

# Checklist for Successful and Optimized Cloud Data Fabric

## CHECKLIST

Cloud data fabric supports quick, effective data-driven design patterns for capturing a wide range of data types: old and new, structured and unstructured, from internal and external sources, and at large scale. Optimized to ingest and automate processing of raw and varied source data for research, innovation and analytics, a data fabric streamlines business operations.

### Data Fabric Planning

A data fabric reduces the up-front effort of getting data organized and stored because we are not required to structure it first. However, that does not eliminate the need for planning. Various issues must to be considered to prevent your data fabric on cloud from becoming the dreaded data swamp.



#### **Develop Ingestion Plan as per Data Sources**

Consider how the data will enter the fabric, whether sources are predictable or require dynamic solutions, what volumes and frequencies are needed, what the raw data retention requirements are, and whether metadata is included or needs to be ingested.

**Assessing these answers will ensure a smooth data onboarding process.**



#### **Set Up Security and Governance**

Determine how you will allow data access, gain understanding of current data retention and archival policies, ensure that encryption and security requirements have been assessed and met. Then, finalize the plan for continued governance and set steps to plan data lineage.

**Addressing these security and governance points up front prevents issues later.**



#### **Ensure Data Quality and Management**

Apply data prep functions after preparation and before analysis, use a universal semantic layer and organize data for optimal retrieval, and provide validity and quality checks. Enable data virtualization. Provide master data and metadata management, and develop and capture metadata as data is ingested or explored.

**Process and repurpose data on the fly at runtime, plus enable more agile development.**



### **Enrich, Standardize, Cleanse and Curate**

Lightly standardize data to provide “just enough structure” for data exploration and SQL-based analytics. Provide different zones of transformation to address varied business needs, such as; staging curation, which includes data cleansing and deduplication; enriched curation, which includes agile analytics or more valuable datasets from derived data; and publishing, which includes business-use case-specific datasets, storage mechanisms or access channels.

**Address varied business needs with varied types of data transformation.**



### **Get on Schedule, Plan Job Management**

Will your cloud strategy embrace cloud-native, multicloud or cloud agnostic, hybrid-cloud or private-cloud options? What level of cloud platform as a service or serverless adoption will you choose? Will you opt for continuous integration and continuous delivery (CI/CD) and modern automation? Will open source technology be used?

**Determining technology choices at the start informs your data fabric path and progress.**

## Organize To Deploy Your Most Effective Data Fabric

The data fabric should be organized for flexible use of data, optimal data retrieval and its consumption. Metadata capabilities of the data fabric will greatly influence how its organization is handled. Consider these steps to discover how the organization of your data fabric can be influenced:



### **Understand Business Challenge, Ingest Relevant Data**

When teams focus on the business problem to be solved, they automatically get the answer to why data fabric is being built. Establishing clearly defined business goals about data usage and analytics adoption help teams make better decisions on which data sources and tools need to be prioritized.

**Having a clear objective in mind as to why data fabric is required will help teams to remain focused on optimizing the data fabric.**



### **Ensure Correct Metadata for Effective Search**

It is very important to tag every bit of data in a data fabric. This approach ensures better organization and understanding of data and prevents a data fabric from turning into a data swamp.

**Correct tagging helps people search and discover different kinds of data faster, saving time and making data usage easier.**



### **Maintain High Quality With Data Governance**

Data fabric should clearly define the way data should be treated, managed and consumed, and determine its life cycle. Excellent data governance equips the organization to “know their data” and maintain a high level of data quality and consistency.

**It is important to assign governance roles to give designated people access to and responsibility for data, to better govern data and reduce redundancies.**



## Apply Automation for Agile Data Fabric

Automation-enabled data fabric allows quicker ingestion and curation, the ability to compose data and analytics pipelines across multiple platforms, and improved reuse of business data. Automated data fabrics are able to generate insights through ad hoc analytics efficiently, helping the business stay innovative and competitive. This can be achieved with the creation of data pipelines that allow data scientists to dynamically run their queries on a broad set of data.

**Enabling automated retraining and optimization of models leads to better business decisions.**

Cloud data fabric supports quick, effective data-driven design patterns for capturing a wide range of data types: old and new, structured and unstructured, from internal and external sources, and at large scale. Optimized to ingest and automate processing of raw and varied source data for research, innovation and analytics, a data fabric streamlines business operations.

[Click here](#) to learn more about how you can optimize your cloud data fabric and streamline your business operations or to contact a Hitachi Vantara representative.



### We Are Hitachi Vantara

We guide our customers from what's now to what's next by solving their digital challenges. Working alongside each customer, we apply our unmatched industrial and digital capabilities to their data and applications to benefit both business and society.

### Hitachi Vantara



Corporate Headquarters  
2535 Augustine Drive  
Santa Clara, CA 95054 USA  
[hitachivantara.com](http://hitachivantara.com) | [community.hitachivantara.com](http://community.hitachivantara.com)

Contact Information  
USA: 1-800-446-0744  
Global: 1-858-547-4526  
[hitachivantara.com/contact](http://hitachivantara.com/contact)

HITACHI is a trademark or registered trademark of Hitachi, Ltd. All other trademarks, service marks, and company names are properties of their respective owners.

CL-032-A BTD July 2020