

Alibaba Cloud Table Store

Data channels

Issue: 20190731

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use

or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).
6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 Danger: Resetting will result in the loss of user configuration data.
	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 Warning: Restarting will cause business interruption. About 10 minutes are required to restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	 Note: You can use Ctrl + A to select all files.
>	Multi-level menu cascade.	Settings > Network > Set network type
Bold	It is used for buttons, menus, page names, and other UI elements.	Click OK .
Courier font	It is used for commands.	Run the <code>cd / d C :/ windows</code> command to enter the Windows system folder.
<i>Italics</i>	It is used for parameters and variables.	<code>bae log list --instanceid Instance_ID</code>
[] or [a b]	It indicates that it is an optional value, and only one item can be selected.	<code>ipconfig [-all -t]</code>

Style	Description	Example
<code>{}</code> or <code>{a b}</code>	It indicates that it is a required value, and only one item can be selected.	<code>swich {stand slave}</code>

Contents

Legal disclaimer.....	I
Generic conventions.....	I
1 OSS.....	1
1.1 Overview.....	1
1.2 Full export (script mode).....	3
1.3 Incremental synchronization (script mode).....	6

1 OSS

1.1 Overview

Table Store is a distributed NoSQL data storage service that is built on Alibaba Cloud Apsara distributed system. It uses data partitioning and load balancing techniques to seamlessly scale up data size and access concurrency, providing storage of, and real-time access to, massive structured data.

Object Storage Service (OSS) is a massive-volume, secure, low-cost, and highly-reliable cloud storage service. It provides 99.999999999% data reliability. You can use RESTful API for storage and access in any place on the Internet. Its capacity and processing capability can be elastically scaled, and multiple storage modes are provided, comprehensively optimizing the storage cost.

Scenarios

Table Store: Provides professional data-persistent storage service and user-oriented real-time read/write operations with high concurrency and low latency.

OSS: Supports backup at an extremely low cost.

Usage

- Write

Data can be directly written to Table Store.

- Read

Data can be directly read from Table Store.

- Back up

Automatic backup is supported.

- Restoration

Data can be re-written to Table Store through Data Integration (OSSReader and OTSWriter).

Constraints

- Write by whole row

Table Store Stream requires that a whole row of data be written to Table Store each time. Currently, the whole-row data write mode is applied to the writing of time sequence data such as IoT data. Therefore, data cannot be modified subsequently.

- Synchronization latency

Currently, periodic scheduling is used and the scheduling interval is 5 minutes. The plugin has a latency of 5 minutes and the total latency of a synchronization task is 5 to 10 minutes.

Activation

- Activate Table Store

1. Go to the [Table Store product details page](#).
2. Click Buy Now.
3. In the [Table Store console](#), create [instances](#) and [data tables](#).



Note:

- To use the incremental tunnel, you must activate the [Stream function](#) for data tables. You can select 24 hours for the validity period.
- Table Store supports the reserved CUs and additional CUs. If the reserved read and write CUs are both set to zero during table creation, then the additional read and write CU is used. You can adjust the reserved read/write CU of each table at any time.
- Table Store offers each registered account 25 GB of free storage per month.

- Activate OSS

1. Go to the [OSS product details page](#).
2. Click Buy Now.

Data tunnel

Offline

- Export the full data to OSS.
 - [Script mode](#)

- Synchronize data to OSS in incremental mode.
 - [Script mode](#)
- Fully import data into Table Store.
 - Script mode

1.2 Full export (script mode)

Data Integration supports data synchronization in wizard mode and script mode. Wizard mode is simpler, while script mode is more flexible.

This section describes how to export full data in Table Store to OSS using the script mode of Data Integration, so that you can download the data as needed or save it as backup data of Table Store to OSS.

Channels

Script mode of Data Integration:

- Reader: OTSReader
- Writer: OSSWriter

Step 1. Create a Table Store data source



Note:

Skip this step if you have created a Table Store data source.

For more information about how to create a data source, see [Create a Table Store data source](#).

Step 2. Create an OSS data source

This operation is similar to Step 1. You only need to select OSS as the data source.



Note:

During parameter configuration of the OSS data source, Endpoint does not contain bucketName.

Step 3. Create an export task

1. Log on to the Data Integration console.
2. On the Sync Tasks page, select Script Mode.

3. In the Import Template dialog box, set Source Type to Table Store (OTS) and Type of Objective to OSS.
4. Click OK to go to the configuration page.

Step 4. Set configuration items

1. On the configuration page, templates for OTSReader and OSSWriter are provided. Complete the configurations by referring to the following annotations.

```
{
  " type ": " job ",      # It cannot be modified .
  " version ": " 1 . 0 ", # It cannot be modified .
  " configurat ion ": {
    " setting ": {
      " errorLimit ": {
        " record ": " 0 " # The import task fails when the
        number of error records exceeds the value .
      },
      " speed ": {
        " mbps ": " 1 ", # Import speed , in Mbps .
        " concurrent ": " 1 " # Concurrent y .
      }
    },
    " reader ": {
      " plugin ": " ots ", # It cannot be modified .
      " parameter ": {
        " datasource ": "", # Name of the data source in
        Data Integratio n , which must be set in advance .
        You can configure data source or write authentica
        tion informatio n such as the AccessKeyI D in
        plaintext . We recommend that you configure data
        source .
        " table ": "", # Table name in Table Store .
        " column ": [ # Name of the column that needs to
        be exported to OSS . If all the columns need to
        be exported to OSS , set this parameter to an
        empty array .
          {
            " name ": " column1 " # Name of the column in
            Table Store , which needs to be imported to OSS
          },
          {
            " name ": " column2 " # Name of the column in
            Table Store , which needs to be imported to OSS
          }
        ],
        " range ": {
          " begin ": [
            {
              " type ": " INF_MIN " # Start position of the
              first primary key column in Table Store . If you
              want to export full data , set this parameter to
              INF_MIN . If you want to export a portion of the
              data , set this parameter as needed . The number of
              configurat ion items in " begin " must be the same
              as the number of primary key columns .
            }
          ],
          " end ": [
            {
```

```

        " type ": " INF_MAX " # End position of the
first primary key column in Table Store . If you
want to export full data , set this parameter to
INF_MAX . If you want to export a portion of the
data , set this parameter as needed .
    }
  ],
  " split ": [ # Used to configure partition
informatio n about the Table Store table , which can
accelerate s the export . In the next version , this
configurat ion is automatica lly processed .
  ]
}
},
" writer ": {
  " plugin ": " oss ",
  " parameter ": {
    " datasource ": "", # Name of the OSS data source
    " object ": "", # Prefix of the object excluding
the bucket name , for example , tablestore / 20171111 /.
If the export is scheduled , a variable , for example
, tablestore /${ date } , must be used , and ${ date } must
be configured when the scheduling parameters are
set .
    " writeMode ": " truncate ", # truncate , append , and
nonConflic t are supported . truncate is used to
clear existing files with the same name , append
is used to add the data to existing files with
the same name , and nonConflic t is used to return
an error when files with the same name exist .
truncate is used during full export .
    " fileFormat ": " csv ", # CSV and TXT are supported .
    " encoding ": " UTF - 8 ", # Encoding mode
    " nullFormat ": " null ", # Defines a string identifier
that represents the null value . It can be an
empty string .
    " dateFormat ": " yyyy - MM - dd HH : mm : ss ", # Time
format
    " fieldDelim iter ": "," # Delimiter of each column
  }
}
}
}
}

```

2. Click Save to save the task.

Step 5. Run the task

1. Click operation to run the task.

If the configurations contain variables, for example, `${date}`, the variable setting page is displayed. You can set only specific values.

2. View logs in the lower part of the page.

If no error is logged, the task is successfully executed, and you can check the data in the target OSS instance.



Note:

Full export is generally a one-time task, and thus you do not need to set automatic scheduling parameters. For more information about how to set the scheduling parameters, see [Incremental synchronization](#).

Step 6. Check the data exported to OSS

1. Log on to the [OSS console](#).
2. Select the bucket and file name, and verify its contents.

1.3 Incremental synchronization (script mode)

Data Integration supports data synchronization in wizard mode and script mode.

Wizard mode is simpler while script mode is more flexible.

This section describes how to synchronize incremental data in Table Store to OpenSearch using the script mode of Data Integration.

Channels

Script mode of Data Integration

- Reader: OTSStream Reader
- Writer: OSSWriter

Configure Table Store

No prior configurations required.

Configure OSS

No prior configurations required.

Configure Data Integration

1. Create a Table Store data source.



Note:

- If you have already created a Table Store data source, skip this step.

- If you do not want to create a data source, you can specify the endpoint, instanceName, AccessKeyID, and AccessKeySecret on the subsequent configuration page.

For more information about how to create a data source, see [Create a Table Store data source](#).

2. Create an OSS data source.

This step is similar to Step 1. You only need to select OSS as the data source.



Note:

During parameter configuration of the OSS data source, Endpoint does not contain bucketName.

3. Create a synchronization task.

- Log on to the [Data Integration console](#).
- On the Sync Tasks page, select Script Mode.
- In the Import Template dialog box that appears, set Source Type to Table Store Stream (OTS Stream) and Type of Objective to OSS.
- Click OK to go to the configuration page.

4. Set configuration items.

- On the configuration page, templates of OTSStreamReader and OSSWriter are provided. Complete the configurations by referring to the following annotations.

```
{
  " type ": " job ",
  " version ": " 1 . 0 ",
  " configurat ion ": {
    " setting ": {
      " errorLimit ": {
        " record ": " 0 " # Allowed number of errors . If
        the number of errors exceeds the value , the
        synchroniz ation task fails .
      },
      " speed ": {
        " mbps ": " 1 ", # Maximum traffic of each synchroniz
        ation task .
        " concurrent ": " 1 " # Number of concurrent synchroniz
        ation tasks each time .
      }
    }
  },
  " reader ": {
    " plugin ": " otsstream ", # Name of the Reader plugin .
    " parameter ": {
      " datasource ": "", # Name of the Table Store data
      source . If this parameter is set , you do not
      need to set endpoint , accessID , accessKey , and
      instanceNa me .
    }
  }
}
```

```

" dataTable ": "", # Name of the table in Table
Store .
" statusTable ": " TableStore StreamRead erStatusTa ble ",
# Table that stores the Table Store Stream status
; using the default value is recommende d
" startTimes tampMillis ": "", # Start time of the
export . In incrementa l export mode , the task
needs to be executed cyclically , and the start
time is different at each execution . Therefore , you
must set a variable , for example , ${ start_time }.
" endTimesta mpMillis ": "", # End time of the export
. You must set a variable , for example , ${ end_time
}.
" date ": " yyyyMMdd ", # Date from which data is
exported . This parameter is the same as startTimes
tampMillis and endTimesta mpMillis , and therefore must
be deleted .
" mode ": " single_ver sion_and_u pdate_only ", # Format
of the data exported from Table Store Stream .
Currently , the parameter must be set to single_ver
sion_and_u pdate_only . Add this parameter if it is
not in the configurat ion template .
" column ":[ # Names of the columns to be exported
from Table Store to OSS . Add this parameter if
it is not in the configurat ion template . Set
this parameter as needed .
{
" name ": " uid " # Name of the column . It
is the primary key column in Table Store .
},
{
" name ": " name " # Name of the column . It
is an attribute column in Table Store .
},
],
" isExportSe quenceInfo ": false , # This parameter can
only be set to false in single_ver sion_and_u
pdate_only mode .
" maxRetries ": 30 # Maximum number of retry times .
}
},
" writer ": {
" plugin ": " oss ", # Name of the Writer plugin
" parameter ": {
" datasource ": "", # Name of the OSS data source
" object ": "", # Prefix of the name of the last
file to be backed up to OSS . The recommende d
value is the Table Store instance name , table
name , or date , for example , " instance / table /{ date
} ".
" writeMode ": " truncate ", # truncate , append , and
nonConflic t are supported . truncate is used to
clear existing files with the same name , append
is used to add the data to existing files with
the same name , and nonConflic t is used to
return an error when files with the same name
exist .
" fileFormat ": " csv ", # File format
" encoding ": " UTF - 8 ", # Encoding mode
" nullFormat ": " null ", # Mode of representa tion in
a TXT file under control
" dateFormat ": " yyyy - MM - dd HH : mm : ss ", # # Time
format
" fieldDelim iter ": " ," # Delimiter of each column

```

```

}
}
}
}
}

```



Note:

For detailed configuration description, see [OTSStreamReader](#) and [OTSStreamWriter](#).

- b. Click Save.
5. Run the task.
 - a. Click operation.
 - b. In the dialog box that appears, set the variable parameters.
 - c. Click OK.
 - d. After the task is completed, log on to the [OSS console](#) to verify whether files are backed up.
 6. Configure scheduling.
 - a. Click Submit.
 - b. In the dialog box that appears, set the scheduling parameters.

The parameters are described as follows.

Parameter	Description
Scheduling type	Select cycle control.
Automatically re-run	This parameter indicates that the task reruns for three times at an interval of 2 minutes if the task fails.
Start date	The default value is recommended, which is from January 1, 1970 to 100 years later.
Scheduling cycle	Select Minute.
Start Time	Select “00:00 to 23:59” , which indicates that scheduling is required for a full day.
Interval	Select 5 Minutes.
start_time	Enter \$[yyyymmddhh24miss-10/24/60], which indicates the time of the scheduling task minus 10 minutes.

Parameter	Description
end_time	Enter <code>\$\$\$[yyyymmddhh24miss-5/24/60]</code> , which indicates the time of the scheduling task minus 5 minutes.
date	Enter <code>\$\$\$[bdp.system.bizdate]</code> , which indicates the scheduling date.
Dependency attributes	Set this parameter if a dependency exists. If no dependency exists, do not set this parameter.
Cross-cycle dependency	Self-dependent: The operation can continue only after the previous scheduling cycle is completed.

c. Click OK.

The periodic synchronization task is configured, the configuration file status is Read-only.

7. Check the task.

a. At the top of the page, click Operation Center.

b. On the left-side navigation pane, click Task List > Cycle Task to view the created synchronization task.

c. The new task begins running at 00:00 on the next day.

- In the left-side navigation pane, choose Task O&M > Cycle Instance to view each pre-created synchronization task of the day. The scheduling interval is 5 minutes and each task processes data from the past 5 to 10 minutes.
- Click the instance name to view its details.

d. You can view the log when a task is running or after it is completed.

8. Check the data exported to OSS.

Log on to the [OSS console](#) to check whether a new file is generated and whether the file content is correct.

Once the preceding settings are completed, data in Table Store can be automatically synchronized to OSS at a latency of 5 to 10 minutes.