Data Lake Governance Center

Data Governance Methodology

Issue 01

Date 2020-11-10



HUAWEI TECHNOLOGIES CO., LTD.

Copyright © Huawei Technologies Co., Ltd. 2020. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: https://www.huawei.com

Email: support@huawei.com

Contents

1 Purpose 1
2 Intended Audience 2
3 Data Governance Framework 3
3.1 Framework 3
3.2 Data Governance Modules 3
3.3 Module Relationships 5
4 Data Governance Organizational Architecture 7
4.1 Framework 7
4.2 Responsibilities 7
5 Measurement and Evaluation System 10
5.1 Data Governance Methods 10
5.2 Measurement Dimensions 11
5.3 Measurement and Scoring Rules 12
6 Huawei Data Governance Cases 14
6.1 Thoughts 14
6.2 Practices 15
6.3 Effects 16
7 Thoughts on Data Governance and the Impact of COVID-19 18
8 Implementation of DAYU Data Governance Methods 19
A Change History 21

 Copyright © Huawei Technologies Co., Ltd.

1 Purpose

Data is a core enterprise asset, and enterprises need to establish data dictionaries to effectively manage their increasingly important data and information resources. They need to establish mechanisms to continuously improve data quality. Effective data management helps enterprises mine data value and reduce operational risks. It simplifies enterprise management, integrates business processes, improves operational efficiency, and makes operations more transparent. Accurate data is the foundation of scientific decision-making. Data architecture and standards need to be unified to ensure efficient operations and clear communications throughout the whole process.

Currently, enterprise data is a mess. There are no unified standards, so key data cannot be identified, sharing between business systems is limited, and data tends to remain siloed. To effectively manage enterprise data, maximize data value, and lay a solid foundation for digital transformation, enterprises are in urgent need of a comprehensive, well developed system of data governance.

DAYU data governance methods have evolved from Huawei's wealth of successful experience with digital transformation. The data governance methods integrate the essence of Huawei's data management methods and draws on industry best practices for data governance. This document aims to describe how enterprises can continuously improve their data management systems, streamline information chains and data flows based on primary business processes, clean data, improve data quality and operational efficiency, make operational results more transparent, drive the growth of intelligent data, and more fully explore data value.

This document first briefly describes the data governance framework, organization architecture, and measurement and evaluation system. Then, it describes the application of data governance based on Huawei's data governance cases and how it has been applied to the recent coronavirus outbreak (COVID-19). Finally, it introduces how the DAYU data governance methods are incorporated into DGC, the data operations platform, with a detailed implementation guide attached.

 Copyright © Huawei Technologies Co., Ltd.

2 Intended Audience

This document is intended for management personnel at all levels in enterprises and institutions, and any personnel involved in data analysis and tool platform development and maintenance.

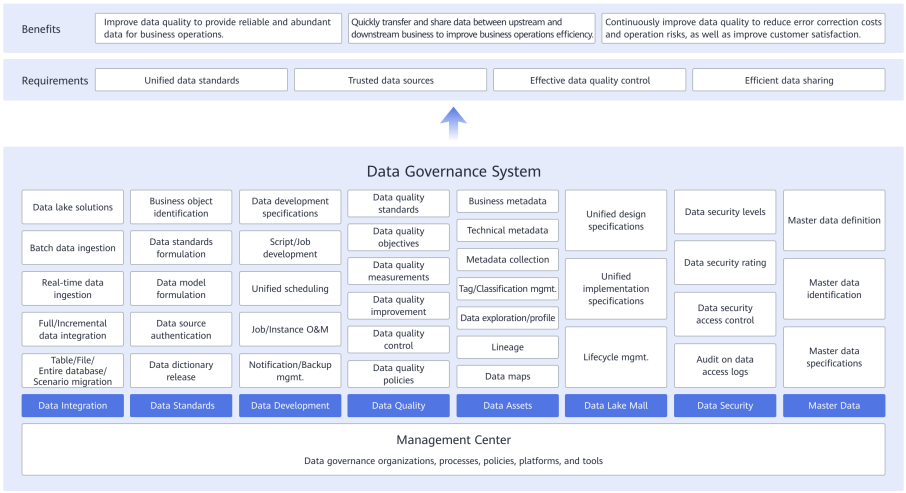
 Copyright © Huawei Technologies Co., Ltd.

3 Data Governance Framework

3.1 Framework

This section introduces the data governance framework. See **Figure 3-1** for details.

Figure 3-1 Data governance framework



3.2 Data Governance Modules

The data governance modules are as follows:

- **Data integration**

  Data integration refers to the importation of data into data lakes. It is not just a data migration. The data needs to be backed up in accordance with a particular method. Before data can be imported into a data lake, there are six items that must be specified about the data: the owner, the release criteria, a security level, the source, an estimation of its quality, and the registration of its metadata. Only after these conditions are met can the data be stored in a data lakes and the data assets registered on the data operations platform.

 Copyright © Huawei Technologies Co., Ltd.

- **Data standards**

  The management of data standards is central to establishing a consistent data language. In keeping with Huawei's method for establishing data standards, data standards come in the form of L1 to L5 data objects. The corresponding publication includes the L1 to L5 standards. The IT systems corresponding to each object must publish corresponding data dictionaries and authenticate the data sources. For objects that are sorted but not incorporated into the IT system, the developers will have to digitize them later.

- **Data development**

  Data development is at the center of orchestration, scheduling, and O&M. It is a one-stop data solution that includes analysis, design, implementation, deployment, and maintenance. It involves the processing and conversion of data to improve data quality. Data development hides the differences between diverse data storage modes and includes the entire process of integration, cleansing and conversion, and data quality monitoring. Data development is the primary field of action for data governance.

- **Data quality**

  The objective of data quality management is to ensure that the data meets requirements for use. Data standards are the basic criteria for data quality. Each business department takes full responsibility for the quality of the data corresponding to their domain. Data quality standards need to be based on business requirements, and quality control objectives need to be established and data quality evaluated based on enterprise data governance requirements. Data quality policies and improvement plans need to meet business requirements, and data quality needs to be continuously managed and controlled.

- **Data assets**

  Data assets include business assets, technical assets, and metrics. Data asset management is an important tool for data governance. The core idea is to build enterprise metadata management centers, establish data asset catalogs and data search engines, visualize data lineages, and create visualized overviews of data assets. Metadata includes business metadata, technical metadata, and operational metadata. All the conceptual data models, logical data models, and physical data models of an enterprise must be systematically managed, and an enterprise data map and data lineage must be established to provide powerful support for invoking data, providing data services, and for O&M.

- **Data lake mall**

  The design and the standards use for data lake mall need to be unified for effective lifecycle management. Intensive management of data services in data lake mall helps reduce the cost of invocation and integration throughout the development process.

- **Data security**

  Data resources used by enterprises include data from both internal and external service systems. Therefore, data security needs to be integrated into data governance. All enterprise data must be assigned a security level. Data access needs to be monitored and controlled whenever data is generated, transmitted, stored, or used. In addition, logs must be generated for creation, retrieval, update, and deletion activities (CRUD) to complete security audit.

 Copyright © Huawei Technologies Co., Ltd.

- **Master data**

  Proper management of master data is critical to establishing data standards and improving data quality. Management of master data is extremely important for effective data governance. The goal of master data management is to ensure that the data definitions of the most important business entities are consistent with the actual physical data. The master data needs to be identified first, so that data governance and IT reconstruction can be carried out based on the specifications of the master data that has been identified. This process streamlines and strengthens business flows and tool chains.

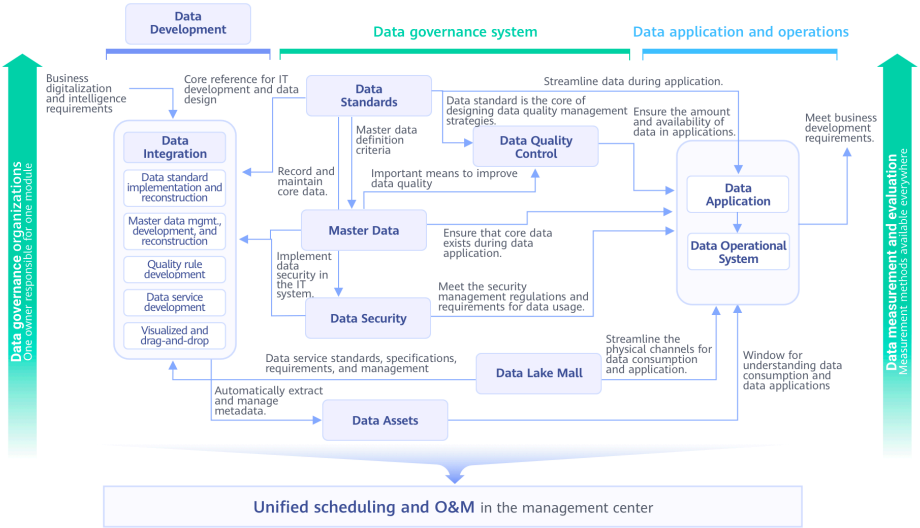
- **Management center**

  The construction of organizations, processes, and policies is an indispensable part of data governance. A management center allows for central management of public and core data sources and cockpits, enabling users assigned different roles to have personalized workspaces.

3.3 Module Relationships

This section introduces the relationships between main modules of data governance. See **Figure 3-2** for details.

Figure 3-2 Module relationships



- **Data standards** provide the core reference for data development and design. They are the result of data development. Data standards help ensure the consistency of data languages. They are the references for defining data, and are an important part of master data management. Data standards serve as a foundation for designing and formulating data quality management policies and rules. Security level classifications and owners specified in data standards facilitate data security management and are also important inputs for the management of data assets.
- **Master data management** is one of the main ways that we can use to improve the quality of our data. In the course of data development, master

 Copyright © Huawei Technologies Co., Ltd.

data can be recorded, updated, and maintained in a unified manner. Master data management ensures the existence and consistency of core data for data applications and operations.

- **Data quality management** is an important way to guarantee data accuracy, consistency, integrity, timeliness, uniqueness, and validity for various data applications and operations. It is also an important prerequisite for enterprises that use data to create value.
- **Data asset management** requires that metadata be collected and registered. This management of data assets makes it possible to use data. Users are able to understand data through the data asset module as well.
- **Data service control** means that standards, specifications, and requirements all need to be managed or controlled in data service development. These controls streamline the physical channels for data applications and data consumption.
- **Data security** means that the IT system needs to keep data secure to meet the compliance requirements of various regulations relevant to data applications during data development.

To effectively carry out data governance, there has to be effective organizational management, specified owners, appraisal systems, process regulations, data governance strategies and platforms.

Figure 3-3 DGC modules



DGC provides the functional modules to meet the data management requirements for ingestion, modeling, design, quality control, and service generation.

 Copyright © Huawei Technologies Co., Ltd.

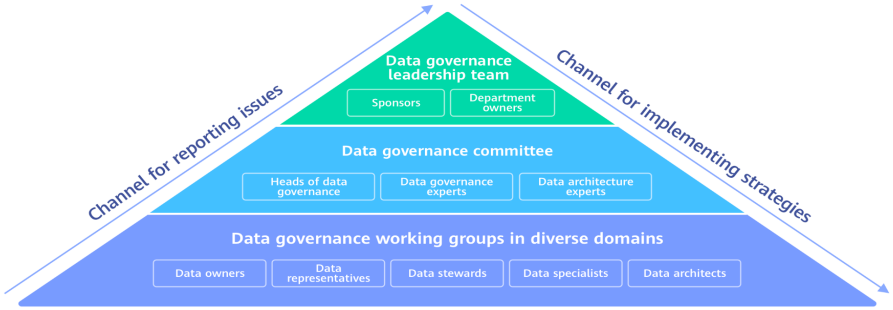
4 Data Governance Organizational Architecture

4.1 Framework

Data governance involves a hybrid organizational structure that includes centralization (full-time input) and virtualization (partial input). This framework involves full-time data governance personnel with professional skills and personnel familiar with businesses and IT systems. Professional business, IT development, and data teams work closely with each other to quickly identify priorities.

The different responsibilities specific to this three-level organization are shown in **Figure 4-1**.

Figure 4-1 Data governance work organization



4.2 Responsibilities

The **data governance leadership team** is comprised of the data governance sponsors and department owners. This team determines the strategic direction of the data governance. The leadership team has ultimate responsibility for how data governance is implemented, and for the promotion and execution of various

 Copyright © Huawei Technologies Co., Ltd.

policies. They have the final say when there are disputes; they monitor the data governance, and manage the data governance budget.

The data governance committee and the data governance working groups of the different domains are the execution teams of the data governance strategies.

- The **data governance committee** consists of those responsible for data governance, data governance experts, and data architecture experts. It arranges data governance work for the enterprises, provides guidance, and conducts regular discussions on data governance within the enterprise, with the goal being more refined and better controlled data. To achieve the long-term goals established by the data governance leadership team, it is necessary to establish and manage data governance processes, formulate time based goals and create plans to achieve them, design data governance methods and general principles, maintain tools and platforms, assist working groups from various domains to conduct data governance, evaluate and report on the overall progress of data governance projects, and resolve cross-domain data governance issues and disputes.
- The **data governance working groups for different domains** handle data governance within a given domain. These working groups formulate data governance objectives and work plans based on the data governance methods and general principles designed by the data governance committee, maintain the data assets of the domain, maintain and update the corresponding data standards and related metadata, design the data measurement rules of the domain, monitor and record data quality issues, continuously improve data quality, and proactively resolve data-related issues, to safely manage data assets of the domain and finally achieve the goal of data governance.

The data governance working group of a given domain is usually composed of the data owner, data representative, data steward, data specialist, and data architect.

- **Data owners** are in charge of data governance at the domain level. A data owner must:
  - Formulate goals and work plans for the data governance in their respective domains and prioritize the execution of various tasks.
  - Create accountability for data governance, dividing up the work in their domain and assigning it to specific group members, follow up, and take responsibility for management of data governance within the group.
  - Design data quality specifications, collect and fulfill requirements on data, resolve data issues and adjudicate disputes.
  - Build and maintain the information architecture of the domain.
  - Cultivate a culture or atmosphere of domain-based data governance.
- **Data representatives** are the leading expert for data governance within a domain. A data representative must:
  - Know the goals of data governance and be familiar with the methods and rules of data governance. They must know how to use data governance tools, identify key business processes and necessary IT systems. Data representatives need to understand how to further refine the data governance objectives within their domain, prioritize the refined objectives, and manage and execute objectives.
  - Manage and resolve issues and disputes as experts within their domain, and submit issues to data owners for final decisions if necessary.

 Copyright © Huawei Technologies Co., Ltd.

  - Ensure the integrity, timeliness, accuracy, consistency, uniqueness, and effectiveness of business data; ensure that every user behavior is recorded, and each record is a piece of data; and report and evaluate data in the domain according to data quality requirements.
  - Build an information architecture for the domain and govern data in accordance with that architecture, watch over the data assets in the domain; and maintain, update, or publish corresponding data standards and data catalogs.
  - Meet upstream and downstream data requirements, and proactively identify these data requirements based on the application scenarios and business requirements of the domain. Data managers need to manage and push forward progress on these requirements.
  - Define the data security level of the domain based on related regulations and manage data access.
- **Data stewards** assist in data governance in their domain. A data steward must:
  - Ensure that the data governance process and content of the domain comply with data governance requirements.
  - Assist the data representatives in tracking and resolving problems.
  - Sort out, maintain, and update metadata and data (including business objects, data standards, and data models) of the domain.
  - Promote the application of data governance tools and platforms in the domain, and maintain the data governance tools and platforms.
- **Data specialists** form an expert team for data governance in a given domain.

  They use professional skills to assist the data representatives to complete diverse data governance tasks assigned by data owners based on the data governance plans for the domain.
- **Data architects** carry out the data governance work at the IT level. A data architect must:
  - Develop and maintain data systems or sub-systems in their domain to ensure that data is recorded in the system and that data standards, quality guidelines, security, master/reference data management, and data services are all promoted throughout the system.
  - Provide data system information such as metadata, data dictionary, and data lineage.
  - Assist in IT-related data governance.
  - Ensure that the technical solutions for the data system comply with the information architecture of the domain and that selected technology can assist in the fulfillment of requirements during long-term data development.

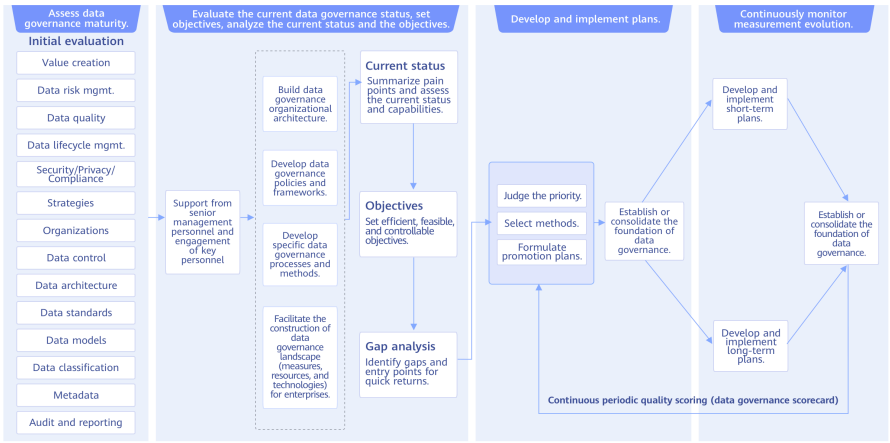
 Copyright © Huawei Technologies Co., Ltd.

5 Measurement and Evaluation System

5.1 Data Governance Methods

Measurement and evaluation of data governance follows a 4 step process: (1) Assess the current state of data governance maturity. (2) Evaluate the current data governance status, set objectives, and analyze the gap between the current status and the objectives. (3) Develop and execute plans for improvement. (4) Continuously monitor and measure progress.

Figure 5-1 Data governance process



Data governance follows the Plan-Do-Check-Act (PDCA) process. The following measurement and evaluation tools have been designed based on the specific requirements of data governance:

Tool	Object	Method	Frequency
Data governance maturity assessment	Enterprise	Questionnaires	Yearly

 Copyright © Huawei Technologies Co., Ltd.

Tool	Object	Method	Frequency
Data governance scorecard	Business and IT departments	Score data governance teams and business and IT departments.	Quarterly

Data Governance Measurement and Evaluation Tools

An annual assessment of data governance maturity allows you to understand every aspect of current data governance, set reasonable objectives, analyze the gap between the current status and the objectives, and formulate practical plans. Quarterly scorecards can be used to inspect the data governance of the various business and IT departments, and to drive further improvements. The annual assessment helps encourage the development of a mature system.

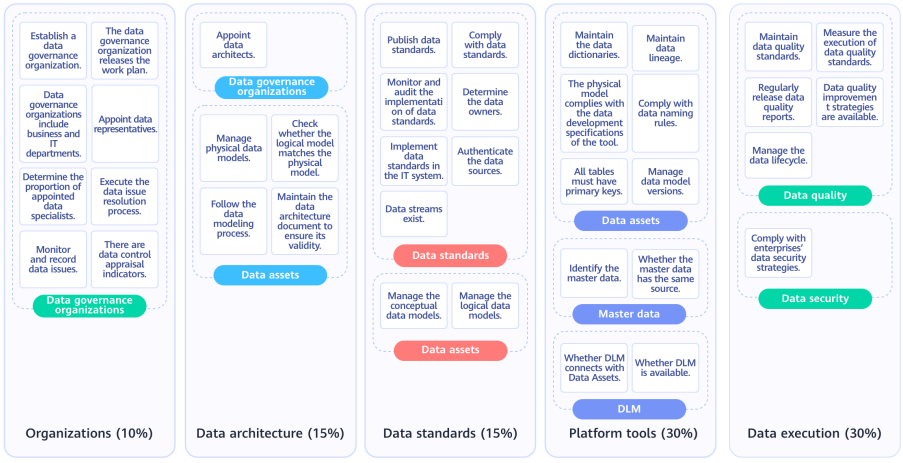
The annual assessment can be conducted based on subjective and objective surveys to comprehensively reflect the current level of maturity for data governance throughout the enterprise and to help identify reasonable objectives. The quarterly data governance scorecard is an important tool to help business and IT departments continuously improve data governance based on key dimensions. Quarterly iterative scoring can help improve data governance for all of the departments in an organization.

5.2 Measurement Dimensions

Assessment of data governance maturity is conducted using questionnaires involving 11 governance modules and more than 60 survey items. The maturity of data governance is comprehensively measured from multiple angles.

A data governance scorecard is used by the data governance organizations and business/IT departments to record quarterly scores for the data governance of each department. The data governance scorecard is an important tool for continuously promoting data governance of each department, improving data quality, and generally strengthening data governance throughout the organization.

Figure 5-2 Data governance scorecard



 Copyright © Huawei Technologies Co., Ltd.

5.3 Measurement and Scoring Rules

There are five levels of data governance maturity:

- Level 1, the lowest level, is when there is no data governance at all.
- Level 2 is when there is a certain amount of data governance, but there is no unified method available.
- Level 3 is when a proactive data management process and formal and consistent data governance methods have been clearly defined and are being implemented, but there is still much room for improvement in data governance.
- Level 4 is when there is quantitative management, which indicates that data governance is already fairly well developed. At level 4, formal and consistent data governance methods are being applied throughout the whole organization and those methods can be quantified and controlled.
- Level 5 is the highest level. At level 5, data governance is already quite mature and the priority should be continuous improvement.

Figure 5-3 Data governance maturity levels



Scorecards are used to score a range of aspects of data governance on a scale of one to five, and each aspect has a sub-item that is scored either 0 or 1. A score of 0 means that performance in that sub-item does not meet requirements, and a score of 1 means that it does.

Figure 5-4 Example

	Data Quality	Data Quality	Data Quality	Data Quality	Data Quality	Data Security	
Process domain	Whether to release data quality standards	Whether to implement data quality standards	Whether to release weekly data quality reports	Whether to formulate strategies to improve data quality	Whether to manage data lifecycle	Whether to comply with enterprises' data security strategies	Score
7.10 Delivery Project Management	1	1	1	0	0	1	3.3

Cover | Measurement Results | Organizations | Data Architectures | Data Standards | Platform Tools | Data Execution | ...

 Copyright © Huawei Technologies Co., Ltd.

In the example shown here, there are six measurement sub-items for "7.10 delivery project management". The calculation formula is as follows: (Total score of sub-items x 5)/Total sub-items. For in this example, the actual score is (1 + 1 + 1 + 0 + 0 + 1) x 5/6 = 3.3.

 Copyright © Huawei Technologies Co., Ltd.

6 Huawei Data Governance Cases

6.1 Thoughts

Huawei has encountered the following problems with data governance:

- Unclear data management responsibilities mean that no one is responsible when issues occur.
- A variety of different data sources result in data that is inconsistent and untrustworthy.
- Massive data migration causes repeated IT investment.
- A lack of standard definitions makes data difficult to understand and use.
- Inconsistent statistical rules for the reports released by each department make it difficult for business departments to make decisions.
- A diversity of data forms, rapid data growth, and complicated processing logic, increase the cost of investment.

During the process of digital transformation, Huawei has been able to resolve these issues. Huawei recognizes that only a complete data governance system can ensure qualified data, effective exploration of data value, and improved competitiveness.

High-Quality Data Powers Business Innovation

Nowadays, enterprises not only compete for market share within a given domain but also compete to develop new domains, which means innovation is even more important than before. To innovate, enterprises need to use technologies to develop and take advantage of their data assets. Data is a powerful driving force for enterprises seeking effective innovation.

However, problems faced by enterprises include data redundancy, a lack of unified standards, and low data quality.

Enterprise IT systems have seen their data expand rapidly over time, and the massive amount of data scattered across different locations makes utilization and management difficult. No unified business perspective is available for enterprises to view their data. There are only independent systems. Enterprises typically do

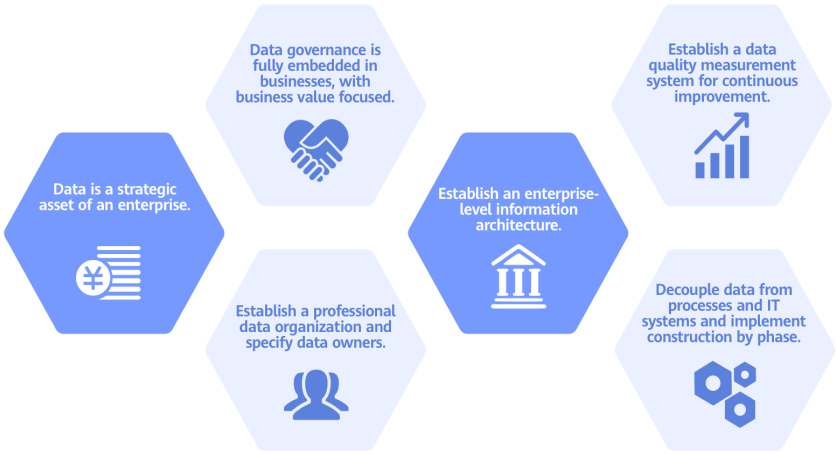
 Copyright © Huawei Technologies Co., Ltd.

not even know where to obtain standardized data or how each system connects to other systems.

Data Is the Key to Compete for High-Quality Customers

Data is the lifeblood of an enterprise. Whoever has the best data will gain market opportunities. In an increasingly competitive market, all enterprises are competing for high-quality customers. Senior executives are racking their brains, thinking about how to formulate the right business strategies. Accurate data analysis determines the source and utilization of every budget and every management decision. If data analysis is not accurate, it cannot deliver competitive advantages that can help win the day in today's fiercely competitive environment.

Figure 6-1 Thoughts on Huawei data governance



6.2 Practices

The data governance process includes generation, integration, analysis, and consumption. Data governance requires rules for this process. Huawei's data governance is focused on establishing organizations to manage data, on dotted line reporting to the corporate data management department, and on establishing cross-domain joint operation teams. Huawei has established a unified data management framework to guide data management by classifications in each domain. Huawei has also established a unified information architecture by specifying the definitions and compositions of the information architecture in policy documents. To manage data quality and data operations, Huawei has created a framework based on ISO 8000 standards. The design and execution of the data standards and operations is assessed twice a year, and the data owners regularly release quality reports to drive continuous improvement in their various domains.

 Copyright © Huawei Technologies Co., Ltd.

Figure 6-2 Visions and objectives

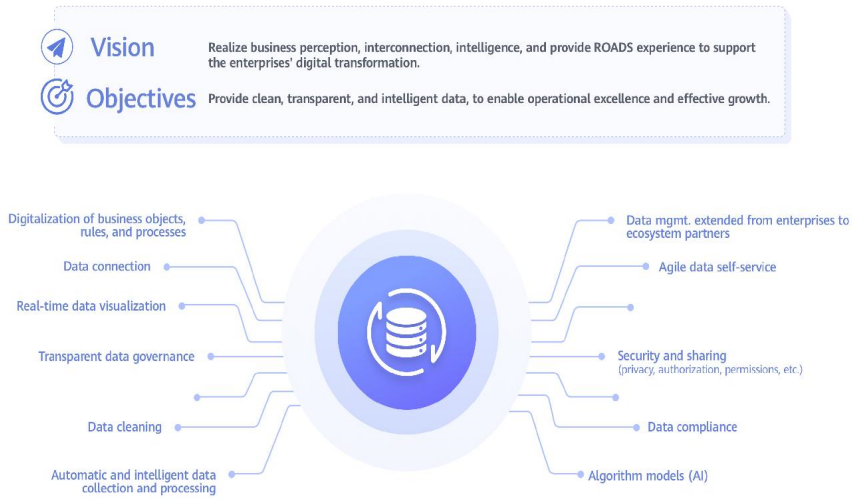
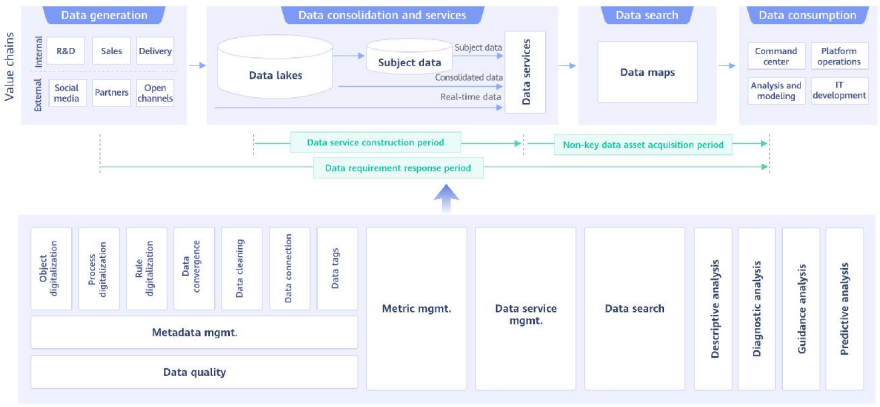


Figure 6-3 Data governance capabilities



6.3 Effects

Take finance as an example. Without data governance, the financial system suffers from a wide range of problems. For example, when IT systems are siloed, a single goal may require operations across multiple systems, which is far too inefficient. In one example, revenue management required data to be exported from 5 different systems and took 11 persons and 50 hours to process. Without data governance, obtaining data can be extremely time-consuming and labor-intensive.

With data governance, an initial draft of a monthly financial report can be generated within 3 days, and the final draft can be ready in 5. An annual report draft can be prepared in 11 days. Though it is not easy to get the ledger right for a large international organization, especially one that operates in more than 170 countries and regions, data governance can make it possible. To efficiently generate accurate and complete financial reports, a highly integrated global financial management system is indispensable, and so is data governance. By automating transaction accounting, optimizing ERP and data scheduling, monitoring data quality, and improving the performance of data analysis

 Copyright © Huawei Technologies Co., Ltd.

platforms, Huawei has made its international finances transparent in real time. The accounting process is traceable and manageable.

Figure 6-4 Practice



Copyright © Huawei Technologies Co., Ltd.

7 Thoughts on Data Governance and the Impact of COVID-19

On February, 3, 2020, the Chinese government held a meeting to highlight the major challenge that COVID-19 would present to China's system of governance. They emphasized that without the support of good data, it would be impossible to respond quickly and allocate appropriate medical resources and take other appropriate measures to contain the epidemic.

With data governance, standard open APIs make data collection more efficient; unified data standards accelerate data convergence; cross-platform data extraction and tracing eliminate silos and facilitate openness and sharing; private data is more secure, and reliability is improved.

Multiple e-government mini-apps with in-app QR "health codes" are accelerating the informatization of the citizenry. With data governance and sharing, in the near future, we may see a single QR code for all applications. Regularizing the online epidemic emergency response system can reduce the daily administrative management costs, enable more data to flow, and create more value. People voluntarily declare their health information to obtain the health QR codes that reflect their health status, and they can then use these codes to gain access to public places such as housing communities, office buildings, or public transit. In the early stages of an epidemic, the use of QR codes enhances safety by removing the need to gather together when waiting in line to fill out various forms related to the epidemic. Government agencies also no longer need to arrange for large groups of people to gather in place for screening. As the epidemic is gradually brought under control, the QR code can be used for contact tracing, which makes it easier to get people back to work. As work resumes in many regions, the cross-regional travel has introduced new challenges to epidemic control. Currently, the priority of data governance is to unify data standards of government agencies at all levels and across different departments to share data such as confirmed cases, suspected cases, close contacts, and regional risk levels.

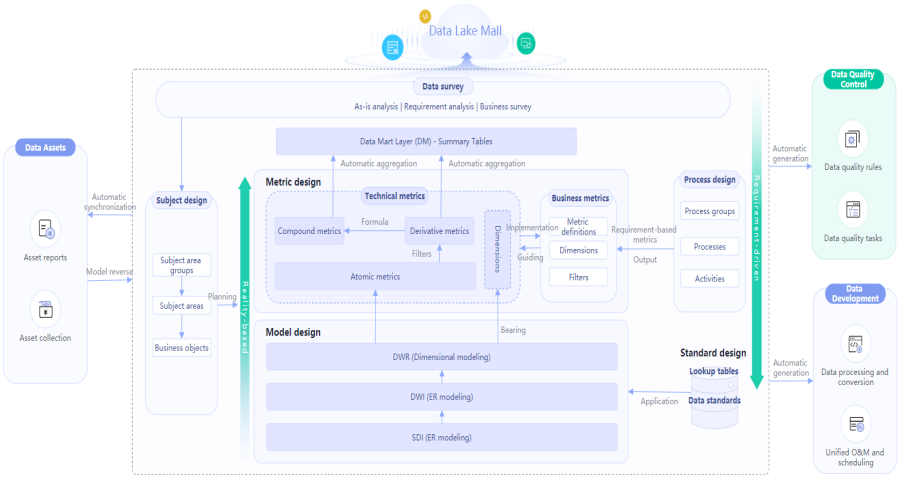
 Copyright © Huawei Technologies Co., Ltd.

8 Implementation of DAYU Data Governance Methods

DGC has integrated the process and functions of the data governance methods. The data governance methods integrated into DGC are a set of detailed specifications (used in the requirement raising, design, implementation, verification, and publishing phases) that guide users using DGC for data governance. The data governance functions refer to the automatic and intelligent tools provided by DGC that can help users efficiently complete data governance.

Refer to **Ask Us** for more details.

Figure 8-1 Process implementation



 Copyright © Huawei Technologies Co., Ltd.

Figure 8-2 Function implementation



 Copyright © Huawei Technologies Co., Ltd.

A Change History

Released On	Description
2020-03-25	This is the first official release.

 Copyright © Huawei Technologies Co., Ltd.