DataArts Studio

Service Overview

Issue 01

Date 2022-09-30

Huawei Cloud Computing Technologies Co., Ltd.

Address: Huawei Cloud Data Center Jiaoxinggong Road
            Qianzhong Avenue
            Gui'an New District
            Gui Zhou 550029
            People's Republic of China

Website: https://www.huaweicloud.com/intl/en-us/

Contents

1 What Is DataArts Studio?

Challenges to Enterprise Digital Transformation

Enterprises often face challenges in the following aspects when managing data:

- Governance
  - Inconsistent data system standards impact data exchange and sharing between different departments.
  - There are no great search tools to help service personnel locate the data they need when they need it.
  - If metadata fails to define data in business terms that are familiar to data consumers, the data is difficult to understand.
  - When there are no good methods to evaluate and control data quality, it makes the data hard to trust.
- Operations
  - Data analysts and decision makers require efficient data operations. There is no efficient data operations platform to address the growing and diversified demands for analytics and reporting.
  - Repeated development of the same data wastes time, slows down development, and results in too many data copies. Inconsistent data standards waste resources and drive up costs.
- Innovation
  - Data silos prevent data from being shared and circulated across departments in enterprises. As a result, cross-domain data analysis and data innovation fail to be stimulated.
  - Currently, most enterprises still utilize their data for analytics and reporting. There is a long way to go before enterprises have widespread, data-driven service innovation.

What Is DataArts Studio?

DataArts Studio is a one-stop data operations platform that drives digital transformation. It allows you to perform many operations, such as integrating and developing data, designing data architecture, controlling data quality, managing data assets, creating data services, and ensuring data security. Incorporating big data storage, computing and analytical engines, it can also construct industry

knowledge bases and help your enterprise build an intelligent end-to-end data system. This system can eliminate data silos, unify data standards, accelerate data monetization, and accelerate your enterprise's digital transformation.

Figure 1-1 shows the architecture.

Figure 1-1 Architecture
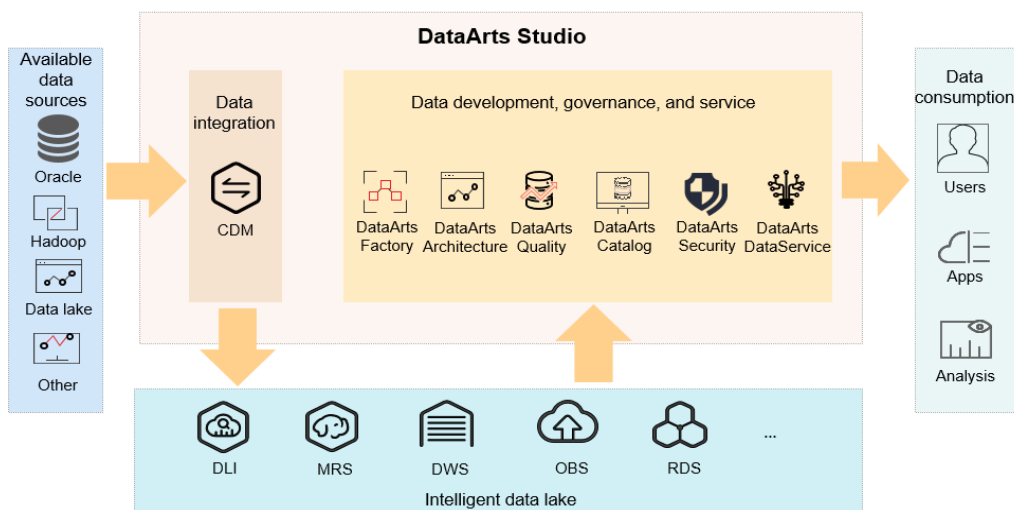
As shown in the figure, DataArts Studio is based on the data lake base and provides capabilities such as data integration, development, and governance. DataArts Studio can connect to data lakes and cloud database services, such as MRS Hive and GaussDB(DWS). These data lakes and cloud database services are used as the data lake base. DataArts Studio can also connect to traditional enterprise data warehouses, such as Oracle and MySQL.

DataArts Studio consists of the following functional modules:

● **Management Center**

Management Center supports data connection management and connects to the data lake base for activities such as data developmentand data governance.

● **DataArts Migration**

DataArts Migration supports data migration between 20+ data sources and integration of data sources into the data lake. It provides wizard-based configuration and management and supports single table, entire database, incremental, and periodic data integration.

● **DataArts Architecture**

DataArts Architecture helps you plan the data architecture, customize models, unify data standards, visualize data modeling, and label data. DataArts Architecture defines how data will be processed and utilized to solve business problems and enables you to make informed decisions.

● **DataArts Factory**

DataArts Factory helps you build a big data processing center, create data models, integrate data, develop scripts, and orchestrate workflows.

- **DataArts Quality**

  DataArts Quality monitors the data quality in real time with data lifecycle management and generates real-time notifications on abnormal events.

- **DataArts Catalog**

  DataArts Catalog provides enterprise-grade metadata management to help you better know your data assets. A data map shows the lineage of your data and allows you to have a global view of your data assets. Data search, operations, and monitoring are smarter than before.

- **DataArts DataService**

  DataArts DataService is a platform where you can develop, test, and deploy your data services. It ensures agile response to data service needs, easier data retrieval, better experience for data consumers, higher efficiency, and better monetization of data assets.

- **DataArts Security**

  DataArts Security provides all-round protection for enterprises' data. It provides access permission management, sensitive data identification, and privacy protection management to help you establish a security warning mechanism, improve the overall security protection capability, and ensure data availability and security compliance.

2 Basic Concepts

DataArts Studio Instance

A DataArts Studio instance is the minimum unit of compute resources provided for users. You can create, access, and manage multiple DataArts Studio instances at the same time. A DataArts Studio instance allows you to access the following modules: Management Center, DataArts Architecture, DataArts Migration, DataArts Factory, DataArts Quality, and DataArts Catalog. You can obtain DataArts Studio instances with specifications tailored to your service requirements.

Workspace

A workspace enables admins to manage member permissions, resources, and configurations of the underlying compute engines.

The workspace is a basic unit for member management as well as role and permission assignment. Each team must have an independent workspace.

You can access the Management Center, DataArts Catalog, DataArts Quality, DataArts Architecture, DataArts DataService, DataArts Factory, and DataArts Migration modules, but only after your account is added to a workspace and assigned the permissions required to perform such operations.

Member and Role

A member is a account that has been assigned the permissions required to access and use a workspace. As an admin, when you add a workspace member, you must set a role.

A role is a predefined combination of permissions. Different roles have different permission sets. After a role is assigned to a member, the member has all the permissions of that role. Each member must have at least one role, and they can have multiple roles at the same time.

CDM Cluster

A CDM cluster run on an ECS. You can create data migration tasks in a CDM cluster and migrate data between homogeneous or heterogeneous data sources in the cloud and on-premises data center.

Data Source

A data source is a medium for storing or processing data, such as a relational database, data warehouse, and data lake. Different data sources use different data storage, transmission, processing, and application modes, as well as different scenarios, technologies, and tools.

Source Data

Source data is the data that is not processed after created. In data management, source data refers to the data directly from source files (such as service system databases, offline files, and IoT files) or copies of source files.

Data Connection

A data connection is a collection of details required for accessing where data is stored, including the connection type, name, and login information.

Concurrency

Concurrency refers to the maximum number of threads that can be concurrently read from the source in a data integration job.

Dirty Data

Dirty data refers to the data meaningless to business or in invalid format. For example, if the source data of the VARCHAR type is not properly converted, it cannot be written to the destination column of the INT type.

Job (DataArts Factory)

A job is composed of one or more nodes that run together to complete data operations.

Node

A node is a definition for the actions to be performed on your data. For example, you can use the MRS Spark node to execute predefined Spark jobs in MRS.

Solution

A solution is a series of convenient and systematic management operations that meet service requirements and objectives. Each solution can contain one or more business-related jobs, and each job can be reused by multiple solutions.

Resource

A resource is the self-defined code or text file that you upload. It is invoked when nodes run.

Expression Language (EL)

Node parameters in data development jobs can be dynamically generated based on the running environment using ELs. An EL often uses simple arithmetic and

calculation logic and references embedded objects including job objects and tool objects.

Environment Variable

An environmental variable is an object with a specific name in the operating system. It contains information to be used by one or more applications.

PatchData

PatchData is an instance that was generated in the past by a repeatedly scheduled job.

Data Governance

Data governance is the process by which you can manage, utilize, and protect your enterprise data throughout the data lifecycle. It includes access control, data quality management, and risk management.

Data Survey

A data survey involves collecting data that is generated when sorting business requirements, creating business processes, and classifying data subjects based on the existing business data and industry status.

Subject Design

Subject design provides hierarchical architectures that help you define and classify data assets, helping you better understand your data assets and clarify the relationship between business domains and business objects.

Subject Area Group

A subject area group is a collection of subject areas that have the same business features.

Subject Area

A subject area is a high-level, non-overlapping classification of data used to manage business objects.

Business Object

A business object includes important information about people, events, and things that are indispensable to your enterprise's operations and management.

Process Design

Process design is to generate a structured framework of data processing process, including the categories, levels, boundaries, scope, and input/output relationships, and reflect the business models and characteristics of your enterprise.

Data Standard

A data standard is the description of data meanings and business rules that must be complied with by your enterprise. It describes the common understanding of certain data at the company level.

Lookup Table

A lookup table includes a series of allowed values and additional text descriptions that are generally associated with data standards to generate a range of values for the verification of quality monitoring rules.

SDI

Source Data Integration (SDI) copies data from source systems.

DWI

Data Warehouse Integration (DWI) integrates and cleanses data from multiple source systems, and builds ER models based on the third normal form (3NF).

DWR

Data Warehouse Report (DWR) is based on multi-dimensional models and its data granularity is the same as that of DWI.

DM

Data Mart (DM) is where multiple types of data are summarized and displayed.

ER Modeling

Entity Relationship (ER) modeling describes business activities of an enterprise. ER models are compliant with the third normal form (3NF). You can use ER models for data integration, which merges and classifies data from different systems by similarity or subject. However, you cannot use ER models for decision-making.

Dimensional Modeling

A dimensional model is generally created for data analysis and decision-making. Its aim is to complete the analysis of complex and multiple user requirements at full speed.

A multidimensional model is a fact table that consists of numeric measure metrics. The fact table is associated with a group of dimensional tables that contain description attributes through primary or foreign keys.

In the DataArts Architecture module of DataArts Studio, dimensional modeling involves constructing bus matrices to extract business facts and dimensions for model creation. You need to sort out business requirements for constructing metric systems and creating summary models.

Metric (DataArts Architecture)

A metric is a statistical value that measures the overall characteristic of a target and indicates the business situation in a business activity of an enterprise. A metric consists of its name and value. The metric name and its meaning reflect the quality and quantity of the metric. The metric value reflects the quantifiable values of the specified time, location, and condition of the metric.

Measure

A measure is a quantifiable value used to measure business situations. It usually refers to a number, for example, an amount, quantity, or period. Measures are numerical values that do not have explicit business relevance, but they can be converted into metrics in a business context.

Dimension

A dimension is used to observe and analyze business data. It supports data aggregation, drilling, and slicing analysis and is used as the GROUP BY condition in SQL statements. Most dimensions have a hierarchical structure, for example, geographic dimension (including country, region, province, and city levels) and time dimension (including annually, quarterly, and monthly levels).

Atomic Metric

An atomic metric is generated based on dimension tables and fact tables of a multidimensional model. The business objects and the finest data granularity of an atomic metric are consistent with those of the multidimensional model. An atomic metric usually consists of measures and attributes related with measures and business objects, all of which aim to support agile self-service consumption of derivative metrics, for example, the number of retail stores (including the store names and levels).

Derivative Metric

A derivative metric is derived from the combination of modifiers, standards, dimensions, and atomic metrics. Modifiers, standards, and definitions are usually the attributes of an atomic metric. An example is the in-store promoter coverage.

Compound Metric

A compound metric is generated by derivative metrics. The dimensions and modifiers of a compound metric are the same as those of the derivative metric. (No new dimensions and modifiers for a compound metric can be generated if its derivative metric has no dimensions and modifiers.)

Data Quality Rule

A data quality rule is a logical unit used to determine whether the data meets business requirements.

Data Asset

A data asset is a resource that is owned or controlled by your enterprise and can be monetized in the future. The data resource is recorded in physical or electronic mode. Not all the data of your enterprise can be considered as a data asset. A data asset must be a data resource that can generate value for your enterprise.

Data Map

A data map is a data search-driven tool that displays the source, quality, distributions, standards, flow directions, and relationships of data in graphical forms. You can use a data map to easily find, read, and consume data.

Metadata

Metadata is data about data. Specifically, it is information about the organization, domain, and relationships of data. Metadata includes metadata entities and metadata elements. A metadata element is a basic unit of metadata, and several related metadata elements form a metadata entity.

Metadata Collection

You can customize a collection policy to collect technical metadata from data sources.

Data Asset Report

A data asset report provides an overview of the data asset and their statistics.

DataArts DataService

DataArts DataService provides data as a product based on data distribution and release frameworks. The product provided meets your requirements for real-time data and industry standards. It can be reused and shared securely.

API Gateway

API Gateway provides API hosting services through the API gateway, covering the full life-cycle management of API release, management, O&M, and sales. It helps you easily implement microservice aggregation, frontend and backend separation, system integration, and open functions and data to partners and developers in a quick, cost-effective, but low risky way.

3 Functions

DataArts Migration: Efficient Ingestion of Multiple Heterogeneous Data Sources

DataArts Migration can help you seamlessly migrate batch data between 30+ homogeneous or heterogeneous data sources. You can use it to ingest data from both on-premises and cloud-based data sources, including file systems, relational databases, data warehouses, NoSQL databases, big data services, and object storage.

DataArts Migration uses a distributed compute framework and concurrent processing techniques to help you migrate enterprise data in batches without any downtime and rapidly build desired data structures.

Figure 3-1 DataArts Migration
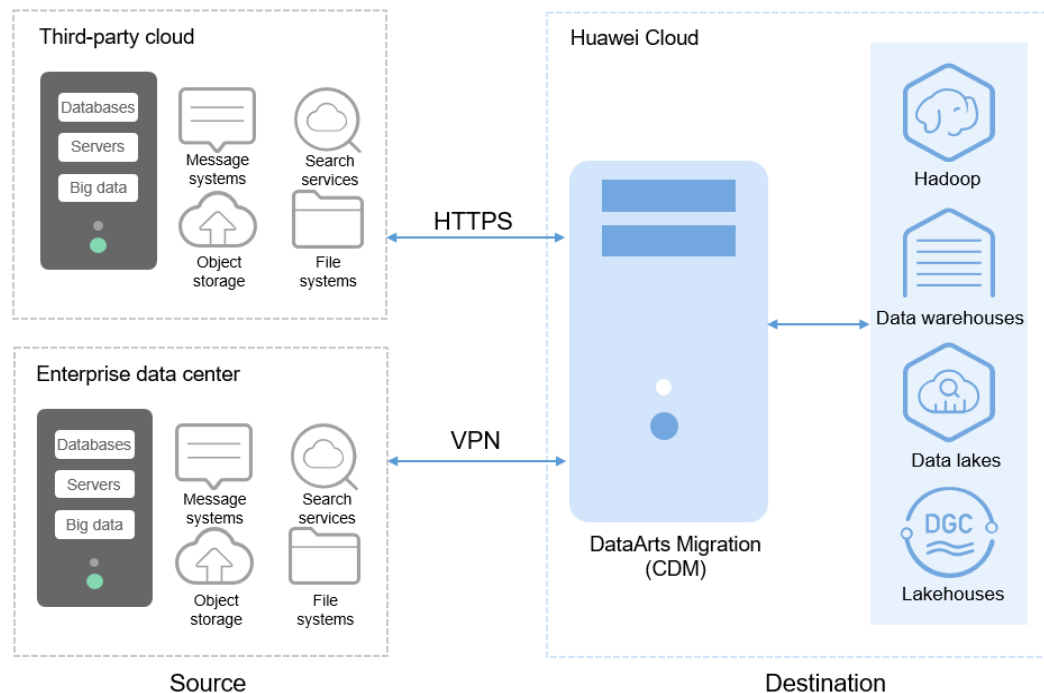
You can manage data on the wizard-based task management page. You can easily create data migration tasks that meet your requirements. DataArts Migration provides the following functions:

- **Table/File/Entire DB migration**

  You can migrate tables or files in batches, and migrate an entire database between homogeneous and heterogeneous database systems. You can include hundreds of tables in a single job.

- **Incremental data migration**

  You can migrate files, relational databases, and HBase in an incremental manner. You can perform incremental data migration by using WHERE clauses and variables of date and time.

- **Migration in transaction mode**

  When a batch data migration job fails to be executed, data will be rolled back to the state before the job started and data in the destination table will be automatically deleted.

- **Field conversion**

  Field conversion includes anonymization, character string operations, and date operations.

- **File encryption**

  You can encrypt files that are migrated to a cloud-based file system in batches.

- **MD5 verification**

  MD5 is used to check file consistency from end to end.

- **Dirty data archiving**

  Data that fails to be processed during migration, is filtered out and is not compliant with conversion or cleansing rules is recorded in dirty data logs. You can easily analyze abnormal data. You can also set a threshold for the dirty data ratio to determine whether a task is successful.

DataArts Architecture: Visualized, Automated, and Intelligent Data Modeling

DataArts Architecture incorporates data governance methods. You can use it to visualize data governance operations, connect data from different layers, formulate data standards, and generate data assets. You can standardize your data through ER modeling and dimensional modeling. DataArts Architecture is a good option for unified construction of metric platforms. With DataArts Architecture, you can build standard metric systems to eliminate data ambiguity and facilitate communications between different departments. In addition to unifying computing logic, you can use it to query data and explore data value by subject.

Figure 3-2 DataArts Architecture
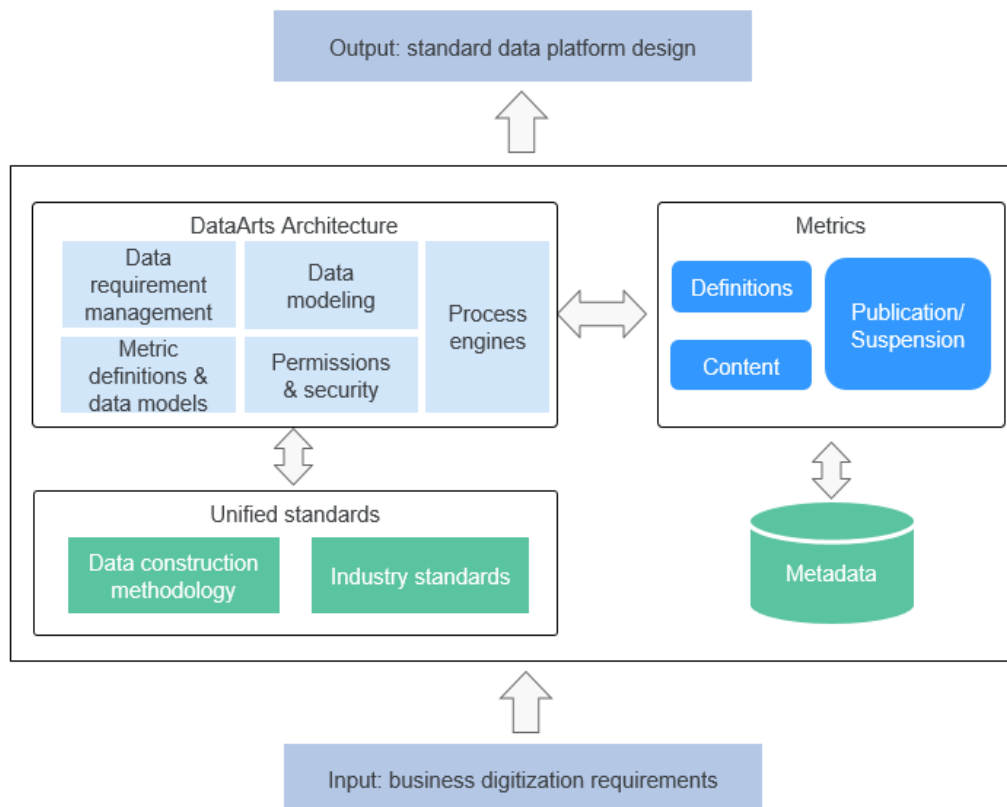
DataArts Architecture offers the following major functions:

- **Subject design**

  You can use DataArts Architecture to build unified data classification systems for directory-based data management. Data classification, search, evaluation, and usage are easier than ever before. DataArts Architecture provides hierarchical architectures that help you define and classify data assets, allowing data consumers to better understand and trust your data assets.

- **Data standards**

  DataArts Architecture can help you create process-based and systematic data standards that fit your needs. Peered with the national and industry standards, these standards enable you to standardize your enterprise data and improve data quality, ensuring that your data is trusted and usable.

- **Data modeling**

  Data modeling involves building unified data model systems. You can use DataArts Architecture to build a tiered, enterprise-class data system based on data specifications and models. The system incorporates data from the public layer and subject libraries, significantly reducing data redundancy, silos, inconsistency, and ambiguity. This allows freer flow of data, better data sharing, and faster innovation.

  The following data modeling methods are supported:

  - **ER modeling**

    ER modeling involves describing the business activities of an enterprise, and ER models are compliant with the third normal form (3NF). You can

use ER models for data integration, which merges and classifies data from different systems by similarity or subject. However, you cannot use ER models for decision-making.

– **Dimensional modeling**

Dimensional modeling involves constructing bus matrices to extract business facts and dimensions for model creation. You need to sort out business requirements for constructing metric systems and creating summary models.

DataArts Factory: One-stop Collaborative Development

DataArts Factory provides an intuitive UI and built-in development methods for script and job development. DataArts Factory also supports fully hosted job scheduling, O&M, and monitoring, and incorporates industry data processing pipelines. You can create data development jobs in a few steps, and the entire process is visual. Online jobs can be jointly developed by multiple users. You can use DataArts Factory to manage big data cloud services and quickly build a big data processing center.

Figure 3-3 DataArts Factory architecture
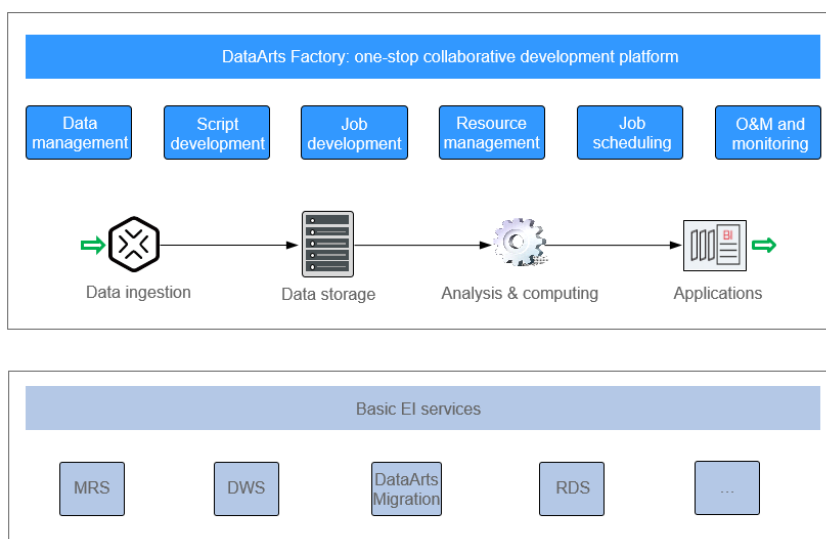
DataArts Factory allows you to manage data, develop scripts, and schedule and monitor jobs. Data analysis and processing are easier than ever before.

● **Data management**

– You can manage multiple types of data warehouses, such as GaussDB (DWS), DLI, and MRS Hive.

– You can use the graphical interface and data definition language (DDL) to manage database tables.

● **Script development**

– Provides an online script editor that allows more than one operator to collaboratively develop and debug SQL, Python, and Shell scripts online.

– You can use Variables.

- **Job development**
  - DataArts Factory provides a graphical designer that allows you to rapidly develop workflows through drag-and-drop and build data processing pipelines.
  - DataArts Factory is preset with multiple task types such as data integration, SQL, Spark, machine learning, and Shell. Data is processed and analyzed based on task dependencies.
  - You can import and export jobs.
- **Resource management**

  You can centrally manage file, jar, and archive resources used during script and job development.
- **Job scheduling**
  - You can schedule jobs to run once or recursively and use events to trigger scheduling jobs.
  - Job scheduling supports a variety of hybrid orchestration tasks. The high-performance scheduling engine has been tested by hundreds of applications.
- **O&M and monitoring**
  - You can run, suspend, restore, or terminate a job.
  - You can view the operation details of each job and each node in the job.
  - You can use various methods to receive notifications when a job or task error occurs.

DataArts Quality: Verifiable and Controllable Data Quality

DataArts Quality can monitor your metrics and data quality, and screen out unqualified data in a timely manner.

- **Metric monitoring**

  You can use DataArts Quality to monitor the quality of data in your databases. You can create metrics, rules, or scenarios that meet your requirements and schedule them in real time or recursively.
- **Data quality monitoring**

  You can create data quality rules to check whether the data in your databases is accurate in real time.

  Qualified data must meet the following requirements: integrity, validity, timeliness, consistency, accuracy, and uniqueness. You can standardize data and periodically monitor data across columns, rows, and tables based on quality rules.

Figure 3-4 Data quality rule system
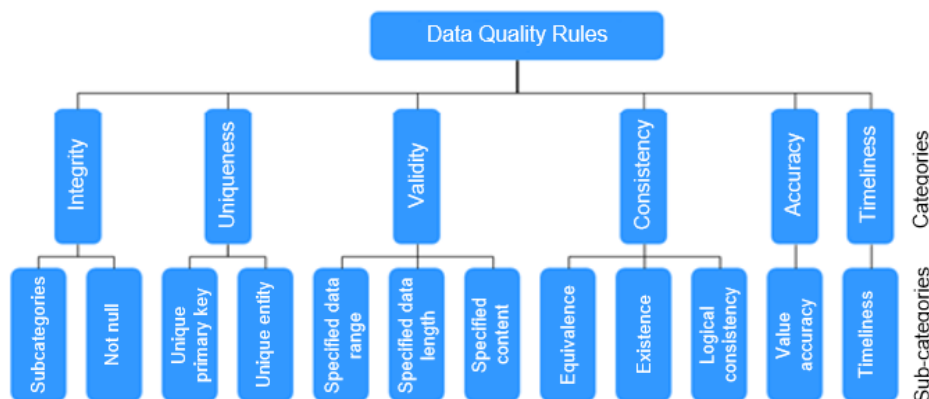
DataArts Catalog: End-to-End Data Asset Visualization

With enterprise-class metadata management, you can define your data assets in business terms familiar to data consumers. Data drilling and source tracing are also supported. A data map shows data lineage and a global view of your data assets. Data search, operations, and monitoring are more intelligent than before.

- **Metadata management**

  Metadata management is vital for data lake governance. You can create policies to collect metadata from your data lake, and customize metadata models to import metadata in batches, associate business data with technical data, manage and use full-link data lineages.

Figure 3-5 Full-link data lineages
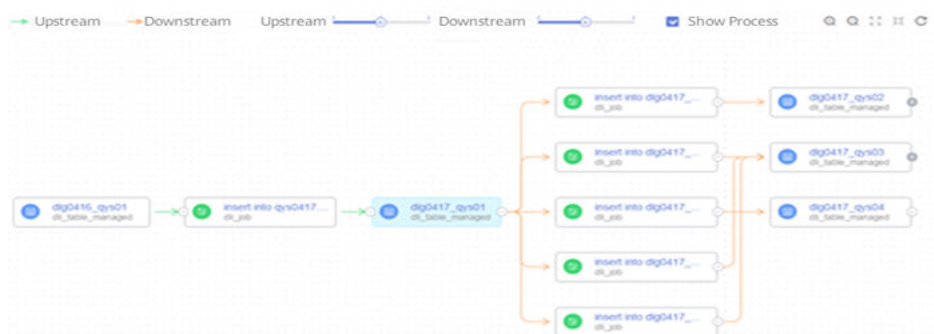
- **Data map**

  Data maps facilitate data search, analysis, development, mining, and operations. They provide lineage information and impact analysis. Data maps make data search easier and faster than before.

  - Keyword search and fuzzy search are supported, helping you quickly locate the data you need.
  - You can search for tables by name. Table details are displayed as soon as the matching table is found. You can also add more descriptions for the searched table.
  - Data maps display the source, destination, and processing logic of a table field.

– You can classify and tag data assets as required.

DataArts DataService: Improved Access, Query, and Search Efficiency

DataArts DataService enables you to manage your enterprise APIs centrally, and controls the access to your subjects, profiles, and metrics. It helps improve the experience for data consumers and the efficiency of data asset monetization. You can use DataArts DataService to generate APIs and register the APIs with DataArts DataService for unified management and publication.

DataArts DataService uses a serverless architecture. You only need to focus on the API query logic, without worrying about infrastructure such as the runtime environment. DataArts DataService supports elastic scaling of compute resources, significantly reducing O&M costs.

Figure 3-6 DataArts DataService architecture
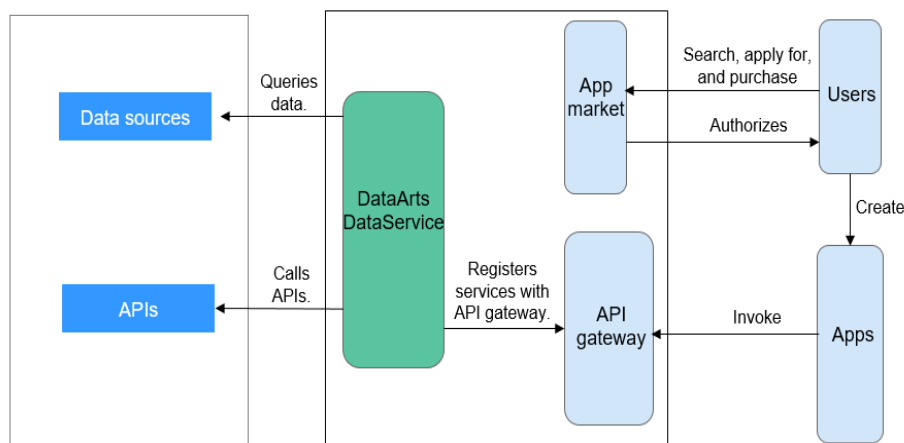
DataArts Security: All-Round Protection

- **Cyber security**

  Tenant isolation and access permissions control are implemented to protect the privacy and data security of systems and users based on preset network isolation, security group, and security hardening rules.

- **User permissions control**

  Role-based access control involving associating roles with permissions and supports fine-grained permission policies to meet different authorization requirements. DataArts Studio provides four roles: admin, developer, deployer, operator, and viewer. Each role has different permissions.

- **Data security**

  DataArts Studio provides the review mechanism for key processes.

  Data is managed by level and category throughout the lifecycle, ensuring data privacy compliance and traceability.

4 Advantages

One-Stop Data Operations Platform

DataArts Studio is a one-stop data operations platform that allows you to perform many operations, including integrating data from every domain, designing data architecture, monitoring data quality, managing data assets centrally, developing data services, and connecting data from different data sources. In a word, it can help you build a comprehensive data governance solution.

Comprehensive Data Control and Governance

DataArts Studio enables you to monitor your data quality in the full data lifecycle, provides you with standard data definitions, generates data processing code, and notifies you immediately when anomaly events occur.

Diverse Data Development Types

DataArts Studio has a wide range of scheduling configuration policies and powerful job scheduling. It supports online collaborative development among multiple users, online editing and real-time query of SQL and shell scripts, and job development via data processing nodes such as CDM, SQL, MRS, Shell, and Spark.

Unified Scheduling and O&M

Fully hosted scheduling is supported. Time- and event-based triggering mechanisms are available. You can schedule a task by minute, hour, day, week, or month.

The visualized task O&M center monitors all tasks and supports notification settings, enabling you to obtain real-time task status and ensuring normal running of services.

Reusable Industrial Knowledge Bases

DataArts Studio provides vertical industries with reusable knowledge bases, including data standards, domain models, subject libraries, algorithm libraries, and metric libraries, and supports fast customization of E2E data operations solutions for industries such as smart government, smart taxation, and smart campus.

Unified Data Asset Management

DataArts Studio allows you to have a global view of your data assets, facilitating fast asset query, intelligent asset management, data source tracing, and data openness. In addition, it enables you to define your business data catalog, terms, and classifications, as well as access to your assets in a unified manner.

Visualized Data Operations in All Scenarios

The data governance and operations process is visual. You can perform configurations using a drag-and-drop interface without coding. The processing result is also visual, facilitating interaction and exploration. Data asset management is also visual and allows you to perform data drilling and source tracing.

All-Round Security Assurance

Unified security authentication, tenant isolation, data grading and classification, and data lifecycle management ensure data privacy, auditability, and traceability.

Role-based access control allows you to associate roles with permissions and supports fine-grained permission policies, meeting different authorization requirements.

5 Application Scenarios

One-Stop Data Operations and Governance Platform

You can use the one-stop data lake operations and governance platform for data collection, architecture design, monitoring, cleansing, modeling, connection, integration, consumption, and intelligent analysis. It helps you rapidly grow your enterprise's big data operations.

Advantages

- Job orchestration for multiple cloud services
- Comprehensive data control and governance
- Diverse data engines

  Support for interconnection with all HUAWEI CLOUD data lake and database services, and with traditional data warehouses, such as Oracle
- Ease of use

  GUI-based orchestration and out-of-the-box availability

Figure 5-1 One-stop data operations platform
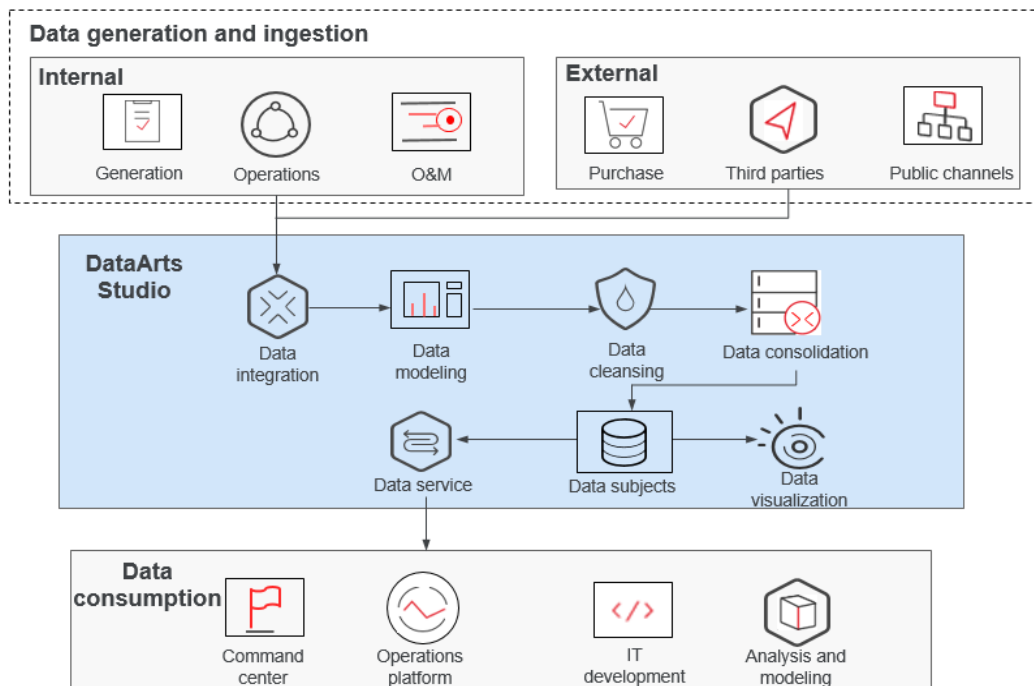
Building Cloud-based Data Platforms with Speed

You can use DataArts Studio to migrate offline data to the cloud and integrate the data into big data services. On the DataArts Studio management console, you can use the integrated data to quickly start developing jobs and easily build enterprise data systems.

Advantages

- Quick data integration

  On the GUI, you can migrate offline or real-time data to cloud warehouses in just a few steps.

- Multiple warehouse services

  You can choose GaussDB (DWS), MRS, or any other warehouses to meet your service needs.

- Secure, stable, and cost-saving

  Data on the cloud is secure owing to one-stop data service capabilities and stable data warehouse services; you no longer need to build and maintain big data clusters, significantly reducing costs.

Figure 5-2 Cloud data platform
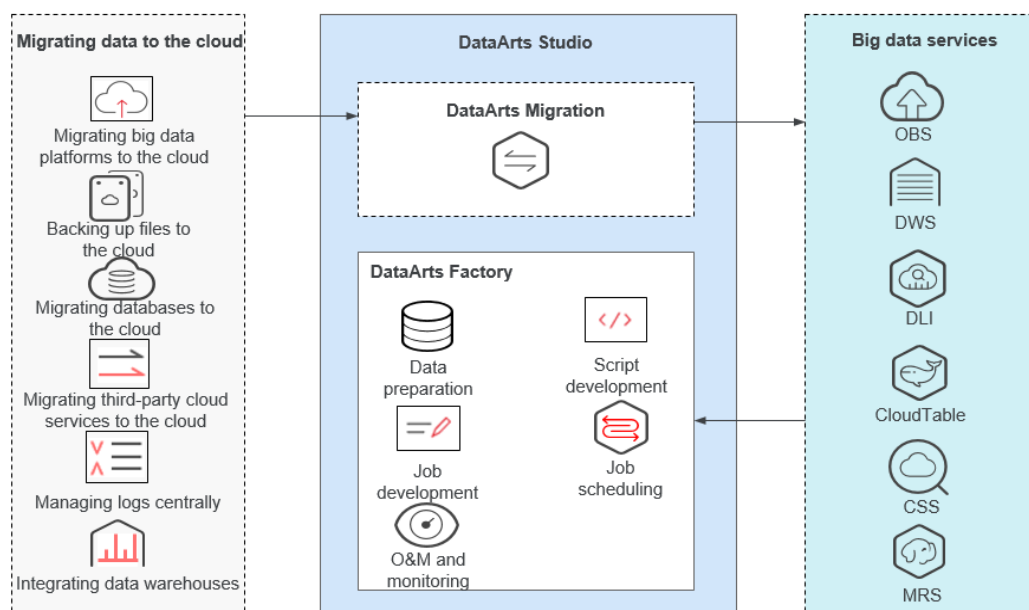
Building Data Lake Governance Platforms Powered by Industry Know-How

Incorporating Huawei's technological expertise in industry models and algorithms, DataArts Studio can help you build a data governance platform to quickly grow your enterprise's data operations capabilities.

Advantages

- Industry-tailored solutions

  Custom solutions for government, taxation, smart city, smart transportation, and smart campus

- Standards compliance

  Compliance of layered industry data standards

- Various domain models

  A variety of industry domain models developed from eight types of data, which are people, organization, event, spatio-temporal, vehicle, asset, device, and resource data, and their relationships

- Quick utilization of industry libraries

  Quick utilization of industry-specific subject libraries, algorithm libraries, and metric libraries

Figure 5-3 Data governance platform
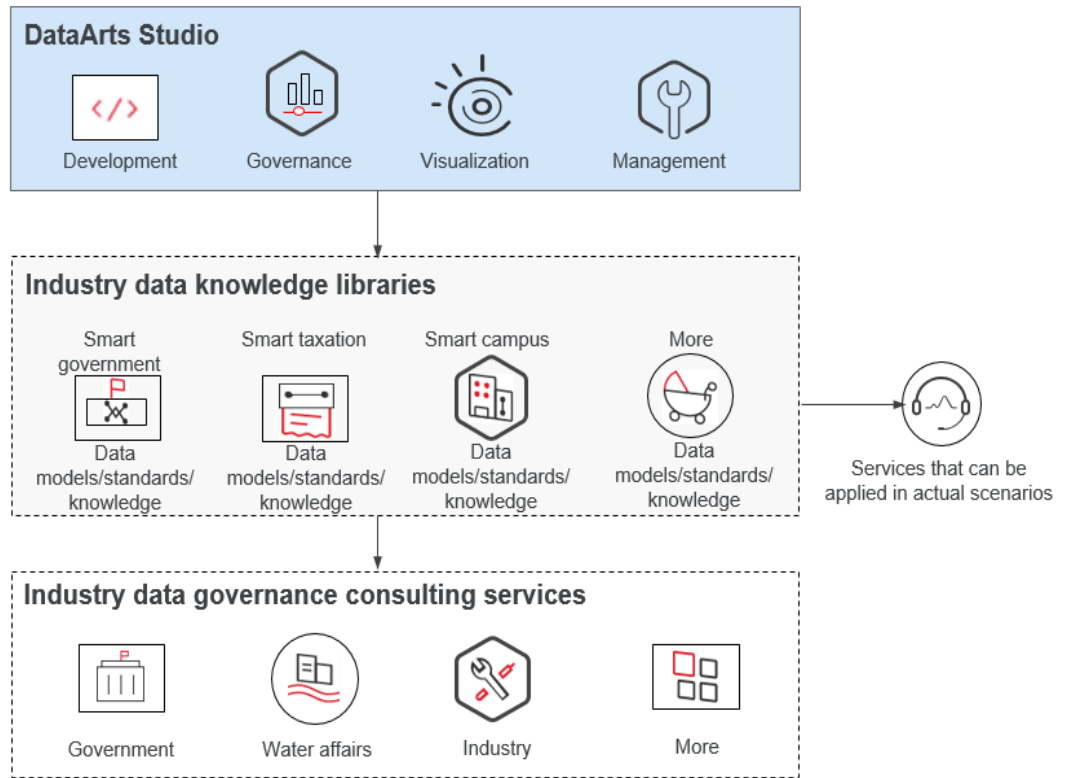
6 DataArts Studio Versions

Select a DataArts Studio version with caution based on the functions and specifications you need.

- After you buy an instance of a specified version, you cannot directly downgrade the version. For example, if you have bought an instance of the enterprise version, you cannot directly downgrade the instance to the starter version. Instead, you will need to back up data of the instance, unsubscribe from it, buy a new instance, and migrate the backup data to the new instance.

- If your business volume keeps increasing and the instance version you have bought cannot meet your requirements, you can upgrade the instance version. To upgrade the instance version, log in to the DataArts Studio console, locate the target DataArts Studio instance, click **Upgrade**, and buy a package with higher specifications.

Application Scenarios of DataArts Studio Versions

Table 6-1 Recommended application scenarios for each DataArts Studio version

Version	Application Scenario
Starter	A primary data lake project with no full-time data development engineers and no data governance needs
Basic	One or two full-time data development engineers, and up to 1,000 data tables
Advanced	Five to ten full-time data development engineers, clear data standards and efficient data quality management, and up to 2,000 data tables
Professional	Large or medium enterprises with a team of 10 to 30 full-time data development engineers and well-designed systems
Enterprise	Large enterprises and enterprises with multiple branches

Specifications of DataArts Studio Versions

Table 6-2 Components supported by DataArts Studio

DataArts Studio Component	Starter	Basic	Advanced	Professional	Enterprise
DataArts Migration	√	√	√	√	√
Management Center	√	√	√	√	√
DataArts Architecture	x	√	√	√	√
DataArts Factory	√	√	√	√	√
DataArts Quality	x	√	√	√	√
DataArts Catalog	x	√	√	√	√
DataArts DataService	x	√	√	√	√
DataArts Security (OBT)	x	√	√	√	√

Table 6-3 DataArts Studio version specifications (a single instance)

Specification	Starter	Basic	Advanced	Professional	Enterprise
DataArts Studio CDM cluster[1]	Number of clusters: 1 Name: cdm.medium vCPUs | memory: 4 vCPUs | 8 GB	Number of clusters: 1 Name: cdm.medium vCPUs | memory: 4 vCPUs | 8 GB	Number of clusters: 1 Name: cdm.large vCPUs | memory: 8 vCPUs | 16 GB	Number of clusters: 1 Name: cdm.xlarge vCPUs | memory: 16 vCPUs | 32 GB	Number of clusters: 1 Name: cdm.xlarge vCPUs | memory: 16 vCPUs | 32 GB

Specification	Starter	Basic	Advanced	Professional	Enterprise
Job node scheduling times/day[2]	5,000/day	20,000/day	40,000/day	80,000/day	200,000/day
Number of technical assets[3]	Not supported	1,000	2,000	4,000	10,000
Number of data models[4]	Not supported	1,000	2,000	4,000	10,000

Notes:

[1] DataArts Studio CDM cluster: This is a free cluster provided together with the DataArts Studio instance. It can be used as an agent for the data connections in Management Center. However, you are not advised to use the node in a data migration job when the node is used as an agent. To buy a CDM cluster used to run CDM jobs, buy a CDM incremental package. For details, see **Buying a DataArts Studio Incremental Package**.

[2] Job node scheduling times/day: It refers to the total number of scheduling times of the data development jobs, quality jobs, comparison jobs, scenarios, and metadata collection jobs per day. The number of scheduling times of data development job per day is measured by node (including the Dummy node), covering PatchData tasks but not test or retry upon failures. For example, if a job contains two DWS SQL nodes and one Dummy node, starts to be executed at 00:00 every day, is scheduled every 10 hours, and a PatchData task is performed on the current day to patch data of the last 10 days, then the number of scheduling times of the job is 66 (2 x 3 + 2 x 3 x 10) for the current day and 6 (2 x 3) for every following day.

In addition, if the number of used scheduling times, scheduling times in use, and scheduling times to be used for job nodes on the current day exceeds the specifications of this version, a message is displayed, indicating that the number of job node scheduling times/day exceeds the quota.

[3] Number of technical assets: number of tables and OBS files in DataArts Catalog

[4] Number of data models: number of logical models, physical models, dimension tables, fact tables, and summary tables in DataArts Architecture.

7 Billing

DataArts Studio provides a basic package and an incremental package. The total price of a DataArts Studio package is automatically calculated on the management console. You can buy a DataArts Studio package in just a few steps.

> **NOTICE**
>
> When you use DataArts Studio, you may need to pay for the following services:
> - OBS: During batch data migration or data development, DataArts Studio may write dirty data or logs to OBS, which incurs data storage fees.
> - EIP: If you use an EIP for your DataArts Migration cluster or DataArts DataService Exclusive cluster, you need to pay for the EIP.
> - SMN: If you enable SMN notifications for your DataArts Studio modules, you need to pay for the notifications.
> - DEW: If you enable KMS when creating a link in DataArts Migration or creating a connection in Management Center, you will be billed for key management.

Billing Items

DataArts Studio provides a basic package and an incremental package.

- A basic package provides a DataArts Studio instance. To use DataArts Studio, you must first buy a basic package. The components and specifications of the instances vary depending on the DataArts Studio version you purchase. For details, see **How Do I Select a DataArts Studio Version?**
- If the basic package cannot meet your requirements, you can buy an incremental package. DataArts Studio provides the following incremental packages:
  - CDM incremental package: CDM can migrate data to the cloud and integrate data into the data lake. It provides wizard-based configuration and management and can integrate data from a single table or an entire database incrementally or periodically. The DataArts Studio basic package contains a CDM cluster. If the cluster cannot meet your requirements, you can buy a CDM incremental package in addition to a DataArts Studio instance of the basic version.

– DataArts DataService Exclusive incremental package: DataArts DataService is a platform where you can develop, test, and deploy your data services. DataArts DataService ensures agile response to data service needs, easier data retrieval, better experience for data consumers, higher efficiency, and better monetization of data assets. To use DataArts DataService Exclusive, you need to purchase an exclusive cluster incremental package in addition to the DataArts Studio basic package.

📖 NOTE

You can use 10 APIs for free in DataArts DataService Exclusive, and you will be charged CNY 1 every day for each extra API.

The following table lists the DataArts Studio billing modes.

Table 7-1 DataArts Studio billing mode

Package	Category	Billing Mode	Description
DataArts Studio basic package	Starter Basic Advanced Professional Enterprise	Yearly/ Monthly	When using DataArts Studio, you also need to buy other cloud services that DataArts Studio depends on, such as EIPs/bandwidth and OBS. The DataArts Studio basic package will not include dependent services. For example, if a GaussDB(DWS) data connection, database, and table are created during data development, you will be billed separately for GaussDB(DWS) based on GaussDB(DWS) pricing.
DataArts Studio incremental package (optional)	DataArts Migration incremental package	● Pay-per-use ● Package (Yearly/ Monthly)	The package is billed based on actual fees. You can use 10 APIs for free in DataArts DataService Exclusive, and you will be charged CNY 1 every day for each extra API.
	DataArts DataService incremental package	Package (Yearly/ Monthly)	The package is billed based on actual fees.

 NOTE

> Pay attention to the following aspects when purchasing a DataArts Migration incremental package:
>
> - If you buy a pay-per-use incremental package, the system automatically creates a CDM cluster based on the specifications you select.
> - If you buy a yearly/monthly incremental package, the system will not automatically create a CDM cluster. If you only have this package, the quota in the package will not be deducted because no CDM cluster exists.
> - If you want to enjoy the preferential price of the yearly/monthly incremental package, you can buy a yearly/monthly incremental package and then buy a pay-per-use incremental package which is in the same region and has the same specifications as the yearly/monthly incremental package.
> - If you buy a pay-per-use incremental package and then buy a yearly/monthly incremental package which is in the same region and has the same specifications as the pay-per-use incremental package, you will be billed in pay-per-use mode first, and then in yearly/monthly mode.

Billing Mode

- **DataArts Studio basic package**

  The DataArts Studio basic package is billed on a monthly/yearly basis. You can select a DataArts Studio basic package based on your service requirements and pay for the package based on the subscription duration. The minimum subscription duration is one month.

- **DataArts Studio incremental package**

  - **Pay per use**: Incremental package is billed every hour based on the actual usage duration.

  - **Yearly/Monthly**: You can select a DataArts Studio incremental package based on your service requirements and pay for the package based on the subscription duration. The minimum subscription duration is one month.

Changing Specifications

DataArts Studio provides multiple specifications in its basic or incremental package. You can select a specification when you buy a package. You can also change the specifications after buying a package in the following ways:

- **DataArts Studio basic package upgrade**: You can upgrade the basic package. During the upgrade, you will be billed daily. After the upgrade is complete, you will be billed based on the new package.

  To upgrade a package, log in to the DataArts Studio console, locate the target DataArts Studio instance, click **Upgrade**, and buy a package with higher specifications.

  After the package is upgraded, the system creates a DataArts Migration cluster. The DataArts Migration cluster in the original basic package will be reserved, but you will not be billed for it. You must manually migrate data connections and jobs from the original CDM cluster to the new one. For details, see **Batch Job Management**.

If these methods do not meet your requirements, you can subscribe to a new DataArts Studio basic or incremental package.

Renewal

- **DataArts Studio basic package**

  DataArts Studio basic packages are billed in yearly/monthly mode. When a package expires, there will be a grace period and a retention period (see **Expiration and Overdue Payment** for details). You are advised to renew the package before it expires.

  The DataArts Studio basic package supports automatic renewal. The default renewal period is as follows:

  – Your package will be renewed each month for monthly billing.

  – Your package will be renewed each year for yearly billing.

  You can enable automatic renewal in either of the following ways:

  – Log in to the management console. On the page for buying DataArts Studio instances, select **Auto-renew**.

  – Go to the **Renewals** page, select the DataArts Studio instance to be renewed in the list, click **Enable Auto-Renew**, and do as prompted.

- **DataArts Studio incremental package**

  – **Pay-per-use DataArts Studio incremental package**

    Pay-per-use billing is calculated by hour. If your account is in arrears, the grace period and retention period will start in sequence (see **Expiration and Overdue Payment** for details). Top up your account to ensure sufficient account balance.

  – **Yearly/Monthly DataArts Studio incremental package**

    When a yearly/monthly DataArts Studio incremental package expires, its billing mode will automatically change to pay-per-use. If you want to continue to use the package, buy it again before it expires.

Expiration and Overdue Payment

- **DataArts Studio basic package**

  DataArts Studio basic packages are billed in yearly/monthly mode. When a package expires, it enters a grace period. During this period, you can access and use DataArts Studio, and the billing is normal. You need to extend the package duration in the grace period. Otherwise, the package will enter the retention period.

  During the retention period, data will be retained, but you cannot access the DataArts Studio instance, perform operations on the DataArts Studio console, or call related APIs. If you do not have sufficient balance in your account or do not renew the resource package before the retention period expires, DataArts Studio will become unavailable and data stored in DataArts Studio will be deleted and cannot be recovered.

- **DataArts Studio incremental package**

  – **Pay-per-use DataArts Studio incremental package**

    Pay-per-use billing is calculated by hour. You can use packages as long as your account balance is sufficient. If your account balance is insufficient to pay the fees for the previous hour, the DataArts Studio incremental package will enter the grace period. During this period, you can access and use the DataArts Studio incremental package, and resources in this

grace period are still charged. You need to top up your account within the grace period. Otherwise, the package will enter the retention period.

During the retention period, data is retained, but you cannot access the DataArts Studio incremental package, perform related operations, or call APIs. If you do not have sufficient balance in your account or do not renew the resource package before the retention period expires, DataArts Studio will become unavailable and data stored in DataArts Studio will be deleted and cannot be recovered.

– **Yearly/Monthly DataArts Studio incremental package**

When the validity period of a package ends, the package's billing mode will automatically change to pay-per-use.

Unsubscription

- **DataArts Studio basic package**

During the validity period of a yearly/monthly DataArts Studio basic package, you can unsubscribe from the package by referring to **Unsubscriptions**.

After unsubscribing from DataArts Studio, you must unsubscribe from the resources of the following services that may incur fees:

– OBS: During batch data migration or data development, DataArts Studio may write dirty data or logs to OBS, which incurs data storage fees.

– EIP: If you use an EIP for your DataArts Migration cluster or DataArts DataService Exclusive cluster, you need to pay for the EIP.

– SMN: If you enable SMN notifications for your DataArts Studio modules, you need to pay for the notifications.

– DEW: If you enable KMS when creating a link in DataArts Migration or creating a connection in Management Center, you will be billed for key management.

- **DataArts Studio incremental package**

– **Pay-per-use DataArts Studio incremental package**

After a DataArts Migration cluster is deleted, the pay-per-use DataArts Migration incremental package will no longer incur fees.

– **Yearly/Monthly DataArts Studio incremental package**

After deleting a DataArts Migration cluster or DataArts DataService Exclusive cluster, you must unsubscribe from the corresponding package by referring to **Unsubscriptions**.

8 DataArts Studio Permissions Management

If you need to assign different permissions to employees in your enterprise to access your DataArts Studio resources, IAM is a good choice for fine-grained permissions management. IAM provides identity authentication, permissions management, and access control, helping you secure access to your resources.

With IAM, you can use your account to create IAM users for your employees, and assign permissions to the users to control their access to specific resource types. For example, if you want to allow some software developers in your enterprise to use DataArts Studio resources but disallow them to delete workspaces or perform any high-risk operations, you can create IAM users for the software developers and grant them only the permissions required for using DataArts Studio resources.

DataArts Studio Permissions

By default, new IAM users do not have any permissions. To assign permissions to a user, add the user to one or more groups and assign permissions policies or roles to these groups. The user then inherits permissions from the groups it is a member of After authorization, the users can perform specified operations.

DataArts Studio is a project-level service deployed in specific physical regions. To assign ServiceStage permissions to a user group, specify the scope as region-specific projects and select projects for the permissions to take effect. If **All projects** is selected, the permissions will take effect for the user group in all region-specific projects. When accessing DataArts Studio, users need to switch to a region where they are authorized to use cloud services.

- **IAM Roles**: IAM initially provides a coarse-grained authorization mechanism to define permissions based on users' job responsibilities. This mechanism provides only a limited number of service-level roles for authorization. However, IAM roles are not an ideal choice for fine-grained authorization and secure access control.

  DataArts Studio provides more flexible, fine-grained authorization based on IAM system roles and workspace roles for specific operations.

  The system-defined roles supported by DataArts Studio include DAYU Administrator and DAYU User. Workspace roles are based on the DAYU User. **DataArts Studio Permissions** lists the common operations supported by

DataArts Studio and the permissions granted to each role. You can select roles as required.

Table 8-1 DataArts Studio system-defined roles

Role	Description	Category
DAYU Administrator	User who has all permissions of DataArts Studio and workspaces. Only DAYU Administrator has the permission to configure default items of DataArts Factory (including the periodic scheduling, multi-IF policy, and hard and soft lock policy). DAYU User does not have this permission. **NOTE** Users assigned the **Tenant Administrator** role have all permissions for all services except IAM. In other words, users with the **Tenant Administrator** role can perform all operations in DataArts Studio.	System-defined role

Role	Description	Category
DAYU User	Common DataArts Studio user Users with the DAYU User policy have the permissions of the role assigned to them in a workspace. Workspace roles include the preset admin, developer, operator, and viewer, and other custom roles. For details about the operation permissions of each role, see **DataArts Studio Permissions**. ● Admin: Users with this role have the permissions to perform all operations in a workspace. You are advised to assign this role to the project owner, development owner, and O&M administrator. ● Developer: Users with this role have the permissions to create and manage work items, but cannot perform operations on workspaces, clusters, and reviewers. You are advised to assign this tole to users who develop and process tasks. ● Operator: Users with this role have the permissions to perform operations such as O&M and scheduling, but cannot modify work items or configurations. You are advised to assign this role to users for O&M management and status monitoring. ● Viewer: Users with this role can only read data from DataArts Studio, but cannot perform operations on workspaces or modify work items or configurations. You are advised to assign this role to users who only want to view information in the workspace but do not perform any operation. **NOTE** Different from DAYU Administrator or Tenant Administrator, the admin does not have permissions to buy DataArts Studio instances or DataArts Studio incremental packages, or create and manage workspaces.	System-defined role

Reference

- **IAM Service Overview**

- **Creating an IAM User and Assigning DataArts Studio Permissions**

9 DataArts Studio Permissions

Five preset roles are available for workspace members: admin, developer, deployer, operator, and viewer.

- Admin: Users with this role have the permissions to perform all operations in a workspace. You are advised to assign this role to the project owner, development owner, and O&M administrator.

- Developer: Users with this role have the permissions to create and manage work items, but cannot perform operations on workspaces, clusters, and reviewers. You are advised to assign this tole to users who develop and process tasks.

- Operator: Users with this role have the permissions to perform operations such as O&M and scheduling, but cannot modify work items or configurations. You are advised to assign this role to users for O&M management and status monitoring.

- Viewer: Users with this role can only read data from DataArts Studio, but cannot perform operations on workspaces or modify work items or configurations. You are advised to assign this role to users who only want to view information in the workspace but do not perform any operation.

This section describes the permissions of the preset roles.

NOTICE

Operation permissions in this section refer to the permissions required for performing resource operations except addition, deletion, modification, and query, such as importing and exporting data, and executing, canceling, starting, and scheduling tasks.

Management Center

Permission	Admin	Developer	Operator	Viewer
Querying the MRS, DWS, or CDM cluster list	Y	Y	Y	Y

Permission	Admin	Developer	Operator	Viewer
Creating databases	Y	Y	N	N
Deleting databases	Y	Y	N	N
Querying databases	Y	Y	Y	Y
Modifying databases	Y	Y	N	N
Creating data tables	Y	Y	N	N
Deleting data tables	Y	Y	N	N
Querying data tables	Y	Y	Y	Y
Editing data tables	Y	Y	N	N
Creating resource migration tasks	Y	Y	N	N
Operating resource migration tasks	Y	Y	Y	N
Querying resource migration tasks	Y	Y	Y	Y
Creating data connections	Y	Y	N	N
Deleting data connections	Y	Y	N	N
Operating data connections	Y	Y	Y	N
Querying data connections	Y	Y	Y	Y
Editing data connections	Y	Y	N	N
Deleting RDS driver packages	Y	N	N	N
Operating RDS driver packages	Y	N	N	N

Permission	Admin	Developer	Operator	Viewer
Querying RDS driver packages	Y	Y	Y	Y
Creating DLI resource mapping configurations	N	N	N	N
Deleting DLI resource mapping configurations	N	N	N	N
Querying DLI resource mapping configurations	N	N	N	N

DataArts Architecture

Permission	Admin	Developer	Operator	Viewer
Creating atomic metrics	Y	Y	N	N
Deleting atomic metrics	Y	Y	N	N
Querying atomic metrics	Y	Y	Y	Y
Editing atomic metrics	Y	Y	N	N
Creating logical entities or physical tables	Y	Y	N	N
Deleting logical entities or physical tables	Y	Y	N	N
Querying logical entities or physical tables	Y	Y	Y	Y
Editing logical entities or physical tables	Y	Y	N	N

Permission	Admin	Developer	Operator	Viewer
Creating configuration centers	Y	N	N	N
Deleting configuration centers	Y	N	N	N
Querying configuration centers	Y	Y	Y	Y
Editing configuration centers	Y	N	N	N
Creating subject designs	Y	Y	N	N
Deleting subject designs	Y	Y	N	N
Querying subject designs	Y	Y	Y	Y
Editing subject designs	Y	Y	N	N
Creating business metrics	Y	Y	N	N
Deleting business metrics	Y	Y	N	N
Querying business metrics	Y	Y	Y	Y
Editing business metrics	Y	Y	N	N
Creating summary tables	Y	Y	N	N
Deleting summary tables	Y	Y	N	N
Querying summary tables	Y	Y	Y	Y
Editing summary tables	Y	Y	N	N
Creating general configurations	Y	Y	N	N

Permission	Admin	Developer	Operator	Viewer
Deleting general configurations (deleting the drafts of published logical entities or tables)	Y	Y	N	N
Operating general configurations (importing, exporting, publishing, suspending, synchronizing, and reversing logical entities or tables)	Y	Y	Y	N
Querying general configurations (querying the drafts of published logical entities or tables)	Y	Y	Y	Y
Editing general configurations (editing the drafts of published logical entities or tables)	Y	Y	N	N
Deleting dimension tables	Y	Y	N	N
Querying dimension tables	Y	Y	Y	Y
Creating process designs	Y	Y	N	N
Deleting process designs	Y	Y	N	N
Querying process designs	Y	Y	Y	Y
Editing process designs	Y	Y	N	N

Permission	Admin	Developer	Operator	Viewer
Creating lookup tables	Y	Y	N	N
Deleting lookup tables	Y	Y	N	N
Querying lookup tables	Y	Y	Y	Y
Editing lookup tables	Y	Y	N	N
Creating models	Y	Y	N	N
Deleting models	Y	Y	N	N
Querying models	Y	Y	Y	Y
Editing models	Y	Y	N	N
Creating derivative or compound metrics	Y	Y	N	N
Deleting derivative or compound metrics	Y	Y	N	N
Operating derivative or compound metrics	Y	Y	N	N
Querying derivative or compound metrics	Y	Y	Y	Y
Editing derivative or compound metrics	Y	Y	N	N
Creating associated quality rules	Y	Y	N	N
Deleting associated quality rules	Y	Y	N	N
Querying associated quality rules	Y	Y	Y	Y

Permission	Admin	Developer	Operator	Viewer
Editing associated quality rules	Y	Y	N	N
Creating fact tables	Y	Y	N	N
Deleting fact tables	Y	Y	N	N
Querying fact tables	Y	Y	Y	Y
Editing fact tables	Y	Y	N	N
Creating directories	Y	Y	N	N
Deleting directories	Y	Y	N	N
Querying directories	Y	Y	Y	Y
Editing directories	Y	Y	N	N
Creating dimensions	Y	Y	N	N
Deleting dimensions	Y	Y	N	N
Querying dimensions	Y	Y	Y	Y
Editing dimensions	Y	Y	N	N
Creating time filters	Y	Y	N	N
Deleting time filters	Y	Y	N	N
Querying time filters	Y	Y	Y	Y
Editing time filters	Y	Y	N	N
Creating data standards	Y	Y	N	N

Permission	Admin	Developer	Operator	Viewer
Deleting data standards	Y	Y	N	N
Querying data standards	Y	Y	Y	Y
Editing data standards	Y	Y	N	N

DataArts Migration

Permission	Admin	Developer	Operator	Viewer
Creating clusters	Y	Y	N	N
Deleting clusters	Y	Y	N	N
Operating clusters	Y	Y	Y	N
Querying clusters	Y	Y	Y	Y
Editing clusters	Y	Y	N	N
Operating links	Y	Y	Y	N
Querying links	N	N	N	N
Operating jobs	Y	Y	Y	N
Querying jobs	N	N	N	N

DataArts Factory

Permission	Admin	Developer	Deployer	Operator	Viewer
Creating schemas	Y	Y	N	N	N
Deleting schemas	Y	Y	N	N	N
Querying schemas	Y	Y	N	Y	Y
Editing schemas	Y	Y	N	N	N
Operating backups	Y	Y	N	Y	N

Permission	Admin	Developer	Deployer	Operator	Viewer
Querying backups	Y	Y	N	Y	Y
Creating PatchData tasks	Y	Y	N	N	N
Operating PatchData tasks	Y	Y	N	Y	N
Querying PatchData tasks	Y	Y	N	Y	Y
Operating dirty data	Y	Y	N	Y	N
Operating backups used for restoration	Y	N	N	Y	N
Querying backups used for restoration	Y	Y	N	Y	Y
Creating directories	Y	Y	N	N	N
Deleting directories	Y	Y	N	N	N
Querying directories	Y	Y	N	Y	Y
Editing directories	Y	Y	N	N	N
Creating notifications	Y	Y	N	N	N
Deleting notifications	Y	Y	N	N	N
Querying notifications	Y	Y	N	Y	Y
Editing notifications	Y	Y	N	N	N
Creating databases	Y	Y	N	N	N
Deleting databases	Y	Y	N	N	N
Querying databases	Y	Y	N	Y	Y

Permission	Admin	Developer	Deployer	Operator	Viewer
Editing databases	Y	Y	N	N	N
Creating solutions	Y	Y	N	N	N
Deleting solutions	Y	Y	N	N	N
Operating solutions	Y	Y	N	Y	N
Querying solutions	Y	Y	N	Y	Y
Editing solutions	Y	Y	N	N	N
Querying IAM agencies	Y	Y	Y	Y	Y
Updating IAM agencies	Y	N	N	N	N
Operating environment variables	Y	Y	N	N	N
Querying environment variables	Y	Y	N	Y	Y
Editing environment variables	Y	Y	N	N	N
Operating job nodes	Y	Y	N	Y	N
Viewing release packages	Y	Y	Y	Y	Y
Operating release packages	Y	N	Y	Y	N
Creating data connections	Y	Y	N	N	N
Deleting data connections	Y	Y	N	N	N
Operating data links	Y	Y	N	Y	N

Permission	Admin	Developer	Deployer	Operator	Viewer
Querying data connections	Y	Y	N	Y	Y
Editing data connections	Y	Y	N	N	N
Canceling release	Y	Y	Y	Y	N
Creating data tables	Y	Y	N	N	N
Deleting data tables	Y	Y	N	N	N
Querying data tables	Y	Y	N	Y	Y
Editing data tables	Y	Y	N	N	N
Operating job instances	Y	Y	N	Y	N
Querying job instances	Y	Y	N	Y	Y
Creating resources	Y	Y	N	N	N
Deleting resources	Y	Y	N	N	N
Operating resources	Y	Y	N	Y	N
Querying resources	Y	Y	N	Y	Y
Editing resources	Y	Y	N	N	N
Editing environment variable mappings	N	N	N	N	N
Operating script editing locks	Y	Y	N	Y	N
Creating scripts	Y	Y	N	N	N
Deleting scripts	Y	Y	N	N	N

Permission	Admin	Developer	Deployer	Operator	Viewer
Operating scripts	Y	Y	N	Y	N
Querying scripts	Y	Y	N	Y	Y
Editing scripts	Y	Y	N	N	N
Adding job tags	Y	Y	N	Y	N
Deleting job tags	Y	Y	N	Y	N
Querying job tags	Y	Y	N	Y	Y
Creating jobs	Y	Y	N	N	N
Deleting jobs	Y	Y	N	N	N
Operating jobs	Y	Y	N	Y	N
Querying jobs	Y	Y	N	Y	Y
Editing jobs	Y	Y	N	Y	N
Querying details about job editing locks	Y	Y	N	Y	Y
Operating job editing locks	Y	Y	N	Y	N
Creating baselines	Y	N	N	Y	N
Querying baselines	Y	Y	N	Y	N
Deleting baselines	Y	N	N	Y	N
Modifying baselines	Y	N	N	Y	N
Querying baseline instances	Y	Y	N	Y	N
Obtaining the list of summary information about assurance jobs	Y	Y	N	Y	N
Querying events	Y	Y	N	Y	N

Permission	Admin	Developer	Deployer	Operator	Viewer
Updating events	Y	N	N	Y	N

DataArts Quality

Permission	Admin	Developer	Operator	Viewer
Data quality monitoring				
Querying the dashboard	Y	Y	Y	Y
Operating instances	Y	Y	Y	N
Querying instances	Y	Y	Y	Y
Creating rule templates	Y	N	N	N
Deleting rule templates	Y	N	N	N
Operating rule templates	Y	N	N	N
Querying rule templates	Y	Y	Y	Y
Editing rule templates	Y	N	N	N
Querying the execution result	Y	Y	N	N
Creating rules	Y	Y	N	N
Deleting rules	Y	Y	N	N
Operating rules	Y	Y	Y	N
Querying rules	Y	Y	Y	Y
Editing rules	Y	Y	N	N
Editing quality scores	Y	N	N	N
Creating directories	Y	Y	N	N

Permission	Admin	Developer	Operator	Viewer
Deleting directories	Y	Y	N	N
Querying directories	Y	Y	Y	Y
Editing directories	Y	Y	N	N
Business metric monitoring				
Querying the dashboard	Y	Y	Y	Y
Operating instances	Y	Y	Y	N
Querying instances	Y	Y	Y	Y
Creating scenarios	Y	Y	N	N
Deleting scenarios	Y	Y	N	N
Operating scenarios	Y	Y	Y	N
Querying scenarios	Y	Y	Y	Y
Editing scenarios	Y	Y	N	N
Creating metrics	Y	Y	N	N
Deleting metrics	Y	Y	N	N
Querying metrics	Y	Y	Y	Y
Editing metrics	Y	Y	N	N
Creating rules	Y	Y	N	N
Deleting rules	Y	Y	N	N
Querying rules	Y	Y	Y	Y
Editing rules	Y	Y	N	N
Creating directories	Y	Y	N	N
Deleting directories	Y	Y	N	N

Permission	Admin	Developer	Operator	Viewer
Querying directories	Y	Y	Y	Y
Editing directories	Y	Y	N	N

DataArts Catalog

Permission	Admin	Developer	Operator	Viewer
Querying data sources	Y	Y	Y	N
Operating task instances	Y	Y	Y	N
Querying task instances	Y	Y	Y	Y
Creating collection tasks	Y	Y	N	N
Deleting collection tasks	Y	Y	N	N
Operating collection tasks	Y	Y	Y	N
Querying collection tasks	Y	Y	Y	Y
Editing collection tasks	Y	Y	N	N
Editing approvals	Y	Y	N	N
Editing asset reports	Y	Y	N	N
Creating tags	Y	Y	N	N
Deleting tags	Y	Y	N	N
Querying tags	Y	Y	Y	Y
Editing tags	Y	Y	N	N
Creating assets	Y	Y	N	N
Deleting assets	Y	Y	N	N
Operating assets	Y	Y	Y	N
Querying assets	Y	Y	Y	Y

Permission	Admin	Developer	Operator	Viewer
Editing assets	Y	Y	N	N
Creating directories	Y	Y	N	N
Deleting directories	Y	Y	N	N
Querying directories	Y	Y	Y	Y
Editing directories	Y	Y	N	N
Creating classifications	Y	Y	N	N
Deleting classifications	Y	Y	N	N
Querying classifications	Y	Y	Y	Y
Editing classifications	Y	Y	N	N
Creating data permission rules	Y	N	N	N
Deleting data permission rules	Y	N	N	N
Querying data permission rules	Y	Y	Y	Y
Editing data permission rules	Y	N	N	N

DataArts DataService

Permission	Admin	Developer	Operator	Viewer
Creating throttling policies	Y	Y	N	N
Deleting throttling policies	Y	Y	N	N
Operating throttling policies	Y	Y	Y	N
Querying throttling policies	Y	Y	Y	Y

Permission	Admin	Developer	Operator	Viewer
Editing throttling policies	Y	Y	N	N
Creating applications	Y	Y	N	N
Deleting applications	Y	Y	N	N
Operating applications	Y	Y	Y	N
Querying applications	Y	Y	Y	Y
Editing applications	Y	Y	N	N
Operating reviews	Y	Y	Y	N
Querying reviews	Y	Y	Y	Y
Creating API catalogs	Y	Y	Y	N
Deleting API catalogs	Y	Y	Y	N
Querying API catalogs	Y	Y	Y	Y
Editing API catalogs	Y	Y	Y	N
Operating clusters	Y	Y	N	N
Querying clusters	Y	Y	Y	Y
Adding reviewers	Y	N	N	N
Deleting reviewers	Y	N	N	N
Operating reviewers	Y	Y	Y	N
Querying reviewers	Y	Y	Y	N
Creating APIs	Y	Y	N	N
Deleting APIs	Y	Y	N	N
Operating APIs	Y	Y	Y	N

Permission	Admin	Developer	Operator	Viewer
Querying APIs	Y	Y	Y	Y
Editing APIs	Y	Y	N	N
Querying data sources	Y	Y	N	N
Querying the dashboard	Y	Y	Y	Y

DataArts Security

Permission	Admin	Developer	Operator	Viewer
Querying the dashboard	Y	Y	Y	Y
Creating data source tracing tasks	Y	Y	N	N
Deleting data source tracing tasks	Y	Y	N	N
Operating data source tracing tasks	Y	Y	N	N
Querying data source tracing tasks	Y	Y	Y	Y
Editing data source tracing tasks	Y	Y	N	N
Operating security task scheduling	Y	Y	N	Y
Creating sensitive data discovery tasks	Y	Y	Y	N
Deleting sensitive data discovery tasks	Y	Y	N	N
Operating sensitive data discovery tasks	Y	Y	Y	N

Permission	Admin	Developer	Operator	Viewer
Querying sensitive data discovery tasks	Y	Y	Y	Y
Editing sensitive data discovery tasks	Y	Y	N	N
Querying data sources	Y	Y	Y	Y
Creating access permissions management tasks	Y	Y	N	N
Deleting access permissions management tasks	Y	Y	N	N
Querying access permissions management tasks	Y	Y	Y	Y
Editing access permissions management tasks	Y	Y	N	N
Querying resource permission configuration	Y	Y	N	N
Creating data masking policies	Y	Y	N	N
Deleting data masking policies	Y	Y	N	N
Operating data masking policies	Y	Y	Y	Y
Querying data masking policies	Y	Y	Y	Y
Editing data masking policies	Y	Y	N	N
Creating security levels	Y	Y	N	N

Permission	Admin	Developer	Operator	Viewer
Deleting security levels	Y	Y	N	N
Querying security levels	Y	Y	Y	Y
Editing security levels	Y	Y	N	N
Creating rule groups	Y	Y	Y	N
Deleting rule groups	Y	Y	N	N
Operating rule groups	Y	Y	Y	N
Querying rule groups	Y	Y	Y	Y
Editing rule groups	Y	Y	Y	N
Creating data masking tasks	Y	Y	N	N
Deleting data masking tasks	Y	Y	N	N
Operating data masking tasks	Y	Y	N	N
Querying data masking tasks	Y	Y	Y	Y
Editing data masking tasks	Y	Y	N	N
Creating data watermarking tasks	Y	Y	N	N
Deleting data watermarking tasks	Y	Y	N	N
Operating data watermarking tasks	Y	Y	N	N
Querying data watermarking tasks	Y	Y	Y	Y

Permission	Admin	Developer	Operator	Viewer
Editing data watermarking tasks	Y	Y	N	N

10 Constraints and Restrictions

Browser Restrictions

The following table lists the recommended browser for logging in to DataArts Studio.

Table 10-1 Browser compatibility

Browser	Recommended Version	Recommended OS	Remarks
Google Chrome	115, 114, and 113	Windows 10	The resolution ranges from 1366x768 px to 1920x1080 px. 1920x1080 px is the optimal resolution for the best display of the console.

Use Restrictions

Before using DataArts Studio, you must read and understand the following restrictions:

Table 10-2 Restrictions for using DataArts Studio

Compo nent	Restriction
Public	1. DataArts Studio must be deployed based on HUAWEI CLOUD. If resources are isolated, DataArts Studio can be deployed in a full-stack DeC. In addition, DataArts Studio can be deployed on Huawei Cloud Stack or Huawei Cloud Stack Online. 2. DataArts Studio is a one-stop platform that provides data integration, development, and governance capabilities. DataArts Studio has no storage or computing capability and relies on the data lake base. 3. Only one DataArts Studio instance can be bound to an enterprise project. If an enterprise project already has an instance, no more instance can be added. 4. Different components of DataArts Studio support different data sources. You need to select a data lake foundation based on your service requirements. For details about the data lakes supported by DataArts Studio, see **Data Sources Supported by DataArts Studio**.
Manage ment Center	1. Due to the constraints of Management Center, other components (such as DataArts Architecture, DataArts Quality, and DataArts Catalog) do not support databases or tables whose names contain Chinese characters or periods (.). 2. The free CDM cluster provided by a DataArts Studio instance has limited specifications. You are advised to use it only as an agent for a data connection in Management Center. 3. You are advised to use different CDM clusters for a data connection agent in Management Center and a CDM migration job. If an agent and CDM job use the same cluster, they may contend for resources during peak hours, resulting in service unavailability. 4. If a CDM cluster functions as the agent for a data connection in Management Center, the cluster cannot connect to multiple MRS security clusters. You are advised to plan multiple agents which are mapped to MRS security clusters one by one. 5. The number of concurrent active threads of an agent is 200. If multiple data connections share an agent, a maximum of 200 SQL jobs and Shell and Python scripts submitted through the connections can run concurrently. Excess tasks will be queued. You are advised to select different agents for different connections to prevent your tasks from being affected by this constraint. 6. A maximum of 200 data connections can be created in a workspace. 7. The concurrency restriction for APIs in Management Center is 100 QPS.

Compo nent	Restriction
DataArt s Migrati on	1. You can enable automatic backup and restoration of CDM jobs. Backups of CDM jobs are stored in OBS buckets. For details, see **Automatic Backup and Restoration of CDM Jobs**. 2. The DataArts Migration cluster is deployed in standalone mode. A cluster fault may cause service and data loss. You are advised to use the CDM Job node of DataArts Factory to invoke CDM jobs and select two CDM clusters to improve reliability. For details, see **CDM Job**. For more constraints on DataArts Migration, see **CDM Constraints**.
DataArt s Factory	1. You can enable backup of assets such as scripts and jobs to OBS buckets. For details, see **Managing Backups**. 2. A maximum of 10,000 jobs can be created in a workspace. 3. A maximum of 1,000 execution results can be displayed for RDS SQL, DWS SQL, Hive SQL, DLI SQL, and Spark SQL scripts, and the data volume is less than 3 MB. If the number of execution results exceeds 1,000, you can dump them. A maximum of 10,000 execution results can be dumped.

DataArt s Architec ture	1. DataArts Architecture supports ER modeling and dimensional modeling (only star models).
	
	2. The maximum size of a file to be imported is 4 MB. A maximum of 3,000 metrics can be imported. A maximum of 500 tables can be exported at a time.
	
	3. The quotas for the objects in a workspace are as follows:
	  • Subjects: 5,000
	  • Data standard directories: 500; data standards: 20,000
	  • Atomic, derivative, and compound metrics: 5,000 for each
	
	4. The quotas for different custom objects are as follows:
	  • Custom subjects: 10
	  • Custom tables: 10
	  • Custom attributes: 10
	  • Custom business metrics: 50

DataArt s Quality	1. The execution duration of data quality jobs depends on the data engine. If the data engine does not have sufficient resources, the execution of data quality jobs may be slow. 2. A maximum of 50 rules can be configured for a data quality job. If necessary, you can create multiple quality jobs.

Compo nent	Restriction
DataArt s Catalog	1. Metadata collection tasks can be obtained through DDL SQL statements of the engine. You are not advised to collect more than 1,000 tables through a single task. If necessary, you can create multiple collection tasks. In addition, you need to set the scheduling time and frequency properly based on your requirements to avoid heavy access and connection pressure on the engine. The recommended settings are as follows: ● If your service requires a metadata validity period of one day, set the scheduling period to max(one day, one-off collection period). This rule also applies to other scenarios. ● If your service mainly runs in the daytime, set a scheduling time in the night during which the pressure on the data source is minimum. This rule also applies to other scenarios. 2. Only the jobs that are scheduled and executed in DataArts Factory generate data lineages. Tested jobs do not generate data lineages.
DataArt s DataSer vice	1. The shared edition is designed only for development and testing. You are advised to use the exclusive edition which is superior to the shared edition. 2. DataArts DataService clusters are bound to workspaces. After a cluster is created, its specifications cannot be modified, and its edition cannot be upgraded.

11 Related Services

IAM

DataArts Studio uses Identity and Access Management (IAM) for authentication and authorization.

CTS

DataArts Studio uses Cloud Trace Service (CTS) to audit users' non-query operations on the management console to ensure that no invalid or unauthorized operations have been performed. CTS enhances security.

ECS

CDM and DataArts DataService clusters of DataArts Studio consist of Elastic Cloud Servers (ECSs). In addition, DataArts Studio can use host connections to connect to ECSs and run Shell or Python scripts.

VPC

Virtual Private Cloud (VPC) provides isolated network environments for DataArts Studio.

EIP

Elastic IP (EIP) enables DataArts Studio to communicate with the Internet.

OBS

DataArts Studio uses Object Storage Service (OBS) buckets to store logs.

SMN

DataArts Studio uses Simple Message Notification (SMN) to send push notifications based on your subscription requirements, so that you can receive immediate notifications when specific events occur.

Direct Connect

Direct Connect enables DataArts Studio to communicate with third-party data centers.

MRS

MapReduce Service (MRS) can be used as the data lake for DataArts Studio and enables data integration, development, and governance.

GaussDB(DWS)

GaussDB(DWS) can be used as the data lake for DataArts Studio and enables data integration, development, governance, and provisioning.

RDS

Relational Database Service (RDS) provides data sources for DataArts Studio and enables data integration, development, and provisioning.