

Alibaba Cloud

Data Transmission Service Replication Modes

Document Version: 20200817

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 because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should 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 this document based on the "status quo", "being defective", and "existing functions" of its products and services. 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 take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, 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 contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, 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 this document 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 directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
 Danger	A danger notice 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.
 Warning	A warning notice 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 restart an instance.
 Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: If the weight is set to 0, the server no longer receives new requests.
 Note	A note indicates supplemental instructions, best practices, tips, and other content.	 Note: You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click Settings> Network> Set network type .
Bold	Bold formatting is used for buttons, menus, page names, and other UI elements.	Click OK .
Courier font	Courier font is used for commands	Run the <code>cd /d C:/window</code> command to enter the Windows system folder.
<i>Italic</i>	Italic formatting is used for parameters and variables.	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] or [a b]	This format is used for an optional value, where only one item can be selected.	<code>ipconfig [-all -t]</code>
{ } or {a b}	This format is used for a required value, where only one item can be selected.	<code>switch {active stand}</code>

Table of Contents


1.Data migration mode	05
2.Data integration mode	08
3.Data synchronization mode	11
4.Change tracking mode	15


1.Data migration mode

The data migration mode migrates data between data stores. This mode is typically used for one-time migrations that require minimized downtimes. The data migration mode also provides transform, and load (ETL) features such as object name mapping and data filtering.

Data migration phases

A data migration task consists of several phases, namely schema migration, full data migration, and incremental data migration. You can select the required phases when creating a data migration task. However, to conduct a minimized-downtime migration, you must select all the three phases.

Migration phase	Description
Schema migration	<p>In the schema migration phase, Data Transmission Service (DTS) migrates schema objects from the source database to the target database, including tables, views, triggers, and stored procedures.</p> <p>To migrate schema objects across heterogeneous databases, DTS converts schema objects into data types that work with the target database. For example, it converts the NUMBER data type in Oracle databases to the DECIMAL data type in MySQL databases.</p>
Full data migration	<p>In the full data migration phase, DTS replicates the existing data from the source database to the target database. If you select only schema migration and full data migration for the task, updates that occur during the full data migration phase will not be captured and replicated to the target database.</p> <div style="background-color: #e6f2ff; padding: 10px; border: 1px solid #d9e1f2;"> <p> Note To ensure data consistency, we recommend that you suspend all updates to the source database during the full data migration phase. To migrate data with minimized downtimes, you must select schema migration, full data migration, and incremental data migration when creating the data migration task.</p> </div>

Migration phase	Description
Incremental data migration	<p>DTS tracks the transaction log of the source database to capture the changes that occur during the full data migration phase. After the full data migration phase, DTS re-applies these data changes in the target database.</p> <div style="border: 1px solid #ADD8E6; padding: 5px; margin-top: 10px;"> <p> Note The incremental data migration phase is an ongoing replication process. Therefore, a migration task with incremental data migration enabled does not automatically stop. You must manually stop the migration task.</p> </div>

ETL features

Data migration supports the following ETL features:

- **Object name mapping:** You can configure name mappings between source and target columns, tables, or databases that have different names.
- **Data filtering:** You can use SQL-like queries to filter the source table to migrate only records that match certain conditions. For example, you can specify a time range to migrate only the latest data.

Alerting

The creator of a migration task can choose to receive alerts about errors that occur during migration and take immediate action.

Throttling

To minimize the impact on your normal operations, you can set transmission rates for the full data migration phase of a data migration task so that the data migration workloads do not overwhelm your source database.

Migration task lifecycle

To fulfill a data replication process in any of the three modes, you must create a data replication task. A data replication task holds all the configurations of the replication process, such as the source database, target database, selected migration phases, and objects to be migrated.

The following table describes the statuses of a data migration task.

Task status	Description	Available operations
Not Started	The migration task is configured but the precheck is not performed.	Perform the precheck and delete the migration task
Prechecking	A precheck is being performed.	Delete the migration task

Task status	Description	Available operations
Precheck Passed	The migration task has passed the precheck but is not started.	Start or delete the migration task
Migrating	Data is being migrated.	Pause, stop, or delete the migration task
Migration Failed	An error occurred. You can determine the phase during which the error occurred according to the progress of the migration task.	Delete the migration task
Paused	The migration task is paused.	Resume or delete the migration task
Completed	The migration task is completed, or you have clicked End to manually stop the migration task.	Delete the migration task

A migration task in the **Migration Failed** status will be retried within seven days of the failure. To discard the data migration, you must manually stop or delete the migration task.

References

- To start data migration, you can log on to the [DTS console](#) to create a data migration task.
- To migrate applications, you can use the Alibaba Cloud Migration Tool. For more information, see [Alibaba Cloud Migration Tool](#).

2.Data integration mode

The data integration mode of Data Transmission Service (DTS) allows you to schedule data migration tasks to run on a regular basis. This mode is typically used for recurring migrations in a large data warehouse system. For example, you can schedule a data migration that recurs every night and transfers the transactional data generated during the day to a data warehouse.


Features

- Provides multiple orchestration methods that suit your varied use cases. For example, you can choose to migrate data to a new database every time or always into the same database.
- Supports weekly and monthly recurrence patterns.


Supported sources and targets

The supported source and target data stores are the same as the ones supported by the data migration mode. The only difference is that the data integration mode does not provide incremental data migration. For more information, see [Supported sources and targets](#).

Orchestration methods

Orchestration method	Description	Example use case
New database every time	<p>For each occurrence of your migration task, DTS creates a new database in the target database system and replicates the selected schema objects and data objects into the new database. For each new database, DTS appends a timestamp-based suffix (<code>_yyymmdd_HH_mm_ss</code>) to the name.</p> <p> Note Using this orchestration method, DTS may take a large amount of storage in the target system. Therefore, you must make sure that the target system has enough storage for the data to be replicated.</p>	You may need to replicate transactional data from a production system to multiple test systems on a regular basis.

Orchestration method	Description	Example use case
Always to the same database	For each occurrence of your migration task, DTS always replicates the selected schema objects and data objects into the same database. Between two occurrences, you must clear the schema objects and data objects in the target database. Otherwise, the next occurrence of the data integration task fails.	You may need to replicate the transactional data from an operational database to an analytics system on a regular basis. This deployment is a useful option when you want to deploy near-real time replications between two databases that are not supported by the data synchronization mode of DTS.

Orchestration method	Description	Example use case
Timestamp-based differentials	<p>For each occurrence of your migration task, DTS replicates only the differentials over the last occurrence. The differentials are determined based on a timestamp column that you specify.</p> <div data-bbox="616 551 976 1288" style="background-color: #e0f2f7; padding: 10px;"> <p> Note</p> <ul style="list-style-type: none"> • This orchestration method supports only migrations between MySQL databases. • Each source table must have a timestamp column. • Each source table must not have foreign key constraints. • Updates that physically remove data objects, such as DELETE and TRUNCATE operations, cannot be replicated. </div>	<p>You may need to schedule a data migration that recurs every night and transfers the transactional data generated during the day to a data warehouse so that your analytics system can work on one-day-old data.</p>

3.Data synchronization mode

The data synchronization mode of Data Transmission Service (DTS) replicates ongoing changes between two data stores. This mode is typically used for OLTP-to-OLAP replications and disaster recovery-related replications.

Supported objects

Objects that can be replicated include databases, tables, and columns. You can specify objects that you want to replicate.

Advanced features

Feature	Description	References
Reselect objects to be synchronized	You can add and remove objects to be synchronized while a data synchronization task is running.	<ul style="list-style-type: none"> Add an object to a data synchronization task Remove an object from a data synchronization task
Track synchronization performance	You can add and remove objects to be synchronized while a data synchronization task is running.	Check the synchronization performance
ETL features	<ul style="list-style-type: none"> Object name mapping: You can define mappings to name target objects differently from their source objects. Supported objects include databases, tables, and columns. Data filtering: You can use SQL-like queries to filter the source table to replicate only records that match certain conditions. 	<ul style="list-style-type: none"> Object name mapping Data filtering

Data synchronization task lifecycle

To fulfill a data replication process in any of the three modes, you must create a data replication task. A data replication task holds all the configurations of the replication process, such as the source database, target database, and objects to be replicated.

The following table describes the statuses of a synchronization task.

Task status	Description	Applicable operations
-------------	-------------	-----------------------


Task status	Description	Applicable operations
Prechecking	A precheck is being performed and the synchronization task has not started.	<ul style="list-style-type: none"> • View the synchronization configurations • Delete the synchronization task • Duplicate synchronization configurations
Precheck Failed	The synchronization task failed to pass the precheck.	<ul style="list-style-type: none"> • Perform the precheck • View the synchronization configurations. • Reselect objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Duplicate synchronization configurations
Not Started	The synchronization task has passed the precheck but is not started.	<ul style="list-style-type: none"> • Perform the precheck • Start the synchronization task • Reselect objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Duplicate synchronization configurations
Performing Initial Synchronization	The initial synchronization is in progress.	<ul style="list-style-type: none"> • View the synchronization configurations. • Delete the synchronization task • Duplicate synchronization configurations

Task status	Description	Applicable operations
Initial Synchronization Failed	Data replication has failed during initial synchronization.	<ul style="list-style-type: none"> • View the synchronization configurations. • Reselect objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Duplicate synchronization configurations
Synchronizing	The task is replicating data.	<ul style="list-style-type: none"> • View the synchronization configurations. • Reselect objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Duplicate synchronization configurations
Synchronization Failed	An error occurred during synchronization.	<ul style="list-style-type: none"> • View the synchronization configurations. • Reselect objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Duplicate synchronization configurations
Paused	The synchronization task is paused.	<ul style="list-style-type: none"> • View the synchronization configurations. • Reselect objects to be synchronized • Modify the synchronization speed • Delete the synchronization task • Duplicate synchronization configurations

A synchronization task in the **Synchronization Failed** or **Initial Synchronization Failed** status will be automatically retried multiple times within seven days of the failure occurring. To discard the data synchronization, you must manually pause or stop the synchronization task.


4. Change tracking mode

The change tracking replication mode of Data Transmission Service (DTS) captures data updates made to your source data store and exposes them as a publisher/subscriber stream. The source data store can be either a user-created MySQL database or an ApsaraDB for MySQL database. The change tracking mode allows you to build a distributed application with decoupled data systems, so you can implement your asynchronous replications between these data systems.


 **Note** The change tracking mode supersedes the former "data subscription" mode. The change tracking mode is more cost-effective than its predecessor in that it supports consumption by multiple consumer groups in the same instance.

Features

- You can allow multiple consumers to subscribe to the same source. For more information, see [Manage consumer groups](#).
- You can enable track changes for selected data tables.
- You can track data changes from user-created MySQL databases.
- You can expose the stream to consumer applications either located in the classic network or in a VPC. The VPC access mode provides lower latency and higher security.

 **Note** You can select a network type when you create a change tracking task. For more information, see [Track data changes from ApsaraDB RDS for MySQL](#).

- You can consume data in a change tracking task by using a standard Kafka client.

 **Note** Only Kafka versions 0.10.0.x to 1.1.x are supported. For more information, see [Use a Kafka client to consume tracked data](#).


- You can reselect objects for change tracking. For more information, see [Reselect objects for change tracking](#).
- You can modify consumption checkpoints.

 **Note** You must use your change tracking client to modify consumption checkpoints.

- You can monitor the status of the change tracking task. You can set a threshold for the consumer latency, depending on how sensitive your application is to latency. You receive an alert when the threshold is reached.


Consumer groups

You can create consumer groups to allow multiple consumer applications to subscribe to the same change tracking task independently. Therefore, applications in different consumer groups can consume the same update more than once.

 **Note** Currently, each consumer group can include only one consumer.

Supported clients

You can consume data updates by using a Kafka client, which can be coded in your preferred programming language. For more information, see the Kafka clients documentation available on the Apache Kafka website. DTS provides an example of using a Kafka client coded in Java to consume data updates. For more information, see [Use a Kafka client to consume tracked data](#).

 **Note** If you use the official Kafka client to consume data, you must specify the username in the following format: <Consumer group account>-<Consumer group ID>, for example, dtstest-dtsae*****@bpv. Otherwise, the connection fails.

Supported data change types and objects

The change tracking mode supports the following types of updates:

- **Schema updates:**

Schema updates are also known as data definition language (DDL) changes. You can track operations that create, delete, or modify the schema objects in instance database. In your application code, you need to implement filtering on the data updates to identify schema updates.

- **Data updates:**

Data updates are also known as data manipulation language (DML) changes. You can track data updates of a selected object, including the INSERT, DELETE, and UPDATE operations.

You can enable change tracking for the following types of objects:

- Databases
- Tables

Change tracking tasks

A change tracking task pulls data updates from the source database in real time and keeps the updates for 24 hours. You can use a Kafka client to consume the data updates captured by the task. You can create, manage, and delete change tracking channels in the DTS console.

Change tracking task lifecycle

The following table describes the statuses of a change tracking task:

Task status	Description	Applicable operations
Prechecking	The change tracking task has been configured and a precheck is in process.	<ul style="list-style-type: none"> • Delete the change tracking instance
Not Started	The change tracking channel has passed the precheck, but is not started.	<ul style="list-style-type: none"> • Delete the change tracking instance • Reselect objects for change tracking

Task status	Description	Applicable operations
Performing Initial Change Tracking	The initial change tracking is in progress. This process takes about 1 minute.	<ul style="list-style-type: none"> • Delete the change tracking instance • Reselect objects for change tracking
Normal	Data updates are being pulled from the data source.	<ul style="list-style-type: none"> • View sample code • Configure monitoring and alerts • Delete the change tracking instance • Reset the change tracking instance • Reselect objects for change tracking
Error	An error occurs when the change tracking task is pulling data updates from the data source.	<ul style="list-style-type: none"> • View sample code • Configure monitoring and alerts • Delete the change tracking instance • Reset the change tracking instance • Reselect objects for change tracking

 **Note**

- A change tracking task in the **Error** status will automatically retry within seven days of the failure. To discard the change tracking progress and configurations, you must delete the change tracking instance.
- After a change tracking instance is released, all the instance and task configurations are permanently cleared. Use caution when you perform this operation.