

Alibaba Cloud MaxCompute

Release Note

Issue: 20190920

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

Contents

Legal disclaimer.....	I
Generic conventions.....	I
1 Announcements.....	1
2 Document updates.....	5
3 History.....	19

1 Announcements

This topic lists updates to MaxCompute in descending chronological order.

August 29, 2019 (Beijing time): External table is upgraded.

An upgrade will be carried out on MaxCompute at 14:00-23:00, 29 August 2019 (Beijing time). During this upgrade, when you use external table custom storage handler to implement the outputer interface, the job may fail if column data is obtained through column names rather than digital subscriptions.

Regions: Asia Pacific SE 1, West USA 1

August 21, 2019 (Beijing time): External table is upgraded.

An upgrade will be carried out on MaxCompute at 14:00-23:00, 21 August 2019 (Beijing time). During this upgrade, when you use external table custom storage handler to implement the outputer interface, the job may fail if column data is obtained through column names rather than digital subscriptions.

Regions: Asia Pacific NE 1, Central Europe 1, China(Hong Kong), Asia Pacific SE 2

The details are as follows: The record passed in `Outputer . output (Record record)` is the record generated by the last operator of Outputer. The column name may change, and the system cannot guarantee the fixed column name. For example, the column name generated by the expression `some_function (column_a)` is a temporary column name, so the use of `record . get (column name)` to obtain column content may be affected. You may use `record . get (index)` instead. To get the column name of the table in Outputer, please call `DataAttributes . getFullTableColumns ()`.

For any questions, you can submit the ticket to contact us.

July 24, 2019(Beijing time): MaxCompute Spark is released.

Region: West USA 1, China(Hong Kong), Central Europe 1, Asia Pacific SE 1, Asia Pacific SOU 1.

March 26, 2019: MaxCompute SQL is upgraded.

- The `GROUPING SETS` clause (together with the `CUBE` and `ROLLUP` subclauses) can be used to aggregate and analyze data of multiple dimensions. For more information, see [#unique_4](#).
- The `INTERSECT`, `MINUS`, and `EXCEPT` clauses gained support. For more information, see [#unique_5](#).
- When the system reads files in ORC or Parquet format by using external tables, it can crop the columns in the files, which can reduce I/O and resource usage, thereby lowering overall computing costs.
- Systems that run in the Java UDX framework are enhanced to support writable parameters. For more information, see [#unique_6](#).

March 1, 2019: External tables of MaxCompute begin to incur charges.

Starting from March 1, 2019, SQL external tables (which are used to process OSS data and Table Store data) of MaxCompute begin to incur charges.

The charging policy is as follows:

$$\text{One SQL} - \text{time complexity} \times \text{SQL computing} \times \text{fee price} = \text{Input data volume} \times$$

The SQL price is 0.0044 USD/GB/Complexity. The complexity coefficient is 1. All the fees are charged on the next day, and you will receive an account bill.

For more information, see [#unique_7](#).

If you have any questions, open a ticket.

January 15, 2019: The underlying structure of MaxCompute in China (Hong Kong) is optimized from 16:00 to 20:00.

The underlying metadata warehouse of MaxCompute in China (Hong Kong) is optimized from 16:00 to 20:00 on January 15, 2019 to improve the performance and stability of MaxCompute. During the release window, users in the Hong Kong region may encounter submission delays or failures for tasks, which may last about one minute. In the worst cases, the application may be unavailable for 30 minutes (or half an hour). Therefore, we recommend that you do not submit any tasks during the release window. If you have any questions, contact us through DingTalk or by opening a ticket. Users in other regions are not affected.

December 24, 2018: MaxCompute supports time zone configuration.

MaxCompute project uses the UTC+8 time zone by default. The time-related built-in functions and datetime, timestamp, date fields are calculated based on UTC+8. From December 24, 2018, users can configure time zones in MaxCompute using either of the following methods:

- **Session level:** Submit the `set odps . sql . timezone =< timezoneid >;` SQL statement and a calculation statement. The following is an example:

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

- **Project level:** The project owner runs the `setProject odps . sql . timezone =< timezoneid >;` SQL statement using a CLI. After a project is configured, the corresponding time zone is used automatically, which will affect the data of existing tasks. Therefore, we recommend that you do not configure the existing projects. Instead, you can configure new projects as needed.

Note:

- The time zone configuration supports SQL built-in date functions, UDF, UDT, UDJ, and select transform.
- The time zone format such as Asia/Shanghai (daylight saving is considered) is supported. The GMT+9 format is not supported.
- If the time in the SDK time zone differs from that in the project time zone, you need to configure the GMT time zone so as to convert the date data to a string.
- After the time zone is configured, there might be difference between the real time and the output time when you run the related SQL statements through MaxCompute. Between the years of 1900 and 1928, the time difference is 352 seconds. Before the year of 1900, the time difference is 9 seconds.
- To ensure the datetime data accuracy in different regions, we will upgrade MaxCompute and the Java SDK and related console versions with the `-oversea` suffix. After the upgrade, the display of existing datetime data (before the year of 1928) stored in MaxCompute might be affected.

- When you upgrade MaxCompute, we recommend that you upgrade the Java SDK and console versions if the local time zone is not UTC+8, so as to ensure the accuracy and consistency between the SQL computing result and the Tunnel transferred data after '1900-01-01'. For the datetime data before '1900-01-01', the SQL computing output and the Tunnel transferred data might differ by 343 seconds. For the existing datetime data before '1928-01-01', the time difference is 352 seconds.
- If you continue using the SDK and console versions without the `-oversea` suffix, you might encounter time difference between the SQL output and the Tunnel transferred data. The time difference before '1900-01-01' is 9 seconds, and the time difference between '1900-01-01' and '1928-01-01' is 352 seconds.

**Note:**

When you update or configure the time zone of Java SDK and console versions, the time zone of DataWorks remains unchanged. Therefore, there might be time difference and you need to evaluate the impact of task scheduling in DataWorks. In the Japan region, the time zone of DataWorks is GMT+9. In the Singapore region, the time zone of DataWorks is GMT+8.

- If you are using a third-party client connected by JDBC, you need to set the time zone on the client to ensure the time consistency.
- MapReduce supports time zone configuration.
- Spark supports time zone configuration.
 1. For tasks that are submitted to ODPS computing clusters, the project time zone can be automatically obtained.
 2. For settings that are made through the yarn-client method, such as spark-shell, spark-sql, and pyspark, you need to configure the spark-defaults.conf parameter of the driver and add `spark.driver.extraJavaOptions -Duser.timezone = America / Los_Angeles`. The 'timezone' in the preceding statement is the time zone to be used.
- PAI supports time zone configuration.
- Graph supports time zone configuration.

2 Document updates

This topic describes the updates to the functions, syntaxes, and permissions of MaxCompute 2.0 product documentation.

Updates in 2019

Date	Feature	Category	Description	Document
2019-07-25	Spark on MaxCompute	New service	Spark on MaxCompute is introduced to Alibaba Cloud regions.	#unique_9 and #unique_7
2019-04-26	<ul style="list-style-type: none">Project lifecycle-related commandsIPv6 address whitelists	New feature	You can run commands to change the lifecycle of a project and can configure an IP address whitelist that contains IPv6 addresses.	#unique_10
2019-04-18	Script Mode SQL	New feature	The SQL engine of MaxCompute supports Script Mode SQL.	#unique_11
2019-04-16	Automatic routing for tunnel endpoints	New rule	If you do not configure a tunnel endpoint, you will be routed to the tunnel endpoint that serves the network where the service endpoint is located.	#unique_12
2019-04-15	Data outflow with data protection enabled	New method	When data protection is enabled, you can configure exception policies to allow for data outflow.	#unique_13

Date	Feature	Category	Description	Document
2019-04-11	InstanceTunnel	New feature	You can use InstanceTunnel to call an SQL instance that starts with the <code>SELECT</code> keyword and is used for querying data.	#unique_14 and #unique_15
2019-03-26	UDF writable parameters for Java UDX systems	New specification	Java UDX systems are enhanced to support UDF writable parameters.	#unique_6
2019-03-26	<code>INTERSECT</code> , <code>MINUS</code> , and <code>EXCEPT</code> statements	New specification	You can run these statements to operate data.	UNION , INTERSECT , and EXCEPT
2019-03-26	<code>GROUPING SETS</code> statement	New specification	The <code>GROUPING SETS</code> statement provides the Cube and Rollup modules. With these modules, you can aggregate and analyze data of multiple dimensions without running multiple <code>UNION ALL</code> clauses.	#unique_4
2019-01-09	Lightning	Experience optimization	The "Lightning" section has been changed to the "Overview" section. In addition, the topic that provides pricing information has been brought offline.	#unique_16

Date	Feature	Category	Description	Document
2019-01-09	Spark on MaxCompute	Experience optimization	Spark on MaxCompute is an open-source framework that functions on the service level to support data processing and analysis operations . Equipped with unified computing resources and data set permissions, Spark on MaxCompute allows you to submit and run jobs while using your preferred development methods .	#unique_9

Updates in 2018

Date	Feature	Category	Description	Document
2018-12-20	Java UDF development through IntelliJ IDEA	Experience optimization	You can use IntelliJ IDEA to develop Java UDFs.	#unique_17
2018-12-17	Java UDF development through Eclipse	Experience optimization	You can use Eclipse to develop Java UDFs.	#unique_18
2018-12-04	MaxCompute for EU West 1 (London)	New region/zone	MaxCompute is rolled out to provision services in EU West 1 (London). You can purchase resources and activate projects as needed on the Alibaba Cloud console of the Intl, China, or Japan edition.	#unique_12

Date	Feature	Category	Description	Document
2018-11-26	JSON data extraction from MongoDB to MaxCompute	Experience optimization	You can use the data synchronization feature of DataWorks to extract JSON data from MongoDB to MaxCompute.	#unique_19
2018-11-15	External tables	Experience optimization	The topic that describes how to process unstructured data is replaced with the "Overview of external tables" topic.	#unique_20
2018-11-13	Migration of JSON data from OSS to MaxCompute	Experience optimization	You can use the data synchronization feature of DataWorks to migrate JSON data from OSS to MaxCompute. In addition, you can use the <code>GET_JSON_OBJECT</code> function provided by MaxCompute to extract JSON data.	#unique_21
2018-11-08	MaxCompute limits	Experience optimization	The original "Reading guidance" topic is replaced with the "What you must know" topic. The new topic contains limits of MaxCompute SQL, data upload and download, MapReduce, security configuration, lighting, and PyODPS, which are not provided in the original topic.	#unique_22

Date	Feature	Category	Description	Document
2018-11-05	<code>cume_dist</code> window function of MaxCompute SQL	New feature	You can use the <code>cume_dist</code> window function to calculate the accumulative distribution of data in a group. That is, the proportion of rows with values that are smaller than or equal to the current value in all rows in the group.	#unique_23
2018-11-02	Dynamic partitioning based on data migration from RDS to MaxCompute	Experience optimization	You can use the data synchronization feature of DataWorks to automatically create partitions and dynamically migrate data from RDS to MaxCompute.	#unique_24
2018-09-29	User Defined Type (UDT)	New feature	You can reference third-party classes or objects in SQL scripts to obtain the data or call the methods of these classes or objects.	#unique_25

Date	Feature	Category	Description	Document
2018-09-29	User Defined Join (UDJ)	New feature	UDJ is an extension to the UDF framework of MaxCompute. With UDJ, you can define join operations and perform these join operations across tables flexibly. In addition, UDJ reduces the need for performing operations at the bottom layer of the distributed system, making the system run faster.	#unique_26
2018-07-30	Solutions to long-tail traffic	Experience optimization	The documentation is updated to describe typical long-tail traffic issues and the corresponding solutions.	#unique_27
2018-05-25	GZIP compression	New feature	MaxCompute can use its built-in extractor to read CSV or TSV data that is compressed by using GZIP from OSS. An uncompressed file differs from a compressed file in their <code>SERDEPROPERTIES</code> attributes.	#unique_28

Date	Feature	Category	Description	Document
2018-05-25	Hive	New specification	With Hive, you can use the unstructured data framework of MaxCompute to process the data that is stored on OSS in a variety of open-source formats such as ORC, PARQUET, SEQUENCEFILE, RCFILE, AVRO, and TEXTFILE.	#unique_29
2018-05-25	<code>SELECT TRANSFORM</code> statement	New feature	The <code>SELECT TRANSFORM</code> statement simplifies code reference and script compilation. It supports a wide range of programming languages such as Java, Python, Shell, and Perl, helping you to develop ad hoc functions.	#unique_30
2018-05-25	<code>SEMI JOIN</code> statement	New feature	In a <code>SEMI JOIN</code> statement, the right-side table is used only to filter the data from the left-side table and will not be presented in the result set. This statement supports two clauses: <code>LEFT SEMI JOIN</code> and <code>LEFT ANTI JOIN</code> .	#unique_31
2018-05-25	Implicit type conversion in <code>CASE WHEN</code> expressions	New specification	MaxCompute SQL allows for implicit type conversion in <code>CASE WHEN</code> expressions.	#unique_32

Date	Feature	Category	Description	Document
2018-05-16	Full table scan	New feature	You can enable and set the full table scan function.	#unique_33
2018-04-14	Renewal of expired instances through Auto-Renew	New specification	When your Subscription instance expires, you can renew it through the Auto-Renew option on the MaxCompute console.	None
2018-04-10	MaxCompute for US East 1 (Virginia)	New region/zone	MaxCompute is rolled out to provision services in US East 1 (Virginia). You can purchase resources and activate projects as needed on the Alibaba Cloud console of the Intl, China, or Japan edition.	#unique_12
2018-04-02	MaxCompute Studio 2.9.0	New feature	MaxCompute Studio 2.9.0 enables you to check and rectify code, develop and debug graphs, and fix errors. For more information, see MaxCompute Studio 2.9.0 .	#unique_34

Date	Feature	Category	Description	Document
2018-03-28	MaxCompute Manager for China East 1 (Hangzhou)	New specification	Enterprise customers who use the Subscription billing method in China East 1 (Hangzhou) can get their computing resource usage and task queuing information. In addition, MaxCompute Manager can manage computing tasks by group based on task priorities.	#unique_35
2018-03-23	Tunnel endpoint customization in MaxCompute Studio	Defect rectification	When you use Instance Tunnel to download SQL query results, you can specify a tunnel endpoint.	#unique_12
2018-03-22	MaxCompute for China East 1 (Hangzhou)	New region/zone	MaxCompute is rolled out in China East 1 (Hangzhou) to provision services. You can purchase resources and activate projects as needed on the Alibaba Cloud console of the Intl, China, or Japan edition.	#unique_12

Date	Feature	Category	Description	Document
2018-02-09	PyODPS 0.7.15	New version /New specification	<p>The relevant features in PyODPS 0.7.15 are optimized or rectified and new features are introduced to enhance development experience and efficiency:</p> <ul style="list-style-type: none">• You can use DataFrame to define functions based on Python 3.6.• The defect due to which the functions that you define through DataFrame in Python 3.5 incur errors is fixed.• You can set the parameters that determine the running of XFlow when XFlow is operating.• When you create a PyODPS DataFrame from a Pandas DataFrame, you can use the <code>as_type</code> parameter to specify the combination type, for example, LIST or DICT.	#unique_36

Date	Feature	Category	Description	Document
2018-02-09	MaxCompute for Asia Pacific SE 5 (Jakarta)	New region/zone	MaxCompute is rolled out to provision services in Asia Pacific SE 5 (Jakarta). You can purchase resources and activate projects as needed on the Alibaba Cloud console of the Intl , China, or Japan edition.	#unique_12
2018-02-03	MaxCompute Studio 2.8.2.3	Experience optimization	<p>In the left-side pane of the MaxCompute Studio Editor, you can directly run an SQL statement without clicking any button in the toolbar.</p> <ul style="list-style-type: none"> In the MaxCompute Studio Project Explorer, built-in functions are classified by type to help you easily find a specific function. 	None
2018-02-03	top instance command	New feature	The documentation is updated to describe the top instance command.	#unique_37
2018-02-03	external info query	New feature	The documentation is updated to describe the external info element, including LOCATION and StorageHandler , of external tables.	#unique_38 and #unique_28

Date	Feature	Category	Description	Document
2018-02-01	top instance command	New specification	In MaxCompute Console 0.29.0, you can run the top instance command to obtain information about the jobs that running or waiting in the queue.	#unique_37
2018-01-31	MaxCompute Manager	New feature	Enterprise customers who use the Subscription billing method can get their computing resource usage and task queuing information. In addition, MaxCompute Manager can manage computing tasks by group based on task priorities.	#unique_35
2018-01-22	DataHub for four regions	New region/zone	DataHub is introduced to four regions: China East 2 (Shanghai), China North 2 (Beijing), China South 1 (Shenzhen), and Asia Pacific SE 1 (Singapore).	None

Date	Feature	Category	Description	Document
2018-01-18	MaxCompute Studio 2.8.2.2	New version /New specification	MaxCompute Studio 2.8.2.2 provides the following functions: <ul style="list-style-type: none"> · Develop Python UDFs. · Compile unstructured unit tests. · Operate complex data types while running UDFs on your computer. · Define SQL statement format rules. 	#unique_34
2018-01-18	MaxCompute for Asia Pacific SOU 1 (Mumbai)	New region/zone	MaxCompute is rolled out in Asia Pacific SOU 1 (Mumbai) to provision services . You can purchase resources and activate projects as needed on the Alibaba Cloud console of the Intl , China, or Japan edition.	#unique_12
2018-01-10	Full table scan	New feature	The full table scan function is forbidden by MaxCompute so as to avoid resource waste and reduce costs.	#unique_10

Date	Feature	Category	Description	Document
2018-01	Project data protection	Experience optimization	The documentation is updated to describe the background information and mechanism of data protection as well as how to allow for data outflow when the data protection mechanism is enabled.	#unique_13
2018-01	Hive UDF	Experience optimization	An example is added to describe how to use Hive UDFs in MaxCompute.	#unique_6
2018-01	TableTunnel	Experience optimization	The documentation is updated to describe TableTunnel, its limits, and how it works at the bottom layer.	#unique_39
2018-01	Full table scan	New feature	The documentation is updated to describe the full table scan function.	#unique_40 , #unique_33 , and SQL Common Errors

3 History

When Alibaba Cloud was founded in September 2009, our vision was to be the first platform for data processing and sharing. In April 2010, that vision came a little closer to reality with the Open Data Processing Service (ODPS) supporting the newly launched loan business area of Ant Financial. From there, things went from strength to strength. In 2012, a Unified Data Platform was established. By the end of 2013, we had the ability to process massive volumes of data on a large scale. From 2014 to 2015, the big data platform was refined, and in 2016, MaxCompute 2.0 was born, providing massive data warehousing solutions and big data modeling capabilities.

Key milestones

Time	Development
2010.04	Named ODPS, the service is released as an operational component of Alibaba Group' s Ant Financial
2013.05	ODPS is released for beta testing
2013.07	ODPS v1.0 is released as a commercially available service. A single cluster contains 5,000 servers, with support for multi-level clusters available.
2016.09	Renamed MaxCompute, v2.0 is released as a commercially available service.