HITACHI
Inspire the Next

The edge-to-multicloud data fabric puts all of the needed capabilities for acquiring, managing, enhancing, governing, and securing data in one layer that both enhances data and optimizes performance as guided by policy directives.

WHITE PAPER

Edge-to-Cloud Data Fabric

How a Comprehensive Data Fabric from Edge to Core to Cloud Will Power the Next Generation of Applications

By Hitachi Vantara

September 2020

Contents

How a Comprehensive Data Fabric from Edge to Core to Cloud Will Power the Next Generation of Applications

Introduction

Frequently, technology trends in consumer industries offer previews of innovation in the business world. New advancements around data fabrics are a case in point.

Broadly speaking a data fabric goes beyond the traditional definition of a database, which generally allows data to be stored and retrieved from a specific repository in a fixed location. In a data fabric, the user of the data really doesn't know or care where the data is stored. The data just shows up when needed. It may be moved around or it may not be. The system takes care of any such details. In more advanced data fabrics, the data is enhanced and performance is optimized to maximize goals according to policies.

Consumer products abstract data for their users to provide a unified and consistent user experience. Think of Netflix or Spotify or Gmail: users don't know how much of what they do occurs in the cloud versus what is being done on the device itself. The service works as desired, providing access to media and data at any given time.

The abstraction in these consumer-facing products may be implemented in different ways, but it conveys the essence of what data fabrics do in their simplest form. They eliminate the differences between the way data is collected, used, and analyzed at the edge, in the core datacenter, and in the cloud. Data fabrics allow applications to function well everywhere, making use of data wherever the users may be. For this type of seamless accessibility of data to occur, data has to be well managed and governed in an abstract fashion.

At Hitachi Vantara we have developed an innovative vision for a comprehensive data fabric that meets the needs of highly complex enterprise data landscapes. This type of advanced data fabric not only provisions data and related computing resources where it makes most sense (at the edge, in core data centers, and across multiple cloud environments) but also enables data to be enhanced, have its performance optimized, and have policies automatically applied for governance and security.

This paper examines how edge-to-multicloud data fabrics will change businesses going forward by supporting more powerful business solutions, conquering complexity using AI and automation and enabling industrial applications to grow in both power and scope.

Note: This is the third paper in this series. The first two papers provide additional context on the vision for the edge-to-multicloud data fabric and the technology that is used to bring it to life. The edge-to-multicloud data fabric includes systems from the edge to core data centers to multiple cloud environments.

More Powerful Business Solutions

The impact of the data fabric in business solutions stems from the power of data to create broader, more detailed, and more sophisticated models of business processes and activity and to put those models to work to make predictions and prescribe optimal next steps.

If architects and developers of business solutions have to re-invent and re-implement capabilities to manage, govern, and secure data, if data and computing resources are permanently stuck in one location, organizations will drown in a sea of complexity. Creating, maintaining, and integrating solutions will cost far too much.

The edge-to-multicloud data fabric puts all of the needed capabilities for acquiring, managing, enhancing, governing, and securing data in one layer that both enhances data and optimizes performance as guided by policy directives.

Here are several ways the edge-to-multicloud data fabric will help business solutions become more powerful:

- **Automate data collection, integration, governance, and security.** When data arrives into the data fabric, it will be analyzed to determine its fingerprint, cataloged, and associated with existing models used for business solutions. Governance policies and security controls will be automatically applied. Based on enhanced metadata, data integration can be done in an automated or assisted manner.
- **Automate data movement and provisioning of computing resources for workloads.** Business solutions can set forth policies about the performance needed for the data versus the importance of reducing costs. The data fabric can then move the data to the most appropriate form of storage at the edge, in the core or in any cloud.
- **Automate construction / refinement of metadata models (digital twins).** New data will be matched with existing models. Construction of new models will be accelerated because the fingerprints and the catalog will allow all relevant data to be found. More data enables the expansion of models of business activity. As the data fabric grows, these models will become more accurate and broader, enabling business solutions and predictive models to become more powerful.
- **Automate business processes based on advanced analytics, ML and AI.** Predictive and prescriptive models can recognize opportunities for automated responses or provide guidance to staff. As the scope of the model of business activity grows, the scope of analytics and automation can also expand to provide larger scale forms of automation.

All of these advancements in the power of applications are based on the edge-to-multicloud data fabric's ability to make an ever larger collection of data useful to the business.

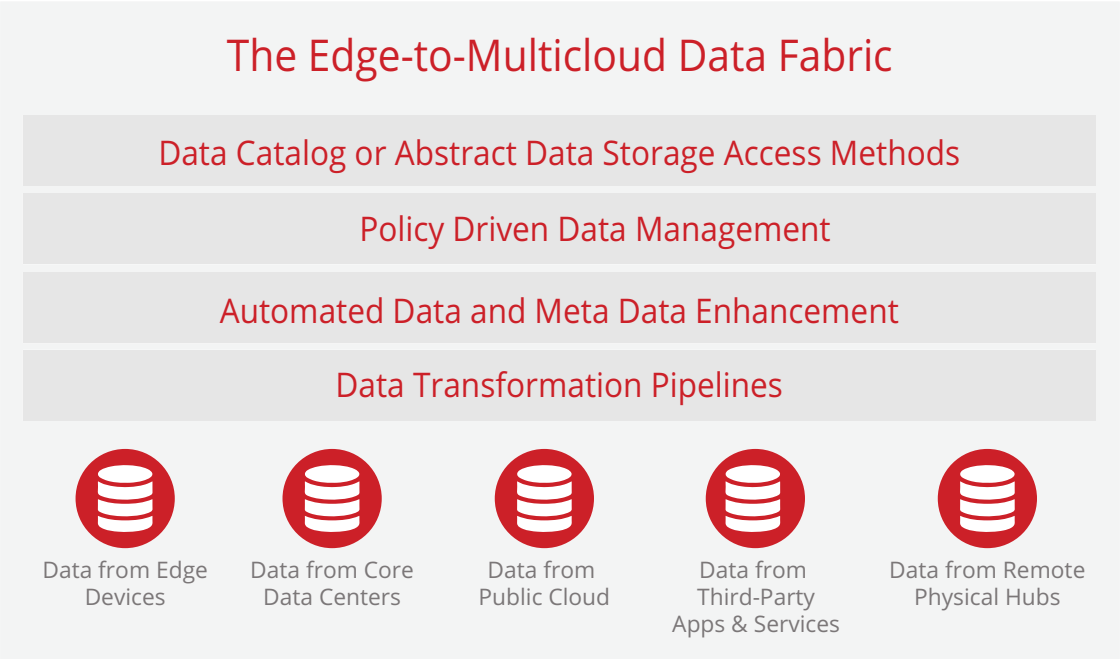
How the Edge-to-Multicloud Data Fabric Conquers Complexity

Perhaps the largest problem the edge-to-multicloud data fabric solves is conquering the complexity of bringing industrial and commercial business solutions to life. To work properly, an edge-to-multicloud data fabric must solve the following problems at scale:

- Accessing and onboarding distributed, heterogeneous data sources coming from OT and commercial systems at the edge, the core, and the cloud.
- Providing data discovery and abstract access for business solutions through a data catalog that has a complete set of both business and technical metadata.
- Providing policy-based capabilities for data management, governance, and security.
- Orchestrating data pipelines at scale that can be operated at the edge, in the core, or in any cloud as dictated by policies.
- Supporting automatic, AI- and ML-driven enhancement of data at scale that adds both metadata and new fields derived from existing data.

With the amount of data now available to companies, and the ever-expanding amounts coming in the future, this type of complexity cannot be addressed manually. It requires vastly more automation than has ever been needed before. To be useful, the data must be curated and extracted at a layer of abstraction that relies on active metadata (possibly provisioned through APIs from data catalogs) that can automatically manage and change the data, and the models based on them. DataOps is the larger architectural concept and methodology underpinning for this type of automation.

The edge-to-multicloud data fabric is able to handle this based on the following architecture:



This architecture employs metadata-driven virtualization to separate access to data from the way the data is stored. Once this is accomplished, the system can expand automated data management and enhancement functions without disrupting workloads. To make this work, metadata must automatically be created and refined and in turn partner with policy engines that conduct data management tasks. With such a system, it is possible to create data pipelines at scale that don't require enormous amounts of care and feeding by highly skilled staff.

The Automation of Analytics and AI

The more companies come to rely on the data fabric, and the more data that is integrated into it, the greater the scope of what this data can be used for. New business solutions and outcomes can be achieved. This process will reduce the burden on staff such as data engineers, data architects, and even data scientists to try to build data models themselves. This will be a gradual development, but a powerful one.

In fact, with the amount of data now available, it's simply not possible for humans to create the type of complex models needed. AI is required to examine and understand many of the underlying patterns in the data. Once it has offered up results, human expertise can be used to refine and embolden the models further. As time goes on, the predictive power of these models will also become broader.

Consider, for example, how this would work in an industrial oil & gas setting. A company may start by optimizing a compressor's maintenance and runtime. Eventually, learnings can be extrapolated to other compressors, from the same models as well as other models, through the use of transfer learning. Then, learnings from multiple compressors can be correlated and applied to oil exploration across different oil fields, helping to optimize overall production. This is the expansive predictive power of models powered by an advanced data fabric.

How Industrial Applications Will Expand in Scope

The integrated, wider array of data available through the data fabric allows for more detailed models of business activity to be created. For instance, in the industrial realm, these models are called digital twins. But these incredibly complex models exist in other industries as well, ranging from financial services to healthcare. Autonomous vehicles perhaps represent the highest level of achievement in this regard where vast amounts of data are integrated to create real-time models that guide the behavior of the vehicles.

Growing Deeper

Ultimately, with the use of edge-to-multicloud data fabrics, the power of business applications will grow because the models will become richer and more detailed. The following framework describes how simpler applications with a more narrow scope gradually expand into larger more powerful applications with much larger scope.

IoT Solution Enhancement Maturity Model

Level		Equipment	Factory	Company	Industry
6	**Symbiotic**	-	-	Optimize supply chain between factories	Crowd manufacturing
5	**Prescriptive**	Optimize control to avoid anomaly status	Optimize production plan / methodology	Optimize production plan / methodology between factories	Optimize production plan in value chain
4	**Predictive**	Predict anomaly status	Predict delay of shipment / total quality	Predict delay / quality of all factories	Predict risks in whole value chain
3	**Analysis**	Detect anomaly status	Detect bottleneck of production / risk of quality	Find issues of all factories	Find issues in whole value chain
2	**Integration**	Monitor all equipment	Monitor whole line (man, machine, material)	Integrate KPI of all factories	Visualize whole value chain
1	**Visualization**	Monitor each equipment	Monitor each production line	Monitor KPI of each factory	Visualize each status

Fed by an advanced data fabric, applications can progress in power from the lowest level, visualization, to the highest level, symbiotic. Growth begins with greater visibility into what is going on, moving from individual equipment to a fleet of equipment. At the next level, beyond just monitoring, analytics start to detect anomalies. At the next level, AI is used to create predictive models. These models will be built into analytics platforms that are ubiquitous across the cloud and available anywhere the company needs applications to be run.

From that point, predictive models can be deployed and become increasingly prescriptive (for example, directly impacting the physical world through controllers). Finally, it is possible to harness these models as the foundation for enhanced automation, a symbiotic process where applications interact and learn from each other.

Growing Wider

As the data models in the edge-to-multicloud data fabric expand, the models of business activity can also expand, making the applications wider so they encompass more business activity. The progression is described in the four columns of the previous diagram.

Equipment

The first form of most applications in this scenario addresses the equipment level, using data and analytics to predict the future state of the equipment. Data and maintenance managers at factories will be able to use data in real-time to assess problems. Quality supervisors will be able to use analytics to see that many products may not be meeting quality standards and stop their production lines before too many resources are wasted.

The data fabric will improve visualization and monitoring of every piece of equipment by creating an integrated history of all relevant data. This data can then be used for advanced analytics, predictive analysis of pitfalls, and finally prescriptive control over equipment that ensures systems are in place for maintenance and upkeep that allow problems to be avoided before they even begin.

Factory

At the factory level, with the data fabric, managers can judge whether necessary tools and assembly processes may be soon unavailable due to maintenance issues, allowing them to reschedule production processes more efficiently.

Entire production lines can be monitored and the data drawn from this can be used to identify potential bottlenecks and factory-wide challenges. Predictive analysis can be used to avoid production and distribution delays, while prescriptive analysis can then be added to better optimize the entire functioning of the factory.

Company

Next, at the company and supply chain levels, businesses will be able to use their new access to quality information to gain insight into their suppliers and customers. Whereas the data analyzed at the equipment and factory level will be more real-time, or within the scope of days or weeks, at the company and supply chain levels, data can be examined at a longer timescale from, stretching to years and beyond.

As more and more data accumulates and models are built with the data fabric, the overall operations and effectiveness of all factories within a business can be evaluated. Common issues can then be identified. Predictive analysis can be used to pinpoint delays and improve the quality of all factories, while prescriptive analysis can be used to optimize and streamline the way all factories work individually and in harmony. The end result is an improvement in the way the company's entire supply chain functions.

Industry

Finally, data fabrics allow companies to expand their insights to the entire supply chain of their industry, predicting risks and identifying opportunities.

The symbiotic layer of this diagram illustrates how data fabrics bring diffuse areas of a business together, from marketing to sales to operations, to improve outcomes by relying on improved company- and industry-wide intelligence.

Because no individuals, or even organizations, can truly understand all of the patterns in the data on their own, AI helps to build and unearth connections that make greater transparency and comprehension possible.

Conclusion

This paper concludes a tour of the edge-to-multicloud data fabric that began with a description of the vision, then showed how technology would bring that vision to life, and then lands at this paper, which explained how the edge-to-multicloud data fabric will transform industries.

From this analysis it is clear that the edge-to-multicloud data fabric describes a vision that must be built one way or another to address the complexities of the modern world and unlock the power of data.

To find out more about how to bring the edge-to-multicloud data fabric to life in your business, please see these sites at HitachiVantara.com:

Data Management and Analytics Solutions

Lumada Software for DataOps

Advisory and Implementation Services

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact