ALIBABA CLOUD

Alibaba Cloud

MaxCompute

Release Note

Document Version: 20200928

Alibaba Cloud

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 `cd /d C:/window` command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	This format is used for an optional value, where only one item can be selected.	`ipconfig [-all|-t]`
{} or {a|b}	This format is used for a required value, where only one item can be selected.	`switch {active|stand}`

Table of Contents

1.Announcements

This topic describes the updates to MaxCompute in reverse chronological order.

October 13, 2020 (UTC+08:00): The SQL engine is updated for better compatibility

The following table describes the schedule for updating the SQL engine. If a change occurs, the new schedule prevails.

Sequence	Region	Date
1	India (Mumbai), Indonesia (Jakarta), and UK (London)	October 13, 2020
2	US (Virginia), UAE (Dubai), China (beijing), and China (Shanghai)	October 15, 2020
3	Japan (Tokyo), Australia (Sydney), US (Silicon Valley), and Malaysia (Kuala Lumpur)	October 20, 2020
4	Singapore (Singapore), China (Hong Kong) and Germany (Frankfurt)	October 22, 2020

The `URL_DECODE` and `CONV` functions in MaxCompute SQL are updated. This section describes the update details.

- `URL_DECODE` function
  - Before the update: If two parameters are specified for the `URL_DECODE` function, the function ignores the second parameter and decodes the value of the first parameter in UTF-8. For example, if you specify `URL_DECODE(url, "gbk")` in code, `URL_DECODE(url)` is executed.

- After the update: If two parameters are specified for the `URL_DECODE` function, the function first performs URL decoding on the value of the first parameter. Then, the function decodes the URL decoding result in the format that is specified by the second parameter. The following examples show how the URL_DECODE function works in DataWorks DataStudio before and after the function is updated:

```
SELECT URL_DECODE("%CD%F5", "gbk");
-- Before the update, the function returns garbled characters. The function ignores gbk and deco
des %CD%F5 in UTF-8.
-- After the update, the function returns the Chinese character. The function performs URL decodi
ng on %CD%F5. \xCD\xF5 is obtained, which is the GBK-encoded string of the Chinese character.

SELECT URL_DECODE("%E7%8E%8B", "gbk");
-- Before the update, the function returns the Chinese character. %E7%8E%8B is the UTF-8-encod
ed string of the Chinese character. The function ignores gbk and decodes %E7%8E%8B in UTF-8.
-- After the update, the function returns NULL. The function performs URL decoding on %E7%8E%8
B. \xE7\x8E\x8B is obtained, which is an invalid GBK-encoded string.

SELECT URL_DECODE("%E7%8E%8B");
-- The function returns the Chinese character both before and after the update. %E7%8E%8B is th
e UTF-8-encoded string of the Chinese character. In this example, no decoding format is specified
for the function, and the function decodes %E7%8E%8B in UTF-8 by default.
```

> **Note** The command-line interface (CLI) of Windows has the following issue: If you use the odpscmd client of MaxCompute to run commands in Windows, the GBK decoding result of the URL_DECODE function may be decoded in another format.

- `CONV` function
  - In a project that uses the Hive-compatible data type edition, the CONV function returns 0 both before and after the update if the input parameters are invalid.
  - In a project that uses the MaxCompute V1.0 or MaxCompute V2.0 data type edition:
    - Before the update: The CONV function returns garbled characters if the input parameters are invalid.
    - After the update: The CONV function returns NULL if the input parameters are invalid.

      For example, if you specify `CONV("00e04c9d034a", 2, 10)` in code, NULL is returned.

July 29, 2020 (UTC+08:00): The default data type edition for a new project is changed from the MaxCompute V1.0 data type edition to the MaxCompute V2.0 data type edition

When you create a MaxCompute project in the DataWorks console, the default data type edition of the project is changed from the MaxCompute V1.0 data type edition to the MaxCompute V2.0 data type edition. This feature is gradually available for new projects in all regions at the international site (alibabacloud.com) from July 29, 2020 to August 6, 2020. If existing projects are available under your Alibaba Cloud account, make sure that the data type edition you select for a new project is the same as that of the existing projects. Incompatibility issues may occur when projects of different data type editions interact with each other.

MaxCompute provides the following three data type editions: MaxCompute V1.0 data type edition, MaxCompute V2.0 data type edition, and Hive-compatible data type edition. These editions are different in definitions and usage. MaxCompute provides three attributes to configure data type editions. You can configure these attributes as required to enable an edition. For more information, see Date types.

> **Note** This feature has no impact on the data type editions of existing MaxCompute projects. You can change the data type editions of existing MaxCompute projects. For more information, see Date types.

June 29, 2020 (UTC+08:00): Data type editions can be selected for new projects

The feature of selecting data type editions for new projects is gradually available in all regions at the international site (alibabacloud.com) from June 29, 2020 to July 15, 2020. After the feature is available, you must select initial data type editions for new MaxCompute projects.

MaxCompute provides the following three data type editions: MaxCompute V1.0 data type edition, MaxCompute V2.0 data type edition, and Hive-compatible data type edition. These editions are different in definitions and usage. MaxCompute provides three attributes to configure data type editions. You can configure these attributes as required to enable an edition. For more information, see Date types.

> **Note** This feature has no impact on the data type editions of existing MaxCompute projects. You can change the data type editions of existing MaxCompute projects. For more information, see Date types.

March 15, 2020 (UTC+08:00): The storage price of MaxCompute is reduced

From March 15, 2020, the storage price of MaxCompute is reduced. For more information, see Storage pricing (pay-as-you-go). The price is reduced based on the following rules:

- The number of pricing tiers is reduced from five to three to simplify storage billing.
- The unit price in each tier is reduced to lower the storage price.

The following table lists the tiered pricing method that was used before March 15, 2020.

Volume of stored data	Tiered unit price	Fixed price
0 < Data volume ≤ 1 GB	N/A	USD 0.00 per day
1 GB < Data volume ≤ 100 GB	USD 0.0028 per GB per day	N/A

Volume of stored data	Tiered unit price	Fixed price
100 GB < Data volume ≤ 1 TB	USD 0.0014 per GB per day	N/A
1 TB < Data volume ≤ 10 TB	USD 0.0013 per GB per day	N/A
10 TB < Data volume ≤ 100 TB	USD 0.0011 per GB per day	N/A
Data volume > 100 TB	USD 0.0009 per GB per day	N/A

The following table lists the tiered pricing method that is used from March 15, 2020.

Volume of stored data	Tiered unit price	Fixed price
0 < Data volume ≤ 1 GB	N/A	USD 0.00 per day
1 GB < Data volume ≤ 10 TB	USD 0.0011 per GB per day	N/A
10 TB < Data volume ≤ 100 TB	USD 0.0009 per GB per day	N/A
Data volume > 100 TB	USD 0.0006 per GB per day	N/A

The storage billing method remains unchanged. For more information, see Storage pricing (pay-as-you-go).

- You are charged for storing data, including tables and resources, in MaxCompute by tier based on the data volume on a daily basis.
- MaxCompute records the volume of data that is stored in each project on an hourly basis and then calculates the average volume of stored data for each day. The storage fee is equal to the average volume of stored data multiplied by the unit price. MaxCompute calculates the average volume of stored data in each project in a day. Therefore, storing more data in a specific project means a lower storage fee.

Assume that the daily average data volume of a project is 1 PB. The following formula is used to calculate the daily fee based on the original tiered pricing method:

```
(100 - 1) GB × USD 0.0028 per GB per day
+ (1,024 - 100) GB × USD 0.0014 per GB per day
+ (10,240 - 1,024) GB × USD 0.0013 per GB per day
+ (102,400 - 10,240) GB × USD 0.0011 per GB per day
+ (10,240 × 10,240 - 102,400) GB × USD 0.0009 per GB per day
= USD 966.486 per day
```

The following formula is used to calculate the daily fee based on the new tiered pricing method:

```
(10,240 - 1) GB × USD 0.0011 per GB per day
+ (102,400 - 10,240) GB × USD 0.0009 per GB per day
+ (10,240 × 10,240 - 102,400) GB × USD 0.0006 per GB per day
= USD 661.9125 per day
```

February 24, 2020 (UTC+08:00): The SQL engine is updated for better compatibility

The following table describes the schedule for updating the SQL engine. If a change occurs, the new schedule prevails.

Sequence	Region	Date
1	Indonesia (Jakarta), UK (London), and India (Mumbai)	February 24, 2020
2	UAE (Dubai), US (Virginia), China North 2 Ali Gov, and China (Hong Kong)	February 26, 2020
3	Malaysia (Kuala Lumpur), Japan (Tokyo), and Germany (Frankfurt)	March 2, 2020
4	US (Silicon Valley), Singapore (Singapore), and Australia (Sydney)	March 4, 2020

- The execution rule of the GET_IDCARD_AGE function has changed.
  - Based on the original execution rule of the GET_IDCARD_AGE function, if the difference between the current year and the year of birth is greater than 100, NULL is returned. After the new rule is applied, the difference between the current year and the year of birth is returned even if the difference exceeds 100. For example, the execution result of `get_idcard_age('110101190001011009')` is NULL before the change and 120 after the change.
  - If you need to apply the original execution rule to a query statement after the change, you must find the query statement and modify it as required. For example, you can add the IF function or CASE WHEN expression for processing the return result of `GET_IDCARD_AGE` to the query statement.

Original query statement	Modified query statement
GET_IDCARD_AGE(idcardno)	if(GET_IDCARD_AGE(idcardno) > 100, NULL, GET_IDCARD_AGE(idcardno))
GET_IDCARD_AGE(idcardno)	CASE WHEN GET_IDCARD_AGE(idcardno) > 100 THEN NULL ELSE GET_IDCARD_AGE(idcardno) END

- The execution rule of the CONCAT_WS function has changed.
  - Before the change, if the CONCAT_WS function that is used in a query does not support Hive and has three or more parameters, including at least one parameter of the ARRAY type, the `array items` will not appear in the final result. For example, the expected result of the `concat_ws(',', array('a'), array('b', 'c'))` function is `"a,b,c"`, but the actual result is `",,,"`.

  - After the change, the parameters of the STRING and ARRAY types can coexist in the CONCAT_WS function without Hive support enabled. For example, the return result of the `concat_ws(',', array('a'), array('b', 'c'))` function is `"a,b,c"`.

- The execution rule of the `Like%%` function has changed in scenarios where the input value is an empty string.

  Before the change, if the input value for the character matching function LIKE is an empty string and its pattern is `%%`, the return value is False. After the change, the return value is True.

```
-- Create a table and insert an empty string into the table.
create table if not exists table_test (a string) lifecycle 3;
insert into table table_test values ('');

select a like '%%' from table_test;

-- The following result is returned before the change:
+------+
| _c0 |
+------+
| false |
+------+

-- The following result is returned after the change:
+------+
| _c0 |
+------+
| true |
+------+
```

December 25, 2019 (UTC+08:00): MaxCompute is compatible with open-source geospatial UDFs.

MaxCompute is compatible with open-source geospatial UDFs, which are implemented by ESRI for Apache Hive. You can register open-source geospatial UDFs with MaxCompute so that the functions can be called like open-source Hive UDFs. For more information, see Open source geospatial UDFs.

October 11, 2019 (UTC+08:00): New features of MaxCompute SQL are introduced

- You can force the JOIN operations or set operators including UNION, INTERSECT, and EXCEPT, inside parentheses () to run first.

```
SELECT * FROM src JOIN (src2 JOIN src3 on xxx) ON yyy;
SELECT * FROM src UNION ALL (SELECT * FROM src2 UNION ALL SELECT * FROM src3);
```

For more information, see JOIN and UNION, INTERSECT, and EXCEPT.

- MaxCompute SQL supports the hive.orderby.position.alias and hive.groupby.position.alias flags.

  If the two flags are enabled, the constants of the INTEGER type in the ORDER BY and GROUP BY clauses of the SELECT statements are processed as column numbers.

```
The columns in the src table are key and value.
SELECT * FROM src ORDER BY 1;
-- Equivalent to
SELECT * FROM src ORDER BY key;
```

  For more information, see SELECT syntax.

- MaxCompute supports the following built-in functions:
  - `STRING JSON_TUPLE(STRING json,STRING key1,STRING key2,...)`: converts a JSON string to a tuple based on a group of keys. The `JSON_TUPLE()` function supports multi-level nesting. It can parse JSON data that contains nested arrays. To parse the same JSON string multiple times, you must call the GET_JSON_OBJECT() function multiple times. However, the JSON_TUPLE() function allows you to enter multiple keys at a time and needs to parse the JSON string only once. Compared with GET_JSON_OBJECT(), JSON_TUPLE() is more efficient. For more information, see String functions.
  - `INT EXTRACT(datepart from timestamp)`: extracts a part from the date that is specified by the datepart parameter, such as YEAR, MONTH, and DAY. The value of timestamp is a date value of the TIMESTAMP type. For more information, see Date functions.

- MaxCompute allows you to specify default values for columns in a table.

  The DEFAULT VALUE clause allows you to specify a default value for a column when you create a table. If you do not specify a value for the column during an INSERT operation, this default value is inserted. The following code provides an example:

```
CREATE TABLE t (a bigint default 10, b bigint);
INSERT INTO TABLE t(b) SELECT 100;
-- Equivalent to
INSERT INTO TABLE t(a, b) SELECT 10, 100;
```

- MaxCompute SQL supports natural joins.

  A NATURAL JOIN operation is a process where two tables are joined on the basis of their common columns. Common columns are columns that have the same name in both tables. MaxCompute supports `OUTER NATURAL JOIN`. You can use the `USING` clause so that the JOIN operation returns common columns only once. The following code provides an example:

```
-- To join the src table that contains the key1, key2, a1, and a2 columns and the src2 table that cont
ains the key1, key2, b1, and b2 columns, you can execute the following statement:
SELECT * FROM src NATURAL JOIN src2;
-- Both the src and src2 tables contain the key1 and key2 columns. In this case, the preceding state
ment is equivalent to the following statement:
SELECT src.key1 as key1, src.key2 as key2, src.a1, src.a2, src2.b1, src2.b2 FROM src INNER JOIN src2 ON
src.key1 = src2.key1 AND src.key2 = src2.key2;
```

For more information, see JOIN.

- The LIMIT and OFFSET clauses are supported.

  The OFFSET clause can be used together with the ORDER BY LIMIT clause to skip a number of rows whose quantity is specified by OFFSET. For example, execute the following statement to sort the rows of the `src` table in ascending order by `key`, and return the 11th to 20th rows. OFFSET 10 indicates that the first 10 rows are skipped. LIMIT 10 indicates that a maximum of 10 rows can be returned.

  ```
  SELECT * FROM src ORDER BY key LIMIT 10 OFFSET 10;
  ```

  For more information, see SELECT syntax.

- MaxCompute supports built-in operators.
  - The IS DISTINCT FROM operator is supported. `a is distinct from b` is equivalent to `not(a <=> b)`, and `a is not distinct from b` is equivalent to `a <=> b`.
  - The || operator is supported to concatenate strings. For example, `a || b || c` is equivalent to `concat(a, b, c)`.

  For more information, see Operators.

- MaxCompute supports partition merging.

  In MaxCompute, a maximum of 60,000 partitions can be created in a table. If excessive partitions exist, you can archive data in the data warehouse and merge partitions to reduce the number of partitions. When you trigger partition merging, MaxCompute merges multiple partitions in the same table into one partition, migrates their data to the merged partition, and then drops the previous partitions. The following code shows the syntax for merging partitions. For more information, see 分区和列操作.

  ```
  ALTER TABLE <tableName> MERGE [IF EXISTS] PARTITION(<predicate>) [, PARTITION(<predicate2>) ...]
  OVERWRITE PARTITION(<fullPartitionSpec>) ;
  ```

- Add/Drop Partitions

  MaxCompute allows you to add or drop multiple partitions at a time by using the following syntax:

```
ALTER TABLE t ADD [IF NOT EXISTS] PARTITION (p = '1') PARTITION (p = '2');
ALTER TABLE t DROP [IF EXISTS]  PARTITION (p = '1'), PARTITION (p = '2');
-- Note that no commas (,) exist between partitions in the ADD clause, whereas commas (,) are used to separate partitions in the DROP clause.
```

August 29, 2019 (UTC+08:00): A custom storage handler for a foreign table is used to update the Outputer interface in some regions

On August 29, 2019, MaxCompute is upgraded. The upgrade may fail if you use a custom storage handler for a foreign table to update the Outputer interface and the column data is obtained by column name instead of array subscript.

Upgrade time: 14:00 to 23:00 on August 29, 2019 (UTC+08:00)

Regions: US (Silicon Valley) and Singapore (Singapore)

August 21, 2019 (UTC+08:00): A custom storage handler for a foreign table is used to update the Outputer interface in some regions

On August 21, 2019, MaxCompute is upgraded. The upgrade may fail if you use a custom storage handler for a foreign table to update the Outputer interface and the column data is obtained by column name instead of array subscript.

Upgrade time: 14:00 to 23:00 on August 21, 2019 (UTC+08:00)

Regions: Japan (Tokyo), Germany (Frankfurt), China (Hong Kong), and Australia (Sydney)

Impact: In `Outputer.output(Record record)`, the input record is generated by the last operator of Outputer. Column names are not fixed.

For example, the column name that is generated by the `some_function(column_a)` expression is a temporary column name. Therefore, we recommend that you use `record.get(index)` instead of `record.get(Column name)` to obtain the content of a column. To obtain column names from a table in Outputer, call `DataAttributes.getFullTableColumns()`.

If you have questions, submit a ticket.

July 24, 2019 (UTC+08:00): Spark on MaxCompute is supported

Regions: China (Hangzhou), China (Beijing), China (Shenzhen), US (Silicon Valley), China (Hong Kong), Germany (Frankfurt), Singapore (Singapore), and India (Mumbai)

March 26, 2019 (UTC+08:00): MaxCompute SQL is updated

- The GROUPING SETS clause as well as the CUBE and ROLLUP subclauses can be used to aggregate and analyze data of multiple dimensions. For more information, see Grouping Sets.
- The INTERSECT, MINUS, and EXCEPT clauses are supported. For more information, see UNION, INTERSECT, and EXCEPT.
- When MaxCompute reads files in the ORC or Parquet format in OSS by using foreign tables, it prunes the columns in the files to reduce I/O, save resources, and lower overall computing costs.
- Systems that run in the Java UDX framework are enhanced to support writable parameters. For more information, see Java UDF.

Optimized SQL performance

- DynamicDAG: a required mechanism for dynamic optimization. DynamicDAG delays optimization of resource allocation or algorithm selection and triggers it at runtime to improve optimizations and reduce the possibility of generating low-performance execution plans.
- ShuffleRemove optimization: optimization for shuffles. MaxCompute supports ShuffleRemove for right tables that have duplicate keys during the execution of the INNER JOIN clause.

March 1, 2019 (UTC+08:00): MaxCompute SQL jobs that process data in foreign tables begin to incur charges

As of March 1, 2019, all MaxCompute SQL jobs that are used to process OSS and Tablestore data begin to incur charges.

Billing standard for SQL computing that involves foreign tables:

```
Fee of an SQL job that involves foreign tables = Input data volume × Unit price for SQL computing that involves foreign tables
```

The unit price for SQL computing that involves foreign tables is USD 0.0044 per GB. All fees are charged the next day, and you will receive a bill in your account. For more information, see Billing. If you have questions, submit a ticket.

16:00 to 20:00 on January 15, 2019 (UTC+08:00): The underlying structure of MaxCompute in the China (Hong Kong) region is optimized

The underlying metadatabase of MaxCompute in the China (Hong Kong) region is optimized from 16:00 to 20:00 on January 15, 2019 to improve the performance and stability of MaxCompute. During the optimization, users in the China (Hong Kong) region may encounter submission delays or job failures for about 1 minute. In the worst cases, applications may be unavailable for up to 30 minutes. We recommend that you do not submit jobs during the optimization. Users in other regions are not affected. If you have questions, contact the MaxCompute team by using DingTalk or submit a ticket.

December 24, 2018 (UTC+08:00): MaxCompute supports time zone configuration

The default time zone of MaxCompute projects is UTC+8. The system executes time-related built-in functions and calculates fields of the DATETIME, TIMESTAMP, and DATE types based on UTC+8. As of December 24, 2018, users can configure time zones in MaxCompute by using one of the following methods:

- At the session level: Execute the `set odps.sql.timezone=<timezoneid>;` statement along with a computing statement. Example:

```
set odps.sql.timezone=Asia/Tokyo;
select getdate();
-- The following result is returned:
output:
+------------+
| _c0      |
+------------+
| 2018-10-30 23:49:50 |
+------------+
```

- At the project level: Execute the `setProject odps.sql.timezone=<timezoneid>;` statement as the project owner. After the time zone of a project is configured, it is used for all time computing, and the data of existing jobs is affected. Therefore, exercise caution when you perform this operation. We recommend that you perform this operation only on new projects.

Limits and usage notes:

- SQL built-in date functions, UDFs, UDT, user-defined joins (UDJs), and the `SELECT TRANSFORM` statement allow you to obtain the time zone attribute of a project to configure the time zone.
- A time zone must be configured in the format such as `Asia/Shanghai`, which supports daylight saving time. Do not configure it in the GMT+9 format.
- If the time zone in the SDK differs from that of the project, you must configure the GMT time zone to convert the data type from DATETIME to STRING.
- After the time zone is configured, differences exist between the real time and the output time of related SQL statements you execute in DataWorks. Between the years of 1900 and 1928, the time difference is 352 seconds. Before the year of 1900, the time difference is 9 seconds.
- MaxCompute, SDK for Java, and the related client are upgraded to ensure that DATETIME data in MaxCompute is correct across time zones. The target versions of SDK for Java and the related client have the `-oversea` suffix. The upgrade may affect the display of DATETIME data that was generated earlier than January 1, 1928 in MaxCompute.
- If the local time zone is not UTC+8 when you upgrade MaxCompute, we recommend that you also upgrade SDK for Java and the related client. This ensures that the SQL-based computing result and data that is transferred by using Tunnel commands after January 1, 1900 are accurate and consistent. For DATETIME data that was generated earlier than January 1, 1900, the SQL-based computing result and data that is transferred by using Tunnel commands might differ up to 343 seconds. For DATETIME data that was generated earlier than January 1, 1928 and was uploaded before SDK for Java and the related client are upgraded, the time in the new version is 352 seconds earlier.
- If you do not upgrade SDK for Java or the related client to versions with the `-oversea` suffix, the SQL-based computing result may differ from data that is transferred by using Tunnel commands. For data that was generated earlier than January 1, 1900, the time difference is 9 seconds. For data that was generated within the period from January 1, 1900 to January 1, 1928, the time difference is 352 seconds.

> **Note** Modifying the time zone configuration in SDK for Java or on the related client does not affect the time zone configuration in DataWorks. Therefore, the time zones are different. You must evaluate how this may affect scheduled jobs in DataWorks. The time zone of a DataWorks server in the Japan (Tokyo) region is GMT+9, and that in the Singapore (Singapore) region is GMT+8.

- If you are using a third-party client that is connected to MaxCompute by using JDBC, you must configure the time zone on the client to ensure that the time of the client and that of the server are consistent.
- MapReduce supports time zone configuration.
- Spark on MaxCompute supports time zone configuration.
  - If jobs are submitted to the MaxCompute computing cluster, the time zone of the project is automatically obtained.
  - If jobs are submitted from spark-shell, spark-sql, or pyspark in yarn-client mode, you must set parameters in the *spark-defaults.conf* file of the driver and add `spark.driver.extraJavaOptions -Duser.timezone=America/Los_Angeles`. The timezone parameter specifies the time zone you want to use.
- Machine Learning Platform for AI (PAI) supports time zone configuration.
- Graph supports time zone configuration.

2.Document updates

This topic describes the latest updates to MaxCompute V2.0 documentation. These updates allow you to understand the new features, syntax, and permissions in MaxCompute and help you improve project development efficiency.

Updates in September 2020

Date	Update item	Category	Description	Documentation
2020-09-08	Description of the LOAD command	New description	Description of the LOAD command is added.	LOAD
2020-09-03	Description of Tunnel Upload	Updated description	Description of Tunnel Upload is updated.	Tunnel commands
2020-09-01	Best practice to migrate data from BigQuery to MaxCompute	New practice	A best practice is added to describe how to migrate data from BigQuery to MaxCompute.	Migrate data from BigQuery to MaxCompute
2020-09-01	Best practice to migrate data from Amazon Redshift to MaxCompute	New practice	A best practice is added to describe how to migrate data from Amazon Redshift to MaxCompute.	Migrate data from Amazon Redshift to MaxCompute

Updates in August 2020

Date	Update item	Category	Description	Documentation
2020-08-07	Best practice to optimize costs	New practice	A best practice is added to describe how to optimize the computing, storage, data upload, and data download costs.	Cost optimization
2020-08-06	Best practice to assign the Super_Administrator role to a RAM user for a MaxCompute project	New practice	A best practice is added to describe how to assign the Super_Administrator role to a RAM user for a MaxCompute project. It also describes how to manage members and permissions by using the Super_Administrator role.	Set a RAM user as the super administrator for a MaxCompute project

Date	Update item	Category	Description	Documentation
2020-08-05	Best practice to segment Chinese text by using Jieba in a PyODPS node	New practice	The following best practices are added: 1. Segment Chinese text by using Jieba, an open source segmentation tool, and write the segmented words and phrases to a new table in a PyODPS node in DataWorks. 2. Use Jieba to segment Chinese text based on a custom dictionary referenced by a closure function.	Use a PyODPS node to segment Chinese text based on Jieba
2020-08-05	Best practice to grant access to a specific UDF only to a specified user	New practice	A best practice is added to describe how to grant access to a specific resource, such as a table or a user-defined function (UDF), only to a specified user.	Grant access to a specific UDF to a specified user
2020-08-05	Best practice to migrate data from Oracle to MaxCompute	New practice	A best practice is added to describe how to use the data synchronization feature of DataWorks to migrate data from Oracle to MaxCompute.	Migrate data from Oracle to MaxCompute
2020-08-05	Best practice to migrate data from a user-created MySQL database on an ECS instance to MaxCompute	New practice	A best practice is added to describe how to use exclusive resource groups for Data Integration to migrate data from a user-created MySQL database on an Elastic Compute Service (ECS) instance to MaxCompute.	Migrate data from a user-created MySQL database on an ECS instance to MaxCompute
2020-08-05	`odps.text.option.use.quote` property supported by SERDEPROPERTIES	New description	The description of the odps.text.option.use.quote property is added. The odps.text.option.use.quote property specifies whether to recognize a double quotation mark (`"`) as the column delimiter in a CSV file.	Access OSS data by using the built-in extractor

Updates in July 2020

Date	Update item	Category	Description	Documentation

Date	Update item	Category	Description	Documentation
2020-07-29	Best practice to migrate data from MaxCompute to Tablestore	New practice	A best practice is added to describe how to migrate data from MaxCompute to Tablestore.	Migrate data from MaxCompute to Tablestore
2020-07-29	Best practice to migrate data from MaxCompute to Object Storage Service (OSS)	New practice	A best practice is added to describe how to use the data synchronization feature of DataWorks to migrate data from MaxCompute to OSS.	Migrate data from MaxCompute to OSS
2020-07-24	Data encryption	New feature	The description of the data encryption feature is added. MaxCompute uses Key Management Service (KMS) to encrypt data for storage. This way, MaxCompute can provide static data protection to meet corporate governance and security compliance requirements.	Data encryption
2020-07-23	Aggregate functions	New description	The description of aggregate functions `APPROX_DISTINCT`, `ANY_VALUE`, `ARG_MAX`, and `ARG_MIN` is added.	Aggregate functions
2020-07-23	New data types supported by Python UDFs	New description	The description of new data types supported by Python UDFs is added.	• Python 2 UDFs • Python 3 UDFs
2020-07-23	SQL functions	New feature	The description of SQL functions that allow you to reference SQL UDFs in SQL scripts is added.	SQL functions
2020-07-23	Code-embedded UDFs	New feature	The description of code-embedded UDFs is added. Code-embedded UDFs allow you to embed Java or Python code into SQL scripts.	• Code-embedded UDFs • Usage examples
2020-07-20	Log auditing	New feature	The description of the features, scenarios, scope, and fields of log auditing is added.	Audit logs

Date	Update item	Category	Description	Documentation
2020-07-15	Best practice to use Tunnel to upload log data to MaxCompute	New practice	A best practice is added to describe how to use Tunnel to upload log data to MaxCompute.	Use Tunnel to upload log data to MaxCompute
2020-07-15	Best practice to use DataHub to migrate log data to MaxCompute	New practice	A best practice is added to describe how to use DataHub to migrate log data to MaxCompute.	Use DataHub to migrate log data to MaxCompute
2020-07-15	Best practice to use of DataWorks Data Integration to migrate log data to MaxCompute	New practice	A best practice is added to describe how to use DataWorks Data Integration to synchronize data collected by LogHub to MaxCompute.	Use DataWorks Data Integration to migrate log data to MaxCompute
2020-07-08	CLONE TABLE	New feature	The description of the CLONE TABLE statement is added. MaxCompute supports the CLONE TABLE statement that allows you to clone data from one table to another. This statement facilitates both data migration and replication.	CLONE TABLE

Updates in June 2020

Date	Update item	Category	Description	Documentation
2020-06-03	Tunnel OVERWRITE command	New description	The description of the Tunnel OVERWRITE command is added.	Tunnel commands
2020-06-01	Optimization of access to instances in a Virtual Private Cloud (VPC) from Spark on MaxCompute	New description and example	The following content is added: • Limits on the VPC whitelist and regions • Examples of merged JSON text when Spark on MaxCompute is used to access different instances	Spark访问VPC实例

Date	Update item	Category	Description	Documentation
2020-06-01	Policy-based access control and download control	New example	Examples on how to use policy-based access control and permission revocation are added.	Policy-based access control and download control

Updates in January 2020

Date	Update item	Category	Description	Documentation
2020-01-14	Improved SQL compatibility	New description	The execution rules of the `GET_IDCARD_AGE`, `CONCAT_WS`, and `LIKE` functions are modified.	• Other functions • String functions • LIKE usage
2020-01-09	Parameter description	New description	The parameters in examples are described in detail.	Project data protection

Updates in December 2019

Date	Update item	Category	Description	Documentation
2019-12-25	Open source geospatial UDFs	New feature	The description of open source geospatial UDFs is added. You can register open source geospatial UDFs with MaxCompute and use them as open source Hive UDFs.	Open source geospatial UDFs

Updates in November 2019

Date	Update item	Category	Description	Documentation

Date	Update item	Category	Description	Documentation
2019-11-06	Description of whether MaxCompute supports partition pruning	New description	The description of whether MaxCompute supports partition pruning is added.	Compare built-in functions of MaxCompute with MySQL and Oracle

Changes in October 2019

Date	Update item	Category	Description	Documentation
2019-10-09	New SQL syntax	New feature	• The syntax to merge partitions is added. • The syntax that uses a pair of parentheses to specify the priorities of operations in JOIN and SETOP statements is added. • The JSON_TUPLE built-in function is added. • The EXTRACT function is added. • Two flags are added. • The LIMIT and OFFSET clauses are supported. • Default values can be specified for columns in a table. • The NATURAL JOIN statement is supported. • New operators are supported. • The syntax to delete partitions is added.	• 分区和列操作 • JOIN • String functions • Date functions • SELECT syntax • SELECT syntax • Table operations • JOIN • Operators • 分区和列操作

Updates in July 2019

Date	Update item	Category	Description	Documentation

Date	Update item	Category	Description	Documentation
2019-07-12	odps.sql.reshuffle.dynamicpt added in SET operations	New command	The command used to configure dynamic partitions is added. This command prevents excessive small files from being generated when dynamic partitions are split.	SET operations

Updates in June 2019

Date	Update item	Category	Description	Documentation
2019-06-17	Description of the VALUES statement	New description	The description of how to create data for simple business testing is added.	VALUES

3.History

This topic provides an overview of the major milestones and achievements of Alibaba Cloud MaxCompute.

Key milestones

Time	Milestone
2009-09	The MaxCompute big data platform was officially launched.
2010-10	MaxCompute ran stably as Alibaba's first-generation cloud computing platform.
2013-08	A single cluster can manage as many as 5,000 servers.
2014-07	MaxCompute was released as a commercial service to completely replace open-source systems.
2015-2018	The MaxCompute platform grew mature with optimized performance and global presence. A single cluster can manage more than 10,000 servers.

Awards

- In November 2018, Alibaba Cloud, represented by MaxCompute, DataWorks, and AnalyticDB, was positioned as a contender in Cloud Data Warehouse (CDW) in Forrester Wave's "Cloud Data Warehouse Report Q4 2018".
- In the public cloud-based BigBench test held in September 2018, MaxCompute scored 18176.71 QPM performance in processing 100 TB data, twice the record in October 2017. In addition, MaxCompute's performance in processing ultra-small 10 TB data was three times that of other competitors.
- In April 2018, MaxCompute's multiple customer cases won the "2017 Big Data Outstanding Product and Application Solution Cases Award".
- In March 2018, MaxCompute made into the big data service list of Forrester's "Cloud Data Warehouse Report Q1 2018 ".
- In March 2018, Gartner released its "Data Management Solutions for Analytics Magic Quadrant (MQ)" report. Alibaba Cloud entered the Gartner's Magic Quadrant as a cloud service provider.
- In October 2017, world's first public cloud-based BigBench test was conducted on MaxCompute. MaxCompute processed more than 100 TB data with a performance of 7830 QPM, making it the first engine in the world to exceed 7000 QPM data processing performance.
- In June 2017, MaxCompute won the Gold Medal at the China International Software Expo.
- In the 2016 CloudSort competition, MaxCompute won the world championships of Indy (special purpose sort) and Daytona (general purpose sort) with a record of $0.82/TB, breaking the $4.51/TB record set by AWS in 2014.
- In the 2015 GraySort benchmark competition, MaxCompute completed data sorting of 100 TB in

377 seconds, breaking the 1406 second record set by Apache Spark.

Certifications

- MaxCompute was certified as the first big data platform in China that supports 10,000 servers in one cluster by the Ministry of Industry and Information Technology (MIIT).
- MaxCompute won certifications from China Academy for Information and Communications Technology (CAICT) and China Electronics Standardization Institute (CESI).

Global recognition for advancements in bigdata standards formulation

- Alibaba is the only company in China that becomes the special member of TPCx-BB.
- As the PMC member of the ORC community of world's top two computing and storage standardization open source systems, MaxCompute has contributed the most code to the community in the past two years. It has become the forerunner for storage standardization.
- MaxCompute is actively involved in Apache Calcite, the world's most famous optimizer project. It has become a special member in this project, making Alibaba one of the two leading companies in the optimizer field in China.