

Alibaba Cloud

大数据计算服务
新功能发布记录

文档版本：20211124

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <code>Instance_ID</code>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1.公告	05
2.文档最新动态	18
2.1. 2021年	18
2.2. 2020年	26
2.3. 2019年	32

1.公告

本文为您提供关于MaxCompute使用功能的各项更新记录。

2021年11月1日-查询加速MCQA查询缓存机制商业化发布

对于每个MCQA查询作业，MaxCompute会在内部创建临时数据集来缓存查询结果。使用相同查询的时候会优先使用缓存的结果。缓存结果按每个查询进行维护，使用缓存结果不会产生任何存储及计算费用，在降低使用费用的同时，还可大幅加快执行速度。

更多查询缓存机制信息，请参见[增强：查询结果缓存机制](#)。

2021年10月15日-MaxCompute运维管家维护通知

MaxCompute运维管家将于中国北京时间（东八区）2021年10月15日22:00~24:00进行后台维护，期间MaxCompute管家的配额管理、作业运维等功能暂不可使用，您可在维护工作完成后再继续使用。维护期间如果有紧急问题，请[提工单](#)联系MaxCompute技术支持团队进行处理，感谢您的配合！

2021年10月15日-备份与恢复功能商业化

自2021年10月15日起，MaxCompute开始对备份的数据按量计费。

计费规则如下：

- MaxCompute项目默认自动备份并保留1天内变化的数据版本，且1天内的备份存储免费。即默认情况下，MaxCompute项目免费提供数据保留周期为1天的自动备份能力。
- 当项目管理员修改备份保留周期超过1天时，MaxCompute将对超过1天的备份数据按量计费，单价为0.0006 USD/GB/天。

更多备份恢复功能信息，请参见[备份与恢复](#)。更多存储计费信息，请参见[存储费用（按量计费）](#)。

2021年9月10日-UPDATE、DELETE和MERGE INTO功能商业化

自2021年9月10日起，MaxCompute SQL的UPDATE、DELETE和MERGE INTO功能结束免费公测，开始正式计费。已经使用UPDATE、DELETE和MERGE INTO功能的按量计费作业，届时会产生计算费用。如果您之前因为担心公测稳定性问题而暂未使用UPDATE、DELETE和MERGE INTO功能，届时可以放心使用。MaxCompute为UPDATE、DELETE和MERGE INTO功能的相关可用性、稳定性等提供与MaxCompute SQL能力同等的保障。

计费规则如下：

- 如果MaxCompute项目的计费模式为包年包月，计算费用不单独收费。
- 如果MaxCompute项目的计费模式为按量计费，计算费用计算规则调整为：单条SQL计算费用=计算输入数据量×SQL复杂度×单价0.0438 USD/GB。其中：SQL复杂度通过SQL关键字个数折算。SQL关键字个数=JOIN个数+GROUP BY个数+ORDER BY个数+DISTINCT个数+窗口函数个数+MAX（INSERT个数|UPDATE个数|DELETE个数-1, 1）。

如果您担心费用超出预期，可以配置消费监控告警控制消费。更多监控告警信息，请参见[消费监控告警消费控制](#)。

您还需要注意如下内容：

- 如果MaxCompute项目的计费模式为按量计费，MaxCompute会根据UPDATE、DELETE和MERGE INTO语句扫描的输入数据量×SQL复杂度×单价计算费用。当您对Transactional表执行更新或删除操作后会生成Delta文件，Delta文件积累较多后会造成数据膨胀并影响读取性能，也会逐渐增大扫描数据量，增加费用。即使MaxCompute项目的计费模式为包年包月，也会占用更多资源，影响后续作业执行效率。您可以通过Compact方法合并Delta文件，更多语法信息请参见[合并Transactional表文件](#)。
- MaxCompute会按照批处理方式执行 delete 、 update 作业，每一条语句都会使用资源并产生费用，建

建议您使用批量方式删除或更新数据，请参见[更新或删除数据（UPDATE | DELETE）](#)。命令示例如下。

```
--推荐方案：使用关联方法，一次改写或删除所有需要操作的数据。
update table1 set col1=(select value1 from table2 where table1.id = table2.id and table1.region = table2.region);
--不推荐方案：逐行操作。
update table1 set col1=1 where id='2021063001'and region='beijing';
update table1 set col1=2 where id='2021063002'and region='beijing';
.....
```

2021年4月20日-新建MaxCompute项目支持标准的OUTER JOIN ON行为

变更前，无论是否是OUTER JOIN，ON条件中针对单个表的过滤条件，都会在JOIN操作之前对数据进行过滤。例如如下代码中Script 1、Script 2和Script 3的结果是一致的。

```
--构造数据。
CREATE TABLE t1 AS SELECT * FROM VALUES (1, 20180101),(2, 20180101),(2, 20180102) t (key, ds);
CREATE TABLE t2 AS SELECT * FROM VALUES (1, 20180101),(3, 20180101),(2, 20180102) t (key, ds);
--Script 1: 过滤条件在ON中。
SELECT t1.*,t2.*
FROM t1
LEFT OUTER JOIN t2
ON t1.key = t2.key and t1.ds='20180101';
--Script 2: 过滤条件在子查询中。
SELECT a.*,t2.*
FROM (SELECT * FROM t1 WHERE t1.ds='20180101') a
LEFT OUTER JOIN t2
ON a.key = t2.key;
--Script 3: 过滤条件在WHERE子句中。
SELECT t1.*,t2.*
FROM t1
LEFT OUTER JOIN t2
ON t1.key = t2.key
WHERE t1.ds='20180101';
```

这是一个非标准的行为。在Script 1中，如果t1.ds中有不是 20180101 的值，在标准情况下，LEFT OUTER JOIN应该保留左表全部的值，左关联t2表的数据，匹配返回对应行的数据或返回NULL。但Script 2或Script 3中，t1.ds中不是 20180101 的值会被过滤掉不再输出，与Script 1的结果不一致。

为兼容标准行为，MaxCompute将修改该默认行为。如果您需要对t1表数据过滤，则应该使用Script 2或Script 3方式。

自2021年4月20日及之后新建的MaxCompute项目中，OUTER JOIN ON中的过滤条件将不再放在JOIN操作之前进行过滤。对于2021年4月20日之前创建的存量MaxCompute项目，会保留变更前的行为，以免变更对存量作业造成影响。对于存量的MaxCompute项目，推荐您使用变更后的行为方式，设置方法如下：

- Session级别：在SQL语句前增加 `set odps.task.sql.outerjoin.ppd=true;` 命令，与SQL语句一起提交。
- Project级别：您可以[提工单](#)联系MaxCompute技术支持团队将您的项目中的 `odps.task.sql.outerjoin.ppd` 配置修改为True。

您可以通过如下示例判断当前MaxCompute项目采用了哪种行为。

```
--基于上文构造数据。
SELECT t1.*,t2.*
FROM t1 FULL OUTER JOIN t2
ON t1.key = t2.key and t1.ds='20180101' and t2.ds='20180101';
--旧的行为返回结果如下。
```

key	ds	key2	ds2
2	20180101	NULL	NULL
1	20180101	1	20180101
NULL	NULL	3	20180101

```
--新的行为返回结果如下。
```

key	ds	key2	ds2
NULL	NULL	2	20180102
2	20180101	NULL	NULL
2	20180102	NULL	NULL
1	20180101	1	20180101
NULL	NULL	3	20180101

2021年4月20日-新建项目的GET_JSON_OBJECT返回行为由JSON保留字符转义变为保留原始字符串

变更前，GET_JSON_OBJECT 函数对于字符串类型的返回结果采用了转义的方式进行输出，导致诸如换行符（\n）、引号（"）等JSON保留字符使用字符串 '\n'、'"' 显示，这是一个非标准的行为。

变更后，GET_JSON_OBJECT 函数会保留原始的字符串，返回行为更标准，处理数据更方便，性能更好。返回值仍是一个JSON字符串，可以继续当做JSON来解析，而不再需要额外使用REPLACE或REGEXP_REPLACE等函数替换反斜线。

自2021年4月20日及之后新创建的MaxCompute项目中，GET_JSON_OBJECT 函数的返回行为默认为保留原始字符串。2021年4月20日之前创建的MaxCompute项目中，GET_JSON_OBJECT 函数的返回行为默认为JSON保留字符转义，以免行为变更对存量作业造成影响。

您也可以通过以下示例判断MaxCompute项目中GET_JSON_OBJECT 函数采用了哪种行为，执行命令如下：

```
select get_json_object('{ "a": "[\\\"1\\\"]" }, '$.a');
--JSON保留字符转义的行为返回：
[\\\"1\\"]
--保留原始字符串的行为返回：
["1"]
```

推荐您使用变更后的行为方式：

- Session级别：在SQL语句前增加 `set odps.sql.udf.getjsonobj.new=true;` 命令，与SQL语句一起提交。
- Project级别：您可以[提工单](#)联系MaxCompute技术支持团队，将您的项目中的GET_JSON_OBJECT 函数返回行为切换为保留原始字符串的行为。

更多GET_JSON_OBJECT 函数信息，请参见[字符串函数](#)。

2020年12月14日-MaxCompute部分日期函数的输入参数校验行为变更

MaxCompute对部分日期函数的输入参数格式进行严格校验，避免异常输入带来的结果正确性或其它问题。该变更将于2020年12月14日到2021年1月30日陆续对国际站各个区域进行分批升级，请您提前评估项目中相关函数的已有用法，并为升级做准备。具体变更如下：

- **CAST(expr AS DATE) 函数行为变更**

- 升级前：DATE日期转换函数对输入的字符串检查不够严格，可以有非日期格式的字符作为输入参数。例如，执行如下查询语句：

```
select cast(a as date) from values ('2020-1x-19') t(a);
```

返回结果如下：

```
+-----+
| a      |
+-----+
| 2020-01-19 |
+-----+
```

- 升级后：DATE日期转换函数对输入的字符串进行严格检查，要求年、月、日3个字段都必须是数字，如果是字母或空格都会被认为是非法输入，返回NULL。例如，执行如下查询语句：

```
select cast(a as date) from values ('2020-1x-19') t(a);
```

返回结果如下：

```
+-----+
| a      |
+-----+
| NULL   |
+-----+
```

- **TO_DATE 函数行为变更**

- 升级前：TO_DATE函数对输入的字符串检查不够严格，可以有非日期格式的字符作为输入参数。例如，执行如下查询语句：

```
select to_date(a) from values ('2020-1x-19') t(a);
```

返回结果如下：

```
+-----+
| _c0    |
+-----+
| 2020-01-19 |
+-----+
```

- 升级后：TO_DATE函数对输入的字符串进行严格检查，要求年、月、日3个字段都必须是数字，如果是字母或空格都会被认为是非法输入，返回NULL。例如，执行如下查询语句：

```
select to_date(a) from values ('2020-1x-19') t(a);
```

返回结果如下：

```
+-----+
| _c0    |
+-----+
| NULL   |
+-----+
```

- **CAST(expr AS TIMESTAMP) 函数行为变更**

- 升级前：TIMESTAMP时间转换函数对输入的字符串检查不够严格，可以有非日期格式的字符作为输入参数。例如，执行如下查询语句：

```
select cast(a as timestamp) from values ('2020-11-19 16:3x:00.001') t(a);
```

返回结果如下：

```
+-----+
| a      |
+-----+
| 2020-11-19 16:03:00.1 |
+-----+
```

- 升级后：TIMESTAMP时间转换函数对输入的字符串进行严格检查，要求年、月、日3个字段都必须是数字，如果是字母或空格都会被认为是非法输入，返回NULL。例如，执行如下查询语句：

```
select cast(a as timestamp) from values ('2020-11-19 16:3x:00.001') t(a);
```

返回结果如下：

```
+-----+
| a      |
+-----+
| NULL   |
+-----+
```

2020年10月13日-SQL兼容性升级

SQL兼容性升级时间预计如下，如有变更，请以实际通知时间为准。

批次	站点	升级时间
第一批	印度（孟买）、印度尼西亚（雅加达）、英国（伦敦）	10月13日
第二批	美国（弗吉尼亚）、阿联酋（迪拜）、华北2政务云（北京）、华东2金融云（上海）	10月15日
第三批	日本（东京）、澳大利亚（悉尼）、美国（硅谷）、马来西亚（吉隆坡）	10月20日
第四批	新加坡（新加坡）、中国（香港）、德国（法兰克福）	10月22日

MaxCompute SQL中的 `URL_DECODE` 和 `CONV` 函数升级，详细变更如下：

- `URL_DECODE` 函数

- 升级前：如果 `URL_DECODE` 函数存在两个参数，第二个编码参数不生效，系统按照UTF-8编码格式解码，即 `URL_DECODE(url, "gbk")` 等于 `URL_DECODE(url)`。

- 升级后：如果 `URL_DECODE` 函数存在两个参数，系统首先按照百分号（%）解码，然后按照第二个编码参数的编码格式解码，返回结果字符串。在DataWorks数据开发环境运行命令的示例如下。

```
SELECT URL_DECODE("%CD%F5", "gbk");
-- 升级前返回乱码。填写的编码参数GBK不生效，使用UTF-8编码格式解码。
-- 升级后返回“王”。\xCD\xF5是“王”的GBK编码。
SELECT URL_DECODE("%E7%8E%8B", "gbk");
-- 升级前返回“王”。%E7%8E%8B是“王”的UTF-8编码。填写的编码参数GBK不生效，仍然使用UTF-8编码格式解码。
-- 升级后返回NULL。 \xE7\x8E\x8B不是合法的GBK编码。
SELECT URL_DECODE("%E7%8E%8B");
-- 升级前后都返回“王”。%E7%8E%8B是“王”的UTF-8编码。不填写编码参数，默认使用UTF-8编码格式解码。
```

 **说明** 在Windows环境下通过MaxCompute客户端（odpscmd）执行命令时，由于Windows环境存在 `cmd` 编码转换问题，可能会将GBK编码再次转换为其他编码。

- **CONV 函数**
 - 如果项目空间为Hive兼容模式版本，对于非法输入参数，升级前后都返回0。
 - 如果项目空间为1.0或2.0数据类型版本：
 - 升级前：如果查询的表字段为非法参数，返回结果为乱码。
 - 升级后：如果查询的表字段为非法参数，返回结果为NULL。

例如，`CONV("00e04c9d034a", 2, 10)` 返回结果为NULL。

2020年7月24日-MaxCompute新增聚合函数

MaxCompute新增聚合函数 `APPROX_DISTINCT(value)`、`ANY_VALUE(value)`、`ARG_MAX(valueToMaximize, valueToReturn)` 和 `ARG_MIN(valueToMinimize, valueToReturn)`。函数功能如下：

- `APPROX_DISTINCT(value)`：返回输入的非重复值的近似数目。
- `ANY_VALUE(value)`：返回取值范围内的任意一个值。
- `ARG_MAX(valueToMaximize, valueToReturn)`：返回valueToMaximize最大值对应行的valueToReturn。
- `ARG_MIN(valueToMinimize, valueToReturn)`：返回valueToMinimize最小值对应行的valueToReturn。

新增聚合函数详情请参见[聚合函数](#)。

2020年7月29日-新增项目的默认数据类型版本从1.0变更为2.0

MaxCompute通过DataWorks控制台新增项目时，项目的默认数据类型版本从1.0变更为2.0。该变更将于2020年7月29日到2020年8月6日陆续对国际站各个区域进行升级。如果您之前创建过存量项目，新建项目时需要选择合适的数据类型版本。不同数据类型版本的项目进行数据交互时可能会引起兼容性问题。

MaxCompute有3个数据类型版本，不同数据类型版本在定义和行为上有一定的差异。MaxCompute将数据类型相关属性组成3个组合，分别对应1.0数据类型版本、2.0数据类型版本和Hive兼容数据类型版本。详情请参见[数据类型版本说明](#)。

 **说明** 此功能对存量项目的数据类型版本无影响。如果存量项目需要更新数据类型版本，请参见[修改项目的数据类型版本](#)。

2020年6月29日-新增项目支持选择数据类型版本功能上线

MaxCompute新增项目初始化数据类型版本的功能将于2020年6月29日到2020年7月15日陆续对国际站各个区域进行升级。升级完成后，您在创建新项目时，将需要选择初始化数据类型版本。

MaxCompute有3个数据类型版本，不同数据类型版本在定义和行为上有一定的差异。MaxCompute将数据类型相关属性组成3个组合，分别对应1.0数据类型版本、2.0数据类型版本和Hive兼容数据类型版本。详情请参见[数据类型版本说明](#)。

 **说明** 此功能对存量项目的数据类型版本无影响。如果存量项目需要更新数据类型版本，请参见[修改项目的数据类型版本](#)。

2020年3月15日-MaxCompute存储降价

MaxCompute于2020年3月15日开始，对现有的[存储费用（按量计费）](#)进行降价，降价原则：

- 将原来的5个阶梯价调成3个阶梯价，减少存储计费复杂度。
- 将新阶梯单价调低，最终达到降价目的。

2020年3月15日前，原存储阶梯价如下。

存储量	阶梯单价	固定价
大于0 GB小于等于1 GB	不涉及	0.00元/天
大于1 GB小于等于100 GB	0.0028 USD/GB/天	不涉及
大于100 GB小于等于1 TB	0.0014 USD/GB/天	不涉及
大于1 TB小于等于10 TB	0.0013 USD/GB/天	不涉及
大于10 TB小于等于100 TB	0.0011 USD/GB/天	不涉及
大于100 TB	0.0009 USD/GB/天	不涉及

2020年3月15日起，新存储阶梯价如下。

存储量	阶梯单价	固定价
大于0 GB小于等于1 GB	不涉及	0.00 USD/天
大于1 GB小于等于10 TB	0.0011 USD/GB/天	不涉及
大于10TB小于等于100 TB	0.0009 USD/GB/天	不涉及
大于100 TB	0.0006 USD/GB/天	不涉及

存储计费统计方式不变，更多信息请参见[存储费用（按量计费）](#)：

- 存储到MaxCompute的数据，包括表（Table）和资源（Resource）等，会按照其数据容量的大小进行阶梯计费，计费周期为天。
- 以小时级别采集您每个项目空间下当前的存储使用情况，并以项目空间为基本单位，计算您当天的存储平均值再乘以单价。由于MaxCompute以项目空间为基本单位计算您当天的存储平均值。因此数据越集中存放在某个项目空间中，存储费用会越低。

假设您的某个项目的每天存储平均值为1 PB，原阶梯单价每天收取的费用如下。

```
(100-1)GB*0.0028 USD/GB/天
+ (1024-100) GB*0.0014 USD/GB/天
+ (10240-1024) GB*0.0013 USD/GB/天
+ (102400-10240) GB*0.0011 USD/GB/天
+ (10240*10240-102400) GB*0.0009 USD/GB/天
=966.486 USD/天
```

新阶梯单价每天收取费用如下：

```
(10240-1)GB*0.0011 USD/GB/天
+ (102400-10240) GB*0.0009 USD/GB/天
+ (10240*10240-102400) GB*0.0006 USD/GB/天
=661.9125 USD/天
```

2020年2月24日-SQL兼容性升级

SQL兼容性升级时间预计如下，如有变更，请以实际通知时间为准。

批次	站点	升级时间
第一批	印度尼西亚（雅加达）、英国（伦敦）、印度（孟买）	2月24日
第二批	中东东部1（迪拜）、美国东部1（弗吉尼亚）、华北2政务云（北京）、中国（香港）	2月26日
第三批	马来西亚（吉隆坡）、日本（东京）、德国（法兰克福）	3月2日
第四批	美国西部1（硅谷）、亚太东南1（新加坡）、亚太东南2（悉尼）	3月4日

● GET_IDCARD_AGE函数行为变更：

- Get_Idcard_Age函数规则中“如果当前年份减去出生年份差值大于100则返回NULL”，升级为“返回年份减去出生年份差值”，即不存在大于100则返回NULL的情况。例如，`get_idcard_age('110101190001011009')` 结果返回NULL，升级后结果为120。
- 如果您希望升级后，该函数返回结果保留现状，则需要找出对应的查询，评估后进行改造。可以通过 `get_idcard_age` 结果加IF函数或者CASE WHEN表达式来解决。

原始查询	修改后查询
<code>GET_IDCARD_AGE(idcardno)</code>	<code>if(GET_IDCARD_AGE(idcardno) > 100, NULL, GET_IDCARD_AGE(idcardno))</code>
<code>GET_IDCARD_AGE(idcardno)</code>	<code>CASE WHEN GET_IDCARD_AGE(idcardno) > 100 THEN NULL ELSE GET_IDCARD_AGE(idcardno) END</code>

● CONCAT_WS函数行为变更：

- 升级前，在查询运行时CONCAT_WS函数没有打开Hive兼容并且有3个及以上参数，其中有一个参数是ARRAY类型时，参数中的 `array item` 不会出现在最终的结果中。例如，`concat_ws(',', array('a'), array('`

b', 'c'))，期望结果为 "a,b,c"，但是当前结果为 ",,,"。

- 升级后，无需打开Hive兼容，CONCAT_WS的参数支持STRING类型和ARRAY类型混合。concat_ws(',', array('a'), array('b', 'c')) 的结果为 "a,b,c"。
- Like%% 查询输入为空时返回值变更。

Like字符匹配函数，当它的输入是空字符串，而Pattern是 %% 时，当前返回为False；升级后，则返回True。

```
--创建表，向表中插入空字符串。
create table if not exists table_test (a string) lifecycle 3;
insert into table table_test values ('');
select a like '%%' from table_test;
--当前返回如下。
+-----+
| _c0 |
+-----+
| false |
+-----+
--升级后返回如下。
+-----+
| _c0 |
+-----+
| true |
+-----+
```

2019年12月25日-开源地理空间UDF

MaxCompute支持兼容ESRI专门为Apache Hive实现的开源地理空间UDF，您可以将此开源地理空间UDF注册到MaxCompute中，以兼容开源Hive UDF的方式实现地理空间函数使用，详情请参见[开源地理空间UDF](#)。

2019年10月11日-MaxCompute SQL新功能

- JOIN与SETOP支持括号指定优先级

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

详细内容请参见[JOIN和交集、并集和补集](#)。

- 支持odps.sql.orderby.position.alias以及odps.sql.groupby.position.alias

当打开这两个Flag的时候，对应ORDER BY和GROUP BY中的整型常量被当做SELECT的列序号处理。

```
表src的列为key和value
SELECT * FROM src ORDER BY 1;
--等同于
SELECT * FROM src ORDER BY key;
```

详细内容请参见[SELECT语法](#)。

- 新增内置函数

- 新增内置函数 `STRING JSON_TUPLE(String ison,String key1,String key2,...)`：传入一组key和一个JSON字符串，返回一个元组。`JSON_TUPLE()` 支持包含中文的JSON数据解析，支持多层嵌套以及包含多重嵌套的数组的JSON数据解析。在需要对同一个JSON字符串多次解析的情况下，相比于多次调用`get_json_object`，`json_tuple`可以一次输入多个key，且JSON字符串只被解析一次，效率更高。详情请参见[字符串函数](#)。
- 新增内置函数 `INT EXTRACT(datepart from timestamp)`：提取日期的一部分，`datepart`支持YEAR、MONTH、DAY等时间描述，`timestamp`为Timestamp类型数据。详细请参见[日期函数](#)。
- 支持指定表的列默认值

DEFAULT VALUE允许您创建TABLE的时候指定默认值，INSERT的时候如果不指定该列，则插入默认值，举例如下。

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

- 支持自然连接

自然连接（Natural join）即参与JOIN的两张表根据字段名字自动决定连接字段。支持 `outer natural join`，支持使用 `using` 字段执行JOIN，输出字段中公共字段只出现一次，举例如下。

```
--表src的列(key1, key2, a1, a2)，表src2的列(key1, key2, b1, b2)
SELECT * FROM src NATURAL JOIN src2;
--由于 src 和 src2 有两个同名字段（key1, key2），所以上面的JOIN相当于：
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;
```

详细请参见[JOIN](#)。

- 支持LIMIT OFFSET

OFFSET语句和ORDER BY LIMIT语句配合，可以指定跳过OFFSET数目的行。如下面的语句将 `src` 按照 `key` 从小到大排序后，输出第11到第20行（OFFSET 10指定跳过前10行，LIMIT 10指定最多输出10行）。

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

详细请参见[SELECT 语法](#)。

- 其他内置语法结构

- 支持IS DISTINCT FROM语法结构：`a is distinct from b` 相当于 `not(a <=> b)`，`a is not distinct from b` 相当于 `a <=> b`。
- 支持字符串连接操作符（||）：`a || b || c` 相当于 `concat(a, b, c)`。

详细请参见[运算符](#)。

- 分区合并

MaxCompute有分区数量上限6万的限制，当分区数量过多时，可使用合并分区功能，对数仓数据进行归档，降低分区数量。合并分区功能会将同一个表下多个分区数据快速合并成一个分区，并删除之前分区，把数据移动到指定的分区下。语法格式如下，详细内容请参见[分区和列操作](#)。

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

- Add/Drop Partitions

支持一次性增加或者删除多个分区，语法格式如下。

```
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');
--注意ADD多个分区之间没有逗号，DROP的多个分区间有逗号。
```

2019年8月29日-外表自定义storagehandler实现Outputer接口升级

北京时间2019年8月29日，MaxCompute进行版本升级。期间，您在使用外表自定义storagehandler实现Outputer接口时，如果通过列名而非数字下标获取列数据，可能会引起作业失败。

升级时间：北京时间，2019年8月29日14:00~23:00

升级Region：美国西部1（硅谷）、亚太东南1（新加坡）

2019年8月21日-外表自定义storagehandler实现Outputer接口升级

北京时间2019年8月21日，MaxCompute进行版本升级。期间，您在使用外表自定义storagehandler实现Outputer接口时，如果通过列名而非数字下标获取列数据，可能会引起作业失败。

升级时间：北京时间，2019年8月21日14:00~23:00

升级Region：亚太东北1（东京）、欧洲中部1（法兰克福）、中国（香港）、亚太东南2（悉尼）

影响原因：`Outputer.output(Record record)` 中，传入的record为Outputer的上一个Operator产生的记录，列名可能发生变化，系统无法保证固定列名。

例如，表达式 `some_function(column_a)` 产生的列名是一个临时列名。因此，使用 `record.get(列名)` 方式来获取列内容的用法都有可能受到影响，建议使用 `record.get(index)`方式 获取。Outputer里如需获取表的列名，请调用 `DataAttributes.getFullTableColumns()`。

如您遇到相关问题，请提交工单咨询。

2019年7月24日-MaxCompute Spark开放

开放Region：华东1、华北2、华南1、美西1、中国（香港）、德国、新加坡、印度。

2019年3月26日-SQL语言功能升级

- 支持Grouping Set多维聚合分析（Cube，Roll up），适配需要对a列做聚合，也要对b列做聚合，或者同时要按照a、b两列做聚合的场景。具体的使用方法请参见[GROUPING SETS](#)。
- 支持INTERSECT/MINUS/EXCEPT，具体使用方法请参见[交集、并集和补集](#)。
- 通过外表读取OSS上的ORC和Parquet格式的文件时，支持对文件的列裁剪，有效减少IO量，节省计算资源和成本。
- Java UDX类型系统增强：UDF支持Writable参数，具体使用方法请参见[Java UDF](#)。

SQL性能优化

- DynamicDAG：动态优化的必要机制，它可以将一个优化（可能是资源配置或算法选择）延迟到运行时，以便提高优化的精准度，降低产生较差计划的可能性。
- ShuffleRemove Optimization：针对Shuffle的优化。从本次版本升级开始，MaxCompute将支持对inner join时右表key相同的场景进行Shuffle Remove优化。

2019年3月1日-MaxCompute 外部表开始收费

从2019年3月1日开始，MaxCompute SQL外部表（处理OSS/Table Store数据）功能开始计费。

计费标准为：

一次SQL计算费用=计算输入数据量*SQL价格

SQL价格为0.0044 USD/GB。当天的所有计量信息在第二天做一次性汇总收费，并直接体现在您的账户账单中。详情请参见[计费方式](#)。如有疑问，您可提交工单咨询。

2019年1月15日16:00~20:00-中国（香港）Region底层优化

为向您提供更好的产品性能和稳定性，MaxCompute将于北京时间2019年1月15日16:00~20:00对中国（香港）Region底层元数据仓库组件进行优化。在优化期间，中国（香港）Region用户的应用可能出现1分钟左右任务无法提交、运行中任务失败的情况。极端情况下，应用不可用时间将延长至30分钟。请您尽量避免在迁移窗口提交作业。其他Region不受影响。如果您有任何问题，可随时通过企业钉钉群或工单反馈。

2018年12月24日-MaxCompute支持时区配置

MaxCompute Project时区默认是中国的东八区，在用到DATETIME、TIMESTAMP或DATE类型相关的字段以及相关时间内置函数都是按东八区进行计算，2018年12月24日MaxCompute开始支持时区配置，有以下两种方式：

- Session级别：`set odps.sql.timezone=<timezoneid>;`，该语句需要与计算语句一起提交。

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

- Project级别：`setProject odps.sql.timezone=<timezoneid>;`，该语句需要Project Owner执行。执行此命令后该Project下相关的时间计算就会取设置后的时区，因此对于原有的任务数据会有影响，所以该操作需谨慎。建议仅对新增Project进行设置，对已有数据的Project不设置。

使用限制及注意事项：

- 时区配置支持范围为：SQL内置日期函数、UDF、UDT、UDJ、`select transform` 都支持获取Project Timezone属性来支持时区。
- 时区支持的格式为 `Asia/Shanghai`（存在夏令时跳变），不支持GMT+9格式。
- 当SDK时区与Project时区不同时，DATETIME类型转化为STRING类型操作需设置GMT时区。
- 时区配置后，通过DataWorks执行相关SQL时，日期显示某些时间段会有时间差异。例如，1900~1928年的日期时间差异5分52秒，1900年之前的日期时间差异9秒。
- 为了保证MaxCompute在多个时区DATETIME类型数据的正确性，MaxCompute服务、Java SDK以及配套的客户端将在近期进行版本更新（-oversea 后缀的Java SDK、客户端版本），更新后可能影响MaxCompute中已经存储的早于1928年的DATETIME类型数据的显示。
- 建议（非中国东八区的Region）同步更新Java SDK、客户端版本，以便保证SQL计算结果及Tunnel传输数据在1900-01-01之后范围的准确性和一致性。而对于早于1900-01-01的DATETIME数据，SQL的计算显示结果和Tunnel传输数据仍旧可能存在343秒的差异。对于新版本SDK、客户端之前已经上传的早于1928-01-01的DateTime数据，在新版本中会减少352秒。
- 如果继续沿用不带有 -oversea 后缀的SDK、客户端，将存在SQL计算结果和Tunnel传输数据存在差异的风险。例如，早于1900-01-01的数据差异为9秒，1900-01-01~1928-01-01的数据差异为352秒。

② 说明 Java SDK、客户端版本更新配置时区不影响DataWorks的时区配置，因此时区会存在差异，需要您对DataWorks中定时任务调度的影响进行计算评估。在日本Region，DataWorks服务器是GMT+9时区，在新加坡Region，DataWorks服务器是GMT+8时区。

- 通过JDBC连接的第三方客户端需要在客户端设置时区，保证与服务端时区设置一致性。
- MapReduce支持时区功能。
- Spark支持时区功能。
 - 对于提交到MaxCompute计算集群的任务形式，可以自动获取Project时区。
 - 通过yarn-client模式启动（例如spark-shell、spark-sql、pyspark等）的设置，需要您手动配置Driver的启动参数（*spark-defaults.conf*），增加 `spark.driver.extraJavaOptions -Duser.timezone=America/Los_Angeles`，Timezone是要使用的时区。
- PAI支持时区功能。
- Graph支持时区功能。

2.文档最新动态

2.1. 2021年

本文为您介绍MaxCompute 2021年内容更新的最新动态，基于此您可以了解MaxCompute中增加了哪些新功能、语法新特性和权限变更，提升项目开发效率。

2021年10月更新记录

时间	特性	类别	描述	产品文档
2021-10-26	支持清空分区表的分区数据。	新说明	MaxCompute支持通过条件筛选方式清空分区数据。如果您希望一次性删除符合某个规则条件的一个或多个分区，可以使用表达式指定筛选条件，通过筛选条件匹配分区并批量清空分区数据。	分区和列操作
2021-10-19	新增日期数据格式转换最佳实践。	新实践	新增STRIMG、TIMESTAMP、DATETIME互相转换实践案例。	日期数据格式转换：STRIMG、TIMESTAMP、DATETIME互相转换

2021年9月更新记录

时间	特性	类别	描述	产品文档
2021-09-30	MaxCompute UDF内容整改。	更新说明	重构MaxCompute UDF目录结构及内容结构，按照UDF、UDTF、UDAF进行整改并补充示例及常见问题。	MaxCompute UDF概述
2021-09-26	补充物化视图使用限制。	更新说明	补充物化视图使用限制。	物化视图操作
2021-09-22	网络开通流程处理方式变更，从提工单变更为填写表单。	更新说明	网络开通流程处理方式变更，从提工单变更为填写表单。	网络开通流程

时间	特性	类别	描述	产品文档
2021-09-16	新增商业智能BI工具连接MaxCompute操作指导。	新说明	MaxCompute支持连接商业智能BI工具以图表形式展现数据分析结果。	- Tableau连接MaxCompute - FineBI连接MaxCompute - FineReport连接MaxCompute - Davinci连接MaxCompute - Superset连接MaxCompute - Yonghong BI连接MaxCompute - Quick BI连接MaxCompute
2021-09-16	新增数据库管理工具连接MaxCompute操作指导。	新说明	MaxCompute支持连接数据库管理工具管理MaxCompute项目。	- DBeaver连接MaxCompute - DataGrip连接MaxCompute - SQL Workbench/J连接MaxCompute
2021-09-15	新增日期函数DATE_SUB。	新说明	按照指定的幅度增减天数，与DATE_ADD逻辑相反。	DATE_SUB
2021-09-15	补充Spark Java示例。	更新说明	补充Spark Java示例。	- Spark-1.x示例 - Spark-2.x示例
2021-09-08	补充核心概念的层次结构。	新说明	补充核心概念的层次结构。	核心概念的层次结构

2021年8月更新记录

时间	特性	类别	描述	产品文档
2021-08-27	支持基于Delta Lake或Hudi存储机制实现湖仓一体能力。	新功能	Delta Lake和Hudi是数据湖方案中常用的存储机制，为数据湖提供流处理、批处理能力。MaxCompute基于阿里云DLF、RDS、OSS产品提供了支持Delta或Hudi存储机制的湖仓一体架构。	基于Delta Lake或Hudi存储机制实现湖仓一体
2021-08-26	数学函数新增CORR函数。	新说明	支持计算两列数据的皮尔逊系数。	数学函数

时间	特性	类别	描述	产品文档
2021-08-25	新增资源规划及规格选型指导。	新说明	帮助您预估购买的资源数量并选择合适的规格类型。	资源规划及规格选型
2021-08-20	聚合函数语法变更。	更新说明	增加聚合函数语法模块，提供语法详细说明。	聚合函数
2021-08-19	其他函数下新增HASH、COMPRESS、DECOMPRESS、NULLIF、FORMAT_NUMBER、SHA、SHA1、SHA2、CRC32函数。	新说明	• HASH: 对输入的参数进行散列运算，返回一个HASH值。 • COMPRESS: 对输入的字符串，按照GZIP算法进行压缩。 • DECOMPRESS: 对于BINARY类型的输入，使用GZIP算法进行解压缩。 • NULLIF: 如果expr1和expr2的值相等，返回NULL；否则返回expr1。 • FORMAT_NUMBER: 将数字转化为指定格式的字符串。 • SHA: 计算字符串或者二进制的SHA-1哈希值，并将值以十六进制字符串返回。 • SHA1: 计算字符串或者二进制的SHA-1哈希值，并将值以十六进制字符串返回。 • SHA2: 计算字符串或者二进制的SHA-2系列哈希值（支持SHA-224, SHA-256, SHA-384和SHA-512）。 • CRC32: 计算字符串或者二进制的循环冗余校验值。	其他函数
2021-08-19	日期函数新增TO_MILLIS函数。	新说明	将指定日期转换为以毫秒为单位的UNIX时间戳。	日期函数
2021-08-18	起售CU数量变更。	更新说明	MaxCompute的CU起售数量从10 CU变更为50 CU。	计费方式
2021-08-11	MaxCompute新增物化视图功能。	新功能	物化视图（Materialized View）本质是一种预计算，即把某些耗时的操作（例如JOIN、AGGREGATE）的结果保存下来，以便在查询时直接复用，从而避免这些耗时的操作，最终达到加速查询的目的。	物化视图操作

时间	特性	类别	描述	产品文档
2021-08-11	UPDATE、DELETE和MERGE INTO功能商业化。	更新说明	自2021年8月10日起，MaxCompute SQL的UPDATE、DELETE和MERGE INTO功能结束免费公测，开始正式计费。已经使用UPDATE、DELETE和MERGE INTO功能的按量计费作业，届时会产生计算费用。如果您之前因为担心公测稳定性问题而暂未使用UPDATE、DELETE和MERGE INTO功能，届时可以放心使用。MaxCompute为UPDATE、DELETE和MERGE INTO功能的相关可用性、稳定性等提供与MaxCompute SQL能力同等的保障。	- 更新或删除数据（UPDATE DELETE） - MERGE INTO
2021-08-10	新增复杂类型函数。	新功能	MaxCompute支持处理MAP、ARRAY、STRUCT等复杂类型数据，提供多种函数支持多样化数据处理。	复杂类型函数
2021-08-04	新增ETL工具连接MaxCompute操作指导。	新说明	MaxCompute支持连接ETL工具管理MaxCompute项目。	- 使用Kettle调度MaxCompute - 使用Apache Airflow调度MaxCompute - 使用Azkaban调度MaxCompute
2021-08-02	in/not in subquery 及 scalar subquery 支持多列。	更新说明	兼容PostgreSQL支持多列的需求。	子查询（SUBQUERY）

2021年7月更新记录

时间	特性	类别	描述	产品文档
2021-07-16	MaxCompute支持定期检查并列出可结合动态过滤器治理的TOP级SQL周期任务。	新功能	如果有该类型任务，列表中会显示最近一次执行的Instance ID，建议您检查作业并参照 动态过滤器（Dynamic Filter） 进行优化，避免产生不必要的资源消耗，并提升作业执行性能。	MaxCompute管家

时间	特性	类别	描述	产品文档
2021-07-15	补充Package的使用限制。	更新说明	补充Package的使用限制。	Package的使用方法

2021年6月更新记录

时间	特性	类别	描述	产品文档
2021-06-07	数据集成导入导出数据内容整改。	更新说明	重构数据集成导入导出数据内容。	• 数据集成导入数据 • 数据集成导出数据

2021年5月更新记录

时间	特性	类别	描述	产品文档
2021-05-31	准备工作内容整改。	更新说明	重构准备工作流程及内容。	准备工作
2021-05-31	快速入门内容整改。	更新说明	按照工具展示快速入门操作。	快速入门

2021年3月更新记录

时间	特性	类别	描述	产品文档
----	----	----	----	------

时间	特性	类别	描述	产品文档
2021-03-25	补全MaxCompute错误码列表信息。	更新说明	完善MaxCompute错误码列表。	● 错误码概述 ● Common错误码 (ODPS-00CCCCX) ● SQL错误码 (ODPS-01CCCCX) ● PL错误码 (ODPS-02CCCCX) ● API错误码 (ODPS-04CCCCX) ● Xlib错误码 (ODPS-05CCCCX) ● MapReduce错误码 (ODPS-07CCCCX) ● ODPS SDK/CLT错误码 (ODPS-08CCCCX) ● Algo错误码 (ODPS-12CCCCX)
2021-03-16	建表语法支持指定列的NOT NULL属性。	更新说明	禁止该列的值为NULL。	表操作

时间	特性	类别	描述	产品文档
2021-03-16	新增FROM_JSON和TO_JSON函数。	新说明	FROM_JSON函数可以将JSON转换成MaxCompute的数据类型，或提取JSON中的信息，根据JSON字符串jsonStr和schema信息，返回ARRAY、MAP或STRUCT类型。TO_JSON函数可以将给定的复杂类型ARRAY、MAP或STRUCT，以JSON字符串格式输出。	字符串函数
2021-03-16	支持Transactional表的ACID属性。	新功能	Transactional表的ACID特性在MaxCompute并发写操作的ACID特性基础上新增了新特性。	ACID语义
2021-03-12	MaxCompute管家支持SQL持续空输出和全表扫描推荐。	新功能	MaxCompute支持定期检查并列出持续执行空输出或全表扫描的TOP级SQL周期任务。如果有该类型任务，列表中会显示最近一次执行的Instance ID，建议您检查作业并进行优化，避免产生不必要的资源消耗。	MaxCompute管家
2021-03-12	支持列级别权限控制。	更新说明	对表列级别进行权限控制。	授权
2021-03-12	支持通过属性打开或关闭Download权限。	更新说明	开启或关闭该功能，需要Project Owner或具备Super_Administrator角色的用户在Project级别配置属性。	Policy和Download权限控制
2021-03-01	新增MaxCompute湖仓一体功能。	新功能	MaxCompute提供湖仓一体方案，该方案可以打破数据湖与数据仓库割裂的体系，并将数据湖的灵活性、生态丰富能力与数据仓库的企业级部署能力进行融合，助力构建数据湖和数据仓库相融合的数据管理平台。	MaxCompute湖仓一体

2021年2月更新记录

时间	特性	类别	描述	产品文档
2021-02-23	新增Python 2 UDTF使用示例。	新示例	新增Python 2 UDTF读取MaxCompute资源示例。	Python 2 UDF
2021-02-23	新增Python 3 UDTF使用示例。	新示例	新增Python 3 UDTF读取MaxCompute资源示例。	Python 3 UDF
2021-02-23	新增Spark作业引用外部文件描述。	新说明	新增Spark作业引用外部文件描述。	搭建开发环境
2021-02-20	新增UDTF与Lateral View结合的使用示例。	新示例	新增UDTF与Lateral View结合的使用示例。	UDTF使用说明
2021-02-09	补充清除用户权限描述。	新说明	补充清除用户权限描述。	用户管理

时间	特性	类别	描述	产品文档
2021-02-07	删除VALUES中只能使用常量的限制。	更新说明	删除VALUES中只能使用常量的限制。	VALUES
2021-02-07	OSS外部表支持分片上传功能。	新说明	支持使用OSS的分片上传（Multipart Upload）功能通过INSERT操作向OSS外部表写入数据。	将数据输出到OSS
2021-02-07	新增 <code>array_intersect</code> 函数，修改 <code>sort_array</code> 函数。	新说明	新增 <code>array_intersect</code> ，用于计算两个 <code>array</code> 的交集。 <code>sort_array</code> 增加 <code>isAsc</code> 参数，对给定数组排序时让用户选择是升序还是降序，默认是升序。	其他函数
2021-02-07	新增过滤条件表达式和 <code>count_if</code> 函数。	新说明	增加过滤条件表达式支持，可以对 <code>select</code> 子句中某个聚合函数指定过滤条件，便于您在同一个聚合语句中单独控制某个聚合函数的数据范围。新增 <code>count_if</code> 函数对满足 <code>if</code> 条件的记录进行计数。	聚合函数
2021-02-07	增强日期函数 <code>date</code> 、 <code>datediff</code> 、 <code>datepart</code> 、 <code>datetrunc</code> 对DATE、TIMESTAMP日期数据类型的支持。	新说明	增强日期函数 <code>date</code> 、 <code>datediff</code> 、 <code>datepart</code> 、 <code>datetrunc</code> 对DATE、TIMESTAMP日期数据类型的支持。	日期函数
2021-02-07	新增 <code>base64</code> 、 <code>unbase64</code> 和 <code>parse_url</code> 函数。	新说明	新增字符串函数 <code>parse_url</code> 对URL进行解析；新增 <code>base64</code> 和 <code>unbase64</code> 函数实现二进制和BASE64格式字符串互转。	字符串函数
2021-02-07	MaxCompute命令行支持关闭表头显示，便于Shell调用。	新说明	在Shell（或Windows命令行）执行窗口，用户会可能需要使用 <code>odpscmd -e SQL</code> 语句执行得到的动态返回值，Shell的变量会获取这个动态返回值，然后在Shell中执行后续作业。此场景需要返回值不包含运行信息、表头等额外信息。您可以通过 <code>set odps.sql.select.output.format={needHeader:false,fieldDelim:''}</code> ；关掉表头显示，将计算结果stdout部分输出到目标句柄。	MaxCompute客户端（odpscmd）
2021-02-07	MaxCompute支持通过UNLOAD命令把数据导出到OSS外部存储。	新功能	MaxCompute支持通过UNLOAD命令导出数据到OSS，方便用户使用OSS存储结构化数据，并方便用户使用OSS上其他计算引擎处理和分析MaxCompute导出的数据。	UNLOAD

时间	特性	类别	描述	产品文档
2021-02-07	支持SkewJoin Hint。	新功能	当两张表Join存在热点，导致出现长尾问题时，您可以通过取出热点key，将数据分为热点数据和非热点数据两部分处理，最后合并的方式，提高Join效率。SkewJoin Hint可以通过自动或手动方式获取两张表的热点key，分别计算热点数据和非热点数据的Join结果并合并，加快Join的执行速度。	SKEWJOIN HINT
2021-02-07	MaxCompute外部表支持通过JDBC方式读写MC-Hologres。	新功能	MaxCompute外部表支持使用数据库JDBC驱动机制访问MC-Hologres数据源数据。您可以使用在MaxCompute上创建MC-Hologres外部表的方式，基于PostgreSQL JDBC驱动查询MC-Hologres数据源的数据。该方式无冗余存储，无需导入导出数据，可实现快速获取查询结果。当前读写MC-Hologres外部表处于公测阶段，不收取计算费用，也暂时不保证SLA。	Hologres外部表
2021-02-01	新增Mars v0.6.3版本变更记录。	新说明	新增Mars v0.6.3版本变更记录。	版本更新记录

2021年1月更新记录

时间	特性	类别	描述	产品文档
2021-01-21	<code>get_json_object</code> 函数返回行为变更。	更新说明	新建MaxCompute项目的 <code>get_json_object</code> 函数的默认返回行为由JSON保留字符转义的方式变更为保留原始字符串的方式	字符串函数
2021-01-19	SELECT 目录整改并优化内容。	更新说明	调整SELECT 模块内容并完善命令示例。	SELECT 语法
2021-01-05	新增数据迁移目录。	新说明	数据上传下载及迁移工具内容调整到新增目录。	数据上传下载

2.2. 2020年

本文为您介绍MaxCompute 2020年内容更新的最新动态，基于此您可以了解MaxCompute中增加了哪些新功能、语法新特性和权限变更，提升项目开发效率。

2020年12月更新记录

时间	特性	类别	描述	产品文档
2020-12-21	新增STACK和GET_USER_ID函数。	新说明	新增STACK函数用于将指定的数据分隔为指定列数。新增GET_USER_ID函数用于获取当前账号的账号ID。	其他函数

时间	特性	类别	描述	产品文档
2020-12-21	新增通过DataWorks管理项目角色权限描述。	新说明	支持通过DataWorks管理项目角色权限。	通过DataWorks管理项目角色权限
2020-12-21	新建MaxCompute项目默认提供Information Schema服务。	更新说明	新建MaxCompute项目默认提供Information Schema服务，无需手动安装Information Schema权限包。	Information Schema概述
2020-12-15	新增Mars版本变更说明。	新说明	提供Mars各版本的变更记录。	Mars版本更新记录
2020-12-15	新增Java SDK版本变更说明。	新说明	提供Java各版本的变更记录。	Java SDK版本更新记录
2020-12-15	新增JDBC版本变更说明。	新说明	提供JDBC各版本的变更记录。	JDBC版本更新记录
2020-12-11	元数据视图列表新增TASK视图。	新说明	TASK视图用于保存运行中作业的实时快照信息。	元数据视图列表
2020-12-10	新增添加项目空间成员并设置角色操作指导。	新说明	如果您使用阿里云账号创建了项目空间，需要RAM用户协助开发时，可以将RAM用户添加为项目空间成员并设置角色，RAM用户才可以协同开发。	添加工作空间成员并设置角色
2020-12-09	新增AdminEvent事件。	新说明	新增CreateProject、UpdateProject、DeleteProject事件，记录项目相关日志。	审计日志
2020-12-08	常用命令整改并优化内容。	更新说明	调整常用命令内容并完善命令示例。	常用命令列表
2020-12-08	Insert语句整改并优化内容。	更新说明	调整Insert语句内容并完善命令示例。	• 插入或覆写数据 (INSERT INTO INSERT OVERWRITE) • 插入或覆写动态分区数据 (DYNAMIC PARTITION) • 多路输出 (MULTI INSERT) • VALUES

2020年11月更新记录

时间	特性	类别	描述	产品文档
2020-11-30	补充国际站的公开数据集信息。	更新说明	补充国际站的公开数据集信息，供测试、试用产品。	公开数据集概述
2020-11-30	新增查询编辑器。	新功能	MaxCompute控制台提供查询编辑器，方便您快速执行SQL语句并进行数据分析操作。	MaxCompute控制台-查询编辑器
2020-11-27	更新升级和降配操作流程。	更新说明	通过新版MaxCompute控制台执行升级和降配操作。	升级和降配
2020-11-26	刷新审计日志字段内容。	更新说明	刷新审计日志字段内容。	审计日志
2020-11-20	支持解除 <code>order by</code> 必须带 <code>limit</code> 的限制。	新说明	支持解除 <code>order by</code> 必须带 <code>limit</code> 的限制。	SELECT 语法
2020-11-17	新增Python UDF使用第三方包操作指导。	新说明	MaxCompute支持您在Python UDF中引用第三方包，例如Numpy包、需要编译的第三方包或依赖动态链接库的第三方包。	Python UDF 使用第三方包
2020-11-04	补充配置JDBC启用MCQA功能下的可选配置参数。	更新说明	补充配置JDBC启用MCQA功能下的可选配置参数。	使用说明

2020年10月更新记录

时间	特性	类别	描述	产品文档
2020-10-30	SEMI JOIN支持MAPJOIN Hint。	更新说明	SEMI JOIN支持MAPJOIN Hint，提高LEFT SEMI或ANTI JOIN的性能，为数据倾斜问题提供解决方案。	SEMI JOIN（半连接）
2020-10-30	MaxCompute支持SORT BY不带DISTRIBUTE BY。	更新说明	MaxCompute支持不带DISTRIBUTE BY的SORT BY，提供数据重排的解决方案，提高SQL执行的过滤性能。	SELECT 语法
2020-10-30	INSERT 命令支持ZORDER BY子句，提升查询时的过滤性能。	新功能	支持把数据相近的行排列在一起，提升查询时的过滤性能，同时降低存储成本。	插入或覆写数据（INSERT INTO INSERT OVERWRITE）
2020-10-30	支持通过条件筛选方式批量删除分区。	新功能	如果用户希望一次性删除符合某个规则条件的一个或多个分区，可以使用条件表达式匹配分区进行批量删除。	分区和列操作

时间	特性	类别	描述	产品文档
2020-10-30	OSS外部表CSV或TSV文件支持GBK编码格式。	更新说明	<code>odps.text.option.encoding</code> 属性支持GBK编码格式。	内置Extractor访问OSS
2020-10-30	时间函数YEAR/QUARTER/MONTH/DAY/HOUR/MINUTE/SECOND增加对DATETIME输入数据类型的支持。	更新说明	时间函数YEAR/QUARTER/MONTH/DAY/HOUR/MINUTE/SECOND增加对DATETIME输入数据类型的支持。	日期函数
2020-10-30	新增WIDTH_BUCKET函数。	新功能	支持获取某个字段值落入的分组编号。	数学函数
2020-10-12	查询加速（MCQA）功能商业化。	更新说明	查询加速（MCQA）功能商业化，正式开始收费。	MCQA概述
2020-10-10	支持修改表的聚簇属性。	新说明	新增修改表的聚簇属性说明。	表操作

2020年9月更新记录

时间	特性	类别	描述	产品文档
2020-09-24	新增PyODPS模块。	新说明	新增PyODPS类型作业说明。	PyODPS
2020-09-17	新增创建RAM用户描述。	新说明	新增创建RAM用户描述。	准备RAM用户
2020-09-08	新增LOAD命令说明。	新说明	新增LOAD命令说明。	LOAD
2020-09-03	更新Tunnel Upload说明。	更新说明	更新Tunnel Upload命令使用说明。	Tunnel命令
2020-09-01	新增BigQuery数据迁移至MaxCompute最佳实践。	新实践	为您介绍如何将BigQuery数据迁移至MaxCompute。	BigQuery数据迁移至MaxCompute
2020-09-01	新增Amazon Redshift数据迁移至MaxCompute最佳实践。	新实践	为您介绍如何将Amazon Redshift数据迁移至MaxCompute。	Amazon Redshift数据迁移至MaxCompute

2020年8月更新记录

时间	特性	类别	描述	产品文档
2020-08-07	新增成本优化最佳实践。	新实践	为您介绍如何优化计算、存储和数据上传下载成本。	成本优化

时间	特性	类别	描述	产品文档
2020-08-06	新增MaxCompute项目设置RAM子账号为超级管理员最佳实践。	新实践	为您介绍如何在MaxCompute项目空间中RAM用户设置为超级管理员，并提供了超级管理员在成员管理和权限管理等方面的使用建议。	MaxCompute项目设置RAM子账号为超级管理员
2020-08-05	新增PyODPS节点实现结巴中文分词最佳实践。	新实践	为您介绍如何使用DataWorks的PyODPS类型节点，借助开源结巴中文分词包实现对中文字段的分词并写入新的表，以及如何利用闭包函数使用自定义词典进行分词。	PyODPS节点实现结巴中文分词（使用第三方包）
2020-08-05	新增实现指定用户访问特定UDF最佳实践。	新实践	为您介绍如何实现将资源（表或UDF）设置为仅能被指定的用户访问。	实现指定用户访问特定UDF
2020-08-05	新增迁移Oracle数据至MaxCompute最佳实践。	新实践	为您介绍如何通过DataWorks数据集成，迁移Oracle上的数据至MaxCompute。	迁移Oracle数据至MaxCompute
2020-08-05	新增迁移ECS自建MySQL数据库至MaxCompute最佳实践。	新实践	为您介绍如何使用独享数据集成资源，将您在ECS上自建的MySQL数据库中的数据，迁移至MaxCompute。	迁移ECS自建MySQL数据库至MaxCompute
2020-08-05	新增SERDEPROPERTIES支持的属性项 <code>odps.text.option.use.quote</code> 。	新说明	指定是否识别CSV的列分隔符 <code>"</code> 。	内置Extractor访问OSS

2020年7月更新记录

时间	特性	类别	描述	产品文档
2020-07-29	新增迁移MaxCompute数据至OTS最佳实践。	新实践	为您介绍如何将MaxCompute数据迁移至表格存储OTS。	MaxCompute数据迁移至OTS
2020-07-29	新增迁移MaxCompute数据至OSS最佳实践。	新实践	为您介绍如何使用DataWorks的数据同步功能将MaxCompute数据迁移至对象存储OSS。	MaxCompute数据迁移至OSS
2020-07-24	新增数据加密功能。	新功能	MaxCompute支持通过密钥管理服务KMS（Key Management Service）对数据进行加密存储，提供数据静态保护能力，满足企业监管和安全合规需求。	数据加密

时间	特性	类别	描述	产品文档
2020-07-23	新增聚合函数。	新说明	新增 APPROX DISTINCT 、 ANY VALUE 、 ARG_MAX 和 ARG_MIN 函数。	聚合函数
2020-07-23	Python UDF增加支持的数据类型。	新说明	Python UDF增加支持的数据类型。	- Python 2 UDF - Python 3 UDF
2020-07-23	新增SQL语言定义函数。	新功能	MaxCompute支持通过SQL语言定义函数（SQL Function）在SQL脚本中使用SQL定义的UDF。	SQL语言定义函数
2020-07-23	新增代码嵌入式UDF。	新功能	MaxCompute支持通过代码嵌入式UDF（Embedded UDF）将Java或Python代码嵌入SQL脚本。	- 代码嵌入式UDF - UDT 示例
2020-07-20	新增审计日志功能。	新功能	为您介绍审计日志的功能、使用场景、范围以及字段定义。	审计日志
2020-07-15	新增通过Tunnel迁移日志数据至MaxCompute。	新实践	为您介绍如何通过Tunnel上传日志数据至MaxCompute。	通过Tunnel迁移日志数据至MaxCompute
2020-07-15	新增通过DataHub迁移日志数据至MaxCompute最佳实践。	新实践	为您介绍如何通过DataHub迁移日志数据至MaxCompute。	通过DataHub迁移日志数据至MaxCompute
2020-07-15	新增通过DataWorks数据集成迁移日志数据至MaxCompute最佳实践。	新实践	为您介绍如何通过数据集成功能同步LogHub数据至MaxCompute。	通过DataWorks数据集成迁移日志数据至MaxCompute
2020-07-08	新增CLONE TABLE功能。	新功能	MaxCompute提供了CLONE TABLE功能，支持将表数据复制到另一个表中，让数据移动和复制更便捷。	CLONE TABLE

2020年6月更新记录

时间	特性	类别	描述	产品文档
2020-06-03	新增Tunnel overwrite命令。	新说明	支持Tunnel overwrite命令。	Tunnel命令
2020-06-01	优化Spark访问VPC实例。	新说明和示例	新增内容如下： - Spark访问VPC的白名单和地域的限制说明。 - Spark访问不同实例时，合并JSON文本的结果示例。	Spark访问VPC实例
2020-06-01	Policy和Download权限控制。	新示例	新增Policy授权和撤销的用法示例。	Policy和Download权限控制

2020年1月更新记录

时间	特性	类别	描述	产品文档
2020-01-14	SQL兼容性升级。	新说明	get_idcard_age、concat_ws 及 like 函数行为变更说明。	其他函数 字符串函数 LIKE字符匹配
2020-01-09	新增参数说明。	新说明	对示例中的参数进行详细的说明。	项目空间的数据保护

2.3. 2019年

本文为您介绍MaxCompute 2019年内容更新的最新动态，基于此您可以了解MaxCompute中增加了哪些新功能、语法新特性和权限变更，提升项目开发效率。

2019年12月更新记录

时间	特性	类别	描述	产品文档
2019-12-25	新增开源地理空间UDF。	新功能	您可以将开源地理空间UDF注册到MaxCompute中，以兼容开源Hive UDF的方式使用地理空间函数。	开源地理空间UDF

2019年11月更新记录

时间	特性	类别	描述	产品文档
2019-11-06	新增MaxCompute不支持分区剪裁函数说明。	新说明	新增MaxCompute不支持分区剪裁函数说明。	与Hive、MySQL、Oracle内建函数对照表

2019年10月更新记录

时间	特性	类别	描述	产品文档
2019-10-09	SQL语法新增功能。	新功能	- 新增合并分区语法。 - 新增JOIN与SET OP支持括号指定优先级语法。 - 新增内置函数JSON_TUPLE。 - 新增日期函数EXTRACT。 - 新增两个flag。 - 支持LIMIT OFFSET。 - 支持指定表的列默认值。 - 支持自然连接。 - 支持新运算符。 - 新增删除分区语法。	- 分区和列操作 - JOIN - 字符串函数 - 日期函数 - SELECT 语法 - SELECT 语法 - 表操作 - JOIN - 运算符 - 分区和列操作

2019年7月更新记录

时间	特性	类别	描述	产品文档
2019-07-12	新增set命令odps-sql-reshuffle-dynamicpt说明。	新命令	该命令用于设置动态分区，以避免拆分动态分区时产生过多小文件。	SET 操作

2019年6月更新记录

时间	特性	类别	描述	产品文档
2019-06-17	新增VALUES文档引导说明。	新说明	介绍了如何在业务测试阶段，创建数据进行简单的测试。	VALUES