DataArts Studio

Service Overview

Issue   01
Date   2024-12-05

Contents

1 **DataArts Studio Infographics**

2 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 much data replication. 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 2-1 shows the architecture.

Figure 2-1 Architecture
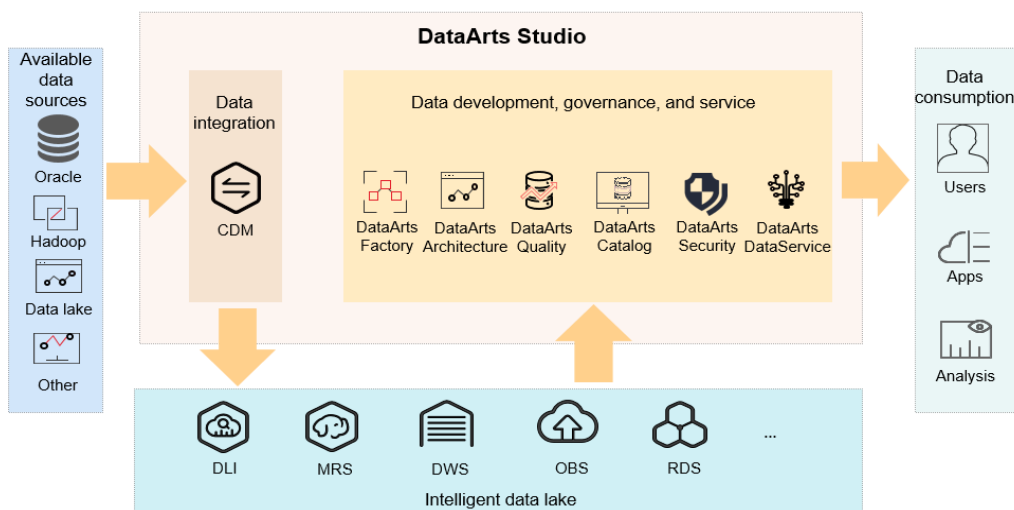
As shown in the figure, DataArts Studio is built on the data foundation and provides capabilities such as data integration, development, and governance. DataArts Studio can connect to Huawei Cloud data lakes and cloud database services, such as MRS Hive and GaussDB(DWS). These data lakes and cloud database services are used as the data foundation. 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 foundation for activities such as data development and 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, ensure data availability, and obtain security certifications.

3 Basic Concepts

Database, Data Warehouse, Data Lake, and Huawei FusionInsight Intelligent Data Lake

For details, see **Database, Data Warehouse, Data Lake, and Huawei FusionInsight Intelligent Data Lake**.

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 a Huawei account 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.

Data Warehouse Planning

DataArts Architecture provides four default data warehouse layers, including SDI, DWI, DWR, and DM. You can customize data warehouse layers. Data warehouse layers and models are centrally managed.

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.

In DataArts Studio, metadata may be used to describe the attributes of data (such as the data connection, type, name, and size) or other related information of data (such as the data owner, tag, category, and security level).

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.

4 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 4-1 DataArts Migration
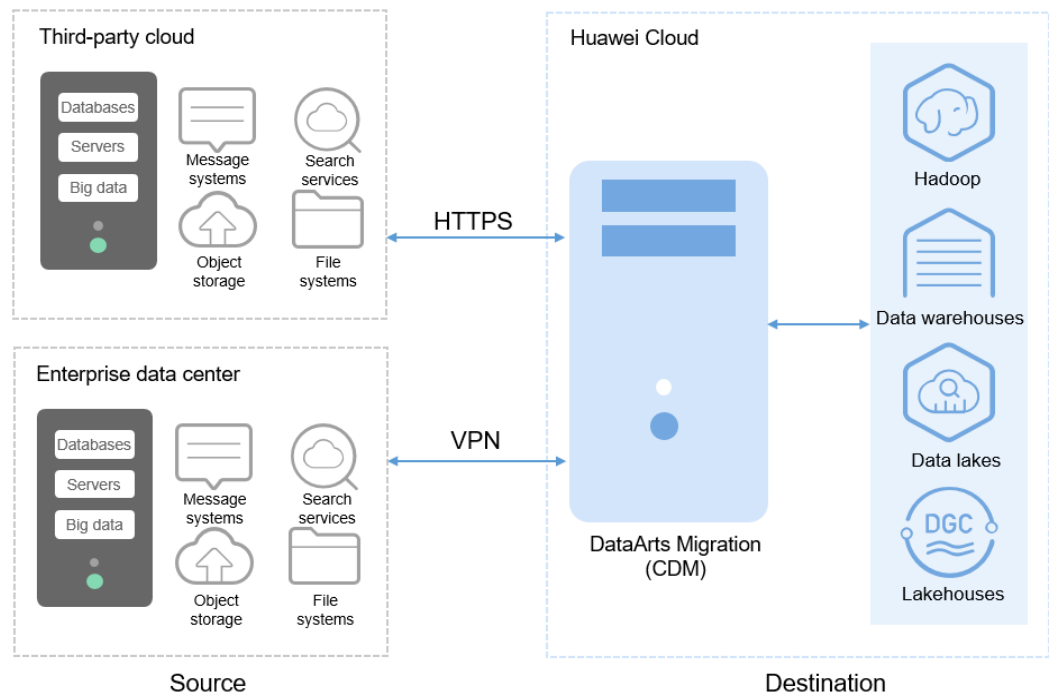
You can manage data on the wizard-based task management page. You can easily create data migration tasks that meet your requirements. DataArts Migration provides the following functions:

- **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 4-2 DataArts Architecture
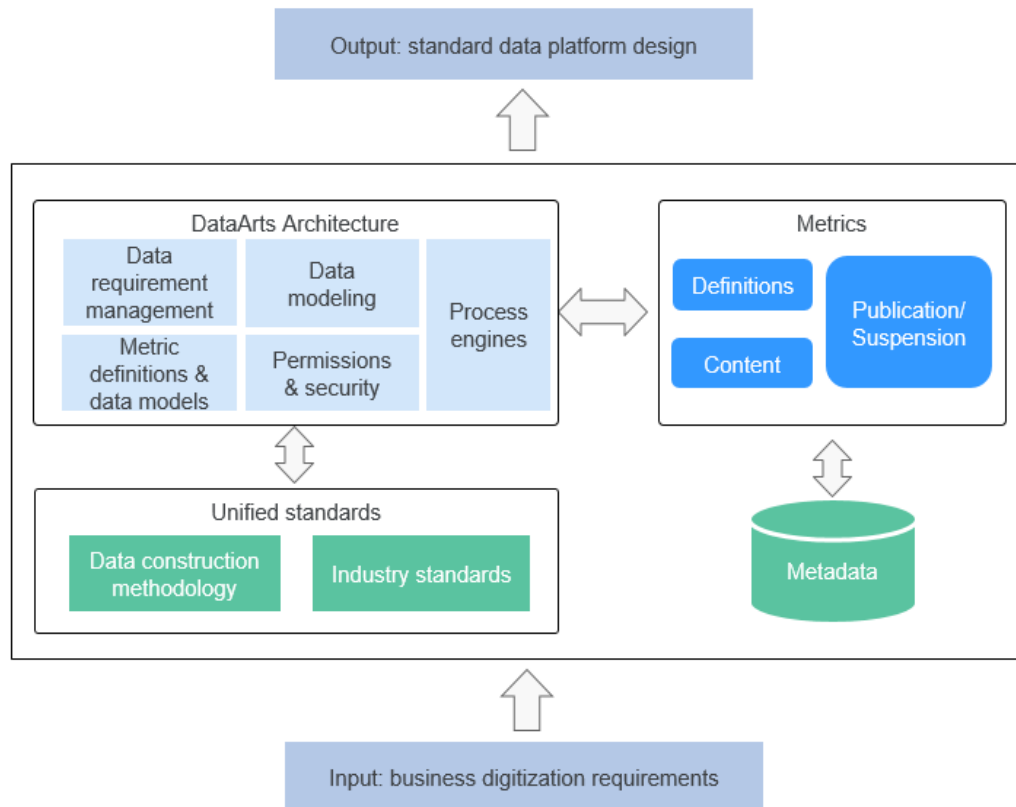
DataArts Architecture offers the following major functions:

- **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.

– **Data mart**

A data mart (DM) aggregates data from multiple layers and consists of a specific analysis object and its related metrics. The DM provides all statistical data by subject.

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 4-3 DataArts Factory architecture
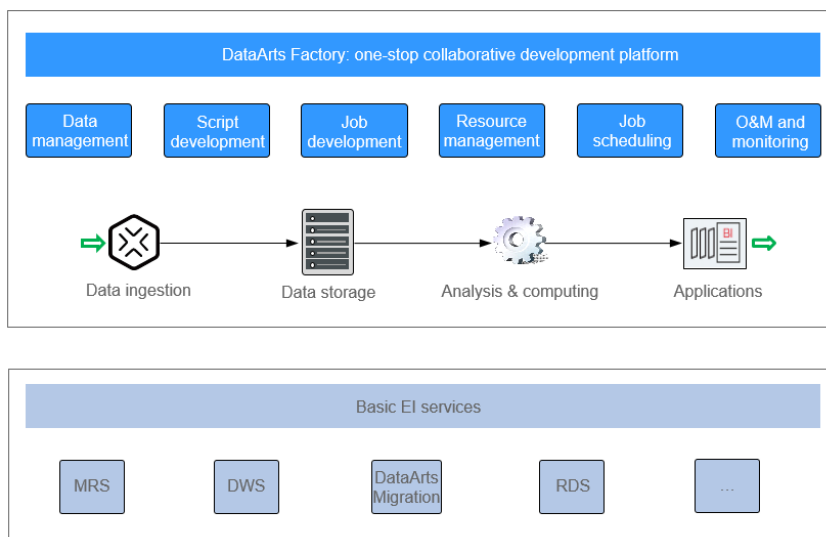
DataArts Factory allows you to manage data, develop scripts, and schedule and monitor jobs. Data analysis and processing are easier than ever before.

● **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 4-4 Data quality rule system
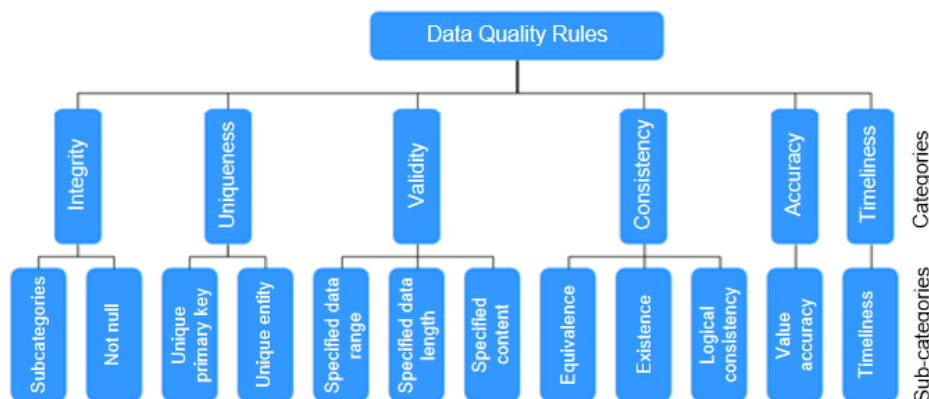
DataArts Catalog: End-to-End Data Asset Visualization

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 4-5 Full-link data lineages
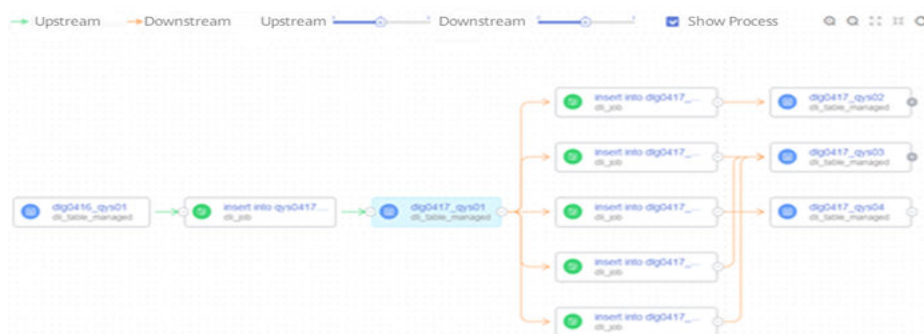
- **Data map**

  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 4-6 DataArts DataService architecture
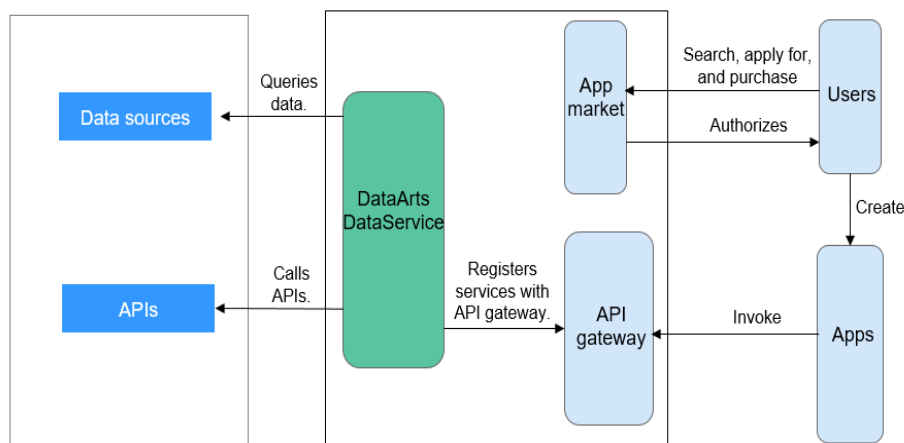
DataArts Security: All-Round Protection

- **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.

5 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.

6 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 6-1 One-stop data operations platform
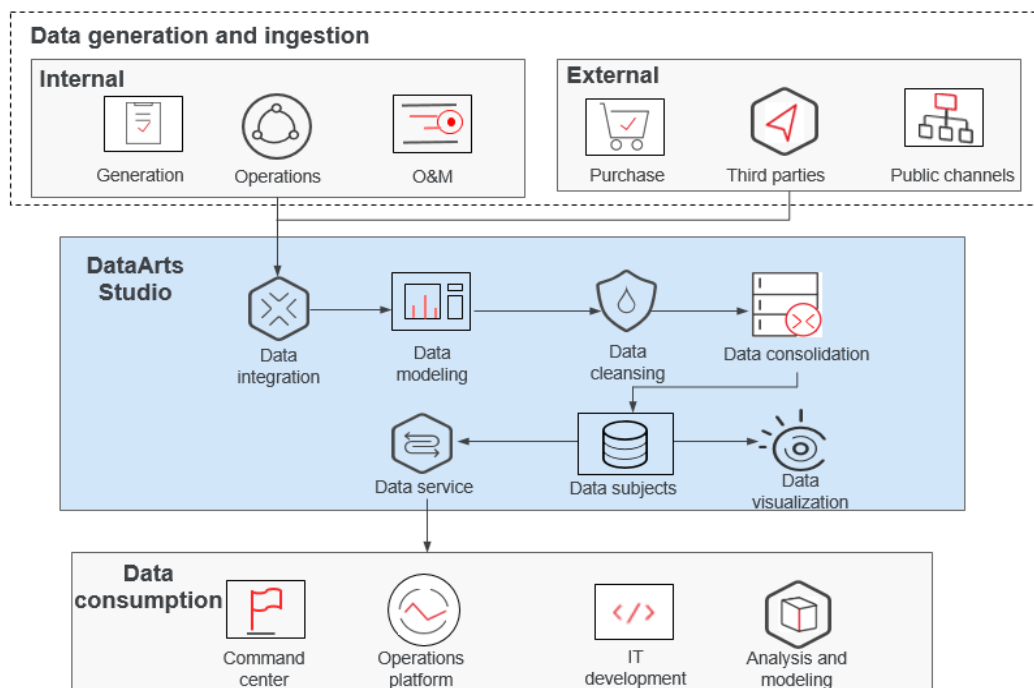
Building Cloud-based Data Platforms with Speed

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 6-2 Cloud data platform
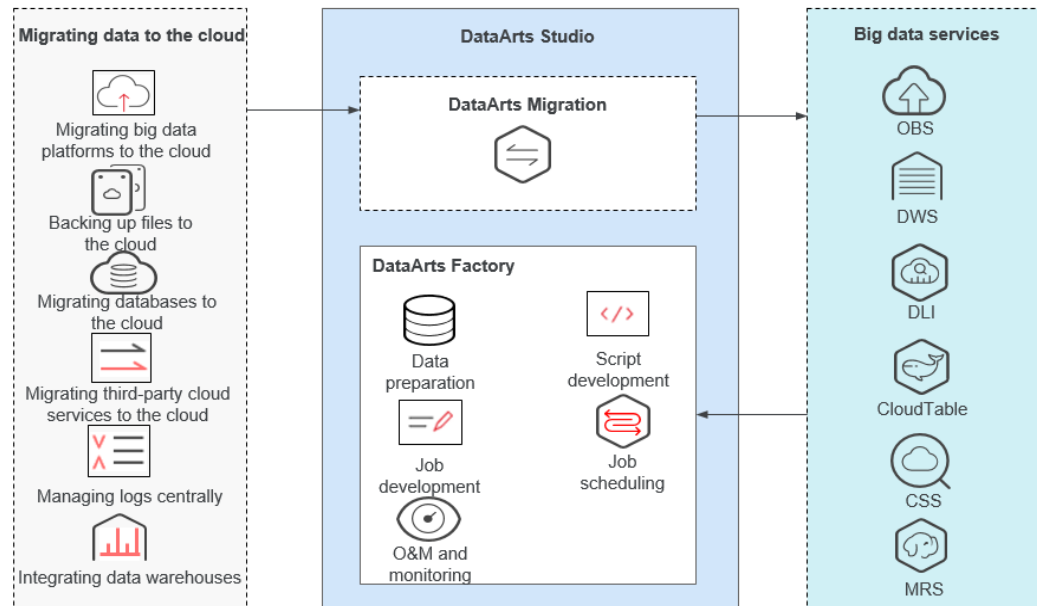
Building Data Lake Governance Platforms Powered by Industry Know-How

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 6-3 Data governance platform
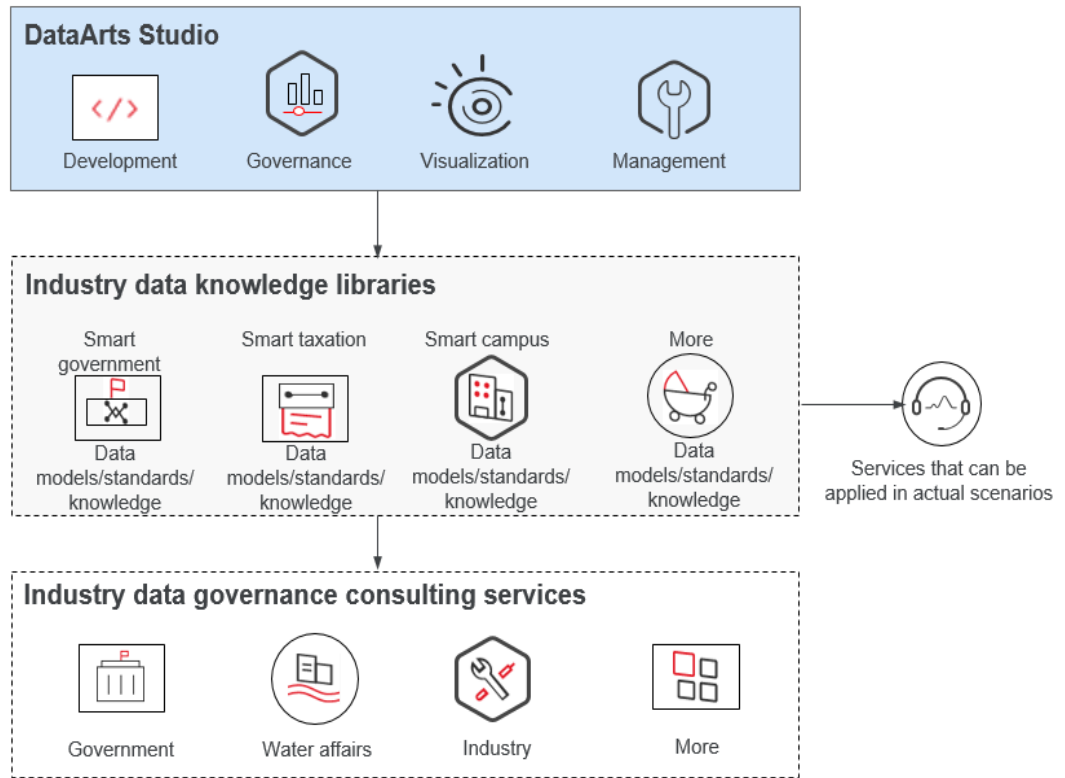
7 Versions

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.

Version Scenarios

The version mode of DataArts Studio has changed in some regions to provide flexible resource configuration and lightweight data governance capabilities. For details about the old and new version modes, see **Version Mode**.

- Now you can purchase DataArts Studio instances of the starter, expert, or enterprise version.

- The version mode change does not affect the DataArts Studio instances you have purchased before, which may be of the starter, basic, advanced, professional, or enterprise version.

Compared with the old version mode, the new version mode provides more favorable prices and more flexible resource scaling. If you want to experience the new version mode, you are advised to buy a new DataArts Studio instance, migrate service data from the original instance to the new instance by referring to **DataArts Studio Data Migration Configuration**, and then unsubscribe from the original instance.

Table 7-1 lists the recommended application scenarios of each version.

Table 7-1 Recommended application scenarios for each DataArts Studio version

Version Mode	Version	Application Scenario
New version mode	Starter	In the initial construction phase, the data lake project mainly manages data ETL tasks in big data development scenarios and does not involve data governance.
	Expert	Small- and medium-sized enterprises (SMEs) have full-time data development and governance personnel and require lightweight data governance capabilities, such as data quality, data assets, and data services. Cost-effectiveness is preferred.
	Enterprise	A complete data management team and system are available. For medium- and large-sized enterprises, the enterprise information architecture, data standards, data models, and data metrics need to be implemented to match the complete DAYU data governance methodology.
Old version mode	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

Version Specifications (New Version Mode)

The new version mode provides the basic, expert, and enterprise version. **Table 7-2** and **Table 7-3** provide the components and specifications of each version, respectively.

Table 7-2 Components supported by DataArts Studio

DataArts Studio Component	Starter	Expert	Enterprise
DataArts Migration	√	√	√

DataArts Studio Component	Starter	Expert	Enterprise
Management Center	√	√	√
DataArts Architecture	x	x	√
DataArts Factory	√	√	√
DataArts Quality	x	This module is available but does not support business metric monitoring, comparison jobs, or quality reports.	√
DataArts Catalog	x	Supported. However, data directories (categories, tags, and collection tasks) cannot be exported through resource migration in the management center.	√
DataArts DataService	x	√	√
DataArts Security	x	This module is available but does not support data watermarking or source tracing.	√

Table 7-3 DataArts Studio version specifications (a single instance)

Specification	Starter	Expert	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.medium vCPUs | memory: 4 vCPUs | 8 GB
Job scheduling times/day[2]	5,000/day	5,000/day	5,000/day

Specification	Starter	Expert	Enterprise
Number of technical assets[3]	Not supported	500	5000
Number of data models[4]	Not supported	Not supported	100

Annotation:

[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 CDM Incremental Package**.

[2] Job scheduling times/day: This specification is calculated based on the total number of scheduling times of data development jobs, quality jobs, reconciliation jobs, service scenarios, and metadata collection jobs executed every day. You can expand the capacity using the job node scheduling times/day incremental package, for details, see **Buying a Job Node Scheduling Times/Day Incremental Package**. 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 total 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 when a batch processing job is scheduled or a real-time job is started.

[3] Number of technical assets: This specification refers to the number of tables and OBS files in the Data Catalog. You can expand the capacity using the incremental package of technology asset quantity. For details about how to purchase the package, see **Buying an Incremental Package for Technical Asset Quantity**.

[4] Number of data models: This specification indicates the number of logical models, physical models, dimension tables, fact tables, and SDR tables in the data architecture. You can expand the capacity using the data model quantity incremental package. For details about how to purchase the package, see **Buying an Incremental Package for Data Model Quantity**.

Version Specifications (Old Version Mode)

Table 7-4 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	x	√	√	√	√

Table 7-5 DataArts Studio version specifications (a single instance)

Specifications	Starter	Basic	Advanced	Professional	Enterprise
DataArts Studio CDM cluster[1]	Number of clusters: 1 Name: cdm.medium vCPUs: 4 | Memory: 8 GB	Number of clusters: 1 Name: cdm.medium vCPUs: 4 | Memory: 8 GB	Number of clusters: 1 Name: cdm.large vCPUs: 8 | Memory: 16 GB	Number of clusters: 1 Name: cdm.xlarge vCPUs: 16 | Memory: 32 GB	Number of clusters: 1 Name: cdm.xlarge vCPUs: 16 | Memory: 32 GB
Job node scheduling times/ day[2]	5,000/day	20,000/day	40,000/day	80,000/day	200,000/day

Specifications	Starter	Basic	Advanced	Professional	Enterprise
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: Due to specifications restrictions, the free CDM cluster provided by a DataArts Studio instance can only be used for informal scenarios such as testing and trial use. To run your migration workloads, buy a CDM incremental package. In addition, you are not advised to use a CDM cluster that serves as a data connection agent to run data migration jobs. For details, see **Buying a CDM 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 total 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 when a batch processing job is scheduled or a real-time job is started.

[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.

Version Mode

The version mode of DataArts Studio has changed in some regions to provide flexible resource configuration and lightweight data governance capabilities.

- Now you can purchase DataArts Studio instances of the starter, expert, or enterprise version.
- The version mode change does not affect the DataArts Studio instances you have purchased before, which may be of the starter, basic, advanced, professional, or enterprise version.

Compared with the old version mode, the new version mode provides more favorable prices and more flexible resource scaling. If you want to experience the new version mode, you are advised to buy a new DataArts Studio instance, migrate service data from the original instance to the new instance by referring to **DataArts Studio Data Migration Configuration**, and then unsubscribe from the original instance.

For details about the changes in the mode of the new version compared with that of the old version, see **Table 7-6**.

Table 7-6 Comparison between the old and new AI markets:

Difference	Old Version Mode	New Version Mode
Provided Version	- Junior version: data integration + data development - Data Integration & Data Development & Data Governance - Data Integration & Data Development & Data Governance - Data Integration & Data Development & Data Governance - Data Integration & Data Development & Data Governance	- Junior version: data integration + data development - Expert Edition: Data Integration + Data Development + Lightweight Data Governance - Data Integration & Data Development & Data Governance
Data governance capabilities	No. Except the basic edition, all editions provide full-function data governance capabilities, which are costly.	Yes. The expert edition provides lightweight data governance capabilities to meet data governance requirements of small- and medium-sized enterprises.

Difference	Old Version Mode	New Version Mode
Supported Incremental Packages	Only the function incremental package is provided. ● CDM incremental package ● DataArts DataService incremental package	Function incremental packages and specifications incremental packages are available. (For details about how to buy them, see **Buying a DataArts Studio Incremental Package**. Function incremental packages: ● CDM incremental package ● DataArts Migration resource group incremental package ● DataArts DataService incremental package Specifications incremental packages: ● Job scheduling times/day ● Incremental package of technical asset quantity ● Incremental package of data model quantity

8 Billing

DataArts Studio provides a basic package and an incremental package. The total price of a package is automatically calculated on the management console. You can buy a 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. For details, see **OBS Pricing Details**.
>
> - EIP: If you use an EIP for your DataArts Migration cluster or DataArts DataService Exclusive cluster, you need to pay for the EIP. For details, see **EIP Pricing Details**.
>
> - SMN: If you enable SMN notifications for your DataArts Studio modules, you need to pay for the notifications. For details, see **SMN Pricing Details**.
>
> - 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. For details about the billing standards, see **DEW pricing details**.

Billing Items

DataArts Studio provides a basic package and an incremental package. **Table 8-1** provides the details.

Table 8-1 DataArts Studio billing items

Billing Item	Billing Item Subcategory	Billing Item Description	Billing Mode	Billing Description
Data Arts Studi o basic packa ge	New version mode: ● Star ter ● Expe rt ● Ente rpris e Old version mode: ● Star ter ● Basi c ● Adv ance d ● Prof essi onal ● Ente rpris e	DataArts Studio instance corresponding to a basic package of DataArts Studio. To use DataArts Studio, you must buy a DataArts Studio instance first. 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?**	Yearl y/ Mont hly	The DataArts Studio basic package does not contain fees generated by other cloud services, such as EIP and OBS. For details about how the DataArts Studio basic package is billed, see **DataArts Studio Pricing Details**.

Billing Item	Billing Item Subcategory	Billing Item Description	Billing Mode	Billing Description
DataArts Studio incremental package (optional)	CDM incremental package	The CDM incremental package corresponds to a CDM cluster. CDM clusters 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.	● Pay-per-use ● Pay-per-use resource package	CDM incremental packages can be billed in pay-per-use mode or by package. ● Pay-per-use: You can enable or disable an incremental package as needed and you are billed by the usage duration. After you create a pay-per-use incremental package, a CDM cluster is automatically created. ● Pay-per-use resource package: You can purchase a pay-per-use package with a certain validity period. The package is economical and recommended if you want to use CDM for a long time. After you create an incremental package billed by package, you will get a package matching the instance specifications in a specific region. No CDM cluster will be automatically created. For details, see **DataArts Migration Incremental Package Pricing Details**.

Billing Item	Billing Item Subcategory	Billing Item Description	Billing Mode	Billing Description
	DataArts Migration resource group incremental package	This type of incremental package provides resource groups for real-time jobs in DataArts Migration. DataArts Migration resource groups can be used to migrate data to the cloud and ingest data into and export data out of a data lake. It provides wizard-based configuration and management and can migrate all and incremental data from a single table, entire database, or database or table shard.	● Pay-per-use ● Pay-per-use resource package	DataArts Migration resource group incremental packages can be billed in pay-per-use mode or by package. ● Pay-per-use: You can enable or disable an incremental package as needed and you are billed by the usage duration. When you create a DataArts Migration resource group incremental package, the system automatically creates a resource group required by real-time data integration jobs based on the specifications you set for the incremental package. ● Pay-per-use resource package: You can purchase a pay-per-use package with a certain validity period. The package is economical and recommended if you want to use CDM for a long time. An incremental package billed by package is associated with a specific region and instance specifications. No resource group is automatically

Billing Item	Billing Item Subcategory	Billing Item Description	Billing Mode	Billing Description
				created. Instead, you can use a resource group you have purchased on the DataArts Studio console in a specified region for 745 hours each month within the validity period of the incremental package. The package is billed based on actual fees.
	DataArts DataService incremental package	This package corresponds to a DataArts DataService Exclusive cluster. 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. To use DataArts DataService Exclusive, you need to purchase a DataArts DataService Exclusive package and create a cluster.	Yearly/ Monthly	The package is billed based on actual fees. You will be billed for DataArts DataService exclusive clusters (which support only yearly/monthly billing) created during the open beta test (OBT). There will be a grace period for you to pay for the DataArts DataService exclusive clusters after the OBT ends. If you do not pay for the clusters, they will be frozen after the grace period expires. **NOTE** You can create 10 DataArts DataService Exclusive APIs for free in each DataArts Studio instance, and you will be charged for each extra API.

Billing Item	Billing Item Subcategory	Billing Item Description	Billing Mode	Billing Description
	Job node scheduling times/day incremental package	The job node scheduling times/day incremental package provides resources of the corresponding specifications. DataArts Studio instances of different versions provide different job node scheduling times/day limits. This specification is calculated based on the total number of scheduling times of data development jobs, quality jobs, reconciliation jobs, service scenarios, and metadata collection jobs executed every day. The number of scheduling times of data development jobs per day is measured by node. You can click More > Quota Usage in the card of the new DataArts Studio instance to view the quota usage. If the number of daily job node scheduling times is close to or reaches this specification, you are advised to purchase a job node scheduling times/day incremental package to avoid job scheduling restrictions.	Yearly/ Monthly	The package is billed based on actual fees.

Billing Item	Billing Item Subcategory	Billing Item Description	Billing Mode	Billing Description
	Technical asset quantity incremental package	The technical asset quantity incremental package provides the corresponding specification resources. The number of technical assets varies according to the version of the DataArts Studio instance. This specification is calculated based on the number of tables and OBS files in the data catalog. You can click More > Quota Usage on the DataArts Studio instance card of the new business model to view the usage. If the number of your technology assets approaches or reaches this specification, you are advised to purchase a technology asset quantity incremental package to avoid asset collection restrictions.	Yearly/ Monthly	The package is billed based on actual fees.

Billing Item	Billing Item Subcategory	Billing Item Description	Billing Mode	Billing Description
	Data model quantity incremental package	The number of data models varies depending on the DataArts Studio instance version. This specification is calculated based on the number of logical models, physical models, dimension tables, fact tables, and summary tables in the data architecture. You can view the usage by clicking More > Quota Usage on the DataArts Studio instance card of the new business model. If the number of your data models is close to or reaches this specification, you are advised to purchase a data model quantity incremental package to avoid restrictions on new data models.	Yearly/Monthly	The package is billed based on actual fees.

📖 NOTE

When buying a CDM incremental package (pay-per-use resource package), pay attention to the following:

- When you buy a CDM incremental package which is billed based on a pay-per-use package, the system does not automatically create a CDM cluster. Instead, you can use a CDM cluster you have obtained on the DataArts Studio console for 745 hours each month within the validity period of the incremental package.

- The pay-per-use resource package can only be used to create a CDM cluster on the DataArts Studio console but not a CDM cluster on the CDM console. To create a CDM cluster on the CDM console, you must buy a discount package (pay-per-use resource package) on the CDM console.

- A discount package can be used by one or more qualified CDM clusters in the specified region. Any resource usage beyond the package quotas is billed based on a pay-per-use basis.

  For example, if you purchase a one-month package (745 hours/month) and two CDM clusters are associated with the package, 372.5 hours (about 15.5 days) can be allocated to each cluster within the one-month subscription. Any usage beyond the allocated hours will be charged in pay-per-use mode.

- If you purchase a package and do not associate it with any CDM clusters, the quota in the package will not be consumed and the validity period of the package will not be extended as well. Therefore, you are advised to make a plan before buying a package.

- 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 a yearly/monthly incremental package in the same region and with the same specifications as the pay-per-use incremental package, the fees generated before you buy the yearly/monthly incremental package are charged in pay-per-use mode, and the subsequent fees are charged based on the yearly/monthly incremental package.

Billing Mode

Table 8-2 lists the billing modes supported for the billing items of DataArts Studio.

Table 8-2 Billing modes supported for DataArts Studio

Billing Mode	Description	Application Scenario	Supported Billing Item
Yearly/ Monthly	- You can purchase cloud resources in yearly/ monthly mode by prepaying for them. - After the purchase is successful, the system allocates cloud resources to your account. - You can unsubscribe from the cloud resources in use. - Upon expiration of the resources, if you do not renew them, the resources will enter the grace period and retention period before they are reclaimed.	This mode is suitable if your business demands are stable and you want to use fixed cloud resources for a long time. It is cost-effective in the long term.	- DataArts Studio basic package - DataArts DataService incremental package - Job node scheduling times/day incremental package - Technical asset quantity incremental package - Data model quantity incremental package
Pay-per-Use	- You can directly create and use pay-per-use resources without prepayment. - You are billed by how many resources you use and how long you use them.	This mode is suitable if your business demands go up sharply or if you want to use resources in a short period of time or intermittently. The price is higher than those of the other modes.	CDM incremental package DataArts Migration resource group incremental package

Billing Mode	Description	Application Scenario	Supported Billing Item
Pay-per-use resource package	• Through prepayment, you can purchase cloud resource quotas that are valid within a certain period of time. • The system does not automatically allocate cloud resources after you purchase a package. • A pay-per-use resource package in use cannot be unsubscribed from. For details, see **Unsubscription Not Allowed**. • When a package has expired or the quota of the package has been used up, you can still use cloud resources in pay-per-use mode.	This mode is suitable and cost-effective if you want to use cloud resources for a long but uncertain period of time.	CDM incremental package DataArts Migration resource group incremental package

Changing Specifications

DataArts Studio provides multiple specifications in its basic or incremental package. You can select specifications that meet your needs when buying 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 DataArts Studio 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 and unsubscribe from the current one.

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.

  To renew the order manually, go to **Renewals**, select the target DataArts Studio instance, and click **Renew** in the row that contains the instance. For details about renewal, see **Renewal Rules**

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

    When a yearly/monthly DataArts Studio incremental package expires, it enters the grace period and then retention period (see **Expiration and Overdue Payment** for details). Top up your account before arrears.

    To top up your account, go to the **Renewals** page, choose **Overview** in the left navigation pane, and click **Top Up**.

  - **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.

    To top up your account, go to the **Renewals** page, choose **Overview** in the left navigation pane, and click **Top Up**.

  - **DataArts Studio incremental package billed by pay-per-use resource package**

    When a package of this type expires, it will be billed pay per use. If you want to use the package mode all the time, buy a new package before the current package expires.

    If you want to buy the DataArts Studio incremental package again, go to the **Renewals** page, select the DataArts Studio incremental package, and click **Buy Again**.

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 the system will be deleted and cannot be recovered.

The grace period and retention period vary based on your level. For details, see **Resource Suspension and Release**.

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

    When a package of this type expires, it enters the grace period. During this period, the package can still be used and billed as usual. You need to extend the package duration in the grace period. Otherwise, the package will enter the retention period.

    During the retention period, data is retained, but you can no longer use the package. If you do not renew your subscription within the retention period, the package will be unavailable and data will be permanently deleted when the retention period ends.

    The grace period and retention period vary based on your level. For details, see **Resource Suspension and Release**.

  - **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 the system will be deleted and cannot be recovered.

    The grace period and retention period vary based on your level. For details, see **Grace Period and Retention Period**.

  - **DataArts Studio incremental package billed by pay-per-use resource 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. For details, see **OBS Pricing Details**.

- EIP: If you use an EIP for your DataArts Migration cluster or DataArts DataService Exclusive cluster, you need to pay for the EIP. For details, see **EIP Pricing Details**.

- SMN: If you enable SMN notifications for your DataArts Studio modules, you need to pay for the notifications. For details, see **SMN Pricing Details**.

- 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. For details about the billing standards, see **DEW pricing details**.

- **DataArts Studio incremental package**

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

    Unsubscribe from a package by following the instructions in **Unsubscribing from Cloud Services**.

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

    This type of incremental packages will not be billed on a pay-per-use basis after the corresponding resources are deleted.

  - **DataArts Studio incremental package billed by pay-per-use resource package**

    This type of package cannot be unsubscribed. For details, see **Unsubscription Not Allowed**.

9 Security

9.1 Shared Responsibilities

Huawei guarantees that its commitment to cyber security will never be outweighed by the consideration of commercial interests. To cope with emerging cloud security challenges and pervasive cloud security threats and attacks, Huawei Cloud builds a comprehensive cloud service security assurance system for different regions and industries based on Huawei's unique software and hardware advantages, laws, regulations, industry standards, and security ecosystem.

Figure 9-1 illustrates the responsibilities shared by Huawei Cloud and users.

- **Huawei Cloud**: Ensure the security of cloud services and provide secure clouds. Huawei Cloud's security responsibilities include ensuring the security of our IaaS, PaaS, and SaaS services, as well as the physical environments of the Huawei Cloud data centers where our IaaS, PaaS, and SaaS services operate. Huawei Cloud is responsible for not only the security functions and performance of our infrastructure, cloud services, and technologies, but also for the overall cloud O&M security and, in the broader sense, the security and compliance of our infrastructure and services.

- **Tenant**: Use the cloud securely. Tenants of Huawei Cloud are responsible for the secure and effective management of the tenant-customized configurations of cloud services including IaaS, PaaS, and SaaS. This includes but is not limited to virtual networks, the OS of virtual machine hosts and guests, virtual firewalls, API Gateway, advanced security services, all types of cloud services, tenant data, identity accounts, and key management.

Huawei Cloud Security White Paper elaborates on the ideas and measures for building Huawei Cloud security, including cloud security strategies, the shared responsibility model, compliance and privacy, security organizations and personnel, infrastructure security, tenant service and security, engineering security, O&M security, and ecosystem security.

Figure 9-1 Huawei Cloud shared security responsibility model
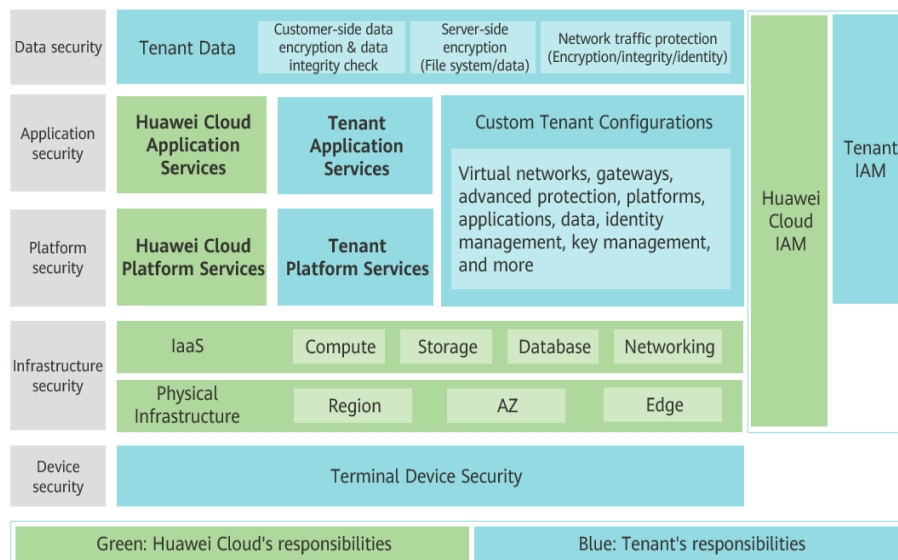
9.2 Asset Identification and Management

You can tag cloud resources to manage them more easily.

Scenario

If your system uses multiple HUAWEI CLOUD services, you can add tags to the resource instances of these services (for DataArts Studio, add tags to instances). The tags will be included in the charging data records (CDRs) generated for the services and instances. If your system is composed of multiple applications, setting the same tag for all resource instances used for each application helps you easily analyze resource usage and costs.

Tags are used to identify and classify purchased DataArts Studio instances. If you add tags to an instance, CDRs generated by the requests for this instance will contain these tags. Using the tags, you can classify CDRs for detailed cost analysis. For example, if a department uses an instance, you can add the department name to the instance as a tag. In this manner, you can analyze the department's development costs based on the CDR that contains the tag.

A tag consists of a key and a value. The key and value can exist in either sequence in a tag. An instance can have a maximum of 20 tags. Each key is unique among all tags of an instance, whereas values can be repetitive or blank.

How to Use

You can create tags for instances on the DataArts Studio console. For details, see **Buying a DataArts Studio Instance**.

9.3 Identity Authentication and Access Control

Identity Authentication

You can access DataArts Studio through the DataArts Studio console or open APIs. In either way, access requests are sent through the RESTful APIs provided by DataArts Studio.

DataArts Studio APIs can be accessed upon successful authentication. Requests sent through the console can be authenticated using tokens, and requests for calling APIs can be authenticated using tokens or AK/SK. For details, see **Authentication**.

Access Control

You can use Identity and Access Management (IAM) to implement fine-grained permissions management. IAM provides identity authentication, permissions management, and access control, helping you secure access to your HUAWEI CLOUD resources. For more information about IAM, see the **IAM Service Overview**.

DataArts Studio **supports only system role-based authorization and does not support policy-based authorization.** To implement fine-grained permission control, DataArts Studio provides the capability of granting permissions to **system roles and workspace roles**. Specific operation permissions are granted to workspace roles, and workspace roles with different permissions can be customized.

As shown in **Figure 9-2** and **Table 9-1**, the system-defined roles supported by DataArts Studio include DAYU Administrator and DAYU User. Workspace roles are based on the DAYU User. **Permissions** lists the common operations supported by DataArts Studio and the permissions granted to each role. You can select roles as required.

Figure 9-2 Permission system
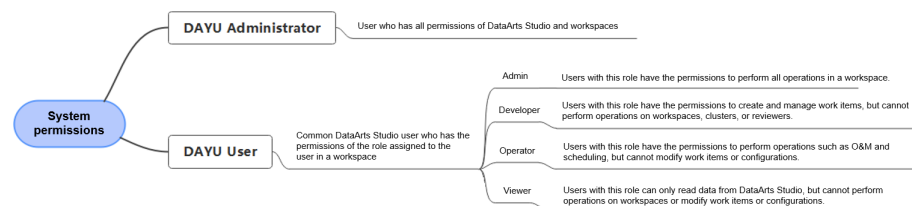
Table 9-1 DataArts Studio system-defined roles

Role	Description	Type
DAYU Administrator	Instance administrator who has all management permissions on a DataArts Studio instance and its workspaces, permissions of dependent services, and service operation permissions in all workspaces. **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	Type
DAYU User	Common user who has permissions to view a DataArts Studio instance and its workspaces, and the permissions of dependent services. After assigned a role, a common user has permissions of the role to perform service operations. Workspace roles include the preset admin, developer, deployer, operator, and viewer. For details about the operation permissions of each role, see **Permissions**. ● Admin: This role has all operation permissions in a workspace. You are advised to assign the admin role to the project owner, development owner, and O&M administrator. ● Developer: This role has permissions to create and manage resources in a workspace. You are advised to assign this role to users who develop and process tasks. ● Operator: This role has the operation permissions of services such as O&M and scheduling in a workspace, but cannot modify resources or configurations. You are advised to assign this role to users responsible for O&M management and status monitoring. ● Viewer: This role can view data in a workspace but cannot perform any other operation. You are advised to assign this role to users who only need to view data in a workspace but do not need to perform operations. ● Deployer: This role is unique to the enterprise mode and has permissions to release task packages in a workspace. In enterprise mode, when a developer submits a script or job version, the system generates a release task. After the developer confirms the release and the deployer approves the release request, the modified job is synchronized to the production environment.	System-defined role

9.4 Data Protection Technologies

Personal Data

To prevent your personal data (such as username, password, and mobile number) from being obtained by unauthorized or unauthenticated entities or individuals,

DataArts Studio encrypts your data during storage and transmission to prevent data leakage.

- Username: Your username is a user asset attribute and is used for approval and notifications, and for obtaining and internationalizing the key value of the agency token.

- IP address: Your IP address is used as the unique identifier of user assets.

- Mobile number: Your mobile number is used for notifications of approval information and other messages.

- Email: Your email address is used for notifications of approval information and other messages.

- Information for creating a database connection, including the database IP address, port number, username, password, and key pair password

- Information for creating a DIS or OBS connection, including the Access Key ID (AK) and Secret Access Key (SK)

Data Storage Security

DataArts Studio uses encryption algorithms to encrypt your personal data before storing it. Your personal data will be deleted 14 days after your service subscription expires.

Data Transmission Security

Your personal data (including passwords) is encrypted using TLS 1.2 or TLS 1.3 during transmission. All the calls made to DataArts Studio APIs use HTTPS to encrypt the transmitted data.

9.5 Audit and Logging

Cloud Trace Service (CTS) records operations performed on cloud resources in your account. The operation logs can be used to perform security analysis, track resource changes, perform compliance audits, and locate faults.

For details about the DataArts Studio operations that can be recorded by CTS, see **Viewing Traces**. After you **enable CTS** and create a tracker, CTS starts to record operations for audit. You can view audit logs of the last seven days in CTS.

In addition, you can **configure key event notifications**. With CTS, you can monitor high-risk and sensitive operations related to IAM in real time. If you perform such an operation when using DataArts Studio, CTS sends a notification to subscribers.

9.6 Service Resilience

DataArts Studio ensures data durability and reliability using technical methods such as AZ fault tolerance, overload protection, and backup and restoration.

For details about the DataArts Studio data that can be backed up, see **Managing Backups**.

9.7 Security Risk Monitoring

Working with Cloud Eye, DataArts Studio helps you monitor your CDM clusters in real time, reports alarms, and sends notifications. You can learn information such as the network inbound rate, network outbound rate, CPU usage, memory usage, disk usage, and job failure rate.

For details about the DataArts Studio monitoring metrics and how to create alarm rules, see **Viewing Metrics**.

9.8 Recovery from Failures

DataArts Studio backs up all data in the database at a scheduled time every day. If the service is faulty, data can be restored using the backup.

For details about the DataArts Studio data that can be backed up, see **Managing Backups**.

9.9 Update Management

Update to Cope with Vulnerabilities

Updates will be made in a timely manner to prevent DataArts Studio from being adversely affected by any vulnerability released in the notices on the HUAWEI CLOUD website, such as *Notice on the Apache Log4j2 Remote Code Execution Vulnerability (CVE-2021-44228)* and *Notice on the Fastjson Remote Code Execution Vulnerability (CNVD-2022-40233)*.

Configuration Update

DataArts Studio configurations can be updated through version update.

9.10 Certificates

Compliance Certificates

Huawei Cloud services and platforms have obtained various security and compliance certifications from authoritative organizations, such as International Organization for Standardization (ISO). You can **download** them from the console.

Figure 9-3 Downloading compliance certificates

Resource Center

Huawei Cloud also provides the following resources to help users meet compliance requirements. For details, see **Resource Center**.

Figure 9-4 Resource center

10 Permission 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 Huawei Cloud resources.

With IAM, you can use your Huawei Cloud 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.

IAM is a free service. You only pay for the resources in your account. For more information about IAM, see **IAM Service Overview**.

DataArts Studio Permissions

New IAM users do not have any permissions assigned by default. You need to first add them to one or more groups and then attach policies or roles to these groups. After authorization, the users can perform specified operations.

DataArts Studio is a project-level service deployed in specific physical regions. To assign 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.

DataArts Studio **supports only system role-based authorization and does not support policy-based authorization.** To implement fine-grained permission control, DataArts Studio provides the capability of granting permissions to **system roles and workspace roles**. Specific operation permissions are granted to workspace roles, and workspace roles with different permissions can be customized.

☐ NOTE

IAM provides the following two authorization mechanisms: Note that DataArts Studio supports only the IAM role-based authorization and does not support the IAM policy-based authorization.

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

● **IAM Policies**: A type of fine-grained authorization mechanism that defines permissions required to perform operations on specific cloud resources under certain conditions. This type of authorization is more flexible and is ideal for least privilege access.

As shown in **Figure 10-1** and **Table 10-1**, the system-defined roles supported by DataArts Studio include DAYU Administrator and DAYU User. Workspace roles are based on the DAYU User. **Permissions** lists the common operations supported by DataArts Studio and the permissions granted to each role. You can select roles as required.

Figure 10-1 Permission system
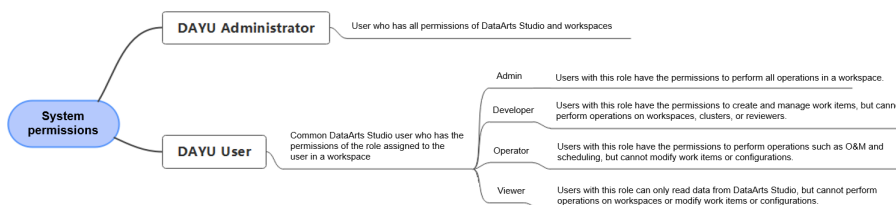
Table 10-1 DataArts Studio system-defined roles

Role	Description	Type
DAYU Administrator	Instance administrator who has all management permissions on a DataArts Studio instance and its workspaces, permissions of dependent services, and service operation permissions in all workspaces. **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	Type
DAYU User	Common user who has permissions to view a DataArts Studio instance and its workspaces, and the permissions of dependent services. After assigned a role, a common user has permissions of the role to perform service operations. Workspace roles include the preset admin, developer, deployer, operator, and viewer. For details about the operation permissions of each role, see **Permissions**. ● Admin: This role has all operation permissions in a workspace. You are advised to assign the admin role to the project owner, development owner, and O&M administrator. ● Developer: This role has permissions to create and manage resources in a workspace. You are advised to assign this role to users who develop and process tasks. ● Operator: This role has the operation permissions of services such as O&M and scheduling in a workspace, but cannot modify resources or configurations. You are advised to assign this role to users responsible for O&M management and status monitoring. ● Viewer: This role can view data in a workspace but cannot perform any other operation. You are advised to assign this role to users who only need to view data in a workspace but do not need to perform operations. ● Deployer: This role is unique to the enterprise mode and has permissions to release task packages in a workspace. In enterprise mode, when a developer submits a script or job version, the system generates a release task. After the developer confirms the release and the deployer approves the release request, the modified job is synchronized to the production environment.	System-defined role

Roles or Policies for DataArts Studio Console

Table 10-2 lists the permissions for the services on which DataArts Studio components depend. You are advised to configure the minimum DataArts Studio-level permissions of dependent services for developers. (For details, see **Authorizing IAM Users to Use DataArts Studio by Complying with the Principle of Least Privilege**.) **Table 10-3** lists the minimum permissions for developers on the services on which DataArts Studio depends.

📖 **NOTE**

> The DAYU Administrator and DAYU User already have the administrator permissions of dependent services. If the DAYU User role is assigned to a common user or user group, the common user or user group may have more permissions of dependent services than required. To prevent this risk, you can manually delete the permissions of dependent services from the user group after assigning the DAYU User role to the user group, and then assign the minimum permissions of dependent services to the user group.

Table 10-2 Roles or policies for the services on which the DataArts Studio console depends on

Console Function	Dependent Service	Role/Policy Required	Function
Management Center	BSS	bss:coupon:view bss:renewal:update bss:discount:view bss:order:view bss:order:pay bss:order:update	Create incremental packages or DataArts Studio instances.
	KMS	kms:cmk:get kms:cmk:list kms:cmk:create kms:cmk:decrypt kms:cmk:encrypt kms:dek:create kms:dek:encrypt kms:dek:decrypt	Encrypt and decrypt data during data connection creation.
	DWS	dws:cluster:list dws:cluster:getDetail dws:openAPICluster:getDetail	Create DWS connections.
	MRS	mrs:cluster:get mrs:cluster:list	Create MRS connections.
	VPC	vpc:publicIps:get vpc:publicIps:list vpc:vpcs:get vpc:subnets:get	Create MRS connections.
	RDS	rds:*:get rds:*:list	Create RDS connections.

Console Function	Dependent Service	Role/Policy Required	Function
DataArts Migration	VPC	vpc:publicIps:get vpc:publicIps:list vpc:vpcs:get vpc:vpcs:list vpc:subnets:get vpc:securityGroups:get vpc:firewalls:list vpc:routeTables:list vpc:subNetworkInterfaces:list	Create CDM clusters or DataArts Studio instances.
	ECS	ecs:flavors:get ecs:cloudServerFlavors:get ecs:availabilityZones:list	Create CDM clusters or DataArts Studio instances.
	CDM	cdm:cluster:create	Create CDM clusters.
	KMS	kms:cmk:get kms:cmk:list kms:cmk:create kms:cmk:decrypt kms:cmk:encrypt kms:dek:create kms:dek:encrypt kms:dek:decrypt	Encrypt and decrypt data during data connection creation.
	MRS	mrs:cluster:get mrs:cluster:list mrs:job:get mrs:job:list	Create MRS connections.
	DWS	dws:cluster:list dws:cluster:getDetail dws:openAPICluster:getDetail	Create DWS connections.
	CDM	cdm:cluster:get cdm:cluster:list cdm:link:operate cdm:job:operate	Perform operations on the CDM console.

Console Function	Dependent Service	Role/Policy Required	Function
	CES	ces:*:get ces:*:list	View monitoring data in Cloud Eye.
	CSS	css:*:get css:*:list	Create CSS connections.
	CloudTable	cloudtable:*:get cloudtable:*:list	Create CloudTable connections.
	RDS	rds:*:get rds:*:list	Create RDS connections.
	Config	rms:resources:list	Create CDM clusters.
DataArts Factory	OBS	obs:object:GetObject obs:object:PutObject obs:bucket:GetBucketLocation obs:bucket:ListAllMyBuckets obs:bucket:ListBucket obs:bucket:CreateBucket	Run scripts, run jobs, and back up jobs.
	SMN	smn:topic:publish smn:topic:list	Send job notifications.
	KMS	kms:cmk:get kms:cmk:list kms:cmk:create kms:cmk:decrypt kms:cmk:encrypt kms:dek:create kms:dek:encrypt kms:dek:decrypt	Encrypt and decrypt data during data connection creation.

Console Function	Dependent Service	Role/Policy Required	Function
	MRS	mrs:cluster:get mrs:cluster:list mrs:job:submit mrs:job:delete mrs:job:stop mrs:sql:execute mrs:sql:cancel mrs:job:get mrs:job:list	Run the following MRS job nodes: MRS Presto SQL, MRS Spark, MRS Spark Python, MRS Flink Job, and MRS MapReduce MRS Spark SQL and MRS Hive SQL
	DLI	dli:queue:submitJob dli:jobs:create dli:jobs:update dli:jobs:get dli:jobs:list dli:jobs:listAll	Run the following DLI job nodes: DLI SQL and DLI Spark
	OBS	obs:object:GetObject obs:object:PutObject obs:object:DeleteObject obs:bucket:GetBucketLocation obs:bucket:ListAllMyBuckets obs:bucket:ListBucket obs:bucket:ListBucketVersions obs:bucket:CreateBucket obs:bucket:DeleteBucket	Run the following OBS job nodes: Create OBS, Delete OBS, and OBS Manager
	DWS	dws:cluster:list dws:cluster:getDetail dws:openAPICluster:getDetail	Create DWS connections.

Console Function	Dependent Service	Role/Policy Required	Function
	CDM	cdm:cluster:get cdm:cluster:list cdm:job:operate	Run the Agent-related scripts and jobs required by data connections and run CDM jobs: RDS SQL, DWS SQL, Hive SQL, SPARK SQL, Shell, and Python
	CES	ces:metricData:list	Query the CPU usage of the DLI queue on the **Overview** page.
	GES	ges:graph:access ges:graph:operate ges:graph:list ges:graph:getDetail ges:metadata:create ges:metadata:operate ges:metadata:delete ges:metadata:list ges:metadata:getDetail ges:jobs:list ges:jobs:getDetail	Run the Import GES job node.
	ECS	ecs:servers:list ecs:servers:get ecs:servers:stop ecs:servers:start ecs:cloudServers:list	Run the Open/Close Resource job node and create host connections.

Console Function	Dependent Service	Role/Policy Required	Function
	DLI	dli:queue:submitJob	Run DLI jobs/scripts.
		dli:queue:cancelJob	
		dli:group:useGroup	
		dli:group:getGroup	
		dli:group:updateGroup	
		dli:group:deleteGroup	
		dli:group:listAllGroup	
		dli:database:createDatabase	
		dli:database:dropDatabase	
		dli:database:displayDatabase	
		dli:database:displayAllDatabases	
		dli:database:explain	
		dli:database:createView	
		dli:database:createTable	
		dli:database:displayAllTables	
		dli:database:createFunction	
		dli:database:describeFunction	
		dli:database:showFunctions	
		dli:database:dropFunction	
		dli:table:select	
		dli:table:update	
		dli:table:delete	
		dli:table:dropTable	
		dli:table:describeTable	
		dli:table:showCreateTable	

Console Function	Dependent Service	Role/Policy Required	Function
		dli:table:showPartitions	
		dli:table:showSegments	
		dli:table:showTableProperties	
		dli:table:insertOverwriteTable	
		dli:table:insertIntoTable	
		dli:table:compaction	
		dli:table:truncateTable	
		dli:table:alterView	
		dli:table:alterTableRename	
		dli:table:alterTableAddColumns	
		dli:table:alterTableDropColumns	
		dli:table:alterTableChangeColumn	
		dli:table:alterTableSetLocation	
		dli:table:alterTableAddPartition	
		dli:table:alterTableRenamePartition	
		dli:table:alterTableSetProperties	
		dli:table:alterTableRecoverPartition	
		dli:table:alterTableDropPartition	
		dli:column:select	
		dli:jobs:create	
		dli:jobs:delete	
		dli:jobs:start	
		dli:jobs:stop	
		dli:jobs:update	
		dli:jobs:export	

Console Function	Dependent Service	Role/Policy Required	Function
		dli:jobs:get	
		dli:jobs:list	
		dli:jobs:listAll	
		dli:resource:useResource	
		dli:resource:updateResource	
		dli:resource:deleteResource	
		dli:resource:getResource	
		dli:resource:listAllResource	
		dli:variable:update	
		dli:variable:delete	
	IAM	iam:agencies:listAgencies	Obtain job agencies.
	DIS	DIS Operator DIS User	Run the following DIS job nodes: DIS Stream, DIS Dump, and DIS Client
	SWR	SWR Admin	The image read permission in SWR is required only when a custom image is selected for a **DLI Spark** node of a job in DataArts Factory. You are advised to add the read permission of the image by referring to **User Permissions**. You are not advised to directly assign the SWR Admin system role to users because this may result in excessive permissions.

Console Function	Dependent Service	Role/Policy Required	Function
DataArts Catalog	OBS	obs:object:GetObject obs:bucket:GetBucketStorage obs:bucket:GetBucketLocation obs:bucket:ListAllMyBuckets obs:bucket:ListBucket	Collect OBS metadata.
	DIS	dis:streams:list dis:transferTasks:list	Collect DIS metadata.
	CSS	css:cluster:list	Collect CSS metadata.
	GES	ges:graph:list ges:graph:getDetail ges:metadata:list ges:metadata:getDetail	Collect GES metadata.
	DLI	dli:database:displayDatabase dli:database:displayAllDatabases dli:table:select dli:table:describeTable dli:table:showPartitions dli:table:showTableProperties dli:jobs:create dli:jobs:get	Collect DLI metadata and analyze data in summary.
	CDM	cdm:cluster:list	Collect CSS metadata.
DataArts Quality	SMN	smn:topic:publish smn:topic:list	Configure job notifications.

Console Function	Dependent Service	Role/Policy Required	Function
	OBS	obs:object:GetObject obs:object:PutObject obs:bucket:GetBucketLocation obs:bucket:ListAllMyBuckets obs:bucket:ListBucket obs:bucket:CreateBucket	Export quality reports.
	MRS	mrs:job:submit mrs:sql:execute mrs:sql:cancel mrs:job:get	Run MRS quality jobs.
	DLI	dli:queue:submitJob dli:jobs:get dli:jobs:listAll	Run DLI quality jobs.
DataArts Security	DLI	dli:queue:submitJob dli:queue:cancelJob dli:database:displayDatabase dli:database:displayAllDatabases dli:database:displayAllTables dli:table:describeTable dli:jobs:create dli:jobs:stop dli:jobs:get dli:resource:deleteResource dli:resource:getResource dli:resource:listAllResource	Manage DLI permissions.

Console Function	Dependent Service	Role/Policy Required	Function
	DWS	dws:cluster:list dws:cluster:getDetail dws:openAPICluster: getDetail	Manage DWS permissions.
	MRS	mrs:cluster:list mrs:job:submit mrs:job:stop	Manage MRS permissions.
	KMS	kms:cmk:list kms:cmk:encrypt kms:cmk:decrypt	Encrypt and decrypt data using KMS.
	CDM	Any CDM permission, for example, cdm:cluster:get	Manage DWS and MRS permissions.

Table 10-3 Minimum permissions for developers on the services on which DataArts Studio depends

Permission Type	Role/Policy-based permissions–system role	Role/Policy-based permissions–custom policy	Role/Policy-based permissions–custom policy
Mandatory	**Mandatory**	**Mandatory**	**Mandatory**

| Per miss ion | • DIS Operat or

• DIS User

• (Optio nal and not recom mende d) SWR Admin | **DataArtsStudio_Permission sOfDependentServi- ces_global**: custom policy for a global dependent cloud service

{
   "Version": "1.1",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": [
            "obs:object:GetObject",
            "obs:object:PutObject",
            "obs:object:DeleteObject",

"obs:bucket:GetBucketStorage",

"obs:bucket:GetBucketLocation",

"obs:bucket:ListAllMyBuckets",
            "obs:bucket:ListBucket",

"obs:bucket:ListBucketVersions",

"obs:bucket:CreateBucket",

"obs:bucket:DeleteBucket",
            "rms:resources:list",

"iam:agencies:listAgencies"
         ]
      }
   ]
} | **DataArtsStudio_Permissions OfDependentServi- ces_region**: custom policy for a regional dependent cloud service

{
   "Version": "1.1",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": [
            "cdm:cluster:get",
            "cdm:cluster:list",
            "cdm:cluster:create",
            "cdm:link:operate",
            "cdm:job:operate",
            "ces:*:get",
            "ces:*:list",
            "cloudtable:*:get",
            "cloudtable:*:list",
      "css:*:get",
            "css:*:list",
            "dis:streams:list",
            "dis:transferTasks:list",
            "dli:queue:submitJob",
            "dli:queue:cancelJob",
            "dli:table:insertOverwriteT- able",
            "dli:table:insertIntoTable",
            "dli:table:alterView",

"dli:table:alterTableRename",
            "dli:table:compaction",
            "dli:table:truncateTable",

"dli:table:alterTableDropColumns",
            "dli:table:alterTableSetPro- perties",
            "dli:table:alterTableChange- Column",
            "dli:table:showSegments",

"dli:table:alterTableRecoverPartition",
            "dli:table:dropTable",
            "dli:table:update",

"dli:table:alterTableDropPartition",
            "dli:table:alterTableAddPar- tition",
            "dli:table:alterTableAddCol- umns",

"dli:table:alterTableRenamePartition",
            "dli:table:delete",
            "dli:table:alterTableSetLo- cation",
            "dli:table:describeTable",
            "dli:table:showPartitions",

"dli:table:showCreateTable",
            "dli:table:showTableProper- ties",
            "dli:table:select",

"dli:resource:updateResource",
            "dli:resource:useResource",
            "dli:resource:getResource", |

		NOTE		"dli:resource:listAllResource",
		The image read permission in SWR is required only when a custom image is selected for a **DLI Spark** node of a job in DataArts Factory. You are advised to add the read permission of the image by referring to **User Permissions**. You are not advised to directly assign the SWR Admin system role to users because this may result in excessive permi		"dli:resource:deleteResource", "dli:database:explain", "dli:database:createDatabase", "dli:database:dropFunction", "dli:database:createFunction", "dli:database:displayAllDatabases", "dli:database:displayAllTables", "dli:database:displayDatabase", "dli:database:describeFunction", "dli:database:createView", "dli:database:createTable", "dli:database:showFunctions", "dli:database:dropDatabase", "dli:group:useGroup", "dli:group:updateGroup", "dli:group:listAllGroup", "dli:group:getGroup", "dli:group:deleteGroup", "dli:column:select", "dli:jobs:start", "dli:jobs:export", "dli:jobs:update", "dli:jobs:list", "dli:jobs:listAll", "dli:jobs:get", "dli:jobs:delete", "dli:jobs:create", "dli:jobs:stop", "dli:variable:update", "dli:variable:delete", "dws:cluster:list", "dws:cluster:getDetail", "dws:openAPICluster:getDetail", "ecs:servers:get", "ecs:servers:list", "ecs:servers:stop", "ecs:servers:start", "ecs:flavors:get", "ecs:cloudServerFlavors:get", "ecs:cloudServers:list", "ecs:availabilityZones:list", "ges:graph:access", "ges:metadata:create", "ges:jobs:list", "ges:graph:operate", "ges:jobs:getDetail", "ges:graph:getDetail", "ges:graph:list", "ges:metadata:list", "ges:metadata:getDetail", "ges:metadata:delete", "ges:metadata:operate", "kms:cmk:get", "kms:cmk:list", "kms:cmk:create", "kms:cmk:decrypt", "kms:cmk:encrypt", "kms:dek:create", "kms:dek:encrypt", "kms:dek:decrypt", "mrs:cluster:get", "mrs:cluster:list", "mrs:job:get", "mrs:job:list", "mrs:job:submit",

Reference

- **IAM Service Overview**
- **Authorizing Other Users to Use DataArts Studio**

11 Permissions

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

- Admin: This role has all operation permissions in a workspace. You are advised to assign the admin role to the project owner, development owner, and O&M administrator.

- Developer: This role has permissions to create and manage resources in a workspace. You are advised to assign this role to users who develop and process tasks.

- Operator: This role has the operation permissions of services such as O&M and scheduling in a workspace, but cannot modify resources or configurations. You are advised to assign this role to users responsible for O&M management and status monitoring.

- Viewer: This role can view data in a workspace but cannot perform any other operation. You are advised to assign this role to users who only need to view data in a workspace but do not need to perform operations.

- Deployer: This role is unique to the enterprise mode and has permissions to release task packages in a workspace. In enterprise mode, when a developer submits a script or job version, the system generates a release task. After the developer confirms the release and the deployer approves the release request, the modified job is synchronized to the production environment.

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
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

Permission	Admin	Developer	Operator	Viewer
Deleting RDS driver packages	Y	N	N	N
Operating RDS driver packages	Y	N	N	N
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
Performing backup management operations	Y	Y	Y	N
Creating logical entities or physical tables	Y	Y	N	N
Deleting logical entities or physical tables	Y	Y	N	N

Permission	Admin	Developer	Operator	Viewer
Querying logical entities or physical tables	Y	Y	Y	Y
Editing logical entities or physical tables	Y	Y	N	N
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
Performing operations in Review Center	Y	Y	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

Permission	Admin	Developer	Operator	Viewer
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
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

Permission	Admin	Developer	Operator	Viewer
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
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

Permission	Admin	Developer	Operator	Viewer
Creating associated quality rules	Y	Y	N	N
Deleting associated quality rules	Y	Y	N	N
Querying associated quality rules	Y	Y	Y	Y
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

Permission	Admin	Developer	Operator	Viewer
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
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

Permission	Admin	Developer	Deployer	Operator	Viewer
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
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

Permission	Admin	Developer	Deployer	Operator	Viewer
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
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

Permission	Admin	Developer	Deployer	Operator	Viewer
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
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

Permission	Admin	Developer	Deployer	Operator	Viewer
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
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

DataArts Quality

Permission	Admin	Developer	Operator	Viewer
Data quality monitoring				
Querying the dashboard	Y	Y	Y	Y

Permission	Admin	Developer	Operator	Viewer
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
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

Permission	Admin	Developer	Operator	Viewer
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
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

Permission	Admin	Developer	Operator	Viewer
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
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

Permission	Admin	Developer	Operator	Viewer
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
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

Permission	Admin	Developer	Operator	Viewer
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
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
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
Creating data classifications	Y	Y	Y	N
Deleting data classifications	Y	Y	Y	N
Querying data classifications	Y	Y	Y	Y
Editing data classifications	Y	Y	Y	N
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

Permission	Admin	Developer	Operator	Viewer
Creating dynamic policies	Y	N	N	N
Deleting dynamic policies	Y	N	N	N
Querying dynamic policies	Y	Y	Y	Y
Editing dynamic policies	Y	N	N	N
Creating security levels	Y	Y	N	N
Deleting security levels	Y	Y	N	N
Querying security levels	Y	Y	Y	Y
Editing security levels	Y	Y	N	N
Creating dynamic masking policies	Y	N	N	N
Deleting dynamic masking policies	Y	N	N	N
Querying dynamic masking policies	Y	Y	Y	Y
Editing dynamic masking policies	Y	N	N	N
Creating dynamic masking subscription policies	Y	N	N	N
Deleting dynamic masking subscription policies	Y	N	N	N
Querying dynamic masking subscription policies	Y	Y	Y	Y
Creating resource permission policies	Y	N	N	N

Permission	Admin	Developer	Operator	Viewer
Deleting resource permission policies	Y	N	N	N
Querying resource permission policies	Y	Y	Y	Y
Editing resource permission policies	Y	N	N	N
Operating security task scheduling	Y	Y	Y	N
Creating permission applications	Y	Y	Y	N
Querying permission applications	Y	Y	Y	Y
Editing permission applications	Y	Y	Y	N
Creating user synchronization tasks	Y	Y	Y	N
Deleting user synchronization tasks	Y	Y	Y	N
Querying user synchronization tasks	Y	Y	Y	Y
Editing user synchronization tasks	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

Permission	Admin	Developer	Operator	Viewer
Querying data masking tasks	Y	Y	Y	Y
Editing data masking tasks	Y	Y	N	N
Operating fine-grained permission control	Y	N	N	N
Querying fine-grained permission control	Y	Y	Y	Y
Editing fine-grained permission control	Y	N	N	N
Creating permission sets	Y	Y	Y	N
Deleting permission sets	Y	Y	Y	N
Querying permission sets	Y	Y	Y	Y
Editing permission sets	Y	Y	Y	N
Querying the dashboard	Y	Y	Y	Y
Creating cross-source permission synchronization policies	Y	N	N	N
Deleting cross-source permission synchronization policies	Y	N	N	N
Querying cross-source permission synchronization policies	Y	Y	Y	Y

Permission	Admin	Developer	Operator	Viewer
Editing cross-source permission synchronization policies	Y	N	N	N
Querying member management tasks	Y	Y	Y	Y
Editing member management tasks	Y	Y	Y	N
Adding permission set members	Y	Y	Y	N
Deleting permission set members	Y	Y	Y	N
Querying permission set members	Y	Y	Y	Y
Querying agencies	Y	Y	Y	Y
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
Querying data access audit	Y	N	N	N
Creating rule groups	Y	Y	Y	N
Deleting rule groups	Y	Y	N	N
Operating rule groups	Y	Y	Y	N

Permission	Admin	Developer	Operator	Viewer
Querying rule groups	Y	Y	Y	Y
Editing rule groups	Y	Y	Y	N
Querying logs of permission synchronization failures	Y	Y	Y	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
Querying sensitive data discovery tasks	Y	Y	Y	Y
Editing sensitive data discovery tasks	Y	Y	N	N
Creating permission sets	Y	Y	Y	N
Deleting permission sets	Y	Y	Y	N
Querying permission sets	Y	Y	Y	Y
Editing permission sets	Y	Y	Y	N
Querying data sources	Y	Y	Y	Y
Querying catalog permission policies	Y	Y	Y	Y
Creating row-level access policies	Y	N	N	N

Permission	Admin	Developer	Operator	Viewer
Deleting row-level access policies	Y	N	N	N
Querying row-level access policies	Y	Y	Y	Y
Editing row-level access policies	Y	N	N	N
Creating queue policies	Y	N	N	N
Deleting queue policies	Y	N	N	N
Querying queue policies	Y	Y	Y	Y
Editing queue policies	Y	N	N	N
Creating security diagnosis tasks	Y	N	N	N
Querying security diagnosis tasks	Y	Y	Y	Y
Querying resource permission configuration	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
Editing data watermarking tasks	Y	Y	N	N

12 Notes and Constraints

Browser Constraints

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

Table 12-1 Browser compatibility

Browser	Recommended Version	Recommended OS	Remarks
Google Chrome	126, 125, and 124	Windows 10	1920x1080 and 2560x1440 are recommended.
Firefox	127 and 126	Windows 10	1920x1080 and 2560x1440 are recommended.
Microsoft Edge	N/A (The version is updated with W10.)	Windows 10	1920x1080 and 2560x1440 are recommended.

Use Constraints

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

Table 12-2 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. For more information about the application scenarios and differences between the full-stack DeC, Huawei Cloud Stack, and Huawei Cloud Stack Online, **contact sales**.
	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**.

Component	Restriction
Management 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. If a CDM cluster functions as the agent for a data connection in Management Center, the cluster supports a maximum of 200 concurrent active threads. If multiple data connections share an agent, a maximum of 200 SQL, Shell, and Python scripts submitted through the connections can run concurrently. Excess tasks will be queued. You are advised to plan multiple agents based on the workload. 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. There is no quota limit for CDM jobs. However, it is recommended that the number of jobs be less than or equal to twice the number of vCPUs in the CDM cluster. Otherwise, job performance may be affected. 3. 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**. 4. If changes occur in the connected data source (for example, the MRS cluster capacity is expanded), you need to edit and save the connection. 5. If you have uploaded an updated version of a driver, you must restart the CDM cluster for the new driver to take effect. 6. The number of concurrent extraction tasks for a job ranges from 1 to 300, and the total number of concurrent extraction tasks for a cluster ranges from 1 to 1,000. The maximum number of concurrent extraction tasks for a cluster depends on the CDM cluster specifications. You are advised to set the maximum number of concurrent extraction tasks to no larger than twice the number of vCPUs. The number of concurrent extraction tasks for a job should not exceed that for a cluster. If the number of concurrent extraction tasks is too large, memory overflow may occur. Exercise caution when changing the maximum number of concurrent extraction tasks. For more constraints on DataArts Migration, see **CDM Constraints**.

Compo nent	Restriction
DataArt s Factory	1. You can enable backup of assets such as scripts and jobs to OBS buckets. For details, see **Managing Backups**. 2. The execution history of scripts, jobs, and nodes is stored in OBS buckets. If no OBS bucket is available, you cannot view the execution history. 3. Resources from a HDFS can be used only by MRS Spark, MRS Flink Job, and MRS MapReduce nodes. 4. A workspace can contain a maximum of 10,000 scripts, a maximum of 5,000 script directories, and a maximum of 10 directory levels. 5. A workspace can contain a maximum of 10,000 jobs, a maximum of 5,000 job directories, and a maximum of 10 directory levels. 6. 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 or 3 MB data can be dumped. 7. Only data of the last six months can be displayed on the **Monitor Instance** and **Monitor PatchData** pages. 8. Only notification records of the last 30 days can be displayed. 9. The download records age out every seven days. When aged out, download records and the data dumped to OBS are both deleted.
DataArt s Architec ture	1. DataArts Architecture supports ER modeling, dimensional modeling (only star models), and data mart. 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 ● Business metrics: 100,000 ● Atomic, derivative, and compound metrics: 5,000 for each 4. The quotas for different custom objects are as follows: ● Custom subjects: 10 ● Custom tables: 30 ● Custom attributes: 10 ● Custom business metrics: 50

Component	Restriction
DataArts 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.
	3. By default, a maximum of 1,000 SQL statements associated with a quality job of a data connection can be executed concurrently. Excess SQL statements will be queued. The value ranges from 10 to 1000.
	4. By default, a maximum of 10,000 SQL statements associated with a quality job in a region can be executed concurrently. Excess SQL statements will be queued.
	5. In the **Instance Running Status** and **Instance Alarm Status** areas on the **Dashboard** page on the **Metric Monitoring** page, data of the last seven days is displayed. On the **Alarms**, **Scenarios**, and **Metrics** pages, data of the last seven, 15, or 30 days can be displayed.
	6. In the **Quantity Changes** area on the **Dashboard** page on the **Quality Monitoring** page, data of 30 days can be displayed. In the **Alarm Trend by Severity** and **Rule Quantity Trend** areas, data of the last seven days can be displayed.
	7. Quality reports are generated in batches on the T+1 day and retained for 90 days.
	8. If you export a quality report to OBS, the report is exported to the OBS path for storing job logs configured for the workspace. The exported record is retained for three months.

Compo nent	Restriction
DataArt s Catalog	1. A maximum of 100 metadata collection tasks can be created in a workspace.
	2. 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 metadata collection has the minimum impact on the data source. This rule also applies to other scenarios.
	3. Only the jobs that are scheduled and executed in DataArts Factory generate data lineages. Tested jobs do not generate data lineages.
	4. Historical data connections of the last seven days, 15 days, or 30 days can be displayed on the **Dashboard** page on the **Metadata Collection** page.

Compo nent	Restriction
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. A maximum of five DataArts DataService Exclusive clusters can be created in a DataArts Studio instance. Each cluster must be associated with a workspace and cannot belong to multiple workspaces. 3. After a DataArts DataService Exclusive cluster is created, its specifications cannot be modified, and its version cannot be upgraded. 4. The maximum number of DataArts DataService Exclusive APIs that can be created in a DataArts Studio instance is the quota of DataArts DataService Exclusive APIs (5,000 by default) or the total API quotas of the clusters in the instance, whichever is smaller. For example, if the quota of DataArts DataService Exclusive APIs for a DataArts Studio instance is 5,000, and two clusters whose API quotas are 500 and 2,000 respectively have been created in the instance, a maximum of 2,500 DataArts DataService Exclusive APIs can be created in the instance. 5. The maximum number of DataArts DataService Exclusive APIs that can be created in a workspace is the quota of DataArts DataService Exclusive APIs (configured in the workspace information) or the total API quotas of the clusters in the instance, whichever is smaller. For example, if the quota of DataArts DataService Exclusive APIs for a workspace is 800, and two clusters whose API quotas are both 500 have been created in the workspace, a maximum of 800 DataArts DataService Exclusive APIs can be created in the workspace. 6. A maximum of 1,000 applications can be created in a workspace. 7. A maximum of 500 throttling policies can be created in a workspace. 8. DataArts DataService allows you to trace and save events. For each event, DataArts DataService records information such as the date, description, and time source (a cluster). Events are retained for 30 days. 9. From the log of a DataArts DataService Exclusive cluster, you can only obtain the last 100 access records of the cluster, evenly from all nodes of the cluster. 10.In the **APIs Called**, **APIs Published**, **Top 5 APIs by Call Rate**, **Top 5 APIs by Call Duration**, and **Top 5 APIs by Call Quantity** areas on the **Overview** page, data of the last 12 hours, one day, seven days, or 30 days can be displayed. The total number of API calls is the sum of the number of APIs made in the last seven days (excluding the current day).

13 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.

API Gateway

API Gateway (APIG) enables DataArts Studio to provision the APIs of its modules.

DLI

Data Lake Insight (DLI) can be used as the data lake for DataArts Studio and enables data integration, development, governance, and provisioning.

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.

CDM

DataArts Studio accesses tenant resources or services through a Cloud Data Migration (CDM) agent. CDM and DataArts Migration are two forms of the same service. The former is an independent service, while the latter is a module of DataArts Studio.