

# Cloud Data Migration

## Service Overview

**Issue** 01  
**Date** 2023-06-21



**Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2024. All rights reserved.**

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

## **Trademarks and Permissions**



HUAWEI and other Huawei trademarks are the property of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

## **Notice**

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

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

---

# Contents

---

|   |           |
|---|-----------|
| <b>1 What Is CDM?</b>                           | <b>1</b>  |
| <b>2 Product Advantages</b>                     | <b>4</b>  |
| <b>3 Supported Data Sources</b>                 | <b>6</b>  |
| 3.1 Supported Data Sources (2.9.3.300)          | 6         |
| 3.2 Supported Data Sources (2.9.2.200)          | 20        |
| 3.3 Supported Data Types                        | 33        |
| <b>4 Application Scenarios</b>                  | <b>65</b> |
| <b>5 Basic Concepts</b>                         | <b>67</b> |
| <b>6 Regions and AZs</b>                        | <b>72</b> |
| <b>7 CDM Migration Principles</b>               | <b>74</b> |
| <b>8 Related Services</b>                       | <b>77</b> |
| <b>9 Constraints</b>                            | <b>80</b> |
| <b>10 Billing</b>                               | <b>86</b> |
| <b>11 Security</b>                              | <b>89</b> |
| 11.1 Shared Responsibilities                    | 89        |
| 11.2 Asset Identification and Management        | 90        |
| 11.3 Identity Authentication and Access Control | 91        |
| 11.4 Data Protection Technologies               | 92        |
| 11.5 Audit and Logging                          | 93        |
| 11.6 Service Resilience                         | 93        |
| 11.7 Security Risk Monitoring                   | 93        |
| 11.8 Recovery from Failures                     | 93        |
| 11.9 Update Management                          | 93        |
| 11.10 Certificates                              | 94        |
| <b>12 Quotas</b>                                | <b>96</b> |
| <b>13 Permissions Management</b>                | <b>97</b> |

# 1 What Is CDM?

---

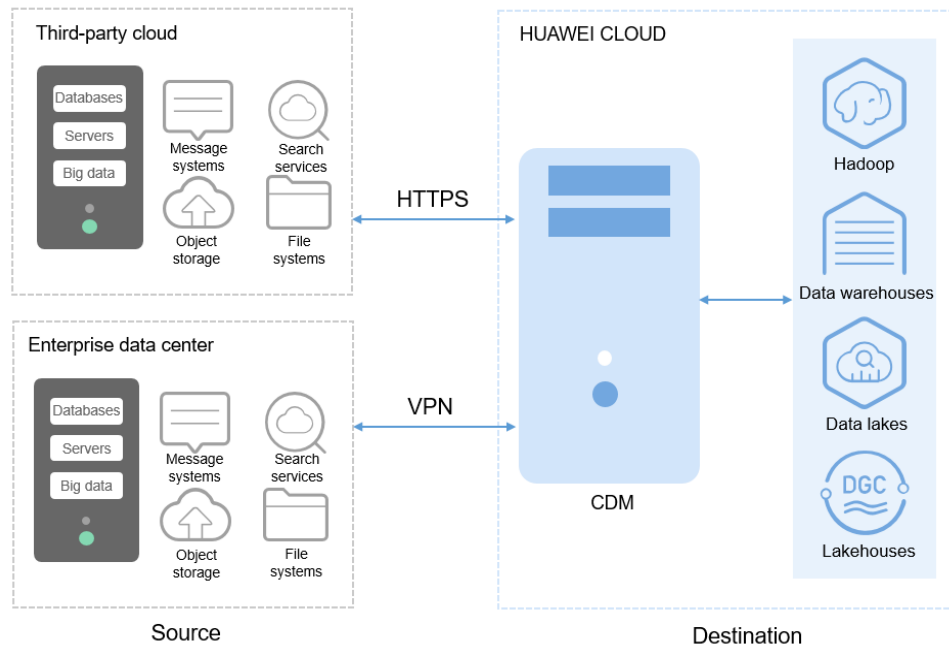
## Product Overview

Cloud Data Migration (CDM) is an efficient and easy-to-use batch data integration service. Based on the big data migration to the cloud and intelligent data lake solution, CDM provides easy-to-use migration capabilities and capabilities of integrating multiple data sources to the data lake, reducing the complexity of data source migration and integration and effectively improving the data migration and integration efficiency.

In the DataArts Studio service, CDM serves as the DataArts Migration component, which provides the same capabilities as the independent CDM service. In later sections of this document, cloud data migration and data integration both refer to CDM.

Based on the distributed computing framework and the parallel processing technology, CDM helps you migrate massive sets of data stably and efficiently. You can migrate data online and quickly construct a desired data structure.

Figure 1-1 CDM positioning



## Functions

- Table/file/entire DB migration**  
 Tables or files can be migrated in batches. An entire database can be migrated between homogeneous and heterogeneous databases. A job can migrate hundreds of tables.
- Incremental data migration**  
 CDM supports incremental migration of files, relational databases, and HBase/CloudTable, as well as with WHERE clauses and macro variables of date and time.
- Migration in transaction mode**  
 When a CDM job fails to be executed, CDM rolls back the data to the state before the job starts and automatically deletes data from the destination table.
- Field conversion**  
 CDM supports field conversion functions, such as anonymization, character string operations, and date operations.
- File encryption**  
 When files are migrated to a file system, CDM can encrypt the files written to the cloud.
- MD5 verification**  
 MD5 verification is supported to check the file consistency from end to end and output verification result.
- Dirty data archiving**  
 CDM can archive the data that fails to be processed during migration, has been filtered out, or is not compliant with conversion or cleaning rules to dirty

data logs. The threshold for dirty data ratio can be set to determine whether a task is successful.

# 2 Product Advantages

Data migration is involved when you consolidate or back up data, or develop new applications on the cloud. Generally, if you want to migrate data, you may develop data migration scripts to read data from the source and write data to the destination. Compared with this method, CDM has the following advantages:

**Table 2-1** CDM advantages

| Item                 | User-Developed Script  | CDM  |
|----------------------|--|--|
| Ease of use          | <p>You need to prepare server resources, and install and configure software, which is time-consuming.</p> <p>Because the data source types are different, the program uses different access interfaces, such as JDBC and native APIs, to read and write data. In this case, various libraries and SDKs are required when you write data migration scripts, resulting in high development and management costs.</p> | <p>CDM provides a web-based management console for enabling services on web pages in real time.</p> <p>You can migrate data by configuring data sources and migration jobs on the GUI and CDM will manage and maintain the data sources and migration jobs for you. In other words, you only need to focus on the data migration logic without worrying about the environment, which greatly reduces development and maintenance costs.</p> <p>CDM also provides RESTful APIs to support third-party system calling and integration.</p> |
| Real-time monitoring | <p>You need to select specific versions to develop as required.</p>  | <p>You can use Cloud Eye to automatically monitor CDM clusters in real time and manage alarms and notifications, so that you can keep track of CDM cluster performance metrics.</p>  |

| Item                           | User-Developed Script   | CDM   |
|--------------------------------|---|---|
| O&M free                       | You need to develop and optimize O&M functions, especially alarm and notification functions, to ensure system availability. Otherwise, manual attendance is required.   | With CDM, you do not need to maintain resources such as servers and VMs. CDM has the log, monitoring, and alarm functions, which send notifications to related personnel in a timely manner to avoid 24/7 hours of manual O&M.  |
| High efficiency                | During data migration, the read and write process is completed in one job. Limited by available resources, the performance is poor and generally cannot meet the requirements of scenarios where massive sets of data need to be migrated.  | Based on the distributed computing framework, CDM jobs are split into independent sub-jobs and executed concurrently, which drastically improves data migration efficiency. In addition, efficient data import interfaces are provided to import data from Hive, HBase, MySQL databases, and Data Warehouse Service (DWS).  |
| Various data sources           | Different tasks must be developed for different data sources, generating a number of scripts.   | Various data sources such as databases, Hadoop, NoSQL, data warehouses, and files are supported. For details, see <a href="#">Supported Data Sources</a> .  |
| Different network environments | As the cloud computing technology develops, user data may be stored in different environments, such as public clouds, on-premises or hosted Internet data centers (IDCs), and hybrid scenarios. In heterogeneous environments, data migration is subject to various factors, for example, network connectivity, which causes inconvenience for development and maintenance. | CDM helps you easily cope with various data migration scenarios, including data migration to the cloud, data exchange on the cloud, and data migration to on-premises service systems, regardless of whether the data is stored on on-premises IDCs, cloud services, third-party clouds, or self-built databases or file systems on Elastic Cloud Servers (ECSs). |



# 3 Supported Data Sources

---

## 3.1 Supported Data Sources (2.9.3.300)

CDM provides the following migration modes which support different data sources:

- **Table/File migration** in the import of data into a data lake or migration of data to the cloud. For details, see [Data Sources Supported by Table/File Migration](#).
- **Entire DB migration** in the import of data into a data lake or migration of data to the cloud. For details, see [Supported Data Sources in Entire DB Migration](#).

 **NOTE**

This section describes the data sources supported by CDM clusters of version 2.9.3.300. The supported data sources vary depending on the CDM cluster version.

### Data Sources Supported by Table/File Migration

Table/File migration can migrate data in tables or files.

[Table 3-1](#) describes the supported data sources.

**Table 3-1** Supported data sources during table/file migration

| Category       | Source                  | Destination  | Description  |
|----------------|-------------------------|--|--|
| Data warehouse | Data Warehouse Service  | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS), Data Lake Insight (DLI), and MRS ClickHouse</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object-based storage: Object Storage Service (OBS)</li> <li>• Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>             | The DWS physical machine management mode is not supported. |
|                | Data Lake Insight (DLI) | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS), Data Lake Insight (DLI), and MRS ClickHouse</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object-based storage: Object Storage Service (OBS)</li> <li>• Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>• NoSQL: CloudTable and MongoDB</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> | Recommended MongoDB version: 4.2                           |
|                | MRS ClickHouse          | Data warehouse: MRS ClickHouse and Data Lake Insight (DLI)   | Recommended MRS ClickHouse version: 21.3.4.X               |

| Category | Source  | Destination  | Description   |
|----------|---|--|---|
| Hadoop   | MRS HDFS  | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> </ul>   | <ul style="list-style-type: none"> <li>• Supported by local storage. Only MRS Hive and MRS Hudi are supported in storage-compute decoupling scenarios.</li> <li>• Only MRS Hive is supported in Ranger scenarios.</li> <li>• Not supported if SSL is enabled for ZooKeeper</li> <li>• Recommended MRS HDFS versions: <ul style="list-style-type: none"> <li>- 2.8.X</li> <li>- 3.1.X</li> </ul> </li> <li>• Recommended MRS HBase versions: <ul style="list-style-type: none"> <li>- 2.1.X</li> <li>- 1.3.X</li> </ul> </li> <li>• MRS Hive and MRS Hudi 2.x versions are not supported. The following versions are recommended: <ul style="list-style-type: none"> <li>- 1.2.X</li> <li>- 3.1.X</li> </ul> </li> </ul> |
|          | MRS HBase   | <ul style="list-style-type: none"> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object-based storage: Object Storage Service (OBS)</li> <li>• Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>  |   |
|          | MRS Hive  | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS), Data Lake Insight (DLI), and MRS ClickHouse</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object-based storage: Object Storage Service (OBS)</li> <li>• Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |   |
| MRS Hudi | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS)</li> <li>• Hadoop: MRS HBase</li> </ul> |  |   |

| Category            | Source  | Destination  | Description  |
|---------------------|---|--|--|
|                     | FusionInsight HDFS  | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> </ul> | <ul style="list-style-type: none"> <li>FusionInsight cannot serve as the destination.</li> <li>Supported only by local storage and not in storage-compute decoupling scenarios</li> <li>Not supported by Ranger</li> <li>Not supported if SSL is enabled for ZooKeeper</li> <li>Recommended FusionInsight HDFS versions:                             <ul style="list-style-type: none"> <li>- 2.8.X</li> <li>- 3.1.X</li> </ul> </li> <li>Recommended FusionInsight HBase versions:                             <ul style="list-style-type: none"> <li>- 2.1.X</li> <li>- 1.3.X</li> </ul> </li> <li>Recommended FusionInsight Hive versions:                             <ul style="list-style-type: none"> <li>- 1.2.X</li> <li>- 3.1.X</li> </ul> </li> </ul> |
| FusionInsight HBase | <ul style="list-style-type: none"> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> </ul>   |  |  |
| FusionInsight Hive  | <ul style="list-style-type: none"> <li>Object-based storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |  |  |

| Category       | Source   | Destination  | Description  |
|----------------|--|--|--|
|                | <ul style="list-style-type: none"> <li>Apache HBase</li> <li>Apache Hive</li> <li>Apache HDFS</li> </ul> | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object-based storage: Object Storage Service (OBS)</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> | <ul style="list-style-type: none"> <li>• Apache cannot serve as the destination.</li> <li>• Supported only by local storage and not in storage-compute decoupling scenarios</li> <li>• Not supported by Ranger</li> <li>• Not supported if SSL is enabled for ZooKeeper</li> <li>• Recommended Apache HBase versions:               <ul style="list-style-type: none"> <li>- 2.1.X</li> <li>- 1.3.X</li> </ul> </li> <li>• Apache Hive 2.x versions are not supported. The following versions are recommended:               <ul style="list-style-type: none"> <li>- 1.2.X</li> <li>- 3.1.X</li> </ul> </li> <li>• Recommended Apache HDFS versions:               <ul style="list-style-type: none"> <li>- 2.8.X</li> <li>- 3.1.X</li> </ul> </li> </ul> |
| Object storage | Object Storage Service (OBS)   | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>   | <ul style="list-style-type: none"> <li>• Object Storage Migration Service (OMS) is recommended for migration between object storage services.</li> <li>• Binary files cannot be imported to a database or NoSQL.</li> </ul>  |

| Category    | Source | Destination  | Description  |
|-------------|--------|--|--|
| File system | FTP    | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> <li>• Object-based storage: Object Storage Service (OBS)</li> </ul> | <ul style="list-style-type: none"> <li>• The file system cannot serve as the destination.</li> <li>• Only text files such as CSV files can be migrated from FTP or SFTP servers to search services. Binary files cannot.</li> <li>• Only binary files can be migrated from FTP or SFTP servers to OBS.</li> <li>• obsutil is recommended for migrating data from HTTP servers to OBS. For details, see <a href="#">Introduction to obsutil</a>.</li> </ul> |
|             | SFTP   |  |  |
|             | HTTP   | Hadoop: MRS HDFS   |  |

| Category             | Source   | Destination  | Description  |
|----------------------|--|--|--|
| Relational database  | RDS for MySQL  | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, MRS Hive, and MRS Hudi</li> <li>Object-based storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> | <ul style="list-style-type: none"> <li>You are advised to use Data Replication Service (DRS) to migrate data between OLTP databases.</li> <li>RDS for MySQL does not support the SSL mode.</li> <li>Recommended Microsoft SQL Server version: 2005 or later</li> <li>The KingBase database and GaussDB can be connected using the PostgreSQL connector. The supported source and destination are the same as those of the PostgreSQL data source.</li> </ul> |
|                      | RDS for SQL Server   | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object-based storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>           |  |
|                      | RDS for PostgreSQL   | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object-based storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>           |  |
|                      | MySQL  | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, MRS Hive, and MRS Hudi</li> <li>Object-based storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>   |  |
|                      | PostgreSQL   |  |  |
|                      | Oracle   |  |  |
| Microsoft SQL Server | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object-based storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |  |  |

| Category | Source   | Destination  | Description  |
|----------|----------|--|--|
|          | SAP HANA | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>• Hadoop: MRS Hive</li> </ul> | <p>SAP HANA data sources have the following restrictions:</p> <ul style="list-style-type: none"> <li>• SAP HANA cannot serve as the destination.</li> <li>• Only the 2.00.050.00.159 2305219 version is supported.</li> <li>• Only the Generic Edition is supported.</li> <li>• BW/4 FOR HANA is not supported.</li> <li>• Only database names, table names, and column names consisting of English letters are supported. Special characters such as spaces and symbols are not allowed.</li> <li>• The following data types are supported: date, digit, Boolean, and character (except SHORTTEXT). Other data types such as binary are not supported.</li> <li>• During migration, tables cannot be automatically created at the destination.</li> </ul> |



| Category | Source                          | Destination   | Description   |
|----------|---------------------------------|---|---|
|          | Database Sharding               | <ul style="list-style-type: none"> <li>Data warehouse: Data Lake Insight (DLI)</li> <li>Hadoop: MRS HBase and MRS Hive</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> <li>Object-based storage: Object Storage Service (OBS)</li> </ul>   | Database shards cannot serve as the destination.  |
|          | ShenTong                        | <ul style="list-style-type: none"> <li>Hadoop: MRS Hive and MRS Hudi</li> </ul>   | -   |
| NoSQL    | Distributed Cache Service (DCS) | Hadoop: MRS HDFS, MRS HBase, and MRS Hive   | NoSQL except CloudTable cannot serve as the destination.<br><br>For how to migrate data from Redis to DCS, see <a href="#">Migrating Data from Self-Hosted Redis to DCS</a> . |
|          | Redis                           |   |   |
|          | Document Database Service       |   |   |
|          | MongoDB                         |   |   |
|          | CloudTable HBase                | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object-based storage: Object Storage Service (OBS)</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |   |

| Category       | Source                       | Destination  | Description   |
|----------------|------------------------------|--|---|
|                | Cassandra                    | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object-based storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>   |   |
| Message system | Data Ingestion Service (DIS) | Search: Cloud Search Service (CSS)   | The message system cannot serve as the destination.   |
|                | Apache Kafka                 |  |   |
|                | DMS Kafka                    |  |   |
|                | MRS Kafka                    | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object-based storage: Object Storage Service (OBS)</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> | <ul style="list-style-type: none"> <li>MRS Kafka cannot serve as the destination.</li> <li>Supported only by local storage and not in storage-compute decoupling scenarios</li> <li>Not supported by Ranger</li> <li>Not supported if SSL is enabled for ZooKeeper</li> </ul> |
| Search         | Elasticsearch                | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> </ul>  | Only the non-security mode is supported.  |
|                | Cloud Search Service (CSS)   | <ul style="list-style-type: none"> <li>Object-based storage: Object Storage Service (OBS)</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>  | You are advised to use Logstash to import data to CSS. For details, see <a href="#">Using Logstash to Import Data to Elasticsearch</a>  |

 NOTE

In the preceding table, the non-cloud data sources, such as MySQL, include on-premises MySQL, MySQL built on ECSs, or MySQL on the third-party cloud.

### Supported Data Sources in Entire DB Migration

Entire DB migration is used when an on-premises data center or a database created on an ECS needs to be synchronized to a database service or big data service on the cloud. It is suitable for offline database migration but not online real-time migration.

**Table 3-2** lists the data sources supporting entire DB migration using CDM.

**Table 3-2** Supported data sources in entire DB migration

| Category   | Data Source            | Read      | Write         | Description  |
|--|------------------------|-----------|---------------|--|
| Data warehouse   | Data Warehouse Service | Supported | Supported     | -  |
| Hadoop (available only for local storage, and not for storage-compute decoupling, Ranger, or ZooKeeper for which SSL is enabled) | MRS HBase              | Supported | Supported     | Entire DB migration only to MRS HBase<br>Recommended versions:<br><ul style="list-style-type: none"> <li>• 2.1.X</li> <li>• 1.3.X</li> </ul>   |
|  | MRS Hive               | Supported | Supported     | Entire DB migration only to a relational database<br>2.x versions are not supported. The following versions are recommended:<br><ul style="list-style-type: none"> <li>• 1.2.X</li> <li>• 3.1.X</li> </ul> |
|  | FusionInsight HBase    | Supported | Not supported | Recommended versions:<br><ul style="list-style-type: none"> <li>• 2.1.X</li> <li>• 1.3.X</li> </ul>  |

| Category | Data Source        | Read      | Write         | Description  |
|----------|--------------------|-----------|---------------|--|
|          | FusionInsight Hive | Supported | Not supported | <p>Entire DB migration only to a relational database</p> <p>2.x versions are not supported. The following versions are recommended:</p> <ul style="list-style-type: none"> <li>• 1.2.X</li> <li>• 3.1.X</li> </ul>                           |
|          | Apache HBase       | Supported | Not supported | <p>Recommended versions:</p> <ul style="list-style-type: none"> <li>• 2.1.X</li> <li>• 1.3.X</li> </ul>  |
|          | Apache Hive        | Supported | Not supported | <p>Entire DB migration only to a relational database</p> <p>2.x versions are not supported. The following versions are recommended:</p> <ul style="list-style-type: none"> <li>• 1.2.X</li> <li>• 3.1.X</li> </ul>                           |
|          | MRS Hudi           | Supported | Supported     | <p>Supported only by local storage and in storage-compute decoupling scenarios</p> <p>2.x versions are not supported. The following versions are recommended:</p> <ul style="list-style-type: none"> <li>• 1.2.X</li> <li>• 3.1.X</li> </ul> |

| Category            | Data Source          | Read      | Write         | Description  |
|---------------------|----------------------|-----------|---------------|--|
| Relational database | RDS for MySQL        | Supported | Supported     | Migration from OLTP to OLTP is not supported. In this scenario, you are advised to use the Data Replication Service (DRS). |
|                     | RDS for PostgreSQL   | Supported | Supported     |  |
|                     | RDS for SQL Server   | Supported | Supported     |  |
|                     | MySQL                | Supported | Not supported |  |
|                     | PostgreSQL           | Supported | Not supported |  |
|                     | Microsoft SQL Server | Supported | Not supported |  |
|                     | Oracle               | Supported | Not supported |  |

| Category | Data Source     | Read      | Write         | Description   |
|----------|-----------------|-----------|---------------|---|
|          | SAP HANA        | Supported | Not supported | <ul style="list-style-type: none"> <li>• Only the 2.00.050.00.15 92305219 version is supported.</li> <li>• Only the Generic Edition is supported.</li> <li>• BW/4 FOR HANA is not supported.</li> <li>• Only database names, table names, and column names consisting of English letters are supported. Special characters such as spaces and symbols are not allowed.</li> <li>• The following data types are supported: date, digit, Boolean, and character (except SHORTTEXT). Other data types such as binary are not supported.</li> <li>• During migration, tables cannot be automatically created at the destination.</li> </ul> |
|          | Dameng database | Supported | Not supported | Only to DWS and Hive  |

| Category | Data Source                     | Read          | Write     | Description                                      |
|----------|---------------------------------|---------------|-----------|--|
| NoSQL    | Distributed Cache Service (DCS) | Not supported | Supported | Only migration from MRS to DCS is supported.     |
|          | Document Database Service (DDS) | Supported     | Supported | Only migration between DDS and MRS is supported. |
|          | CloudTable                      | Supported     | Supported | -  |

## 3.2 Supported Data Sources (2.9.2.200)

CDM provides the following migration modes which support different data sources:

- **Table/File migration** in the import of data into a data lake or migration of data to the cloud. For details, see [Data Sources Supported by Table/File Migration](#).
- **Entire DB migration** in the import of data into a data lake or migration of data to the cloud. For details, see [Supported Data Sources in Entire DB Migration](#).

### NOTE

This section describes the data sources supported by CDM clusters of version 2.9.2.200. The supported data sources vary depending on the CDM cluster version.

### Data Sources Supported by Table/File Migration

Table/File migration can migrate data in tables or files.

[Table 3-3](#) describes the supported data sources.

**Table 3-3** Supported data sources during table/file migration

| Category       | Source                  | Destination  | Description  |
|----------------|-------------------------|--|--|
| Data warehouse | GaussDB(DWS)            | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS), Data Lake Insight (DLI), and MRS ClickHouse</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> </ul>   | The DWS physical machine management mode is not supported. |
|                | Data Lake Insight (DLI) | <ul style="list-style-type: none"> <li>• Object storage: Object Storage Service (OBS)</li> <li>• Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> | -  |
|                | MRS ClickHouse          | Data warehouse: MRS ClickHouse and Data Lake Insight (DLI)   | Recommended MRS ClickHouse version: 21.3.4.X               |



| Category | Source                       | Destination  | Description   |
|----------|------------------------------|--|---|
| Hadoop   | MRS HDFS                     | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> </ul>   | <ul style="list-style-type: none"> <li>• Supported by local storage. Only MRS Hive and MRS Hudi are supported in storage-compute decoupling scenarios.</li> <li>• Only MRS Hive is supported in Ranger scenarios.</li> <li>• Not supported if SSL is enabled for ZooKeeper</li> <li>• Recommended MRS HDFS versions: <ul style="list-style-type: none"> <li>- 2.8.X</li> <li>- 3.1.X</li> </ul> </li> <li>• Recommended MRS HBase versions: <ul style="list-style-type: none"> <li>- 2.1.X</li> <li>- 1.3.X</li> </ul> </li> <li>• MRS Hive and MRS Hudi 2.x versions are not supported. The following versions are recommended: <ul style="list-style-type: none"> <li>- 1.2.X</li> <li>- 3.1.X</li> </ul> </li> </ul> |
|          | MRS HBase                    | <ul style="list-style-type: none"> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object storage: Object Storage Service (OBS)</li> <li>• Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>  |   |
|          | MRS Hive                     | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS), Data Lake Insight (DLI), and MRS ClickHouse</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object storage: Object Storage Service (OBS)</li> <li>• Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |   |
| MRS Hudi | Data warehouse: GaussDB(DWS) |  |   |

| Category            | Source  | Destination  | Description  |
|---------------------|---|--|--|
|                     | FusionInsight HDFS  | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> </ul> | <ul style="list-style-type: none"> <li>FusionInsight cannot serve as the destination.</li> <li>Supported only by local storage and not in storage-compute decoupling scenarios</li> <li>Not supported by Ranger</li> <li>Not supported if SSL is enabled for ZooKeeper</li> <li>Recommended FusionInsight HDFS versions:                             <ul style="list-style-type: none"> <li>- 2.8.X</li> <li>- 3.1.X</li> </ul> </li> <li>Recommended FusionInsight HBase versions:                             <ul style="list-style-type: none"> <li>- 2.1.X</li> <li>- 1.3.X</li> </ul> </li> <li>Recommended FusionInsight Hive versions:                             <ul style="list-style-type: none"> <li>- 1.2.X</li> <li>- 3.1.X</li> </ul> </li> </ul> |
| FusionInsight HBase | <ul style="list-style-type: none"> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> </ul>   |  |  |
| FusionInsight Hive  | <ul style="list-style-type: none"> <li>Object storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |  |  |

| Category       | Source   | Destination  | Description  |
|----------------|--|--|--|
|                | <ul style="list-style-type: none"> <li>Apache HBase</li> <li>Apache Hive</li> <li>Apache HDFS</li> </ul> | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object storage: Object Storage Service (OBS)</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> | <ul style="list-style-type: none"> <li>• Apache cannot serve as the destination.</li> <li>• Supported only by local storage and not in storage-compute decoupling scenarios</li> <li>• Not supported by Ranger</li> <li>• Not supported if SSL is enabled for ZooKeeper</li> <li>• Recommended Apache HBase versions:               <ul style="list-style-type: none"> <li>- 2.1.X</li> <li>- 1.3.X</li> </ul> </li> <li>• Apache Hive 2.x versions are not supported. The following versions are recommended:               <ul style="list-style-type: none"> <li>- 1.2.X</li> <li>- 3.1.X</li> </ul> </li> <li>• Recommended Apache HDFS versions:               <ul style="list-style-type: none"> <li>- 2.8.X</li> <li>- 3.1.X</li> </ul> </li> </ul> |
| Object storage | Object Storage Service (OBS)   | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>   | <ul style="list-style-type: none"> <li>• Object Storage Migration Service (OMS) is recommended for migration between object storage services.</li> <li>• Binary files cannot be imported to a database or NoSQL.</li> </ul>  |

| Category            | Source             | Destination  | Description  |
|---------------------|--------------------|--|--|
| File system         | FTP                | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>   | <ul style="list-style-type: none"> <li>The file system cannot serve as the destination.</li> <li>Only text files such as CSV files can be migrated from FTP or SFTP servers to search services. Binary files cannot.</li> <li>obsutil is recommended for migrating data from HTTP servers to OBS. For details, see <a href="#">Introduction to obsutil</a>.</li> </ul>   |
|                     | SFTP               |  |  |
|                     | HTTP               | Hadoop: MRS HDFS   |  |
| Relational database | RDS for MySQL      | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, MRS Hive, and MRS Hudi</li> <li>Object storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> | <ul style="list-style-type: none"> <li>You are advised to use Data Replication Service (DRS) to migrate data between OLTP databases.</li> <li>RDS for MySQL does not support the SSL mode.</li> <li>Recommended Microsoft SQL Server version: 2005 or later</li> <li>The KingBase database and GaussDB can be connected using the PostgreSQL connector. The supported source and destination are the same as those of the PostgreSQL data source.</li> </ul> |
|                     | RDS for SQL Server | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>           |  |
|                     | RDS for PostgreSQL | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object storage: Object Storage Service (OBS)</li> <li>NoSQL: CloudTable</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>           |  |

| Category | Source               | Destination  | Description |
|----------|----------------------|--|-------------|
|          | MySQL                | <ul style="list-style-type: none"> <li>● Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>● Hadoop: MRS HDFS, MRS HBase, MRS Hive, and MRS Hudi</li> <li>● Object-based storage: Object Storage Service (OBS)</li> <li>● NoSQL: CloudTable</li> <li>● Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |             |
|          | PostgreSQL           |  |             |
|          | Oracle               |  |             |
|          | Microsoft SQL Server |  |             |

| Category | Source   | Destination   | Description  |
|----------|----------|---|--|
|          | SAP HANA | <ul style="list-style-type: none"> <li>• Data warehouse: Data Lake Insight (DLI)</li> <li>• Hadoop: MRS Hive</li> </ul> | <p>SAP HANA data sources have the following restrictions:</p> <ul style="list-style-type: none"> <li>• SAP HANA cannot serve as the destination.</li> <li>• Only the 2.00.050.00.159 2305219 version is supported.</li> <li>• Only the Generic Edition is supported.</li> <li>• BW/4 FOR HANA is not supported.</li> <li>• Only database names, table names, and column names consisting of English letters are supported. Special characters such as spaces and symbols are not allowed.</li> <li>• The following data types are supported: date, digit, Boolean, and character (except SHORTTEXT). Other data types such as binary are not supported.</li> <li>• During migration, tables cannot be automatically created at the destination.</li> </ul> |

| Category  | Source   | Destination   | Description   |
|-----------|--|---|---|
|           | Database sharding  | <ul style="list-style-type: none"> <li>• Data warehouse: Data Lake Insight (DLI)</li> <li>• Hadoop: MRS HBase and MRS Hive</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> <li>• Object-based storage: Object Storage Service (OBS)</li> </ul>   | <p>Database shards cannot serve as the destination.</p> <p>A shard link connects to multiple backend data sources at the same time. The link can be used as the job source to migrate data from those data sources to other data sources.</p> |
| NoSQL     | Redis  | Hadoop: MRS HDFS, MRS HBase, and MRS Hive   | NoSQL except CloudTable cannot serve as the destination.  |
|           | Document Database Service (DDS)  |   |   |
|           | MongoDB  |   |   |
|           | CloudTable HBase   | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object storage: Object Storage Service (OBS)</li> <li>• Relational database: RDS for MySQL, RDS for PostgreSQL, RDS for SQL Server, MySQL, PostgreSQL, Microsoft SQL Server, and Oracle</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |   |
| Cassandra | <ul style="list-style-type: none"> <li>• Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>• Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>• Object storage: Object Storage Service (OBS)</li> <li>• NoSQL: CloudTable</li> <li>• Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> |   |   |

| Category       | Source                       | Destination  | Description   |
|----------------|------------------------------|--|---|
| Message system | Data Ingestion Service (DIS) | Search: Cloud Search Service (CSS)   | The message system cannot serve as the destination.   |
|                | Apache Kafka                 |  |   |
|                | DMS Kafka                    |  |   |
|                | MRS Kafka                    | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> <li>Object storage: Object Storage Service (OBS)</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul> | <ul style="list-style-type: none"> <li>MRS Kafka cannot serve as the destination.</li> <li>Supported only by local storage and not in storage-compute decoupling scenarios</li> <li>Not supported by Ranger</li> <li>Not supported if SSL is enabled for ZooKeeper</li> </ul> |
| Search         | Elasticsearch                | <ul style="list-style-type: none"> <li>Data warehouse: GaussDB(DWS) and Data Lake Insight (DLI)</li> <li>Hadoop: MRS HDFS, MRS HBase, and MRS Hive</li> </ul>  | Only the non-security mode is supported.  |
|                | Cloud Search Service (CSS)   | <ul style="list-style-type: none"> <li>Object storage: Object Storage Service (OBS)</li> <li>Relational database: RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server</li> <li>NoSQL: CloudTable</li> <li>Search: Elasticsearch and Cloud Search Service (CSS)</li> </ul>  | You are advised to use Logstash to import data to CSS. For details, see <a href="#">Using Logstash to Import Data to Elasticsearch</a>  |

 **NOTE**

In the preceding table, the non-cloud data sources, such as MySQL, include on-premises MySQL, MySQL built on ECSs, or MySQL on the third-party cloud.

## Supported Data Sources in Entire DB Migration

Entire DB migration is used when an on-premises data center or a database created on an ECS needs to be synchronized to a database service or big data



service on the cloud. It is suitable for offline database migration but not online real-time migration.

**Table 3-4** lists the data sources supporting entire DB migration using CDM.

**Table 3-4** Supported data sources in entire DB migration

| Category   | Data Source                  | Read      | Write         | Description  |
|--|------------------------------|-----------|---------------|--|
| Data warehouse   | Data Warehouse Service (DWS) | Supported | Supported     | -  |
| Hadoop (available only for local storage, and not for storage-compute decoupling, Ranger, or ZooKeeper for which SSL is enabled) | MRS HBase                    | Supported | Supported     | Entire DB migration only to MRS HBase<br>Recommended versions:<br><ul style="list-style-type: none"> <li>• 2.1.X</li> <li>• 1.3.X</li> </ul>   |
|  | MRS Hive                     | Supported | Supported     | Entire DB migration only to a relational database<br>2.x versions are not supported. The following versions are recommended:<br><ul style="list-style-type: none"> <li>• 1.2.X</li> <li>• 3.1.X</li> </ul> |
|  | FusionInsight HBase          | Supported | Not supported | Recommended versions:<br><ul style="list-style-type: none"> <li>• 2.1.X</li> <li>• 1.3.X</li> </ul>  |
|  | FusionInsight Hive           | Supported | Not supported | Entire DB migration only to a relational database<br>2.x versions are not supported. The following versions are recommended:<br><ul style="list-style-type: none"> <li>• 1.2.X</li> <li>• 3.1.X</li> </ul> |

| Category            | Data Source          | Read      | Write         | Description   |
|---------------------|----------------------|-----------|---------------|---|
|                     | Apache HBase         | Supported | Not supported | Recommended versions: <ul style="list-style-type: none"> <li>• 2.1.X</li> <li>• 1.3.X</li> </ul>  |
|                     | Apache Hive          | Supported | Not supported | Entire DB migration only to a relational database<br>2.x versions are not supported. The following versions are recommended: <ul style="list-style-type: none"> <li>• 1.2.X</li> <li>• 3.1.X</li> </ul> |
| Relational database | RDS for MySQL        | Supported | Supported     | Migration from OLTP to OLTP is not supported. In this scenario, you are advised to use the Data Replication Service (DRS).  |
|                     | RDS for PostgreSQL   | Supported | Supported     |   |
|                     | RDS for SQL Server   | Supported | Supported     |   |
|                     | MySQL                | Supported | Not supported |   |
|                     | PostgreSQL           | Supported | Not supported |   |
|                     | Microsoft SQL Server | Supported | Not supported |   |
|                     | Oracle               | Supported | Not supported |   |

| Category | Data Source     | Read      | Write         | Description   |
|----------|-----------------|-----------|---------------|---|
|          | SAP HANA        | Supported | Not supported | <ul style="list-style-type: none"> <li>• Only the 2.00.050.00.15 92305219 version is supported.</li> <li>• Only the Generic Edition is supported.</li> <li>• BW/4 FOR HANA is not supported.</li> <li>• Only database names, table names, and column names consisting of English letters are supported. Special characters such as spaces and symbols are not allowed.</li> <li>• The following data types are supported: date, digit, Boolean, and character (except SHORTTEXT). Other data types such as binary are not supported.</li> <li>• During migration, tables cannot be automatically created at the destination.</li> </ul> |
|          | Dameng database | Supported | Not supported | Only to DWS and Hive  |

| Category | Data Source                     | Read      | Write     | Description                                      |
|----------|---------------------------------|-----------|-----------|--|
| NoSQL    | Redis                           | Supported | Supported | -  |
|          | Document Database Service (DDS) | Supported | Supported | Only migration between DDS and MRS is supported. |
|          | CloudTable Service (CloudTable) | Supported | Supported | -  |

### 3.3 Supported Data Types

To ensure that data is completely imported to the migration destination, correctly configure field mappings based on data types supported for different data sources. For details, see [Table 3-5](#).

**Table 3-5** Supported data types

| Data Connection Type                     | Data Type  |
|--|--|
| MySQL                                    | <a href="#">Data Types Supported in MySQL Database Migration</a>             |
| SQL Server                               | <a href="#">Data Types Supported in SQL Server Database Migration</a>        |
| Oracle                                   | <a href="#">Data Types Supported in Oracle Database Migration</a>            |
| PostgreSQL                               | <a href="#">Data Types Supported in PostgreSQL Database Migration</a>        |
| ShenTong                                 | <a href="#">Data Types Supported in ShenTong Database Migration</a>          |
| SAP HANA                                 | <a href="#">Data Types Supported in SAP HANA Database Migration</a>          |
| DWS                                      | <a href="#">Data Types Supported in DWS Database Migration</a>               |
| Dameng                                   | <a href="#">Data Types Supported in Dameng Database Migration</a>            |
| DLI                                      | <a href="#">Data Types Supported in DLI Database Migration</a>               |
| Elasticsearch/Cloud Search Service (CSS) | <a href="#">Data Types Supported in Elasticsearch/CSS Database Migration</a> |

## Data Types Supported in MySQL Database Migration

When the source end is a MySQL database and the destination end is a Hive or DWS database, the following data types are supported:

**Table 3-6** Data types supported for the open-source MySQL database

| Category         | Type          | Description  | Storage Format Example | Hive    | DWS  |
|------------------|---------------|--|------------------------|---------|--|
| Character string | CHAR(M)       | A fixed-length string of 1 to 255 characters, for example, CHAR(5). The length limit is not mandatory. It is set to 1 by default.  | 'a' or 'aaaaa'         | CHAR    | CHAR   |
|                  | VARCHAR(M)    | A variable-length string consists of 1 to 255 characters (more than 255 characters for MySQL of a later version). Example: VARCHAR(25).<br>When creating a field of the VARCHAR type, you must define the length.              | 'a' or 'aaaaa'         | VARCHAR | VARCHAR  |
| Value            | DECIMAL(M, D) | Uncompressed floating-point numbers cannot be unsigned. In unpacking decimals, each decimal corresponds to a byte. Defining the number of display lengths (M) and decimals (D) is required. NUMERIC is the synonym of DECIMAL. | 52.36                  | DECIMAL | When D is 0, it corresponds to BIGINT. When D is not 0, it corresponds to NUMERIC. |
|                  | NUMERIC       | Same as DECIMAL  | -                      | DECIMAL | NUMERIC  |

| Category | Type             | Description   | Storage Format Example | Hive   | DWS     |
|----------|------------------|---|------------------------|--------|---------|
|          | INTEGER          | An integer of normal size that can be signed. If the value is signed, it ranges from -2147483648 to 2147483647.<br><br>If the value is unsigned, the value ranges from 0 to 4294967295. Up to 11-bit width can be specified.            | 5236                   | INT    | INTEGER |
|          | INTEGER UNSIGNED | Unsigned form of INTEGER  | -                      | BIGINT | INTEGER |
|          | INT              | Same as INTEGER   | 5236                   | INT    | INTEGER |
|          | INT UNSIGNED     | Same as INTEGER UNSIGNED  | -                      | BIGINT | INTEGER |
|          | BIGINT           | A large integer that can be signed. If the value is signed, it ranges from -9223372036854775808 to 9223372036854775807. If the value is unsigned, the value ranges from 0 to 18446744073709551615. Up to 20-bit width can be specified. | 5236                   | BIGINT | BIGINT  |
|          | BIGINT UNSIGNED  | Unsigned form of BIGINT   | -                      | BIGINT | BIGINT  |

| Category | Type               | Description   | Storage Format Example | Hive     | DWS      |
|----------|--------------------|---|------------------------|----------|----------|
|          | MEDIUMINT          | A medium-sized integer that can be signed. If the value is signed, it ranges from -8388608 to 8388607.<br>If the value is unsigned, it ranges from 0 to 16777215, and you can specify a maximum of 9-bit width. | -128, 127              | INT      | INTEGER  |
|          | MEDIUMINT UNSIGNED | Unsigned form of MEDIUMINT  | -                      | BIGINT   | INTEGER  |
|          | TINYINT            | A very small integer that can be signed. If signed, the value ranges from -128 to 127.<br>If unsigned, the value ranges from 0 to 255, and you can specify a maximum of 4-bit width.                            | 100                    | TINYINT  | SMALLINT |
|          | TINYINT UNSIGNED   | Unsigned form of TINYINT  | -                      | TINYINT  | SMALLINT |
|          | BOOL               | The bool of MySQL is tinyint(1).  | -128, 127              | SMALLINT | BYTEA    |
|          | SMALLINT           | A small integer that can be signed. If the value is signed, it ranges from -32768 to 32767.<br>If unsigned, the value ranges from 0 to 65535, and you can specify a maximum of 5-bit width.                     | 9999                   | SMALLINT | SMALLINT |

| Category | Type              | Description  | Storage Format Example | Hive   | DWS      |
|----------|-------------------|--|------------------------|--------|----------|
|          | SMALLINT UNSIGNED | Unsigned form of SMALLINT  | -                      | INT    | SMALLINT |
|          | REAL              | Same as DOUBLE   | -                      | DOUBLE | -        |
|          | FLOAT(M,D)        | Unsigned floating-point numbers cannot be used. The display length (M) and number of decimal places (D) can be specified. This is not mandatory, and the default value is 10,2. In the preceding information, 2 indicates the number of decimal places and 10 indicates the total number of digits (including decimal places). The decimal precision can reach 24 floating points. | 52.36                  | FLOAT  | FLOAT4   |
|          | DOUBLE(M,D)       | Unsigned double-precision floating-point numbers cannot be used. The display length (M) and number of decimal places (D) can be specified. This is not mandatory.<br><br>The default value is 16,4, where 4 is the number of decimal places. The decimal precision can reach 53-digit. REAL is a synonym of DOUBLE.  | 52.36                  | DOUBLE | FLOAT8   |
|          | DOUBLE PRECISION  | Similar to DOUBLE  | 52.3                   | DOUBLE | FLOAT8   |



| Category      | Type      | Description   | Storage Format Example  | Hive                   | DWS       |
|---------------|-----------|---|-------------------------|------------------------|-----------|
| Bit           | BIT(M)    | Stored bit type value. BIT(M) can store up to <i>M</i> bits of values, and <i>M</i> ranges from 1 to 64.  | B'1111100'<br>B'1100'   | TINYINT                | BYTEA     |
| Time and date | DATE      | The value is in the <i>YYYY-MM-DD</i> format and ranges from <b>1000-01-01</b> to <b>9999-12-31</b> . For example, <b>December 30, 1973</b> will be stored as <b>1973-12-30</b> .   | 1999-10-01              | DATE                   | TIMESTAMP |
|               | TIME      | Stores information about the hour, minute, and second.  | '09:10:21' or '9:10:21' | Not supported (string) | TIME      |
|               | DATETIME  | The date and time are in the <i>YYYY-MM-DD HH:MM:SS</i> format and range from <b>1000-01-01 00:00:00</b> to <b>9999-12-31 23:59:59</b> . For example, <b>3:30 p.m. on December 30, 1973</b> will be stored as <b>1973-12-30 15:30:00</b> .                            | '1973-12-30 15:30:00'   | TIMESTAMP              | TIMESTAMP |
|               | TIMESTAMP | Timestamp type. Timestamp between midnight on January 1, 1970 and a time point in 2037. Similar to the DATETIME format (YYYYMMDDHHMMSS), except that no hyphen is required. For example, <b>3:30 p.m. December 30, 1973</b> will be stored as <b>19731230153000</b> . | 19731230153000          | TIMESTAMP              | TIMESTAMP |

| Category             | Type         | Description  | Storage Format Example   | Hive                   | DWS           |
|----------------------|--------------|--|--------------------------|------------------------|---------------|
|                      | YEAR(M)      | The year is stored in 2-digit or 4-digit number format. If the length is specified as 2 (for example, YEAR(2)), the year ranges from 1970 to 2069 (70 to 69). If the length is specified as 4, the year ranges from 1901 to 2155. The default length is 4. | 2000                     | Not supported (string) | Not supported |
| Multi media (binary) | BINARY(M)    | The number of bytes is <i>M</i> . The length of a variable-length binary string ranges from 0 to <i>M</i> . <i>M</i> is the value length plus 1.   | 0x2A3B4058 (binary data) | Not supported          | BYTEA         |
|                      | VARBINARY(M) | The number of bytes is <i>M</i> . A fixed binary string with a length of 0 to <i>M</i> .   | 0x2A3B4059 (binary data) | Not supported          | BYTEA         |
|                      | TEXT         | The maximum length of the field is 65535 characters. TEXT is a "binary large object" and is used to store large binary data, such as images or other types of files.   | 0x5236 (binary data)     | Not supported          | Not supported |
|                      | TINYTEXT     | A binary string of 0 to 255 bytes in short text  | -                        | -                      | Not supported |
|                      | MEDIUMTEXT   | A binary string of 0 to 167772154 bytes in medium-length text  | -                        | -                      | Not supported |
|                      | LONGTEXT     | A binary string of 0 to 4294967295 bytes in large-length text  | -                        | -                      | Not supported |

| Category     | Type       | Description   | Storage Format Example | Hive          | DWS                  |
|--------------|------------|---|------------------------|---------------|----------------------|
|              | BLOB       | The maximum length of the field is 65535 characters. BLOB is a "binary large object" and is used to store large binary data, such as images or other types of files. BLOB is case-sensitive.  | 0x5236 (binary data)   | Not supported | Not supported        |
|              | TINYBLOB   | A binary string of 0 to 255 bytes in short text   | -                      | Not supported | Not supported        |
|              | MEDIUMBLOB | A binary string of 0 to 167772154 bytes in medium-length text   | -                      | Not supported | Not supported        |
|              | LONG BLOB  | A binary string of 0 to 4294967295 bytes in large-length text   | 0x5236 (binary data)   | Not supported | Not supported        |
| Special type | SET        | SET is a string object that can have no or multiple values. The values come from the allowed column of values specified when the table is created. When specifying the SET column values that contain multiple SET members, separate the members with commas (,). The SET member value cannot contain commas (,). | -                      | -             | Not supported        |
|              | JSON       | -   | -                      | Not supported | Not supported (TEXT) |

| Category | Type | Description   | Storage Format Example | Hive          | DWS           |
|----------|------|---|------------------------|---------------|---------------|
|          | ENUM | When an ENUM is defined, a list of its values is created, which are the items that must be used for selection (or NULL). For example, if you want a field to contain "A", "B", or "C", you can define an ENUM ("A", "B", or "C"). Only these values (or NULL) can be used to fill in the field. | -                      | Not supported | Not supported |

### Data Types Supported in Oracle Database Migration

When the source end is an Oracle database and the destination end is a Hive or DWS database, the following data sources are supported:

**Table 3-7** Data types supported for the Oracle database

| Category         | Type      | Description   | Hive    | DWS     |
|------------------|-----------|---|---------|---------|
| Character string | char      | Fixed-length character string, which is padded with spaces to reach the maximum length.   | CHAR    | CHAR    |
|                  | nchar     | Fixed-length character string contains data in Unicode format.  | CHAR    | CHAR    |
|                  | varchar2  | Synonym of VARCHAR. It is a variable-length string, unlike the CHAR type, which does not pad the field or variable to reach its maximum length with spaces. | VARCHAR | VARCHAR |
|                  | nvarchar2 | Variable-length character string contains data in Unicode format.   | VARCHAR | VARCHAR |
| Value            | number    | Stores numbers with a precision of up to 38 digits.   | DECIMAL | NUMERIC |

| Category            | Type                           | Description   | Hive          | DWS                  |
|---------------------|--------------------------------|---|---------------|----------------------|
|                     | binary_float                   | 2-bit single-precision floating point number  | FLOAT         | FLOAT 8              |
|                     | binary_double                  | 64-bit double-precision floating point number   | DOUBLE        | FLOAT 8              |
|                     | long                           | A maximum of 2 GB character data can be stored.   | Not supported | Not supported        |
| Time and date       | date                           | 7-byte date/time data type, including seven attributes: century, year in the century, month, day in the month, hour, minute, and second.  | DATE          | TIMESTAMP            |
|                     | timestamp                      | 7-byte or 11-byte fixed-width date/time data type that contains decimals (seconds)  | TIMESTAMP     | TIMESTAMP            |
|                     | timestamp with time zone       | 3-byte timestamp, which supports the time zone.   | TIMESTAMP     | TIME WITH TIME ZONE  |
|                     | timestamp with local time zone | 7-byte or 11-byte fixed-width date/time data type. Time zone conversion occurs when data is inserted or read.   | TIMESTAMP     | Not supported (TEXT) |
|                     | interval year to month         | 5-byte fixed-width data type, which is used to store a time segment.  | Not supported | Not supported (TEXT) |
|                     | interval day to second         | 11-byte fixed-width data type, which is used to store a time segment. The time segment is stored in days/hours/minutes/seconds. The value can also contain nine decimal places (seconds). | Not supported | Not supported (TEXT) |
| Multimedia (binary) | raw                            | A variable-length binary data type. Character set conversion is not performed for data stored in this data type.  | Not supported | Not supported        |

| Category | Type     | Description   | Hive          | DWS           |
|----------|----------|---|---------------|---------------|
|          | long raw | Stores up to 2 GB binary information.   | Not supported | Not supported |
|          | blob     | A maximum of 4 GB data can be stored.   | Not supported | Not supported |
|          | clob     | In Oracle 10g and later versions, a maximum of (4 GB) x (database block size) bytes of data can be stored. CLOB contains the information for which character set conversion is to be performed. This data type is ideal for storing plain text information. | String        | Not supported |
|          | nclob    | This type can store a maximum of 4 GB data. When the character set is converted, this type is affected.   | Not supported | Not supported |
|          | bfile    | An Oracle directory object and a file name can be stored in the database column, and the file can be read through the Oracle directory object and file name.  | Not supported | Not supported |
| Others   | rowid    | It is the address of a row in the database table. It is 10 bytes long.  | Not supported | Not supported |
|          | urowid   | It is a common row ID and does not have a fixed rowid table.  | Not supported | Not supported |

## Data Types Supported in SQL Server Database Migration

When the source end is a SQL Server database and the destination end is a Hive, Oracle or DWS database, the following data sources are supported:

**Table 3-8** Data types supported for the SQL Server database

| Category         | Type | Description   | Hive | DWS  | Oracle |
|------------------|------|---|------|------|--------|
| String data type | char | Fixed-length character string, which is padded with spaces to reach the maximum length. | CHAR | CHAR | CHAR   |

| Category          | Type     | Description   | Hive     | DWS      | Oracle       |
|-------------------|----------|---|----------|----------|--------------|
|                   | nchar    | Fixed-length character string contains data in Unicode format.  | CHAR     | CHAR     | CHAR         |
|                   | varchar  | A variable-length string consists of 1 to 255 characters (more than 255 characters for MySQL of a later version). Example: VARCHAR(25). When creating a field of the VARCHAR type, you must define the length.  | VARCHAR  | VARCHAR  | VARCHAR      |
|                   | nvarchar | Stores variable-length Unicode character data, similar to varchar.  | VARCHAR  | VARCHAR  | VARCHAR      |
| Numeric data type | int      | int is stored in four bytes, where one binary bit represents a sign bit, and the other 31 binary bits represent a length and a size, and may represent all integers ranging from $-2^{31}$ to $2^{31} - 1$ .  | INT      | INTEGER  | INTEGER      |
|                   | bigint   | bigint is stored in eight bytes, where one binary bit represents a sign bit, and the other 63 binary bits represent a length and a size, and may represent all integers ranging from $-2^{63}$ to $2^{63} - 1$ .  | BIGINT   | BIGINT   | NUMBER       |
|                   | smallint | Data of the smallint type occupies two bytes of storage space. One binary bit indicates a positive or negative sign of an integer value, and the other 15 binary bits indicate a length and a size, and may represent all integers ranging from $-2^{15}$ to $2^{15}$ . | SMALLINT | SMALLINT | NUMBER       |
|                   | tinyint  | Tinyint data occupies one byte of storage space and can represent all integers ranging from 0 to 255.   | TINYINT  | TINYINT  | NUMBER       |
|                   | real     | The value can be a positive or negative decimal number.   | DOUBLE   | FLOAT4   | NUMBER       |
|                   | float    | The number of digits (in scientific notation) of the mantissa of a float value, which determines the precision and storage size   | FLOAT    | FLOAT8   | binary_float |
|                   | decimal  | Numeric data type with fixed precision and scale  | DECIMAL  | NUMERIC  | NUMBER       |

| Category                       | Type           | Description   | Hive                   | DWS                    | Oracle        |
|--------------------------------|----------------|---|------------------------|------------------------|---------------|
|                                | numeric        | Stores zero, positive, and negative fixed point numbers.  | DECIMAL                | NUMERIC                | NUMBER        |
| Date and time data type        | date           | Stores date data represented by strings.  | DATE                   | TIMESTAMP              | DATE          |
|                                | time           | Time of a day, which is recorded in the form of a character string.   | Not supported (string) | TIME                   | Not supported |
|                                | datetime       | Stores time and date data.  | TIMESTAMP              | TIMESTAMP              | Not supported |
|                                | datetime2      | Extended type of datetime, which has a larger data range. By default, the minimum precision is the highest, and the user-defined precision is optional.   | TIMESTAMP              | TIMESTAMP              | Not supported |
|                                | smalldatetime  | The smalldatetime type is similar to the datetime type. The difference is that the smalldatetime type stores data from January 1, 1900 to June 6, 2079. When the date and time precision is low, the smalldatetime type can be used. Data of this type occupies 4-byte storage space. | TIMESTAMP              | TIMESTAMP              | Not supported |
|                                | datetimeoffset | A time that uses the 24-hour clock and combined with date and the time zone.  | Not supported (string) | TIMESTAMP              | Not supported |
| Multimedia data types (binary) | text           | Stores text data.   | Not supported (string) | Not supported (string) | Not supported |
|                                | netxt          | The function of this type is the same as that of the text type. It is non-Unicode data with variable length.  | Not supported (string) | Not supported (string) | Not supported |



| Category           | Type        | Description   | Hive                   | DWS                    | Oracle        |
|--------------------|-------------|---|------------------------|------------------------|---------------|
|                    | image       | Variable-length binary data used to store pictures, catalog pictures, or paintings.   | Not supported (string) | Not supported (string) | Not supported |
|                    | binary      | Binary data with a fixed length of $n$ bytes, where $n$ ranges from 1 to 8,000.   | Not supported (string) | Not supported (string) | Not supported |
|                    | varbinary   | Variable-length binary data   | Not supported (string) | Not supported (string) | Not supported |
| Currency data type | money       | Stores currency values.   | Not supported (string) | Not supported (string) | Not supported |
|                    | small money | Similar to the money type, a currency symbol is prefixed to the input data. For example, the currency symbol of CNY is ¥.   | Not supported (string) | Not supported (string) | Not supported |
| Data type          | bit         | Bit data type. The value is 0 or 1. The length is 1 byte. A bit value is often used as a logical value to determine whether it is true(1) or false(0). If a non-zero value is entered, the system replaces it with 1. | Not supported          | Not supported          | Not supported |
| Other data types   | rowversion  | Each piece of data has a counter. The value of the counter increases when an insert or update operation is performed on a table that contains the <b>rowversion</b> column in the database.                           | Not supported          | Not supported          | Not supported |

| Category | Type              | Description  | Hive          | DWS           | Oracle        |
|----------|-------------------|--|---------------|---------------|---------------|
|          | unique identifier | A 16-byte globally unique identifier (GUID) is a unique number generated by the SQL Server based on the network adapter address and host CPU clock. Each GUID is a hexadecimal number ranging from 0 to 9 or a to f. | Not supported | Not supported | Not supported |
|          | cursor            | Cursor data type   | Not supported | Not supported | Not supported |
|          | sql_variant       | Stores any valid SQL Server data except the text, image, and timestamp data, which facilitates the development of the SQL Server.  | Not supported | Not supported | Not supported |
|          | table             | Stores the result set after a table or view is processed.  | Not supported | Not supported | Not supported |
|          | xml               | Data type of the XML data. XML instances can be stored in columns or variables of the XML type. The stored XML instance size cannot exceed 2 GB.   | Not supported | Not supported | Not supported |

## Data Types Supported in PostgreSQL Database Migration

When the source end is a PostgreSQL database and the destination end is Hive, DLI, or DWS, the following data types are supported:

**Table 3-9** Data types supported for the PostgreSQL database

| Category  | Type | Description  | Hive | DWS  | DLI                    |
|-----------|------|--|------|------|------------------------|
| Character | char | Fixed-length string, which is padded to a specified length with spaces on the right. | CHAR | CHAR | Not supported (string) |

| Category | Type         | Description  | Hive         | DWS          | DLI                    |
|----------|--------------|--|--------------|--------------|------------------------|
|          | varchar      | Variable-length string. Fields or variables are not padded to the maximum length with spaces.  | CARCHAR      | CARCHAR      | Not supported (string) |
| Value    | smallint     | The extension name int2 is stored in two bytes and ranges from -32768 to 32767.  | SMALLINT     | SMALLINT     | SMALLINT               |
|          | int          | The extension name int4 is stored in four bytes and ranges from -2147483648 to 2147483647.   | INTEGER      | INT          | INT                    |
|          | bigint       | The extension name int8 is stored in eight bytes and ranges from -9223372036854775808 to 9223372036854775807.  | BIGINT       | BIGINT       | BIGINT                 |
|          | decimal(p,s) | The precision p represents the number of valid digits stored in the value, and the scale s represents the number of digits after the decimal point that can be stored. The maximum value of p is 1000. | DECIMAL(P,S) | DECIMAL(P,S) | DECIMAL(P,S)           |

| Category      | Type        | Description  | Hive                   | DWS              | DLI                    |
|---------------|-------------|--|------------------------|------------------|------------------------|
|               | float       | 4-byte or 8-byte storage. float(n): For the single precision, the value of n ranges from 1 to 24, the number of valid precision digits is 6, and the length is four bytes. For the double precision, the value of n ranges from 25 to 53, the number of valid precision digits is 15, and the length is 8 bytes. | FLOAT/<br>DOUBLE       | FLOAT/<br>DOUBLE | FLOAT/<br>DOUBLE       |
|               | smallserial | Sequence data type, which is stored in smallint format   | SMALLINT               | SMALLINT         | SMALLINT               |
|               | serial      | Sequence data type, which is stored in int format  | INTEGER                | INT              | INT                    |
|               | bigserial   | Sequence data type, which is stored in bigint format   | BIGINT                 | BIGINT           | BIGINT                 |
| Time and date | date        | Stores the date.   | DATE                   | DATE             | DATE                   |
|               | timestamp   | Stores date and time data without time zones.  | TIMESTAMP              | TIMESTAMP        | Not supported (string) |
|               | timestamptz | Stores the date and time, including the time zone.   | TIMESTAMP              | TIMESTAMPZ       | Not supported (string) |
|               | time        | Time within one day, excluding the time zone   | Not supported (string) | TIME             | Not supported (string) |
|               | timez       | Time within one day, including the time zone   | Not supported (string) | TIMEZ            | Not supported (string) |

| Category      | Type     | Description   | Hive                   | DWS                    | DLI                    |
|---------------|----------|---|------------------------|------------------------|------------------------|
|               | interval | Time interval   | Not supported (string) | Not supported (string) | Not supported (string) |
| Bit string    | bit      | Fixed-length string, for example, <b>b'000101'</b>  | Not supported (string) | Not supported (string) | Not supported (string) |
|               | varbit   | Variable-length string, for example, <b>b'101'</b>  | Not supported (string) | Not supported (string) | Not supported (string) |
| Currency type | money    | The value is stored in eight bytes and ranges from -922337203685477.5808 to 922337203685477.5807. | DOUBLE                 | MONEY                  | DECIMAL(P,S)           |
| Boolean       | boolean  | The value is stored in one byte and can be <b>1</b> , <b>0</b> , or <b>NULL</b> .                 | BOOLEAN                | BOOLEAN                | BOOLEAN                |
| Text type     | text     | Variable-length text without a length limit   | Not supported (string) | Not supported (string) | Not supported (string) |

## Data Types Supported in DWS Database Migration

If the migration source is a DWS database, the following data types are supported.

**Table 3-10** Data types supported for the DWS database

| Category  | Type    | Description   |
|-----------|---------|---|
| Character | char    | Fixed-length string, which is padded to a specified length with spaces on the right.          |
|           | varchar | Variable-length string. Fields or variables are not padded to the maximum length with spaces. |
| Value     | double  | Stores double-precision floating-point numbers.   |

| Category      | Type         | Description   |
|---------------|--------------|---|
|               | decimal(p,s) | The precision p represents the number of valid digits stored in the value, and the scale s represents the number of digits after the decimal point that can be stored. The maximum value of p is 1000.  |
|               | numeric      | Stores zero, positive, and negative fixed point numbers.  |
|               | real         | Same as double  |
|               | int          | int is stored in four bytes, where one binary bit represents a sign bit, and the other 31 binary bits represent a length and a size, and may represent all integers ranging from $-2^{31}$ to $2^{31} - 1$ .  |
|               | bigint       | bigint is stored in eight bytes, where one binary bit represents a sign bit, and the other 63 binary bits represent a length and a size, and may represent all integers ranging from $-2^{63}$ to $2^{63} - 1$ .  |
|               | smallint     | Data of the smallint type occupies two bytes of storage space. One binary bit indicates a positive or negative sign of an integer value, and the other 15 binary bits indicate a length and a size, and may represent all integers ranging from $-2^{15}$ to $2^{15}$ . |
|               | tinyint      | Tinyint data occupies one byte of storage space and can represent all integers ranging from 0 to 255.   |
| Time and date | date         | Stores the date.  |
|               | timestamp    | Stores date and time data without time zones.   |
|               | time         | Time within one day, excluding the time zone  |
| Bit string    | bit          | Fixed-length string, for example, <b>b'000101'</b>  |
| Boolean       | boolean      | The value is stored in one byte and can be <b>1</b> , <b>0</b> , or <b>NULL</b> .   |
| Text type     | text         | Variable-length text without a length limit   |

## Data Types Supported in ShenTong Database Migration

When the source is a ShenTong database and the destination is MRS Hive or MRS Hudi, the following data types are supported.

**Table 3-11** Data types supported for the ShenTong database

| Category  | Type      | Description   | Storage Format Example | MRS Hive      | MRS Hudi      |
|-----------|-----------|---|------------------------|---------------|---------------|
| Character | VARCHAR   | Stores specified fixed-length character strings.                        | 'a' or 'aaaaa'         | VARCHAR(765)  | STRING        |
|           | BPCHAR    | Stores specified variable-length character strings.                     | 'a' or 'aaaaa'         | VARCHAR(765)  | STRING        |
| Value     | NUMERIC   | Stores zero, positive, and negative fixed point numbers.                | 52.36                  | DECIMAL(10,0) | DECIMAL(18,0) |
|           | INT       | Stores zero, positive, and negative fixed point numbers.                | 5236                   | INT           | INT           |
|           | BIGINT    | Stores signed integers. Integer part: 19 digits; decimal part: 0 digits | 5236                   | BIGINT        | BIGINT        |
|           | TINYINT   | Stores signed integers. Integer part: 3 digits; decimal part: 0 digits  | 100                    | SMALLINT      | INT           |
|           | BINARY    | Stores fixed-length binary data.  | 0x2A3B4058             | Not supported | FLOAT         |
|           | VARBINARY | Stores variable-length binary data.                                     | 0x2A3B4058             | Not supported | BINARY        |
|           | FLOAT     | Stores floating-point numbers with binary precision.                    | 52.36                  | FLOAT         | FLOAT         |
|           | DOUBLE    | Stores double-precision floating-point numbers.                         | 52.3                   | DOUBLE        | DOUBLE        |

| Category      | Type      | Description   | Storage Format Example  | MRS Hive      | MRS Hudi  |
|---------------|-----------|---|---|---------------|-----------|
| Time and date | DATE      | Stores information about the year, month, and day.                                      | '1999-10-01'<br>,<br>'1999/10/01'<br>, or<br>'1999.10.01'   | DATE          | DATE      |
|               | TIME      | Stores information about the hour, minute, and second.                                  | '09:10:21' or<br>'9:10:21'  | STRING        | STRING    |
|               | TIMESTAMP | Stores information about the year, month, day, hour, minute, and second.                | 2002-12-12 09:10:21',<br>'2002-12-12 09:10:21'<br>'2002/12/12 09:10:21' or<br>'2002.12.12 09:10:21' | TIMESTAMP     | TIMESTAMP |
| Multimedia    | CLOB      | Stores variable-length binary large objects with a maximum length of 2 GB minus 1 byte. | 0x5236 (binary data)  | STRING        | STRING    |
|               | BLOB      | Stores variable-length binary large objects with a maximum length of 2 GB minus 1 byte. | 0x5236 (binary data)  | Not supported | BINARY    |
| Boolean       | BOOLEAN   | The value is stored in one byte and can be <b>1</b> , <b>0</b> , or <b>NULL</b> .       | 1   | BOOLEAN       | BOOLEAN   |

## Data Types Supported in SAP HANA Database Migration

If the source is an SAP HANA database, the following data types are supported.



**Table 3-12** Data types supported for the SAP HANA database

| Category      | Type      | Description   |
|---------------|-----------|---|
| Character     | VARCHAR   | Stores specified fixed-length character strings.  |
|               | NVARCHAR  | Variable-length character string contains data in Unicode format.   |
|               | TEXT      | It is used to store long character strings. The maximum length of a string is 2 GB minus 1 byte. Long text strings are stored.  |
| Value         | BIGINT    | Stores signed integers. Integer part: 19 digits; decimal part: 0 digits   |
|               | TINYINT   | Stores signed integers. Integer part: 3 digits; decimal part: 0 digits  |
|               | SMALLINT  | Data of the smallint type occupies two bytes of storage space. One binary bit indicates a positive or negative sign of an integer value, and the other 15 binary bits indicate a length and a size, and may represent all integers ranging from $-2^{15}$ to $2^{15}$ . |
|               | REAL      | The value can be a positive or negative decimal number.   |
|               | DECIMAL   | Numeric data type with fixed precision and scale  |
|               | FLOAT     | Stores floating-point numbers with binary precision.  |
|               | DOUBLE    | Stores double-precision floating-point numbers.   |
| Time and date | DATE      | Stores information about the year, month, and day.  |
|               | TIME      | Stores information about the hour, minute, and second.  |
|               | TIMESTAMP | Stores information about the year, month, day, hour, minute, and second.  |
| Multi media   | CLOB      | Stores variable-length binary large objects with a maximum length of 2 GB minus 1 byte.   |
|               | NCLOB     | This type can store a maximum of 4 GB data. When the character set is converted, this type is affected.   |
| Boolean       | BOOLEAN   | The value is stored in one byte and can be <b>1</b> , <b>0</b> , or <b>NULL</b> .   |

## Data Types Supported in DLI Database Migration

If the migration source is a DLI database, the following data types are supported.

**Table 3-13** Data types supported for the DLI database

| Category      | Type      | Description   |
|---------------|-----------|---|
| Character     | CHAR      | Stores specified fixed-length character strings.  |
|               | VARCHAR   | Same as CHAR  |
|               | STRING    | It is used to store long character strings. The maximum length of a character string is 2 GB minus 1 byte. Long text strings are stored.  |
| Value         | BIGINT    | Stores signed integers. Integer part: 19 digits; decimal part: 0 digits   |
|               | TINYINT   | Stores signed integers. Integer part: 3 digits; decimal part: 0 digits  |
|               | SMALLINT  | Data of the smallint type occupies two bytes of storage space. One binary bit indicates a positive or negative sign of an integer value, and the other 15 binary bits indicate a length and a size, and may represent all integers ranging from $-2^{15}$ to $2^{15}$ . |
|               | INT       | Stores signed integers. Integer part: 10 digits; decimal part: 0 digits   |
|               | DECIMAL   | Numeric data type with fixed precision and scale  |
|               | FLOAT     | Stores floating-point numbers with binary precision.  |
|               | DOUBLE    | Stores double-precision floating-point numbers.   |
| Time and date | DATE      | Stores information about the year, month, and day.  |
|               | TIMESTAMP | Stores information about the year, month, day, hour, minute, and second.  |
| Boolean       | BOOLEAN   | The value is stored in one byte and can be <b>1</b> , <b>0</b> , or <b>NULL</b> .   |

## Data Types Supported in Elasticsearch/CSS Database Migration

If the migration source is an Elasticsearch/CSS database, the following data types are supported.

**Table 3-14** Data types supported for the Elasticsearch/CSS database

| Category  | Type    | Description               | Storage Format Example | MySQL  |
|-----------|---------|---------------------------|------------------------|--------|
| Character | keyword | Stores character strings. | "keyword"              | String |

| Category | Type    | Description   | Storage Format Example  | MySQL    |
|----------|---------|---|---|----------|
|          | text    | Stores long character strings. The maximum length of a character string is 2 GB minus 1 byte. Long text strings are stored. | "long string"   | TEXT     |
|          | string  | Stores long character strings. The maximum length of a character string is 2 GB minus 1 byte. Long text strings are stored. | "a string"  | String   |
| Integer  | short   | Stores 16-bit signed integers ranging from -32768 to 32767.   | 32765   | smallint |
|          | integer | Stores 32-bit signed integers ranging from $-2^{31}$ to $2^{31} - 1$ .  | 3276566   | int      |
|          | long    | Stores 64-bit signed integers ranging from $-2^{63}$ to $2^{63} - 1$ .  | 327656666   | bigint   |
| Value    | double  | 64-bit IEEE 754 double-precision floating-point format  | 21.333  | double   |
|          | float   | 32-bit IEEE 754 single-precision floating-point format  | 21.333  | double   |
| Boolean  | boolean | The value is stored in one byte and can be <b>1</b> , <b>0</b> , or <b>NULL</b> .   | 1   | Boolean  |
| Object   | object  | A string of flat storage objects  | {"users.name": ["John","Smith"],<br>users.age": [26,28],<br>"users.sex": [1,2]} | TEXT     |

| Category     | Type          | Description                               | Storage Format Example  | MySQL              |
|--------------|---------------|---|---|--------------------|
| Nested       | nested        | A string of nested storage objects        | <pre> {"users.name" : "John" , "users.age" : 26, "users.sex" : 1} { "users.name" : "Smith", "users.age" : 28, "users.sex" : 2} </pre> | TEXT               |
| Date         | date          | A string in the date format               | "2018-01-13" or "2018-01-13 12:10:30"   | DATE or time Stamp |
| Special type | ip            | A string in the IP address format         | "192.168.1 27.100"  | String             |
| Array        | string_array  | An array of strings                       | ["str","str"]   | TEXT               |
|              | short_array   | An array of 16-bit integers               | [1,1,1]   | TEXT               |
|              | integer_array | An array of 32-bit integers               | [1,1,1]   | TEXT               |
|              | long_array    | An array of 64-bit integers               | [1,1,1]   | TEXT               |
|              | float_array   | An array of 32-bit floating-point numbers | [1.0,1.0,1.0]   | TEXT               |
|              | double_array  | An array of 64-bit floating-point numbers | [1.0,1.0,1.0]   | TEXT               |
| Value range  | completion    | A string that is automatically completed  | "string"  | TEXT               |

## Data Types Supported in Dameng Database Migration

When the source end is a Dameng database and the destination end is a Hive or DWS database, the following data types are supported.

**Table 3-15** Data types supported for the Dameng database

| Category  | Type      | Description   | Storage Format Example | Hive    | DWS     |
|-----------|-----------|---|------------------------|---------|---------|
| Character | CHAR      | Stores specified fixed-length character strings.                        | 'a' or 'aaaaa'         | CHAR    | CHAR    |
|           | CHARACTER | Same as CHAR  | 'a' or 'aaaaa'         | CHAR    | CHAR    |
|           | VARCHAR   | Stores specified variable-length character strings.                     | 'a' or 'aaaaa'         | VARCHAR | VARCHAR |
|           | VARCHAR2  | Same as VARCHAR   | 'a' or 'aaaaa'         | VARCHAR | VARCHAR |
| Value     | NUMERIC   | Stores zero, positive, and negative fixed point numbers.                | 52.36                  | DECIMAL | NUMERIC |
|           | DECIMAL   | Similar to NUMERIC  | 52.36                  | DECIMAL | NUMERIC |
|           | DEC       | Same as DECIMAL   | 52.36                  | DECIMAL | NUMERIC |
|           | NUMBER    | Same as NUMERIC   | 52.36                  | DECIMAL | NUMERIC |
|           | INTEGER   | Stores signed integers. Integer part: 10 digits; decimal part: 0 digits | 5236                   | INT     | INTEGER |
|           | INT       | Same as INTEGER   | 5236                   | INT     | INTEGER |
|           | BIGINT    | Stores signed integers. Integer part: 19 digits; decimal part: 0 digits | 5236                   | BIGINT  | BIGINT  |

| Category      | Type             | Description  | Storage Format Example                                    | Hive              | DWS                      |
|---------------|------------------|--|---|-------------------|--------------------------|
|               | TINYINT          | Stores signed integers. Integer part: 3 digits; decimal part: 0 digits | 100   | TINYINT           | SMALLINT                 |
|               | SMALLINT         | Stores signed integers. Integer part: 5 digits; decimal part: 0 digits | 9999  | SMALLINT          | SMALLINT                 |
|               | BYTE             | Similar to TINYINT. Integer part: 3 digits; decimal part: 0 digits     | 100   | TINYINT           | SMALLINT                 |
|               | BINARY           | Stores fixed-length binary data.                                       | 0x2A3B4058  | BINARY (NULL)     | BYTEA (NULL)             |
|               | VARBINARY        | Stores variable-length binary data.                                    | 0x2A3B4058  | BINARY (NULL)     | BYTEA (NULL)             |
|               | FLOAT            | Stores floating-point numbers with binary precision.                   | 52.36   | FLOAT             | FLOAT8                   |
|               | DOUBLE           | Similar to FLOAT   | 52.36   | DOUBLE            | FLOAT8                   |
|               | REAL             | Stores binary floating-point numbers.                                  | 52.3  | FLOAT             | FLOAT4                   |
|               | DOUBLE PRECISION | Stores double-precision floating-point numbers.                        | 52.3  | DOUBLE            | FLOAT8                   |
| Bit string    | BIT              | Stores 1, 0, or NULL.  | 1, 0, or NULL   | TINYINT(1 0 NULL) | BOOLEAN(true false NULL) |
| Time and date | DATE             | Stores information about the year, month, and day.                     | '1999-10-01'<br>,<br>'1999/10/01'<br>, or<br>'1999.10.01' | DATE              | TIMESTAMP                |

| Category | Type                           | Description   | Storage Format Example   | Hive                   | DWS                      |
|----------|--------------------------------|---|--|------------------------|--------------------------|
|          | TIME                           | Stores information about the hour, minute, and second.  | '09:10:21' or '9:10:21'  | Not supported (string) | TIME                     |
|          | TIMESTAMP                      | Stores information about the year, month, day, hour, minute, and second.  | 2002-12-12 09:10:21', '2002-12-12 9:10:21'<br>'2002/12/12 09:10:21' or '2002.12.12 09:10:21'                         | TIMESTAMP              | TIMESTAMP                |
|          | TIME WITH TIME ZONE            | Stores a TIME value with a time zone. Add the time zone information to the end of the TIME type.  | '09:10:21 +8:00', '09:10:21+8:00' or '9:10:21+8:00'  | Not supported (string) | TIME WITH TIME ZONE      |
|          | TIMESTAMP WITH TIME ZONE       | Stores a TIMESTAMP value with a time zone. Add the time zone information to the end of the TIMESTAMP type.  | 2002-12-12 09:10:21 +8:00', '2002-12-12 9:10:21 +8:00'<br>'2002/12/12 09:10:21 +8:00' or '2002.12.12 09:10:21 +8:00' | TIMESTAMP              | TIMESTAMP WITH TIME ZONE |
|          | TIMESTAMP WITH LOCAL TIME ZONE | Stores the TIMESTAMP value of a local time zone. The standard time zone type (TIMESTAMP WITH TIME ZONE) can be converted to the local time zone type. | 2002-12-12 09:10:21 +8:00', '2002-12-12 9:10:21 +8:00'<br>'2002/12/12 09:10:21 +8:00' or '2002.12.12 09:10:21 +8:00' | Not supported (string) | Not supported (TEXT)     |

| Category | Type                    | Description  | Storage Format Example  | Hive                   | DWS                      |
|----------|-------------------------|--|---|------------------------|--------------------------|
|          | DATETIME WITH TIME ZONE | Same as TIMESTAMP WITH TIME ZONE   | 2002-12-12 09:10:21 +8:00', '2002-12-12 09:10:21 +8:00'<br>'2002/12/12 09:10:21 +8:00' or '2002.12.12 09:10:21 +8:00' | TIMESTAMP              | TIMESTAMP WITH TIME ZONE |
|          | INTERVAL YEAR           | Interval of years. The leading precision specifies the range of years.                   | INTERVAL '0015' YEAR  | Not supported (string) | Not supported (VARCHAR)  |
|          | INTERVAL YEAR TO MONTH  | Interval of months and years. The leading precision specifies the range of years.        | INTERVAL '0015-08' YEAR TO MONTH  | Not supported (string) | Not supported (VARCHAR)  |
|          | INTERVAL MONTH          | Interval of months. The leading precision specifies the range of months.                 | INTERVAL '0015' MONTH   | Not supported (string) | Not supported (VARCHAR)  |
|          | INTERVAL DAY            | Interval of days. The leading precision specifies the range of days.                     | INTERVAL '150' DAY  | Not supported (string) | Not supported (VARCHAR)  |
|          | INTERVAL DAY TO HOUR    | Interval of hours and days. The leading precision specifies the range of days.           | INTERVAL '9 23' DAY TO HOUR   | Not supported (string) | Not supported (VARCHAR)  |
|          | INTERVAL DAY TO MINUTE  | Interval of minutes, hours, and days. The leading precision specifies the range of days. | INTERVAL '09 23:12' DAY TO MINUTE   | Not supported (string) | Not supported (VARCHAR)  |



| Category | Type                      | Description   | Storage Format Example                 | Hive                   | DWS                     |
|----------|---------------------------|---|--|------------------------|-------------------------|
|          | INTERVAL DAY TO SECOND    | Interval of seconds, minutes, hours, and days. The leading precision specifies the range of days. | INTERVAL '09 23:12:01.1' DAY TO SECOND | Not supported (string) | Not supported (VARCHAR) |
|          | INTERVAL HOUR             | Interval of hours. The leading precision specifies the range of hours.                            | INTERVAL '150' HOUR                    | Not supported (string) | Not supported (VARCHAR) |
|          | INTERVAL HOUR TO MINUTE   | Interval of minutes and hours. The leading precision specifies the range of hours.                | INTERVAL '23:12' HOUR TO MINUTE        | Not supported (string) | Not supported (VARCHAR) |
|          | INTERVAL HOUR TO SECOND   | Interval of seconds, minutes, and hours. The leading precision specifies the range of hours.      | INTERVAL '23:12:01.1' HOUR TO SECOND   | Not supported (string) | Not supported (VARCHAR) |
|          | INTERVAL MINUTE           | Interval of minutes. The leading precision specifies the range of minutes.                        | INTERVAL '150' MINUTE                  | Not supported (string) | Not supported (VARCHAR) |
|          | INTERVAL MINUTE TO SECOND | Interval of seconds and minutes. The leading precision specifies the range of minutes.            | INTERVAL '12:01.1' MINUTE TO SECOND    | Not supported (string) | Not supported (VARCHAR) |
|          | INTERVAL SECOND           | Interval of seconds. The leading precision specifies the range of the integer part of the second. | INTERVAL '51.1' SECOND                 | Not supported (string) | Not supported (VARCHAR) |

| Category   | Type          | Description  | Storage Format Example   | Hive          | DWS           |
|------------|---------------|--|--------------------------|---------------|---------------|
| Multimedia | IMAGE         | IMAGE specifies the image type in the multimedia information.<br><br>An image consists of a pixel lattice with a maximum length of 2 GB minus 1 byte. In addition to storing image data, other binary data can also be stored. | 0x2A3B4058 (binary data) | Not supported | Not supported |
|            | LONGVARBINARY | Same as IMAGE  | 0x2A3B4059 (binary data) | Not supported | Not supported |
|            | TEXT          | Stores the long string type.<br><br>The maximum length of a string is 2 GB minus 1 byte.   | 0x5236 (binary data)     | Not supported | Not supported |
|            | LONGVARCHAR   | Similar to TEXT  | 0x5236 (binary data)     | Not supported | Not supported |
|            | BLOB          | Stores variable-length binary large objects with a maximum length of 2 GB minus 1 byte.  | 0x5236 (binary data)     | Not supported | Not supported |
|            | CLOB          | Stores variable-length binary large objects with a maximum length of 2 GB minus 1 byte.  | 0x5236 (binary data)     | Not supported | Not supported |

| Category | Type  | Description  | Storage Format Example | Hive          | DWS           |
|----------|-------|--|------------------------|---------------|---------------|
|          | BFILE | Specified the binary files stored in the operating systems. Files are stored in the operating systems instead of the databases. They can be read only. | -                      | Not supported | Not supported |

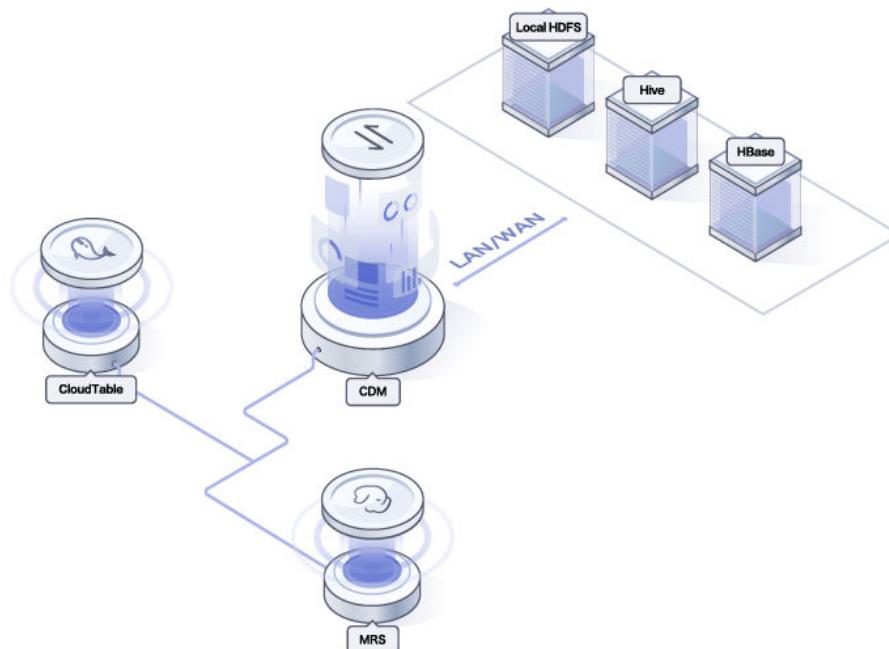
# 4 Application Scenarios

## Migrating Big Data to the Cloud

Local data is stored in the IDC that you have built or rent, or on the third-party cloud, including data stored in relational databases, NoSQL databases, OLAP databases, and file systems.

In this scenario, if you want to use the cloud computing and storage resources, you must migrate local data to the cloud in advance. Ensure that the local network can communicate with the cloud network.

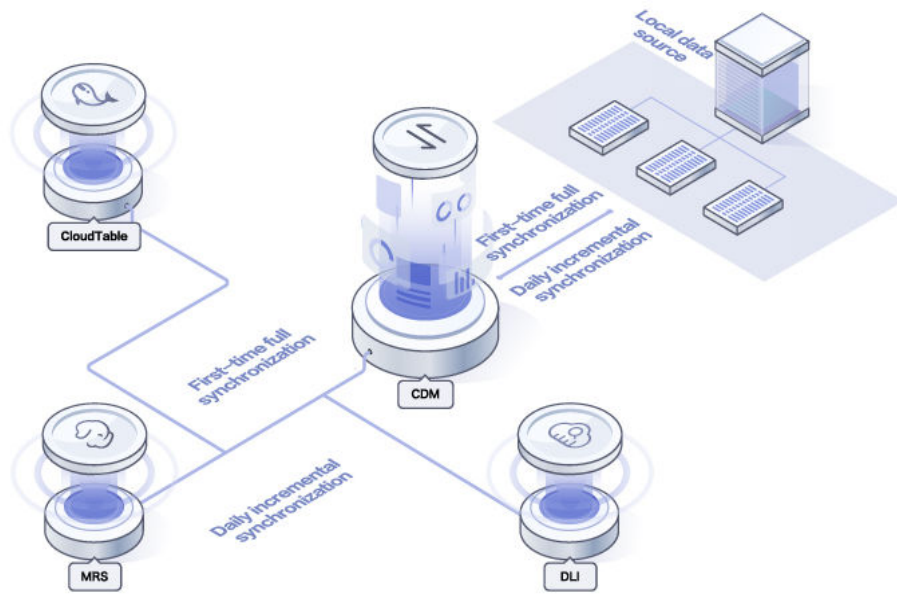
**Figure 4-1** Migrating Local Data to the Cloud



## Batch Importing Data to a Data Lake

You can import all your local data to a data lake, and import incremental data to the data lake one day later.

Figure 4-2 Batch importing data to a data lake



# 5 Basic Concepts

---

## CDM Cluster

A CDM cluster is a CDM instance. It consists of one or more VMs. You can create multiple CDM clusters for different purposes. For example, you can create a CDM cluster for the financial department and the procurement department respectively to isolate data access permissions.

## Local Environment

A local environment is a data storage system in the IDC that you have built or rent, or on the third-party cloud, including relational databases and file systems.

## Local Data

Local data is stored in the IDC that you have built or rent, or on the third-party cloud, including data stored in relational databases, NoSQL databases, OLAP databases, and file systems.

## Connector

A connector is a built-in object template used for connecting to a data source. Currently, CDM uses connectors to connect to OBS, MRS, and databases. New connectors can be added to CDM as well.

## Link

A link is an object set up based on a connector and used to connect to a specific data source.

To create a link, you must specify the link name, connector, data source address, and authentication information. For example, to connect to a MySQL database, you must set the host IP address, port number, username, and password.

After a link is set up, it can be used by multiple jobs as either a source or a destination link.

## Job

A job is a data migration task that you have created to migrate data from a specific data source to another. To create a job, you must specify a source link, destination link, and data mapping rules.

## Source Job Configuration

During job creation, the source link specifies the data source from which data is extracted. The job parameters of different source links vary. For example, the table or directory from which data is exported is specified in the job configuration at the source end.

## Destination Job Configuration

During job creation, the destination link specifies the data source to which data is loaded. The job parameters of different destination links vary. For example, the table or directory to which data is imported is specified in the job configuration of the destination end.

## Field Mapping

During job creation, especially jobs of migrating data between heterogeneous data sources, you must configure the mapping between the source and destination data sources, such as field mapping and field type mapping.

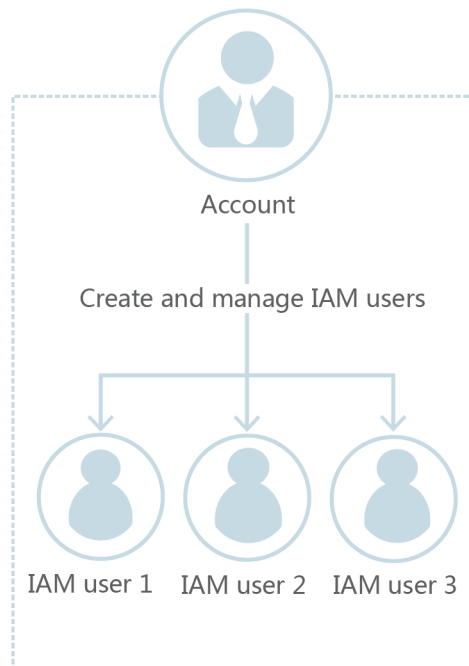
## Account

The account registered with the cloud owns your cloud resources and has full access permissions for the resources. You can use the account to reset user passwords and assign permissions. Your account receives and pays all bills generated by your IAM users' use of resources. To log in to the management console using an account, choose **Account Login**.

## IAM User

IAM users are created by an account to use cloud services. Each IAM user has their own password and access key to access cloud services using the console or APIs. The users manage cloud resources for the account based on assigned permissions. IAM users do not own resources or make payments. It is the account that controls user permissions and pays the bills. To log in to the management console as an IAM user, choose **IAM User Login**.

**Figure 5-1** Relationship between an account and its IAM users



## User Group

User groups facilitate centralized user management and streamlined permissions management. IAM users must be added to a user group to obtain the permissions required for accessing specified resources or cloud services in the account. A user can be added to multiple groups, which allows them to inherit different permissions.

The default user group **admin** has all of the permissions required to use all of the cloud resources. Users in this group can perform operations on all the resources, including but not limited to creating user groups and users, assigning permissions, and managing resources.

## Policy and Authorization

A policy is a set of permissions defined in JSON format. It defines which operations on which cloud resources are allowed. There are two types of policies: system policy and custom policy.

- System policies are pre-defined by IAM and cannot be modified.
- If system policies do not meet your requirements, you can create custom policies for fine-grained access control.

Authorization is the process of granting required permissions for a user to perform a task. After a system or custom policy is assigned to a user group, users in the group inherit the permissions defined by the policy to manage resources.

For example, the content of the **CDM Administrator** policy defining all permissions of CDM is as follows:

```
{  
  "Version": "1.1",
```



```

"Statement": [
  {
    "Action": [
      "cdm:*:*",
      "ecs:*:*",
      "vpc:*:*",
      "evs:*:*",
      "bss:*:*",
      "CTS:*:*",
      "eps:*:*",
      "obs:*:*",
      "CES:*:*"
    ],
    "Effect": "Allow"
  }
]

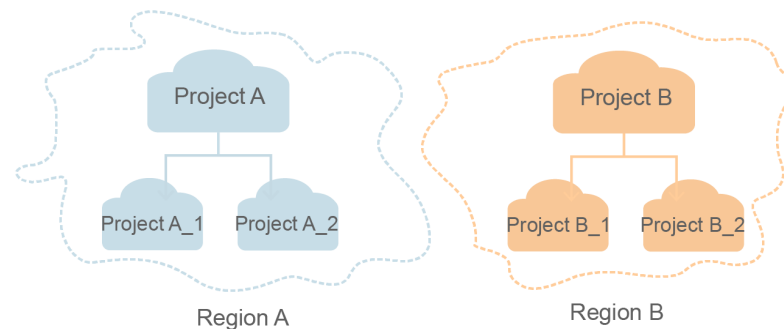
```

## Project

A project corresponds to a service region. Default projects are defined to group and physically isolate resources (including computing, storage, and network resources) across regions.

- Users can be granted permissions in a default project to access all resources in the region associated with the project.
- If you need more refined access control, you can create subprojects under a default project and create resources in subprojects. Then you can assign required permissions for users to access only the resources in specific subprojects.

**Figure 5-2** Project isolation model



## Identity Credentials

Identity credentials are used for authentication when you or your IAM users access cloud services through the console or APIs. Identity credentials include the password and access keys, which can be managed in IAM.

- Password: A common identity credential for logging in to the management console or calling cloud service APIs.
- Access key: An access key ID/secret access key (AK/SK) pair, which is used only for calling cloud service APIs. Each access key provides a signature for

cryptographic authentication to ensure that access requests are secret, complete, and correct.

# 6 Regions and AZs

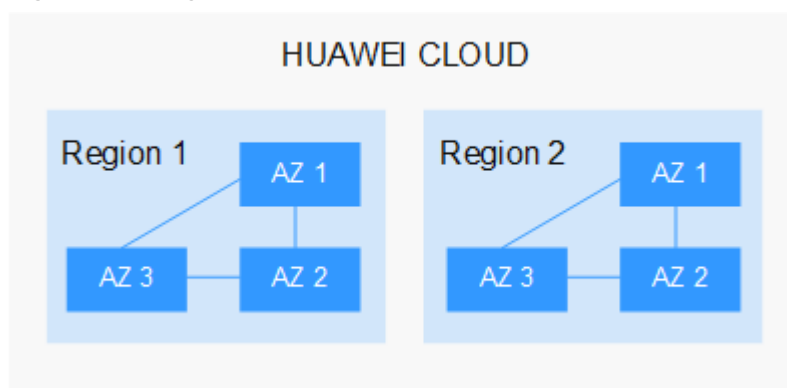
## Concept

A region and availability zone (AZ) identify the location of a data center. You can create resources in a specific region and AZ.

- Regions are divided from the dimensions of geographical location and network latency. Public services, such as Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Object Storage Service (OBS), Virtual Private Cloud (VPC), Elastic IP (EIP), and Image Management Service (IMS), are shared within the same region. Regions are classified as universal regions and dedicated regions. A universal region provides universal cloud services for common tenants. A dedicated region provides services of the same type only or for specific tenants.
- An AZ contains one or more physical data centers. Each AZ has independent cooling, fire extinguishing, moisture-proof, and electricity facilities. Within an AZ, computing, network, storage, and other resources are logically divided into multiple clusters. AZs within a region are interconnected using high-speed optical fibers to allow you to build cross-AZ high-availability systems.

**Figure 6-1** shows the relationship between regions and AZs.

**Figure 6-1** Regions and AZs



HUAWEI CLOUD provides services in many regions around the world. You can select a region and AZ as needed. For more information, see [HUAWEI CLOUD Global Regions](#).

## Region Selection

When selecting a region, consider the following factors:

- Location

You are advised to select a region close to you or your target users. This reduces network latency and improves access rate. However, Chinese mainland regions provide basically the same infrastructure, BGP network quality, as well as operations and configurations on resources. Therefore, if you or your target users are in the Chinese mainland, you do not need to consider the network latency differences when selecting a region.

The countries and regions outside the Chinese mainland, such as Bangkok and Hong Kong, provide services for users outside the Chinese mainland. If you or your target users are in the Chinese mainland, these regions are not recommended due to high access latency.

- If you or your target users are in the Asia Pacific region, except the Chinese mainland, select the **AP-Bangkok** or **AP-Singapore** region.
- If you or your target users are in Africa, select the **AF-Johannesburg** region.
- If you or your target users are in Europe, select the **EU-Paris** region.

- Resource price

Resource prices may vary in different regions. For details, see [Product Pricing Details](#).

## AZ Selection

When determining whether to deploy resources in the same AZ, consider your applications' requirements on disaster recovery (DR) and network latency.

- For high DR capability, deploy resources in different AZs in the same region.
- For low network latency, deploy resources in the same AZ.

## Regions and Endpoints

An endpoint is the **request address** for calling an API. Endpoints vary depending on services and regions. You can obtain endpoints of the service from [Endpoints](#).

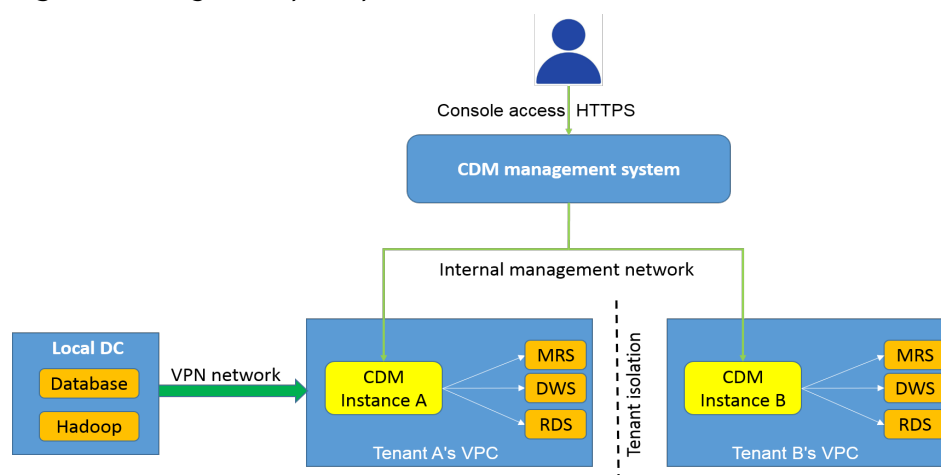
# 7 CDM Migration Principles

## Migration Principles

When a tenant uses CDM, the CDM system provisions a fully-managed CDM instance in the tenant's VPC. The instance allows only console and RESTful API access. Therefore the tenant cannot access the instance through other interfaces (such as SSH). This ensures data isolation between CDM tenants, prevents data leakage, and ensures transmission security during data migration between different cloud services in a VPC. Tenants can also use the VPN to migrate data from the on-premises data center to cloud services to ensure migration security.

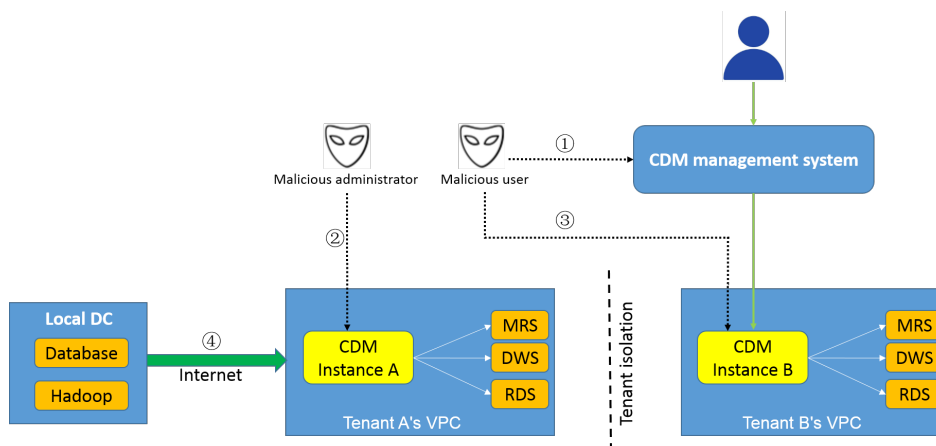
CDM works in push-pull mode. CDM pulls data from the migration source and pushes the data to the migration destination. Data access operations are initiated by CDM. SSL will be used if the data source (such as RDS) supports it. During the migration, the usernames and passwords of the migration source and destination are required. Such information is stored in the database of the CDM instance. Protecting such information is critical to ensure CDM security.

Figure 7-1 Migration principles



## Security Boundary and Risk Mitigation

Figure 7-2 Risk mitigation



As shown in [Figure 7-2](#), CDM may have the following threats:

1. Threats from the Internet: Malicious tenants may attack CDM through the CDM console.
2. Threats from the data center: Malicious CDM administrators obtain tenants' data source access information (usernames and passwords).
3. Threats from malicious tenants: Malicious tenants steal data from other tenants.
4. Data exposure to the public network: Data is exposed when it is migrated from the public network.

CDM offers the following mechanisms to prevent potential security risks:

1. Threats from the Internet: Tenants cannot log in to the CDM console through the public network. CDM provides a two-layer security mechanism.
  - a. On the one hand, the cloud console framework requires user authentication when tenants access management consoles of cloud services.
  - b. On the other hand, Web Application Firewall (WAF) filters requests from all consoles and stops request attack code or content.
2. Threats from the data center: Tenants must provide the usernames and passwords of the migration source and destination to complete data migration. To prevent the CDM administrators from obtaining such information and attacking important data sources of tenants, CDM provides a three-level protection mechanism.
  - a. CDM stores passwords encrypted by AES-256 in the database of the instance to ensure tenant isolation. The database is run by user **Ruby** and listens to only 127.0.0.1. Therefore, tenants cannot remotely access the database.
  - b. After the instance is provisioned, CDM changes the passwords of users **root** and **Ruby** to random passwords and does not store them in any place. This prevents the CDM administrators from accessing tenants' instances and databases containing password information.

- c. CDM instances work in push-pull mode. Therefore, the instances do not have any listening port enabled in the VPC, and tenants cannot access the local database or operating system from the VPC.
3. Threats from malicious tenants: CDM runs instances on independent VMs, so that tenants' instances are completely isolated and secure. Malicious tenants cannot access instances of other tenants.
4. Data exposure to the public network: In push-pull mode, even if elastic IP addresses (EIPs) are bound to the CDM clusters, no port is enabled for the EIPs. In this way, attackers cannot access and attack CDM using the EIPs. However, when data is migrated from the public network, tenants' data sources are exposed to the public network and threatened by third-party attacks. Therefore, tenants are advised to use ACLs or firewalls on the data source server for security. In this case, for example, only the access requests from the EIPs bound to the CDM clusters are allowed.

# 8 Related Services

## IAM

Your registered cloud account has full access to its resources and cloud services. If you need to assign different permissions to employees in your enterprise to access your CDM resources, Identity and Access Management (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 cloud resources.

## VPC

CDM clusters are created in the subnets of a Virtual Private Cloud (VPC). VPCs provide a secure, isolated, and logical network environment for CDM clusters.

## MRS

CDM supports data import and export using MRS.

## OBS

CDM supports data import and export using OBS, which also stores backup files and logs of CDM clusters.

## Cloud Eye

CDM uses Cloud Eye to monitor cluster performance metrics, delivering status information in a concise and efficient manner, as shown in [Table 8-1](#).

**Table 8-1** CDM performance metrics

| Metric   | Description  | Value Range | Monitored Object     |
|----------|--|-------------|----------------------|
| Bytes In | Measures the network inbound rate of the monitored object.<br>Unit: byte/s | ≥ 0 bytes/s | CDM cluster instance |



| Metric       | Description   | Value Range | Monitored Object     |
|--------------|---|-------------|----------------------|
| Bytes Out    | Measures the network outbound rate of the monitored object.<br>Unit: byte/s | ≥ 0 bytes/s | CDM cluster instance |
| CPU Usage    | Measures the CPU usage of the monitored object.<br>Unit: %                  | 0% to 100%  | CDM cluster instance |
| Memory Usage | Measures the memory usage of the monitored object.<br>Unit: %               | 0% to 100%  | CDM cluster instance |

## CTS

CDM uses Cloud Trace Service (CTS) to record operations for later query, audit, and backtrack operations. [Table 8-2](#) displays the recorded CDM operations.

**Table 8-2** Key operations recorded by CTS

| Operation                        | Resource Type | Trace Name       |
|----------------------------------|---------------|------------------|
| Creating a cluster               | cluster       | createCluster    |
| Deleting a cluster               | cluster       | deleteCluster    |
| Modifying cluster configurations | cluster       | modifyCluster    |
| Starting a cluster               | cluster       | startCluster     |
| Restarting a cluster             | cluster       | restartCluster   |
| Importing a job                  | cluster       | clusterImportJob |
| Binding an EIP                   | cluster       | bindEip          |
| Unbinding an EIP                 | cluster       | unbindEip        |
| Creating a link                  | link          | createLink       |
| Modifying a link                 | link          | modifyLink       |
| Testing a link                   | link          | verifyLink       |
| Deleting a link                  | link          | deleteLink       |
| Creating a job                   | job           | createJob        |
| Modifying a job                  | job           | modifyJob        |
| Deleting a job                   | job           | deleteJob        |

| Operation      | Resource Type | Trace Name |
|----------------|---------------|------------|
| Starting a job | job           | startJob   |
| Stopping a job | job           | stopJob    |

## DWS

CDM allows you to import data to and export data from DWS.

## RDS

CDM allows you to import data to and export data from RDS, including RDS for MySQL, RDS for PostgreSQL, and RDS for SQL Server.

## DDS

CDM allows you to export data from DDS, but it does not allow you to import data to DDS.

## CSS

CDM allows you to import data to and export data from CSS.

## DIS

CDM allows you to import data to DIS, and export data from DIS to CSS.

## CloudTable

CDM allows you to import data to and export data from CloudTable.

## DLI

CDM allows you to import data to DLI, but it does not allow you to export data from DLI.

## SFS

CDM allows you to import data to and export data from SFS.

## DMS

CDM allows you to import data to DMS, and export data from DMS to CSS.

## DataArts Studio

CDM can function as the DataArts Migration component of DataArts Studio and work with other components of DataArts Studio to complete data migration and periodic scheduling tasks.

# 9 Constraints

## CDM System Constraints

1. A free CDM cluster provided together with the DataArts Studio instance can be used as an agent for the data connections in Management Center. However, you are not advised to use the cluster as a node for running data migration jobs when the cluster is used as an agent.
2. You can purchase CDM clusters of other specifications on the DataArts Studio console as incremental packages or directly purchase clusters on the CDM console. The differences are as follows:
  - a. Billing: Clusters purchased on the DataArts Studio console only support packages purchased on the DataArts Studio console, while clusters purchased on the CDM console only support the discount packages purchased on the CDM console.
  - b. Permission control: For clusters purchased on the DataArts Studio console, permissions are managed based on the DataArts Studio permission system. For clusters purchased on the CDM console, permissions are managed based on the CDM permission system.
  - c. Application scenarios: Clusters purchased on the DataArts Studio console are isolated by workspace and can be used only in associated workspaces. Clusters purchased on the CDM console are not isolated by workspace and can be used in all DataArts Studio workspaces.
3. You cannot modify the flavor of an existing cluster. If you require a higher flavor, create a cluster with your desired flavor.
4. Arm CDM clusters do not support agents. The CDM cluster version (Arm or x86) is determined by the architecture of underlying resources.
5. CDM does not support the function of controlling the data migration speed. Therefore, do not perform data migration during peak hours.
6. During data migration, CDM imposes pressure on the data source. You are advised to create a database account for data migration and configure an account policy to reduce the resource consumption of the data source. For example, you can configure a policy to delete the connections of the account when the CPU usage exceeds 30% to prevent impact on services.
7. The baseline and maximum bandwidths of the NIC of the cdm.large CDM instance is 0.8 Gbit/s and 3 Gbit/s, respectively. The theoretical maximum volume of data that can be transmitted per instance per day is about 8 TB.

Similarly, the baseline and maximum bandwidths of the NIC of the `cdm.xlarge` instance are 4 Gbit/s and 10 Gbit/s, respectively, and the theoretical maximum volume of data that can be transmitted per instance per day is about 40 TB. The baseline and maximum bandwidths of the NIC of the `cdm.4xlarge` instance is 36 Gbit/s and 40 Gbit/s, respectively, and the theoretical maximum volume of data that can be transmitted per instance per day is about 360 TB. You can use multiple CDM instances if you want faster data transfer.

The actual amount of data that can be migrated in a day depends on the data source type, the read and write performance of the source and destination, and the actual available bandwidth. Typically you can migrate as much as 8 TB per day (large file migration to OBS) using the `cdm.large` instance. It is recommended that you test the speed with a small amount of data before migration.

8. CDM supports incremental file migration (by skipping repeated files), but does not support resumable transfer.  
For example, if three files are to be migrated and the second file fails to be migrated due to the network fault. When the migration task is started again, the first file is skipped. The second file, however, cannot be migrated from the point where the fault occurs, but can only be migrated again.
9. During file migration, a single task supports millions of files. If there are too many files in the directory to be migrated, you are advised to split the files into different directories and create multiple tasks.
10. You can export links and jobs configured on CDM to a local directory. To ensure password security, CDM does not export the link password of the corresponding data source. Therefore, before importing job configurations to CDM, you need to manually input the password in the exported JSON file or configure the password in the import dialog box.
11. The cluster cannot automatically upgrade to a new version. You need to use the job export and import functions to upgrade the cluster to the new version.
12. If OBS is unavailable, CDM does not automatically back up users' job configurations. You need to export and back up configuration data using the export function.
13. If VPC peering connection is configured, the peer VPC subnet may overlap with the CDM management network. As a result, data sources in the peer VPC cannot be accessed. You are advised to use the public network for cross-VPC data migration, or contact the administrator to add specific routes to the VPC peering connection in the CDM background.
14. If the destination of a CDM job is a DWS or NewSQL database, constraints of the source end, such as the primary key and unique index, cannot be migrated together.
15. When performing a CDM job, ensure that the JSON file formats of the two clusters are the same so that jobs can be imported from the source cluster to the destination cluster.
16. If a running job is interrupted unexpectedly, the data that has been written to the destination will not be deleted. You must manually delete the data if needed.
17. The size of a file to be transferred cannot exceed 1 TB.

## General Constraints on Database Migration

1. CDM is mainly used for batch migration. It supports only limited incremental migration but does not support real-time incremental migration. You are advised to use Data Replication Service (DRS) to migrate the incremental data of the database to RDS.
2. The entire DB migration of CDM supports only data table migration but not migration of database objects such as stored procedures, triggers, functions, and views.

CDM applies only to scenarios where databases are migrated to the cloud at a time, including homogeneous and heterogeneous database migrations. CDM is not applicable to data synchronization, for example, disaster recovery and real-time synchronization.

3. If CDM fails to migrate an entire database or table, the data that has been imported to the target table will not be rolled back automatically. If you want to perform migration in transaction mode, configure the **Import to Staging Table** parameter to enable a rollback upon a migration failure.

In extreme cases, the created stage table or temporary table cannot be automatically deleted. You need to manually clear the table (the name of the stage table ends with `_cdm_stage`), for example, `cdmtet_cdm_stage`).

4. If CDM needs to access data sources in the on-premises data center (for example, the on-premises MySQL database), the data sources must support Internet access and the CDM instances must be bound with elastic IP addresses. In this case, the security practice is to configure the firewall or security policies to allow only the EIPs of the CDM instances to access the local data sources.
5. Only common data types are supported, including character strings, digits, and dates. Object types are limited. If objects are too large, migration cannot be performed.
6. Only the GBK and UTF-8 character sets are supported.
7. A field name cannot contain `&` or `%`.
8. `jdbc2hive` and `hive2jdbc` entire DB migration is implemented by field name mapping, and is unavailable if the source and destination field names are inconsistent.

## Permissions Configuration for Relational Database Migration

Common minimum permissions required by relational database migration:

- MySQL: You need to have the read permission on the **INFORMATION\_SCHEMA** database and data tables.
- Oracle: You need to have the **resource** role and have the **select** permissions on the data table in the tablespace.
- Dameng: You need to have the **select any table** permission in the schema.
- DWS: You need to have the **schema usage** permission and the query permission on the data tables.
- SQL Server: You need to have the **sysadmin** permission.
- PostgreSQL: You need to have the **select** permission on schema tables in the database.

## Constraints on FusionInsight HD and Apache Hadoop

If the FusionInsight HD and Apache Hadoop data sources are deployed in the on-premises data center, CDM must access all nodes in the cluster for reading and writing the Hadoop files. Therefore, the network access must be enabled for each node.

You are advised to use [Direct Connect](#) to improve the migration speed while ensuring network access.

## Constraints on DWS and FusionInsight LibrA

1. If the DWS primary key or table contains only one field, the field type must be a common character string, value, or date. When data is migrated from another database to DWS, if automatic table creation is selected, the primary key must be of the following types. If no primary key is set, at least one of the following fields must be set. Otherwise, the table cannot be created and the CDM job fails.
  - INTEGER TYPES: TINYINT, SMALLINT, INT, BIGINT, NUMERIC/DECIMAL
  - CHARACTER TYPES: CHAR, BPCHAR, VARCHAR, VARCHAR2, NVARCHAR2, TEXT
  - DATA/TIME TYPES: DATE, TIME, TIMETZ, TIMESTAMP, TIMESTAMPTZ, INTERVAL, SMALLDATETIME

### NOTE

For clusters of version 2.9.1.200 or earlier, the NVARCHAR2 data type is not supported for DWS.

2. In DWS, the character string " is null. A null character string cannot be inserted into a field with non-null constraints. This is inconsistent with the MySQL behavior. MySQL does not consider that " is null. Migration from MySQL to DWS may fail due to the preceding reason.
3. When the Gauss Data Service (GDS) mode is used to quickly import data to DWS, you need to configure a security group or firewall policy to allow DataNodes of DWS or FusionInsight LibrA to access port 25000 of the CDM IP address.
4. When data is imported to DWS in GDS mode, CDM automatically creates a foreign table for data import. The table name ends with a universally unique identifier (UUID), for example, **cdmtest\_aecf3f8n0z73dsl72d0d1dk4lcir8cd**. If a job fails, it will be automatically deleted. In extreme cases, you may need to manually delete it.

## Constraints on OBS

1. During file migration, the system automatically transfers the files concurrently. In this case, **Concurrent Extractors** in the task configuration is invalid.
2. Resumable transmission is not supported. If CDM fails to transfer files, OBS fragments are generated. You need to clear fragments on the OBS console to prevent space occupation.
3. CDM does not support the versioning control function of OBS.
4. During incremental migration, the number of files or objects in the source directory of a single job depends on the CDM cluster flavor. A `cdm.large`

cluster supports a maximum of 300,000 files; a `cdm.medium` cluster supports a maximum of 200,000 files; and a `cdm.small` cluster supports a maximum of 100,000 files.

If the number of files or objects in a single directory exceeds the upper limit, split the files or objects into multiple migration jobs based on subdirectories.

## Constraints on DLI

- To use CDM to migrate data to DLI, you must have the read permissions of OBS.
- If the destination is DLI, you are advised to set the number of concurrent extractors to 1. Otherwise, data may fail to be written.

## Constraints on Oracle

Real-time incremental data synchronization is not supported for Oracle databases.

## Constraints on DCS and Redis

1. The Redis service of the third-party cloud cannot serve as the migration source. However, the Redis set up in the on-premises data center or on the ECS can be the migration source and destination.
2. Only the hash and string data formats are supported.

## Constraints on DDS and MongoDB

When you migrate MongoDB or DDS data, CDM reads the first row of the collection as an example of the field list. If the first row of data does not contain all fields of the collection, you need to manually add fields.

## Constraints on CSS and Elasticsearch

1. CDM supports automatic creation of indexes and field types. The index and field type names can contain only lowercase letters.
2. You cannot modify the field type under an index after it is created, but only create another field.

If you need to modify the field type, you need to create an index or run the Elasticsearch command on Kibana to delete the existing index and create another index (the data is also deleted).

3. When the field type of the index created by CDM is date, the data format must be `yyyy-MM-dd HH:mm:ss.SSS Z`. For example, **2018-08-08 08:08:08.888 +08:00**.

During data migration to CSS, if the original data of the **date** field does not meet the format requirements, you can use the **field conversion** function of CDM to convert the data to the preceding format.

## Constraints on DIS and Kafka

- The data in the message body is a record in CSV format that supports multiple delimiters. Messages cannot be parsed in binary or other formats.
- If a job is set to run for a long time, the job will fail if the DIS system is interrupted.

- If the source is MRS Kafka, custom fields are not supported in field mapping.
- If the source is DMS Kafka, custom fields are supported in field mapping.

## Constraints on CloudTable and HBase

1. When you migrate data from CloudTable or HBase, CDM reads the first row of the table as an example of the field list. If the first row of data does not contain all fields of the table, you need to manually add fields.
2. Because HBase is schema-less, CDM cannot obtain the data types. If the data is stored in binary format, CDM cannot parse the data.

## Constraints on Hive

- If Hive stores timestamp data in Parquet format, timestamps are accurate to the nanosecond, for example, 2023-03-27 00:00:00.000. If the source data precision is higher than the nanosecond, the data will be truncated during field mapping. For example, if the source data is **2023-03-27 00:00:00.12345**, it will be truncated to **2023-03-27 00:00:00.123** at the destination.
- If Hive serves as the migration destination and the storage format is Textfile, delimiters must be explicitly specified in the statement for creating Hive tables. The following is an example:

```
CREATE TABLE csv_tbl(
  smallint_value smallint,
  tinyint_value tinyint,
  int_value int,
  bigint_value bigint,
  float_value float,
  double_value double,
  decimal_value decimal(9, 7),
  timestmamp_value timestamp,
  date_value date,
  varchar_value varchar(100),
  string_value string,
  char_value char(20),
  boolean_value boolean,
  binary_value binary,
  varchar_null varchar(100),
  string_null string,
  char_null char(20),
  int_null int
)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES (
  "separatorChar" = "\t",
  "quoteChar" = "'",
  "escapeChar" = "\\"
)
STORED AS TEXTFILE;
```



# 10 Billing

This section describes the CDM billing items and billing modes.

## NOTICE

When you use CDM, you may need to pay for the following services:

- OBS: During data migration, CDM may write dirty data to OBS, which incurs data storage fees. For details, see [OBS Pricing Details](#).
- EIP: If you buy an EIP for your CDM cluster, you need to pay for the EIP. For details, see [EIP Pricing Details](#).

## Billing Items

| Billing Item         | Description   |
|----------------------|---|
| CDM cluster instance | <ul style="list-style-type: none"><li>• Billing for the selected instance flavor</li><li>• Available billing modes: pay-per-use billing (hourly) or a discount package (monthly/yearly)</li></ul> |

## Billing Modes

CDM provides pay-per-use and discount packages for you to select. The longer you use, the cheaper you will be.

- Pay-per-use

You can enable or disable CDM as you like and you are charged by the usage duration.

  - The pay-per-use billing accounts for only cluster instance fees but not public network traffic fees.
  - A cluster is bound to a specific region. You can only use the cluster you purchased in the region to which the cluster is bound.
  - After you purchase a CDM cluster, the system automatically creates the cluster you purchased. You can bind an EIP to the cluster on the management console when necessary.

For details about pay-per-use pricing, see [Product Pricing Details](#).

- Discount package (pay-per-use resource package)

You can purchase a discount package for a period of usage in advance. This mode is economical and recommended if you need to use CDM for a long time.

- After you buy a discount package, you need to create CDM clusters of the corresponding instances in the specified region. The package can be used for 745 hours in each billing month within the validity period.
- 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 discount 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 discount 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 properly plan the resource usage before purchasing a discount package.
- If you want to avail the preferential price of the discount package, you can buy such a package and a pay-per-use cluster in the same region and with the same specifications as the package.
- If you purchase a pay-per-use cluster and then a discount package in the same region and with the same specifications as the cluster, the fees generated before you purchase the discount package are charged in pay-per-use mode, and the subsequent fees are charged based on the discount package.

## Changing Billing Mode

Four types of instance flavors are available when you subscribe to CDM. You can select proper instance flavors based on service requirements.

After a cluster is created, you cannot change the cluster specifications. You can change the cluster specifications only by deleting the cluster and creating a cluster with the new specifications.

## Renewal

You can renew a resource package upon its expiration, or you can set auto-renewal rules for a resource package.

## Expiration and Overdue Payment

After a yearly/monthly package expires, you will be billed for subsequently used resources on a pay-per-use basis. If your account is in arrears, a grace period and a retention period will be granted. Such periods depend on the customer tier and subscription type. For details about the fees and resource handling during the grace and retention periods, see [Grace Period and Retention Period](#).

## Unsubscription

Deleted CDM clusters will no longer be billed based on a pay-per-use basis. Discount packages cannot be unsubscribed from.

- Deleted CDM clusters will no longer be billed pay per use, cannot be recovered, and will not consume the associated discount package. For details, see [Deleting a Cluster](#).
- Discount packages are pay-per-use resource packages and cannot be unsubscribed from. For details, see [Unsubscription Not Allowed](#).

After deleting a cluster or unsubscribing from CDM, you must unsubscribe from the resources of the following services that may incur fees:

- OBS: During data migration, CDM may write dirty data to OBS, which incurs data storage fees. For details, see [OBS Pricing Details](#).
- EIP: If you buy an EIP for your CDM cluster, you need to pay for the EIP. For details, see [EIP Pricing Details](#).

# 11 Security

---

## 11.1 Shared Responsibilities

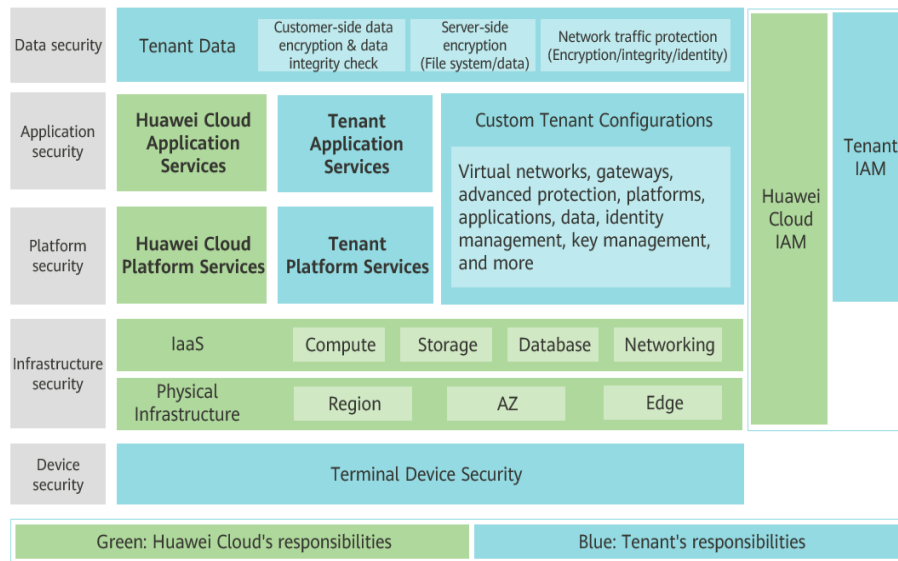
Huawei guarantees that its commitment to cyber security will never be outweighed by the consideration of commercial interests. To cope with emerging cloud security challenges and pervasive cloud security threats and attacks, Huawei Cloud builds a comprehensive cloud service security assurance system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

**Figure 11-1** 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 11-1** Huawei Cloud shared security responsibility model



## 11.2 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 CDM, add tags to clusters). 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.

In CDM, you can add tags to identify and classify clusters. If you add tags to a cluster, CDRs generated by the requests for this cluster will contain these tags. Using the tags, you can classify CDRs for detailed cost analysis. For example, if a department uses a cluster, you can add the department name to the cluster 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. A cluster can have a maximum of 10 tags. Each key is unique among all tags of a cluster, whereas values can be repetitive or blank.

### How to Use

You can add tags to a CDM cluster on the CDM console. For details, see [Managing CDM Cluster Tags](#).

## 11.3 Identity Authentication and Access Control

### Identity Authentication

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

CDM APIs can be accessed upon successful authentication. Requests sent through the CDM console and requests for calling APIs can both be authenticated using tokens.

### 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 [IAM Service Overview](#).

You can grant users permissions by using roles and policies.

- **Roles:** A type of coarse-grained authorization mechanism that defines permissions related to user responsibilities. This mechanism provides only a limited number of service-level roles for authorization. When using roles to grant permissions, you need to also assign other roles on which the permissions depend to take effect. However, roles are not an ideal choice for fine-grained authorization and secure access control.
- **Policies:** A type of fine-grained authorization mechanism that defines permissions required to perform operations on specific cloud resources under certain conditions. This mechanism allows for more flexible policy-based authorization, meeting requirements for secure access control. For example, a specific user group is not allowed to delete a cluster. Only basic CDM operations (such as creating and querying jobs) are allowed.

**Table 1** lists all the system-defined roles and policies supported by CDM.

**Table 11-1** System-defined roles and policies supported by CDM

| Role/Policy Name                 | Description   | Type                  |
|----------------------------------|---|-----------------------|
| CDM Administrator                | Permissions: <ul style="list-style-type: none"> <li>Administrator permissions for all operations on CDM resources. Users granted these permissions must also be granted permissions of the <b>Tenant Guest</b> and <b>Server Administrator</b> policies.</li> <li>Users granted permissions of the <b>VPC Administrator</b> policy can create VPCs and subnets.</li> <li>Users granted permissions of the <b>Cloud Eye Administrator</b> policy can view monitoring information of CDM clusters.</li> </ul> | System role           |
| CDM FullAccess                   | Administrator permissions for CDM. Users granted these permissions can perform all operations on CDM resources.   | System-defined policy |
| CDM FullAccessExcept EIPUpdating | Users granted these permissions can perform all operations on CDM resources except binding and unbinding EIPs.  | System-defined policy |
| CDM CommonOperations             | Users granted these permissions can operate CDM jobs and links.   | System-defined policy |
| CDM ReadOnlyAccess               | Read-only permissions for CDM. Users granted these permissions can only view CDM clusters, links, and jobs.   | System-defined policy |

## 11.4 Data Protection Technologies

### Data Storage Security

To prevent your personal data (such as username, password, and mobile number) from being obtained by unauthorized or unauthenticated entities or individuals, CDM encrypts your data during storage and transmission to prevent data leakage.

### Data Destruction Mechanism

When you delete a CDM cluster, your personal data stored in the cluster will also be deleted.

When you delete your phone number and email address on the console and disable message notifications, the phone number and email address will also be deleted from the database.

## Data Transmission Security

Your personal data is encrypted using TLS 1.2 or TLS 1.3 during transmission. All the calls made to CDM APIs use HTTPS to encrypt the transmitted data.

## 11.5 Audit and Logging

Cloud Trace Service (CTS) records operations on the cloud resources in your account. You can use the logs generated by CTS to perform security analysis, track resource changes, audit compliance, and locate faults.

For details about the CDM operations that can be recorded by CTS, see [Key CDM Operations Recorded by CTS](#). 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 CDM, CTS sends a notification to subscribers.

## 11.6 Service Resilience

CDM ensures data durability and reliability using technical methods such as throttling and backup and restoration.

## 11.7 Security Risk Monitoring

Working with Cloud Eye, CDM 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 CDM monitoring metrics and how to create alarm rules, see [Viewing Metrics](#).

## 11.8 Recovery from Failures

CDM supports [Scheduled Backup/Restoration](#). You can enable this function, and jobs will be backed up to OBS buckets as scheduled. If the service becomes faulty, you can recover jobs by importing their backups.

## 11.9 Update Management

### Update to Cope with Vulnerabilities

Updates will be made in a timely manner to prevent CDM 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

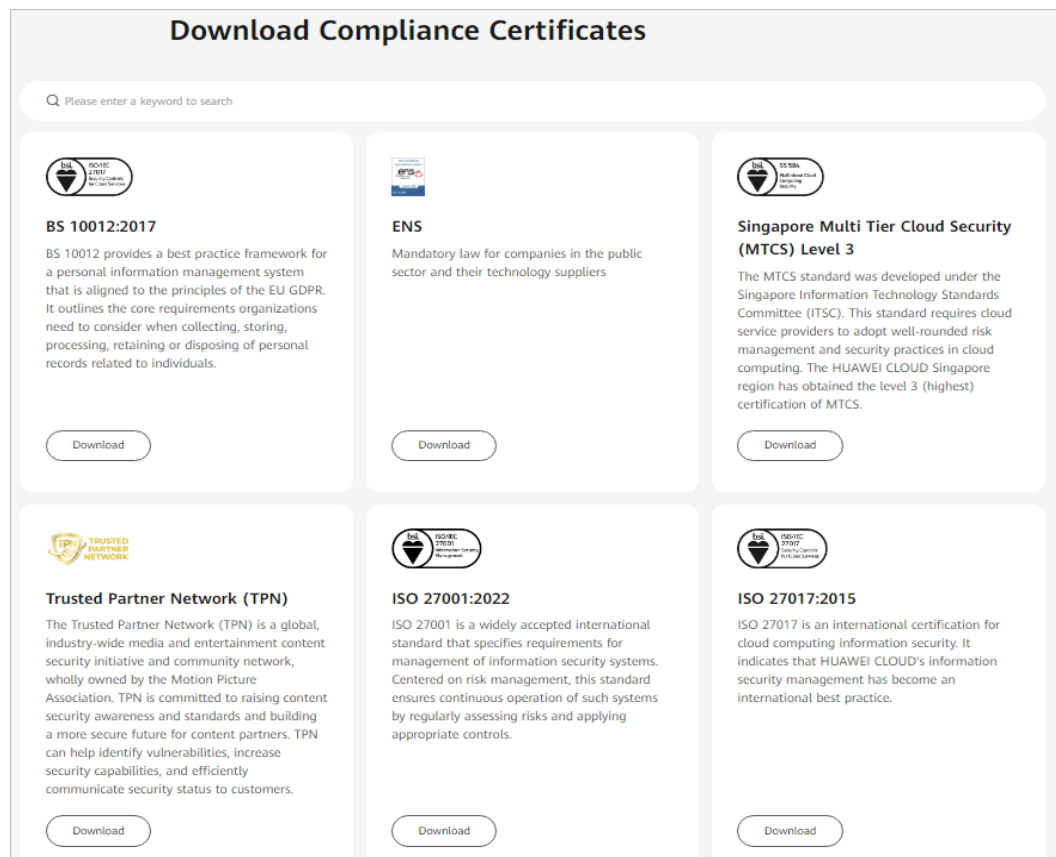
CDM configurations can be updated through version update.

# 11.10 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.

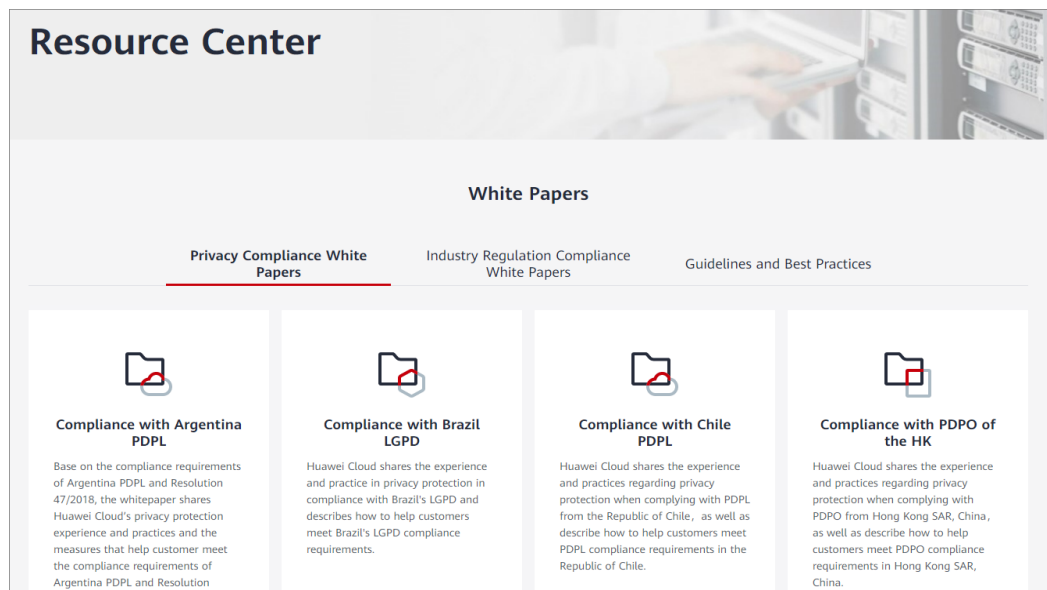
**Figure 11-2** Downloading compliance certificates



## Resource Center

Huawei Cloud also provides the following resources to help users meet compliance requirements. For details, see [Resource Center](#).

Figure 11-3 Resource center



# 12 Quotas

---

CDM uses the following infrastructure resources:

- ECS
- VPC
- EIP
- Simple Message Notification (SMN)
- IAM

For details about how to view and modify the quota, see [Quotas](#).

# 13 Permissions Management

---

If you need to assign different permissions to employees in your enterprise to access your CDM 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 cloud resources.

With IAM, you can use your cloud account to create IAM users, and assign permissions to the users to control their access to specific resources. For example, some employees in your enterprise need to use CDM resources but should not be allowed to delete CDM clusters or perform any other high-risk operations. In this scenario, you can create IAM users for the employees and grant them only the permissions required for using CDM resources.

If your cloud account does not require individual IAM users for permissions management, skip this section.

IAM can be used free of charge. You pay only for the resources in your account. For more information about IAM, see [What Is IAM?](#)

## CDM Permissions

By default, new IAM users do not have permissions assigned. You need to add a user to one or more groups, and attach permissions policies or roles to these groups. Users inherit permissions from the groups to which they are added and can perform specified operations on cloud services based on the permissions.

CDM is a project-level service deployed and accessed in specific physical regions. Therefore, CDM permissions are assigned to users in specific regions and only take effect for these regions. If you want the permissions to take effect for all regions, you need to assign the permissions to users in each region. When accessing CDM, the users need to switch to a region where they have been authorized to use the CDM service.

You can grant users permissions by using roles and policies.

- **Roles:** A type of coarse-grained authorization mechanism that defines permissions related to user responsibilities. This mechanism provides only a limited number of service-level roles for authorization. When using roles to grant permissions, you need to also assign other roles on which the permissions depend to take effect. However, roles are not an ideal choice for fine-grained authorization and secure access control.

- **Policies:** A type of fine-grained authorization mechanism that defines permissions required to perform operations on specific cloud resources under certain conditions. This mechanism allows for more flexible policy-based authorization, meeting requirements for secure access control. For example, a specific user group is not allowed to delete a cluster. Only basic CDM operations (such as creating and querying jobs) are allowed.

**Table 1** lists all the system-defined roles and policies supported by CDM.

**Table 13-1** System-defined roles and policies supported by CDM

| Role/Policy Name                 | Description  | Type                  |
|----------------------------------|--|-----------------------|
| CDM Administrator                | <p>Permissions:</p> <ul style="list-style-type: none"> <li>• Administrator permissions for all operations on CDM resources. Users granted these permissions must also be granted permissions of the <b>Tenant Guest</b> and <b>Server Administrator</b> policies.</li> <li>• Users granted permissions of the <b>VPC Administrator</b> policy can create VPCs and subnets.</li> <li>• Users granted permissions of the <b>Cloud Eye Administrator</b> policy can view monitoring information of CDM clusters.</li> </ul> | System role           |
| CDM FullAccess                   | Administrator permissions for CDM. Users granted these permissions can perform all operations on CDM resources.  | System-defined policy |
| CDM FullAccessExcept EIPUpdating | Users granted these permissions can perform all operations on CDM resources except binding and unbinding EIPs.   | System-defined policy |
| CDM CommonOperations             | Users granted these permissions can operate CDM jobs and links.  | System-defined policy |
| CDM ReadOnlyAccess               | Read-only permissions for CDM. Users granted these permissions can only view CDM clusters, links, and jobs.  | System-defined policy |

**Table 13-2** lists the common operations supported by each system-defined policy or role of CDM. Select the policies or roles as required.

**Table 13-2** Common operations supported by each system-defined policy or role of CDM

| Operation                        | CDM FullAccess | CDM FullAccessExceptEIPUpdating | CDM CommonOperations | CDM ReadOnlyAccess |
|----------------------------------|----------------|---------------------------------|----------------------|--------------------|
| Creating clusters                | √              | √                               | ×                    | ×                  |
| Binding or unbinding EIPs        | √              | ×                               | ×                    | ×                  |
| Querying the cluster list        | √              | √                               | √                    | √                  |
| Querying cluster details         | √              | √                               | √                    | √                  |
| Restarting clusters              | √              | √                               | ×                    | ×                  |
| Modifying cluster configurations | √              | √                               | ×                    | ×                  |
| Deleting clusters                | √              | √                               | ×                    | ×                  |
| Creating links                   | √              | √                               | √                    | ×                  |
| Querying links                   | √              | √                               | √                    | √                  |
| Modifying links                  | √              | √                               | √                    | ×                  |
| Deleting links                   | √              | √                               | √                    | ×                  |
| Creating jobs                    | √              | √                               | √                    | ×                  |
| Querying jobs                    | √              | √                               | √                    | √                  |
| Modifying jobs                   | √              | √                               | √                    | ×                  |
| Starting jobs                    | √              | √                               | √                    | ×                  |
| Stopping jobs                    | √              | √                               | √                    | ×                  |
| Querying job status              | √              | √                               | √                    | √                  |
| Querying job execution history   | √              | √                               | √                    | √                  |
| Deleting jobs                    | √              | √                               | √                    | ×                  |