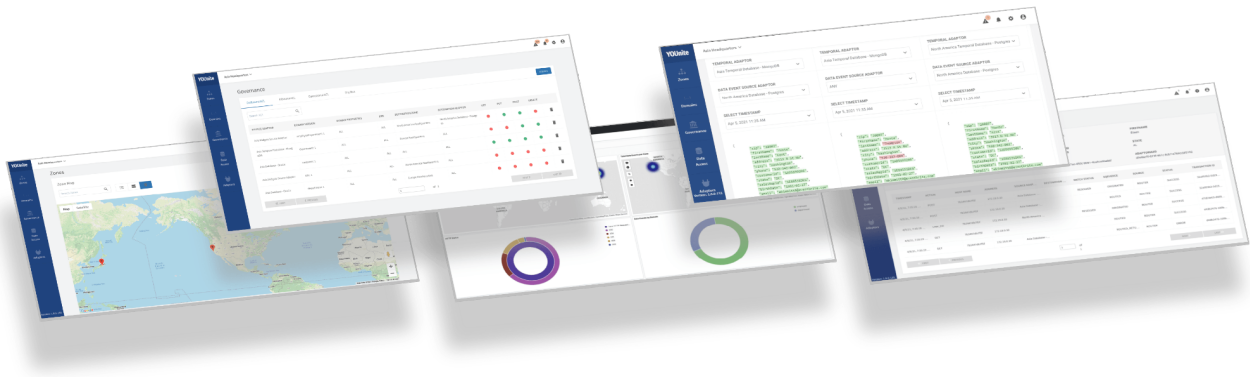
"A key solution to these data storage management challenges has to do with how business owners see the stored data," the report says. "The idea is to see it—all of it—as if through a single pane of glass. It goes beyond data democratization and into storage unification. CIOs should be able to look across the multiple cloud ecosystems in a seamless manner."[5]

Comparing Costs between Data Unification and Data Integration

One of the remarkable outcomes of migrating to data unification is that the overall connectivity across the enterprise data tier increases while integration costs are reduced

The most significant and important design pattern for architecting event-driven data integrations is the use of a hub. Traditional integration designs use point-to-point, or some hybrid ad-hoc approach.   For most enterprises, dozens of integrations are in place with many more planned over the next few years. Unless the data tier is unified, the organization's ability to stay relevant and competitive will only become more difficult.

Take an example of an enterprise data tier that plans to integrate twelve systems. Ignoring the different types of data. The CIO wants every system to have the capability to integrate with

[5] Condon, Stephanie. "Enterprises Are Collecting More Data, but Do They Know What to Do with It?" ZDNet, ZDNet, 15 July 2020, www.zdnet.com/article/enterprises-are-collecting-more-data-but-do-they-know-what-to-do-with-it/.

each other, but to be conservative, we will reduce the requirement so that each system will only need to connect with half of the other systems (6)[6].

---

[6] Reeve, April. "Canonical Modeling - Best Practice Message Modeling for Data Integegration", DATAVERSITY, EMC Consulting, April 11, 2013, https://youtu.be/mRthhbAtJ7o

Traditional Data Integration
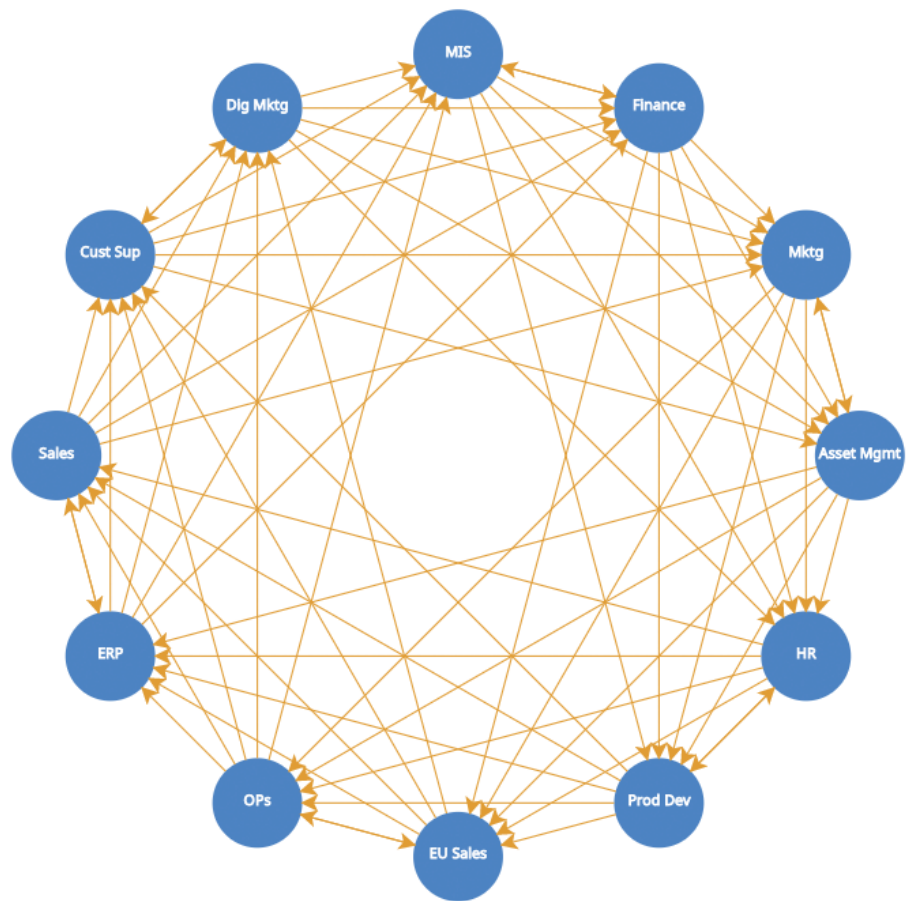
With traditional data integrations, the number of integrations soar -- while only providing a subset of features that data unification delivers

Number of systems: 12

Number of integrations required for <u>half coverage</u>: 66

Data Unification

With Data Unification, all systems are integrated with each other with all the federated enterprise wide features covered above

Number of systems: 12

Number of integrations required for full coverage: 24

Cost Comparison

Assuming that each integration costs $12,000 to implement/deploy and $8,000 to maintain per year. The total integration cost estimates for our 12 system integration over 5 years:

	TRADITIONAL	DATA UNIFICATION	SAVINGS
Cost to Implement and Deploy	$792,000	$288,000	**$504,000**
Annual Costs	$528,000	$192,000	**$336,000**
Total Costs Over 5 Years	$3,432,000	$1,248,000	**$2,184,000**

Open vs Locking into a Vendor

The Freedom of Choosing Best-of-Breed Technology

Historically, if customers wanted to integrate their entire enterprise they had to lock into an overly complex, excessively featured platform from a single proprietary vendor, which effectively locked the customer out of leveraging the latest technologies.

Data unification is tightly coupled with many related technologies such as Graph DBs, semantic modeling, data cleaning, data quality, business process management, and data integration — and — these technologies are constantly evolving and that no single vendor can provide the best overall solution for the customer.

Many vendors take a one-size-fits all approach that locks customers into a proprietary platform, but a federated data unification platform should provide core data integration, governance and, data access capabilities while giving the customer the freedom of choosing technologies that want to integrate into their data fabric architecture.

Conclusion

For organizations to stay relevant and competitive, they must provide decision makers tools that leverage all of the data across the data tier. With the amount of data collected and the number of data silos increasing in the enterprise, organizations must recognize that traditional data integration strategies will be unable to meet these demands and that alternative data unification initiatives must be prioritized.