

Cloud Data Migration

Best Practice

Issue 02

Date 2018-09-10

Copyright © Huawei Technologies Co., Ltd. 2019. All rights reserved.

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

Trademarks and Permissions



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

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

Notice

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

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

Contents

1 IoV Big Data Service Migration to Cloud.....	1
1.1 Scenario Overview.....	1
1.2 Migration Preparation.....	2
1.3 Using CDM to Migrate Data of the Last Month.....	3
1.4 Using DES to Migrate Historical Data Generated One Month Ago.....	8
1.5 Restoring the HBase Table on MRS.....	8
2 Consulting Service Data Migration in Financial Industry.....	11
2.1 Scenario Overview.....	11
2.2 Analysis Process.....	14
2.3 (Optional) Registering an Account.....	14
2.4 Using CDM to Upload Data to OBS.....	16
2.4.1 Uploading Inventory Data.....	16
2.4.2 Uploading Incremental Data.....	21
2.5 Using DLI for Data Analysis.....	21

1 IoV Big Data Service Migration to Cloud

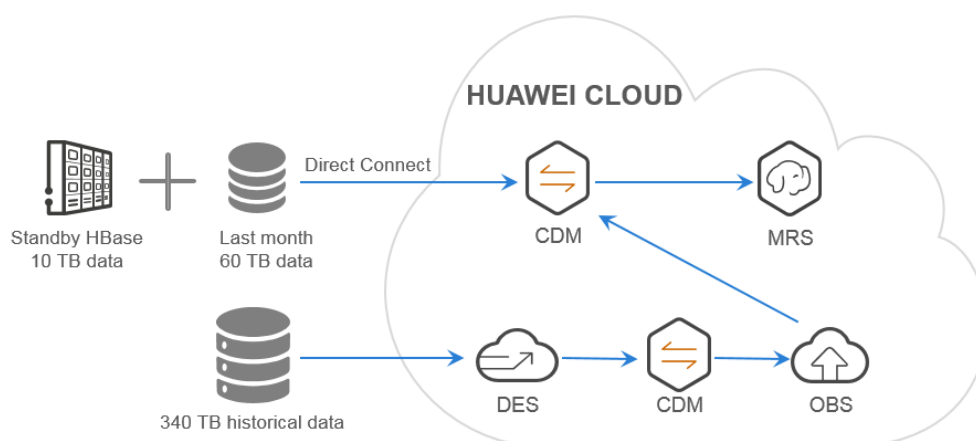
1.1 Scenario Overview

Background

Company *H* intends to build an enterprise-class cloud management platform for its IoV service to centrally manage and deploy hardware resources and common software resources, and implement cloud-based and service-oriented transformation of IT applications. Cloud Data Migration (CDM) helps company *H* build the platform without code modification and data loss.

Migration Scheme

Figure 1-1 Migration scheme



Company *H* stores 854 tables (400 TB) in the Cloudera Hadoop (CDH) HBase cluster and 149 tables (about 10 TB) in the standby HBase cluster. An amount of 60 TB data is increased in the last month.

Use CDM to extract HBase HFiles from the CDH cluster and save the extracted data to MapReduce Service (MRS) HDFS, and run the HBase repair command to rebuild the HBase table. Based on this migration scheme, the following two migration modes are optional:

1. Use CDM to migrate historical data generated one month ago from **Data Express Service (DES)**. The migration path is as follows:

CDH → DES → CDM (HUAWEI CLOUD) → OBS → CDM (HUAWEI CLOUD) → MRS

 **NOTE**

DES is well suited to the scenario where a large amount of data is to be transmitted, no private line is set up between the private cloud and HUAWEI CLOUD, and the bandwidth from the private cloud network to the public network is limited.

- Advantage: The transmission is highly reliable without depending on the private line and network quality.
 - Disadvantage: The migration takes a long time.
2. CDM migrates data of the last month and data of the standby HBase cluster through the private line.

CDH → CDM (HUAWEI CLOUD) → MRS

 **NOTE**

The advantage and disadvantage of direct migration using the private line are as follows:

- Advantage: Data does not need to be migrated multiple times, which shortens the overall migration duration.
- Disadvantage: When a large amount of data is transmitted, the private line bandwidth is heavily occupied, which affects the concurrent services of the customer and crosses multiple switches.

1.2 Migration Preparation

Prerequisites

1. The CDH HBase version is earlier than or equal to the MRS HBase version.
2. You cannot perform the write, split, and merge operations on the tables to be migrated.
3. Use **Direct Connect** to establish a Direct Connect connection between the CDH cluster and HUAWEI CLOUD Virtual Private Cloud (VPC).

Migration Process

1. Estimate the amount of data to be migrated and the migration duration.
2. Output the detailed data tables, the number and sizes of files to be migrated for subsequent verification.
3. Configure migration tasks in batches to ensure the migration progress and speed.
4. Check the number and sizes of files.
5. Restore the HBase table on MRS and verify the restoration.

Required Information

Item	Information	Description	Example Value
DES Teleport	Mount address	Address to which the DES Teleport box is mounted on the customer's VM	//IP address of the VM/ huawei
	DeviceManager	Storage management system of the DES Teleport box, which is related to the management IP address	https://Management IP address:8088/ deviceManager/ devicemanager/login/ login.html
	Username	Username for logging in to DeviceManager	admin
	Password	Password for logging in to DeviceManager	-
CDH cluster	NameNode IP	IP address of the active NameNode in the CDH cluster	192.168.2.3
	HDFS port	The default port number is 9000.	9000
	HDFS URI	NameNode URI of HDFS in the CDH cluster	hdfs://192.168.2.3:9000
OBS	OBS endpoint	Endpoint of OBS	obs.cn-north-1.myhwclouds.com
	OBS bucket	OBS bucket that stores historical data one month ago of the CDH cluster	cdm
	AK/SK	AK and SK for accessing OBS	-
MRS	Manager IP	IP address of MRS Manager	192.168.3.11

1.3 Using CDM to Migrate Data of the Last Month

The standby HBase cluster stores about 10 TB data, and the amount of data increased in the last month is about 60 TB. Therefore, the total amount of data is about 70 TB. Company H's 20GE private line supports the cdm.xlarge cluster of CDM. Considering the migration duration, costs, and performance, the cdm.xlarge cluster is used to perform concurrent migrations.

Table 1-1 CDM cluster specifications

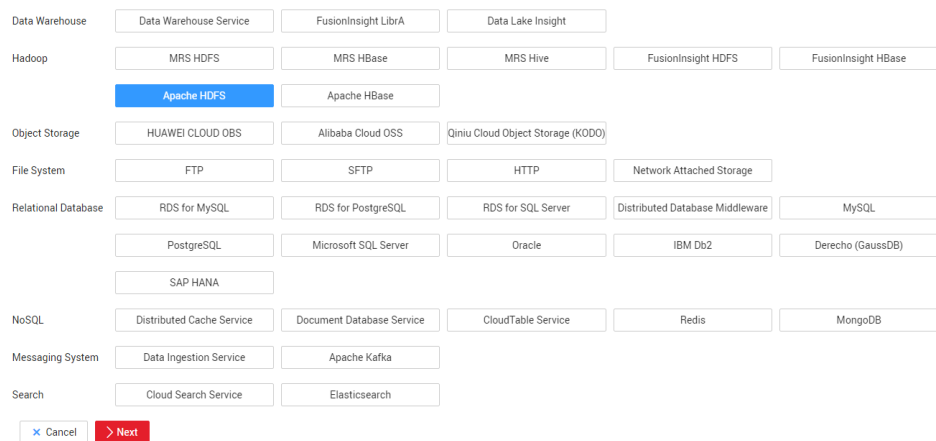
Instance Flavor	vCPUs/ Memory	Maximum/ Assured Bandwidth	Concurrent Extractors	Application Scenario
cdm.small	2 vCPUs and 4 GB memory	0.8/0.2 Gbit/s	10	Proof of Concept (PoC) verification and development test
cdm.medium	4 vCPUs and 8 GB memory	1.5/0.4 Gbit/s	20	A single table with fewer than 10 million pieces of data, applicable to most migration scenarios and multiple migration tasks
cdm.large	8 vCPUs and 16 GB memory	3/0.8 Gbit/s	30	A single table with 10 million or more than 10 million pieces of data
cdm.xlarge	16 vCPUs and 32 GB memory	10/4 Gbit/s	100	TB-level data migration requiring 10GE bandwidth

 **NOTE**

You can use multiple CDM clusters to perform migrations concurrently to improve migration efficiency. The MRS HDFS multi-replica policy occupies network bandwidth and affects the migration efficiency.

Creating Links on HUAWEI CLOUD CDM

1. Log in to [the CDM management console](#) and create two CDM clusters.
 - Select the **cdm.xlarge** flavor.
 - The clusters must reside in the same VPC as MRS and DirectConnect.
 - Configure other parameters as required or retain the default values.
2. Perform the following operations to create an Apache HDFS link:
 - a. In the **Operation** column, click **Job Management**. The job management page is displayed.
 - b. Choose **Link Management > Create Link**. On the page that is displayed, select **Apache HDFS**.



- c. Click **Next** and configure the link parameters. The URI format is *hdfs://NameNode IP address:Port number*. If Kerberos authentication is not enabled in the CDH cluster, set **Authentication Method** to **SIMPLE**.

* Name	<input type="text" value="CDH-hdfs"/>
* Connector	<input type="text" value="HDFS"/>
* Hadoop Type	<input type="text" value="Apache Hadoop"/>
* URI	<input type="text" value="hdfs://10.10.10.1:8020"/>
* Authentication Method	<input type="text" value="SIMPLE"/>
* Run Mode	<input type="text" value="STANDALONE"/>
IP and Host Name Mapping	<input type="text"/>

✕ Cancel < Previous Test Save

- d. Click **Test**. If a test success message is displayed in the upper right corner, the link works properly. Click **Save**.
3. Perform the following operations to create an MRS HDFS link:
- a. Choose **Link Management > Create Link**. On the page that is displayed, select **MRS HDFS**.

Data Warehouse	<input type="text" value="Data Warehouse Service"/>	<input type="text" value="FusionInsight LibrA"/>	<input type="text" value="Data Lake Insight"/>
Hadoop	<input checked="" type="text" value="MRS HDFS"/>	<input type="text" value="MRS HBase"/>	<input type="text" value="MRS Hive"/>
	<input type="text" value="FusionInsight HDFS"/>	<input type="text" value="FusionInsight HBase"/>	<input type="text" value="Apache HDFS"/>
Object Storage	<input type="text" value="HUAWEI CLOUD OBS"/>	<input type="text" value="Alibaba Cloud OSS"/>	<input type="text" value="Qiniu Cloud Object Storage (KODO)"/>
File System	<input type="text" value="FTP"/>	<input type="text" value="SFTP"/>	<input type="text" value="HTTP"/>
Relational Database	<input type="text" value="Network Attached Storage"/>	<input type="text" value="RDS for MySQL"/>	<input type="text" value="RDS for PostgreSQL"/>
	<input type="text" value="RDS for SQL Server"/>	<input type="text" value="Distributed Database Middleware"/>	<input type="text" value="MySQL"/>
	<input type="text" value="PostgreSQL"/>	<input type="text" value="Microsoft SQL Server"/>	<input type="text" value="Oracle"/>
NoSQL	<input type="text" value="IBM Db2"/>	<input type="text" value="Derecho (GaussDB)"/>	<input type="text" value="SAP HANA"/>
	<input type="text" value="Distributed Cache Service"/>	<input type="text" value="Document Database Service"/>	<input type="text" value="CloudTable Service"/>
Messaging System	<input type="text" value="Redis"/>	<input type="text" value="MongoDB"/>	<input type="text" value="Data Ingestion Service"/>
Search	<input type="text" value="Apache Kafka"/>	<input type="text" value="Cloud Search Service"/>	<input type="text" value="Elasticsearch"/>

✕ Cancel Next

- b. Click **Next** and configure the link parameters. Set **Authentication Method** to **SIMPLE** and retain the default run mode.

* Name	<input type="text" value="mrs_hbase"/>	
* Connector	<input type="text" value="HDFS"/>	
* Hadoop Type	<input type="text" value="MRS"/>	
* Manager IP <small>?</small>	<input type="text" value="██████████"/>	Select
* Authentication Method	<input type="text" value="SIMPLE"/>	
* Run Mode <small>?</small>	<input type="text" value="EMBEDDED"/>	

✕ Cancel
< Previous
Test
Save

- c. Click **Test**. If a test success message is displayed in the upper right corner, the link works properly. Click **Save**.

Creating a Migration Job on HUAWEI CLOUD CDM

1. On the job management page of the CDM cluster, choose **Table/File Migration** > **Create Job** to create jobs. Create a migration job for each table file directory.

Job Configuration

*** Job Name**


Source Job Configuration	Destination Job Configuration
* Source Link Name <input type="text" value="CDH-hdfs"/>	* Destination Link Name <input type="text" value="mrs_hbase"/>
* Source Directory/File <small>?</small> <input type="text" value="/hbase/data/default/table_2"/>	* Write Directory <small>?</small> <input type="text" value="/hbase/data/default/table_2"/>
* File Format <small>?</small> <input type="text" value="Binary"/>	* File Format <small>?</small> <input type="text" value="Binary"/>
Show Advanced Attributes	Duplicate File Processing Method <input type="text" value="Replace"/>
	Compression Codec <small>?</small> <input type="text" value="None"/>
	Show Advanced Attributes

✕ Cancel
> Next

– Source Job Configuration

- **Source Link Name:** Select the created [Apache HDFS link](#).
 - **Source Directory/File:** Select the directory where the HBase table of the CDH cluster resides. For example, `/hbase/data/default/table_20180815` indicates that all files in the `table_20180815` directory will be migrated.
 - **File Format:** Select **Binary**.
 - **Destination Job Configuration**
 - **Destination Link Name:** Select the created [MRS HDFS link](#).
 - **Write Directory:** Select the MRS HBase directory, for example, `/hbase/data/default/table_20180815/`. The directory must carry a table name (for example, `table_20180815`). If the directory does not exist, CDM automatically creates it.
 - **File Format:** Select **Binary**.
 - Retain the default values of other parameters.
2. Click **Next** to configure the task. By default, **Concurrent Extractors** is **3**. You can increase the number of concurrent extractors to improve the migration efficiency. Retain the default values of other parameters.

Configure Task

Retry if failed 

Retry 3 times if failed 

Schedule Execution

Yes

No

Hide Advanced Attributes

Concurrent Extractors 


8


Write Dirty Data 

Yes


No


Is Disposable Job After completed

Don't Drop 

 Cancel

 Previous

 Save

 Save and Run

3. Repeat the preceding operations to create migration jobs for other directories. The parameter settings are the same. The number of jobs in the two CDM clusters is evenly allocated and executed concurrently.
4. After a job is executed, you can view the detailed statistics by clicking **Historical Record** in the **Operation** column.

Executed By	Start Time	Last Updated	Duration	Status	Statistics	Schedule	Log
...	5m 34s	Succeeded	Rows read: 0 Bytes read: 14.32 GB Files read: 1 Count of All Files: 1	Written rows: 0 Bytes written: 14.32 GB Written files: 1 Count of All Bytes: 14.32 GB	False Log

[Back](#)

1.4 Using DES to Migrate Historical Data Generated One Month Ago

Migration Process

1. Use a script to import the historical data generated one month ago to the DES Teleport box. For details about the operations related to the DES Teleport box, see the corresponding guide.
2. Use DES to deliver data to HUAWEI CLOUD data center.
3. Use CDM to migrate data from DES to Object Storage Service (OBS).
4. Use CDM to migrate data from OBS to MRS.

The operations on CDM are the same as those described in [Using CDM to Migrate Data of the Last Month](#). File directories are transmitted in binary format and two clusters concurrently execute jobs.

Precautions

1. If the migration affects the source HDFS cluster, manually stop the job.
2. If a large number of jobs fail, perform the following operations:
 - a. Check whether the DES Teleport box is fully written. If the Teleport box is fully written, clear the failed directories to ensure that the data written later is complete.
 - b. Check the network connectivity.
 - c. Check the source HDFS cluster. Check whether indicators are abnormal. If any indicator is abnormal, suspend the migration task.

1.5 Restoring the HBase Table on MRS

After the CDH HBase table directories are migrated to MRS HBase, you can run commands to restore the directories. For data that may change, create snapshots to ensure that the data remains unchanged, and then migrate and restore the data.

Running Commands to Restore the Data Remaining Unchanged

For example, to restore the `/hbase/data/default/table_20180811` table, perform the following operations:

1. Access the node where MRS Client is located, for example, **master1**.
2. Run the following command to switch to user **omm**:

```
su - omm
```

3. Run the following command to load environment variables:
source /opt/client/bigdata_env
4. Run the following command to modify the directory permission:
hdfs dfs -chown omm:hadoop -R /hbase/data/default/table_20180811
 - **omm:hadoop**: Indicates the username. Replace it with the actual username.
 - **/hbase/data/default/table_20180811**: Indicates the path of the table.
5. Run the following command to restore metadata:
hbase hbck -fixMeta table_20180811
6. Run the command to restore regions:
hbase hbck -fixAssignments table_20180811
7. If the message "Status: OK" is displayed, the table is restored successfully.

Using Snapshots to Migrate and Restore the Data that May Change

1. Run the following command in the HBase shell of the source CDH cluster:
flush <table name>
2. Run the following command in the HBase shell of the source CDH cluster:
compact <table name>
3. If the Snap function is not enabled in the table, run the following command to enable the function:
hadoop dfsadmin -allowSnapshot \$path
4. Run the following commands to create an HDFS snapshot named **s0**:
hdfs dfs -createSnapshot <snapshotDir> [s0]
hdfs dfs -createSnapshot test
5. CDM copies files to MRS using the HDFS snapshot. Configure the migration job on CDM as follows:
 - **Source Directory/File**: **/hbase/data/default/src_test/.snapshot/s0**
 - **Write Directory**: **/hbase/data/default/Table name**
6. Run the **fixMeta** and **fixAssignments** commands to restore the table. For details, see [Running Commands to Restore the Data Remaining Unchanged](#).
7. Run the following command to delete the snapshot from the CDH cluster:
hdfs dfs -deleteSnapshot <snapshotDir> s0

Rectifying the Fault That Occurs When Restoring a Table

1. After the **fixMeta** command is executed, the error message "xx inconsistent" is displayed.
The **fixMeta** command is used to check the consistency of metadata between HDFS and HBase. This is a normal situation. Proceed to run the **fixAssignments** command.
2. After the **fixAssignments** command is executed, the error message "xx inconsistent" is displayed.
The **fixAssignments** command is used to restore all regions. Sometimes, some regions go online slowly. You can run the following command to check whether the HBase table is successfully restored:
hbase hbck Table name

If the message "Status: OK" is displayed, the HBase table is restored successfully.

3. After the **fixAssignments** command is executed, an error message is displayed, indicating that multiple regions have the same startkey and some regions overlap.

Run the following command:

hbase hbck -fixHdfsOverlaps *Table name*

Then run the **fixMeta** and **fixAssignments** commands.

2 Consulting Service Data Migration in Financial Industry

2.1 Scenario Overview

Consulting company *H* uses CDM to import local trade statistics to OBS, and Data Lake Insight (DLI) to conduct trade statistics analysis. In this way, company *H* builds its big data analytics platform at an extremely low cost, allowing the company more time to focus on their businesses and make innovations continuously.

Background

Company *H* is a commercial organization in China that engages in collecting trade statistics of major trading nations and buyer data. It has a large-scale trade statistics database. The collected data is widely used in industry research, international trade promotion, and other fields.

In the past, company *H* uses its own big data cluster with maintenance by dedicated personnel. Each year, company *H* purchases the dedicated bandwidth from China Telecom and China Unicom and invests heavily in equipment room, electric power, private networks, servers, and O&M. However, the company cannot satisfy customers' ever-changing service requirements due to insufficient manpower and capability restrictions of its big data cluster. As a result, only 4% of 100 TB inventory data is useful.

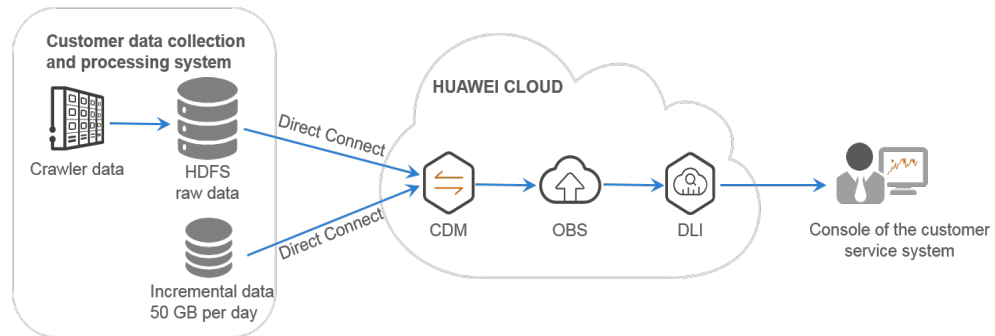
After migrating local trade statistics to HUAWEI CLOUD, company *H* can make full use of the 100 TB inventory data in maximizing asset monetization, without the need of constructing and maintaining infrastructures but relying on HUAWEI CLOUD's big data analysis capabilities.

CDM and DLI use the pay-per-use billing mode, so maintenance personnel are not required and the dedicated bandwidth cost is reduced. Compared with the offline data center, CDM and DLI save the maintenance cost by 70%. In addition, CDM and DLI have low skill demands for personnel and enable smooth migration of existing services, shortening the service rollout period by 50%.

Task

Use CDM, OBS, and DLI to complete trade statistics analysis using the existing data (for example, trade detail records and basic information) of company *H*'s customer data collection and processing system.

Figure 2-1 Scenario scheme



Data Types

- Trade detail records
Trade detail records indicate goods trade statistics of major trading nations.

Table 2-1 Trade detail records

Field Name	Field Type	Field Description
hs_code	string	List of import and export offering code
country	smallint	Basic information about countries
dollar_value	double	Transaction amount
quantity	double	Transaction volume
unit	smallint	Measurement unit
b_country	smallint	Basic information about the target country
imex	smallint	Import or export
y_year	smallint	Year
m_month	smallint	Month

- Basic information
The basic information indicates the dictionary data corresponding to the fields in the trade detail records.

Table 2-2 Basic information about countries (description of **country**)

Field Name	Field Type	Field Description
countryid	smallint	Country code
country_en	string	English name of a country
country_cn	string	Chinese name of a country

Table 2-3 Information about the update time (description of **updatetime**)

Field Name	Field Type	Field Description
countryid	smallint	Country code
imex	smallint	Import or export
hs_len	smallint	Length of the offering code
minstartdate	string	Minimum start time
startdate	string	Start time
newdate	string	Update time
minnewdate	string	Last update time

Table 2-4 Information about import and export offering code (description of **hs246**)

Field Name	Field Type	Field Description
id	bigint	ID
hs	string	Offering code
hs_cn	string	Chinese name of an offering
hs_en	string	English name of an offering

Table 2-5 Information about units (description of **unit_general**)

Field Name	Field Type	Field Description
id	smallint	Measurement unit code
unit_en	string	English name of a measurement unit
unit_cn	string	Chinese name of a measurement unit

2.2 Analysis Process

Introduction

To use CDM, OBS, and DLI to analyze trade statistics, you need to perform the following three steps:

1. **(Optional) Registering an Account**

Before using CDM to migrate and analyze data, you need to register an account with HUAWEI CLOUD and perform real-name authentication. If you already have an account, directly log in.

2. **Using CDM to Upload Data to OBS**

- a. Use CDM to upload the inventory data of company *H* to OBS.
- b. Configure a scheduled task of CDM to automatically upload incremental data to OBS every day.

3. **Using DLI to Analyzing Data**

Interconnect the customer service system with DLI. DLI directly analyzes the service data in OBS to support the customers of company *H* for trade statistics analysis.

2.3 (Optional) Registering an Account

Register an account with HUAWEI CLOUD and perform real-name authentication. If you already have an account, log in directly.

Registering an Account with HUAWEI CLOUD

1. Visit HUAWEI CLOUD at <https://www.huaweicloud.com/en-us/>.
2. Click **Register** in the upper right corner. See **Figure 2-2**.

Figure 2-2 Registration page

Register Your HUAWEI CLOUD Account

Account name

Password

Confirm password

+86(China) Mobile number

Verification code Get Code

I have read and agreed to the [Privacy Statement](#) and [HUAWEI CLOUD Customer Agreement](#).

Register

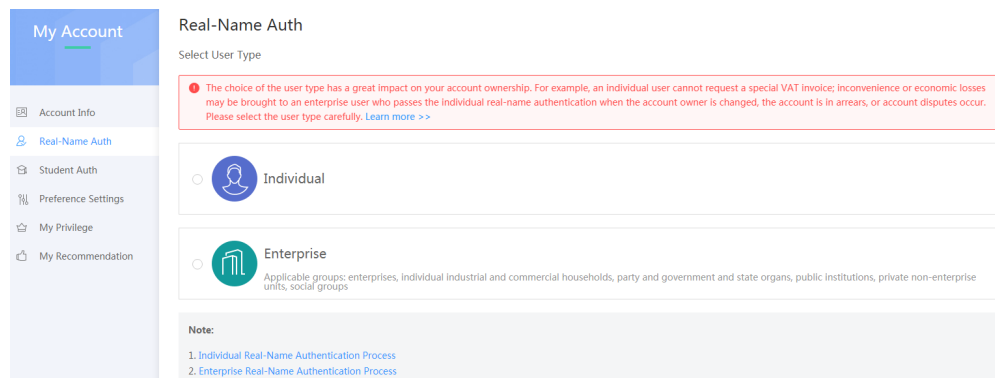
3. Enter an account name, password, and mobile number, and click **Get Code** to obtain the SMS verification code sent by the system. Enter the verification code.
4. Read and select **I have read and agree to the Privacy Statement and HUAWEI CLOUD Customer Agreement**. Click **Register**.

Real-name authentication

According to the laws and regulations of China, cloud service users must finish the real-name authentication before they use any cloud service.

1. After logging in to HUAWEI CLOUD, click the username in the upper right corner, and select **Real Name Authentication**.
2. On the **Real-Name Auth** page that is displayed, select **Individual** or **Enterprise** based on the actual situation. See [Figure 2-3](#).

Figure 2-3 Real-name authentication



3. Perform the authentication as prompted.

2.4 Using CDM to Upload Data to OBS

2.4.1 Uploading Inventory Data

1. Use **Direct Connect** to establish a Direct Connect connection between the local data center and HUAWEI CLOUD Virtual Private Cloud (VPC).
2. Create an OBS bucket and record the access domain name, port number, access key ID (AK), and secret access key (SK) of the OBS bucket.
3. Log in to the **CDM management console** and click **Buy CDM** in the upper right corner to create a CDM cluster. See **Figure 2-4**.
 - **Instance Type**: Select **cdm.medium**, which applies to most migration scenarios.
 - **VPC**: VPC of the CDM cluster. Select the VPC that connects to the local data center through Direct Connect.
 - (Optional) **Subnet** and **Security Group**: You can configure either of them.
 - Auto Shutdown and Scheduled Startup/Shutdown: Retain the default values.

Figure 2-4 Creating a cluster

* Cluster Name ?

* Version 1.5.0

* Instance Type ? ?
Selected Instance Specifications: cdm.medium | 4vCPUs | 8GB

* VPC ? ? [View VPC](#) ?

* Subnet ? ?

* Security Group ? ? [View Security Group](#) ?

* Auto Shutdown ? Auto Shutdown and Scheduled Startup/Shutdown cannot be enabled at the same time.

* Scheduled Startup

* Scheduled Shutdown ?


- After the cluster is created, choose **Job Management > Link Management > Create Link**. The page for selecting a link type is displayed. See [Figure 2-5](#).

Figure 2-5 Selecting a connector

Data Warehouse	<input type="text" value="Data Warehouse Service"/>	<input type="text" value="FusionInsight LibrA"/>	<input type="text" value="Data Lake Insight"/>
Hadoop	<input type="text" value="MRS HDFS"/>	<input type="text" value="MRS HBase"/>	<input type="text" value="MRS Hive"/>
	<input type="text" value="FusionInsight HDFS"/>	<input type="text" value="FusionInsight HBase"/>	
	<input type="text" value="Apache HDFS"/>	<input type="text" value="Apache HBase"/>	
Object Storage	<input type="text" value="HUAWEI CLOUD OBS"/>	<input type="text" value="Alibaba Cloud OSS"/>	<input type="text" value="Qiniu Cloud Object Storage"/>
File System	<input type="text" value="FTP"/>	<input type="text" value="SFTP"/>	<input type="text" value="HTTP"/>
	<input type="text" value="Network Attached Storage"/>		
Relational Database	<input type="text" value="RDS(MySQL)"/>	<input type="text" value="RDS(PostgreSQL)"/>	<input type="text" value="RDS(SQL Server)"/>
	<input type="text" value="DDM"/>	<input type="text" value="MySQL"/>	
	<input type="text" value="PostgreSQL"/>	<input type="text" value="Microsoft SQL Server"/>	<input type="text" value="Oracle"/>
	<input type="text" value="IBM Db2"/>	<input type="text" value="Derecho(GaussDB)"/>	
	<input type="text" value="SAP HANA"/>		
NoSQL	<input type="text" value="Distributed Cache Service"/>	<input type="text" value="Document Database Service"/>	<input type="text" value="CloudTable Service"/>
	<input type="text" value="Redis"/>	<input type="text" value="MongoDB"/>	
Messaging System	<input type="text" value="Data Ingestion Service"/>	<input type="text" value="Apache Kafka"/>	
Search	<input type="text" value="Cloud Search Service"/>	<input type="text" value="Elasticsearch"/>	

- To connect to the local Apache HDFS of company *H*, select **Apache HDFS**, and click **Next**. Set the HDFS link parameters as required. See [Figure 2-6](#).

Figure 2-6 Creating an HDFS link

* Name	<input type="text" value="hdfs_link"/>
* Connector	<input type="text" value="HDFS"/>
* Hadoop Type	<input type="text" value="Apache Hadoop"/>
* URI 	<input type="text" value="hdfs://192.168.1.100:8020/"/>
Authentication Method	<input type="text" value="Kerberos"/>
Principal	<input type="text" value="Select"/>
Keytab File	<input type="button" value="Select"/> No file selected.

<input type="button" value="Cancel"/>	<input type="button" value="Previous"/>	<input type="button" value="Test"/>	<input type="button" value="Save"/>
---------------------------------------	---	-------------------------------------	-------------------------------------

 **NOTE**

- **Name:** Enter a custom link name, for example, **hdfs_link**.
 - **URI:** Enter the NameNode URI of HDFS of company *H*.
 - **Authentication Method:** Select **Kerberos** if Hadoop is in security mode to obtain the **principal** and **keytab** files from the client for authentication.
 - **Principal** and **Keytab File:** Obtain the **principal** account and **keytab** file from the Hadoop administrator.
6. Click **Save**. CDM automatically checks whether the link is available.
 - If the link is available, a message is displayed, indicating that the link is successfully saved, and the link management page is displayed.
 - If the link is unavailable, check whether the link parameters are correctly configured or whether the firewall of company *H* allows the elastic IP address (EIP) of the CDM cluster to access the data source.
 7. Click **Create Link** to create an OBS link. On the page that is displayed, select **HUAWEI CLOUD OBS**, and click **Next**. Set the OBS link parameters as required. See [Figure 2-7](#).

Figure 2-7 Creating an OBS link

* Name	<input type="text" value="obslink"/>
* Connector	<input type="text" value="OBS"/>
Object Storage Type	<input type="text" value="Huawei cloud OBS"/>
OBS Server	<input type="text" value="obs.myhuaweicloud.com"/>
Port	<input type="text" value="443"/>
AK	<input type="text"/>
SK	<input type="text"/>

NOTE

- **Name:** Enter a custom link name, for example, **obslink**.
 - **OBS Server:** Enter the domain name or IP address of OBS, for example, **obs.myhuaweicloud.com**.
 - **Port:** Enter the port number of OBS, for example, **443**.
 - **AK and SK:** Enter the AK and SK used for accessing the OBS database. To obtain the AK and SK, log in to the management console, click the username in the upper right corner, and select **My Credential** from the drop-down list. On the page that is displayed, click the **Access Keys** tab.
8. Click **Save**. The **Link Management** page is displayed.
 9. Choose **Table/File Migration > Create Job** to create a job for migrating trade statistics of company *H* to OBS. See [Figure 2-8](#).

Figure 2-8 Creating a job

Job Configuration

* Job Name

Source Job Configuration	Destination Job Configuration
* Source Link Name <input type="text" value="hdfs_link"/> +	* Destination Link Name <input type="text" value="obslink"/> +
* Source Directory/File <input type="text" value="/data"/> ⊖	* Bucket Name <input type="text" value="Select -cdm"/> ⊖
* File Format <input type="text" value="Binary"/>	* Write Directory <input type="text" value="/program/"/> ⊖
Show Advanced Attributes	* File Format <input type="text" value="Binary"/>
	Duplicate File Processing Method <input type="text" value="Skip"/>
	Show Advanced Attributes

NOTE


- **Job Name:** Enter a user-defined job name.
 - **Source Link Configuration:**
 - **Source Link Name:** Select the HDFS link **hdfs_link** created in 5.
 - **Source Directory/File:** Set this parameter to the local storage path of company *H*'s trade statistics. The value can be either a directory or a file. Set this parameter to a directory. CDM migrates all files in the directory to OBS.
 - **File Format:** Select **Binary**. The file format refers to the format used by CDM to transmit data. The formats of the original files are not changed. **Binary** is recommended for migration between files because the transmission efficiency and performance are optimal.
 - **Destination Link Configuration:**
 - **Destination Link Name:** Select the OBS link **obslink** created in 7.
 - **Bucket Name and Write Directory:** Enter the path for storing trade statistics in OBS. CDM writes the files to this path.
 - **File Format:** Select **Binary**. Similar to the source link, the formats of the original files are not changed.
 - **Duplicate File Processing Method:** Select **Skip**. CDM determines that a file is a duplicate file only when the file name and file size are the same on the source and destination ends. In this case, CDM skips the file and does not migrate the file to OBS.
10. Click **Next** to go to the tab page for configuring the task parameters. For the migration of inventory data, retain the default values of the parameters.
 11. Click **Save and Run**. The **Job Management** page is displayed. You can view the job execution progress and result.
 12. After the job is successfully executed, click **Historical Record** to view the number of written rows, number of read rows, number of written bytes, number of written files, and execution logs.

2.4.2 Uploading Incremental Data

1. After uploading inventory data using CDM, click **Edit** in the **Operation** column to modify a job.
2. Retain the values of the basic parameters, and click **Next** to modify the task parameters. See [Figure 2-9](#).

Figure 2-9 Configuring a scheduled task

Configure Task

Concurrent Extractors 

Schedule Execution

Minute Hour **Day** Week Month

Cycle (days) Executed once every ** days.

Validity Period

Start Time End Time

[Show Advanced Attributes](#)

3. Select **Schedule Execution** and configure the scheduled task.
 - Set **Cycle (days)** to 1 day.
 - Set **Start Time** to 00:01:00 every day.

In this way, CDM automatically performs full migration in the early morning every day. However, because **Duplicate File Processing Method** is set to **Skip**, files with the same name and size are not migrated. Therefore, only new files are uploaded every day.

4. Click **Save**.

2.5 Using DLI for Data Analysis

Interconnect the customer service system with DLI to conduct trade statistics analysis on OBS data.

NOTE

You can access DLI using ODBC APIs on the local PC running a Windows OS.

Prerequisites

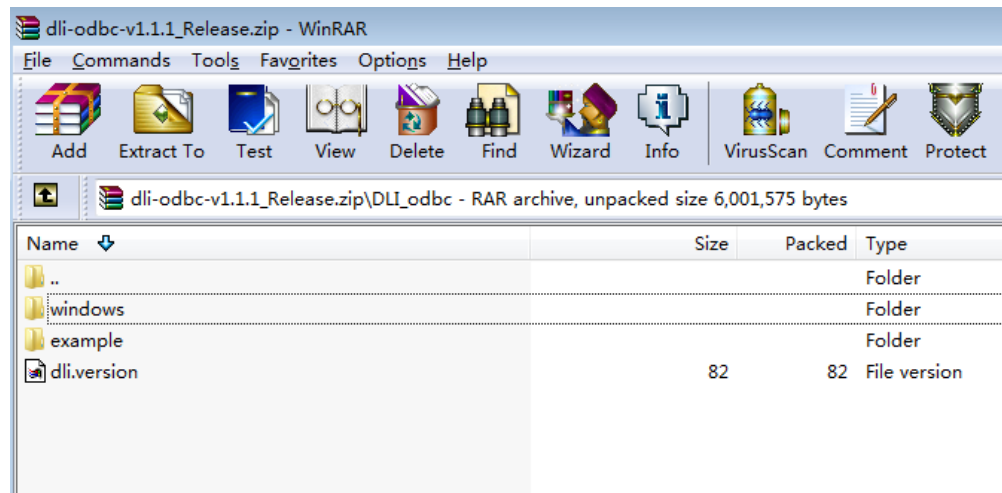
Before using the ODBC interface to access DLI, download the ODBC driver release package as follows:

1. Log in to the DLI management console.
2. Click **Common links** in the upper right corner.
3. In the **Download Driver** area, click **Download**.

ODBC driver release package structure

- The ODBC driver release package of DLI is **dli-odbc-v1.1.1_Release.zip**.
- **Figure 2-10** shows the directory structure of the ODBC driver release package.

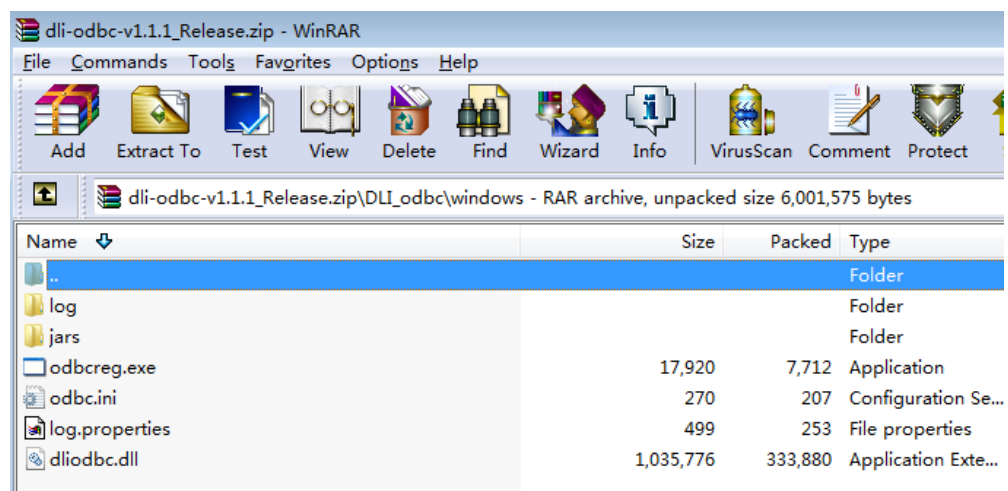
Figure 2-10 Directory structure of the ODBC driver release package



In the preceding figure:

- The **example** directory provides C and C# sample code for client applications.
- The **windows** directory refers to the ODBC driver installation directory.
- The **dli.version** file refers to the driver version.
- **Figure 2-11** shows the directory structure of **DLI_odbc\windows**.

Figure 2-11 Directory structure of **DLI_odbc\windows**



In the preceding figure:

- The **log** directory stores ODBC driver logs.
- The **jars** directory includes the Java package on which the ODBC driver depends.
- **odbcereg.exe** refers to the driver and data source registration tool.

- **odbc.ini** refers to the driver configuration file, which is used to register data sources.
- **log.properties** refers to the log configuration file, which is used to configure the log output level and output path.
- **dliodbc.dll** refers to the ODBC dynamic driver library.

Environment Requirements

- Windows 2012 or later
- Java 1.8
- Visual Studio 2012 (used to compile and run sample programs)

Environment Preparation

1. Install the Java operating environment.

Obtain the 32-bit installation package of Java 1.8 from the Java official website. The download path is as follows:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

For details, see [Figure 2-12](#).

Figure 2-12 Downloading page

Java SE Development Kit 8 Downloads

Thank you for downloading this release of the Java™ Platform, Standard Edition Development Kit (JDK™). The JDK is a development environment for building applications, applets, and components using the Java programming language.

The JDK includes tools useful for developing and testing programs written in the Java programming language and running on the Java platform.

See also:

- [Java Developer Newsletter](#): From your Oracle account, select **Subscriptions**, expand **Technology**, and subscribe to **Java**.
- [Java Developer Day hands-on workshops \(free\) and other events](#)
- [Java Magazine](#)

[JDK 8u161 checksum](#)
[JDK 8u162 checksum](#)

Java SE Development Kit 8u161

You must accept the [Oracle Binary Code License Agreement for Java SE](#) to download this software.

Accept License Agreement Decline License Agreement

Product / File Description	File Size	Download
Linux ARM 32 Hard Float ABI	77.92 MB	jdk-8u161-linux-arm32-vfp-hflt.tar.gz
Linux ARM 64 Hard Float ABI	74.88 MB	jdk-8u161-linux-arm64-vfp-hflt.tar.gz
Linux x86	168.96 MB	jdk-8u161-linux-i586.rpm
Linux x86	183.76 MB	jdk-8u161-linux-i586.tar.gz
Linux x64	166.09 MB	jdk-8u161-linux-x64.rpm
Linux x64	180.97 MB	jdk-8u161-linux-x64.tar.gz
macOS	247.12 MB	jdk-8u161-macosx-x64.dmg
Solaris SPARC 64-bit (SVR4 package)	139.99 MB	jdk-8u161-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	99.29 MB	jdk-8u161-solaris-sparcv9.tar.gz
Solaris x64	140.57 MB	jdk-8u161-solaris-x64.tar.Z
Solaris x64	97.02 MB	jdk-8u161-solaris-x64.tar.gz
Windows x86	198.54 MB	jdk-8u161-windows-i586.exe
Windows x64	206.51 MB	jdk-8u161-windows-x64.exe

NOTE

After the installation is complete, record the Java installation directory, for example: **C:\Program Files (x86)\java\jdk1.8.0_161**.

2. Install the software Visual Studio 2012 required for installing and compiling the example code.

Installing the ODBC Driver

1. Decompress the installation package.

Decompress **dli-odbc-v1.1.1_Release.zip** to the root directory of drive C.

NOTE

The installation directory can be adjusted. You are advised to install it in the root directory of disk C.

2. Register the ODBC driver.
 - a. In the PC running the Windows OS, click **Start**, enter **cmd** in the search box, and press **Enter**. In the command prompt window, access the **windows** directory under the directory where **dli-odbc-v1.1.1_Release.zip** is decompressed. For example: **C:\DLI_odbc\windows**.

- b. Run the following command:

```
.\odbcereg /i DLIODBC dliodbc.dll C:\DLI_odbc\windows
```

 **NOTE**

- **DLIODBC** refers to the name of the registered DLI ODBC driver.
 - **C:\DLI_odbc\windows** refers to the **windows** directory under the directory where **DLI-odbc-v1.1.1_Release.zip** is decompressed. You can change the directory according to the actual conditions.
3. Enter **y** as prompted.

The command output is as follows:

```
ODBCREG - to register or unregister ODBC driver

Proceeding to register...

driver: DLIODBC
dll: dliodbc.dll
path: C:\DLI_odbc\windows

Confirm (y/n)
DLIODBC installed/registered successfully
```

4. View the registry. Click **Start**, enter **cmd** in the search box, and press **Enter**. In the command prompt window, enter **regedit** and press **Enter** to open the registry editor.

 **NOTE**

- Check that driver **DLIODBC** exists in **HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\ODBC\ODBCINST.INI** of the registry.
- The registry path of the 32-bit Windows OS is **HKEY_LOCAL_MACHINE\SOFTWARE\ODBC\ODBCINST.INI**.

Configuring Environment Variables

1. Configure environment variables.

Configure and modify the following system variables:

- Variable: **JAVA_CLIENT_DLL**

Value: C:\Program Files (x86)\java\jdk1.8.0_161\jre\bin\client\jvm.dll

 **NOTE**

Configure the variable based on the actual installation directory.

- Variable: **Path**

Value: C:\Program Files (x86)\java\jdk1.8.0_161\jre\bin\client;%Path%

 **NOTE**

Configure the variable based on the actual installation directory.

2. Restart the system for the settings to take effect.

Configuring Log Parameters

Open the **log.properties** file, for example, **C:\DLI_odbc\windows\log.properties**, in the ODBC installation directory.

Modify the following parameters:

- Log output level
log4cplus.rootLogger=INFO,DLILog

 **NOTE**

You are advised to set the level to **INFO**. To locate a fault, you can change the value to **DEBUG**.

- Log output path:

log4cplus.appender.DLILog.File=C:\DLI_odbc\windows\log\DLI_odbc.log

 **NOTE**

You are advised to set the log output path to the log directory in the installation path. Set this parameter based on the actual installation path.

Configuring Data Sources

1. Prepare the data source configuration file.

Modify the **odbc.ini** file in the **C:\DLI_odbc\windows** directory as follows:

```
[DLI]
DRIVER=DLIODBC
DATABASE=#### Configure the actual DLI database name.
HOST=DLI.cn-north-1.myhuaweicloud.com
PORT=-1
PROJECTID=#### Set this parameter based on the actual account used to log in
to DLI.
QUEUENAME=default #### Set this parameter based on the actual queue created
in DLI.
AK=#### Set this parameter based on the actual account used to log in to DLI.
SK=#### Set this parameter based on the actual account used to log in to DLI.
REGION=dli.cn-north-1.myhuaweicloud.com
AUTHMODE=1
SVCNAME=DLI
CONFFILE=odbc.ini
USEPROXY=0
```

2. Add the user data source.

- a. Run the following command:

.\odbcereg.exe /d DLIODBCUserDS C:\DLI_odbc\windows\odbc.ini

 **NOTE**

DLIODBCUserDS refers to the user data source name, which is used in applications.

- b. After confirming that the print is consistent with the content in the **odbc.ini** file, enter **y**.

The command output is as follows:

```
ODBCREG - to register or unregister ODBC driver

Attribute to be register are:
DRIVER = DLIODBC
DATABASE = ####
HOST = dli.cn-north-1.myhuaweicloud.com
PORT = -1
PROJECTID = ####
QUEUENAME = default
AK = ####
SK = ####
REGION = dli.cn-north-1.myhuaweicloud.com
AUTHMODE = 1
SVCNAME = DLI
CONFFILE = odbc.ini
Confirm(y/n)
Add datasource .
```

3. View the registry. Click **Start**, enter **cmd** in the search box, and press **Enter**. In the command prompt window, enter **regedit** and press **Enter** to open the registry editor.

Check whether **HKEY_CURRENT_USER\Software\ODBC\ODBC.INI\DLIODBCUserDS** exists.

4. Add the system data source.

- a. Run the following command:

```
.\odbcreg.exe /s DLIOBCSysDS C:\DLI_odbc\windows\odbc.ini
```



DLIOBCSysDS refers to the system data source name, which is used in applications.

- b. After confirming that the print is consistent with the content in the **odbc.ini** file, enter **y**.

The command output is as follows:

```
ODBCREG - to register or unregister ODBC driver

Attribute to be register are:
DRIVER = DLIODBC
DATABASE = ####
HOST = dli.cn-north-1.myhuaweicloud.com
PORT = -1
PROJECTID = ####
QUEUEName = default
AK = ####
SK = ####
REGION = dli.cn-north-1.myhuaweicloud.com
AUTHMODE = 0
SVCNAME = DLI
CONFFILE = odbc.ini
Confirm(y/n)
Add datasource .
```

5. View the registry. Click **Start**, enter **cmd** in the search box, and press **Enter**. In the command prompt window, enter **regedit** and press **Enter** to open the registry editor.

Check whether **HKEY_LOCAL_MACHINE\SOFTWARE\Wow6432Node\ODBC\ODBC.INI\DLIOBCSysDS** exists.

Accessing DLI Based on the ODBC Driver Through Code

1. Create databases and tables, including the trade statistics database, trade detail record table, and basic information table.
2. Conduct data query.

A C# code example is as follows:

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Data;
using System.Data.Odbc;

namespace ConsoleApplication1
{
    class Program
    {
        static void Main(string[] args)
        {
            string connectionString = "dsn=DLIOBCds";

            string queryString = "show tables;";//show tables; You can use the
            corresponding DDL or DML statement instead.
            OdbcCommand command = new OdbcCommand(queryString);
```

```
using (OdbcConnection connection = new OdbcConnection(connectionString))
{
    OdbcDataReader reader;
    try
    {
        command.Connection = connection;

        connection.Open();

        reader = command.ExecuteReader();

        while (reader.Read())
        {
            Console.WriteLine("\t{0}\t{1}\t{2}",
                reader[0], reader[1], reader[2]);
        }

        reader.Dispose();
        reader.Close();
    }
    catch (Exception ex)
    {
        Console.WriteLine(ex.Message);
    }
}

command.Dispose();
}
```

 **NOTE**

During the query, replace the name of **dns** specified in **string connectionString** with the DLI ODBC driver.