

阿里云 MaxCompute

产品简介

文档版本：20181213

法律声明

阿里云提醒您在使用或阅读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
	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。	 禁止： 重置操作将丢失用户配置数据。
	该类警示信息可能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，恢复业务所需时间约10分钟。
	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。	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。	单击 确定 。
<code>courier</code> 字体	命令。	执行 <code>cd /d C:/windows</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid 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>swich {stand slave}</code>

目录

法律声明.....	I
通用约定.....	I
1 什么是MaxCompute.....	1
2 通告记录.....	4
3 用户必读.....	12
4 与其它阿里云服务的集成使用.....	15
5 文档最新动态.....	18
6 发展历程.....	30
7 通告记录.....	31

1 什么是MaxCompute

大数据计算服务 (MaxCompute , 原名ODPS) 是一种快速、完全托管的GB/TB/PB级数据仓库解决方案。

当今社会数据收集手段不断丰富，行业数据大量积累，数据规模已增长到了传统软件行业无法承载的海量数据 (百GB、TB乃至PB) 级别。MaxCompute服务于批量结构化数据的存储和计算，提供海量数据仓库的解决方案及分析建模服务。

由于单台服务器的处理能力有限，海量数据的分析需要分布式计算模型。分布式的计算模型对数据分析人员要求较高且不易维护：数据分析人员不仅需要了解业务需求，同时还需要熟悉底层分布式计算模型。MaxCompute为您提供完善的数据导入方案以及多种经典的分布式计算模型，助您快速解决海量数据的计算问题，有效降低企业成本并保障数据安全。您可以不必关心分布式计算和维护细节，便可轻松完成大数据分析。



说明：

MaxCompute已经在阿里巴巴集团内部得到大规模应用，例如大型互联网企业的数据仓库和BI分析、网站的日志分析、电子商务网站的交易分析、用户特征和兴趣挖掘等。

DataWorks和MaxCompute关系紧密：DataWorks为MaxCompute提供一站式的数据同步、业务流程设计、数据开发、管理和运维功能，详情请参见[DataWorks#原大数据开发套件#](#)。

MaxCompute视频简介

MaxCompute学习路径

您可以通过[MaxCompute学习路径](#)快速了解MaxCompute的相关概念、基础操作、进阶操作等。

产品优势

- 大规模计算存储

MaxCompute适用于100GB以上规模的存储及计算需求，最大可达EB级别。

- 多种计算模型

MaxCompute支持SQL、MapReduce、Graph等计算类型及MPI迭代类算法。

- 强数据安全

MaxCompute已稳定支撑阿里全部离线分析业务7年以上，提供多层沙箱防护及监控。

- 低成本

与企业自建私有云相比，MaxCompute的计算存储更高效，可以降低20%-30%的采购成本。

功能概述

- 数据通道

- 批量、历史数据通道

[TUNNEL](#)是MaxCompute为您提供的数据传输服务，提供高并发的离线数据上传下载服务。支持每天TB/PB级别的数据导入导出，特别适合于全量数据或历史数据的批量导入。Tunnel为您提供Java编程接口，并且在MaxCompute的客户端工具中，有对应的命令实现本地文件与服务数据的互通。

- 实时、增量数据通道

针对实时数据上传的场景，MaxCompute提供了延迟低、使用方便的[DataHub](#)服务，特别适用于增量数据的导入。DataHub还支持多种数据传输插件，例如Logstash、Flume、Fluentd、Sqoop等，同时支持日志服务Log Service中的[投递日志到MaxCompute](#)，进而使用DataWorks进行日志分析和挖掘。

- 计算及分析任务

MaxCompute支持多种计算模型，详情如下。

- [SQL](#)：MaxCompute以表的形式存储数据，支持多种[数据类型](#)，并对外提供SQL查询功能。您可以将MaxCompute作为传统的数据库软件操作，但其却能处理TB、PB级别的海量数据。



说明：

- MaxCompute SQL不支持事务、索引及Update/Delete等操作。
- MaxCompute的SQL语法与Oracle、MySQL有一定差别，您无法将其他数据库中的SQL语句无缝迁移到MaxCompute上来。详情请参见[与其他SQL语法的差异](#)。
- 在使用方式上，MaxCompute SQL最快可以在分钟、乃至秒级别完成查询，无法在毫秒级别返回结果。
- MaxCompute SQL的优点是学习成本低，您不需要了解复杂的分布式计算概念。如果您具备数据库操作经验，便可快速熟悉MaxCompute SQL的使用。

- [UDF](#)：即用户自定义函数。

MaxCompute提供了很多[内建函数](#)来满足您的计算需求，同时您还可以通过创建自定义函数来满足不同的计算需求。

- **MapReduce** : MaxCompute MapReduce是MaxCompute提供的Java MapReduce编程模型，它可以简化开发流程，更为高效。若您使用MaxCompute MapReduce，需要对分布式计算概念有基本了解，并有相对应的编程经验。MaxCompute MapReduce为您提供Java编程接口。
- **Graph** : MaxCompute提供的Graph功能是一套面向迭代的图计算处理框架。图计算作业使用图进行建模，图由点（Vertex）和边（Edge）组成，点和边包含权值（Value）。通过迭代对图进行编辑、演化，最终求解出结果，典型应用：[PageRank](#)、[单源最短距离算法](#)、[K-均值聚类算法](#)等。
- SDK
SDK是MaxCompute提供给开发者的工具包，当前支持[Java SDK](#)及[Python SDK](#)。
- 安全
MaxCompute提供了功能强大的安全服务，为您的数据安全提供保护，详情请参见[安全指南](#)。

后续步骤

现在，您已经学习了MaxCompute的产品优势、功能特性等相关简介，您可以继续学习下一个教程。在该教程中您将了解MaxCompute的相关收费情况，详情请参见[产品定价](#)。

2 通告记录

本文为您提供关于MaxCompute使用功能的各项更新记录。

2018年11月9日- Hash Clustering功能发布

Hash Clustering通过允许在建表时设置表的Shuffle和Sort属性，使得MaxCompute可以根据数据已有的存储特性来优化执行计划，提高效率，节省资源消耗。如果在初始表数据生成时，按照Hash Shuffle和Sort的方式存储，则后续查询中将避免对数据的再次Shuffle和Sort。具体如何在建表时设置、查看、管理Shuffle和Sort属性请参考[DDL语句](#)。



说明：

目前Hash Clustering还存在限制：

- 不支持Insert Into，只能通过Insert Overwrite来添加数据。
- 不支持Tunnel直接Upload到Range Cluster表，因为Tunnel上传数据是无序的。

2018年11月3日-底层服务软件缺陷修复

2018年11月3日10:00到19:00将对MaxCompute底层服务进行软件缺陷修复。正常情况下，您不会感知到升级过程。极端情况下，可能会产生约20分钟的作业运行失败或作业提交失败。一旦出现作业失败，请您重新提交失败任务或提交工单反馈。

2018年10月31日-SQL外部表功能开启计费

从2018年10月31日开始，MaxCompute SQL[外部表](#)功能开始计费。采用的计费标准为一次SQL计算费用=计算输入数据量*SQL复杂度*SQL价格。SQL价格是0.03元/GB/复杂度，复杂度系数为1。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。

2018年10月16日-Python UDF全面开放

从2018年10月16日开始，MaxCompute公共云环境下的Python UDF使用权限将全面开放，新增的project默认可用Python UDF，具体的Python UDF使用方式请参考文档[Python UDF](#)。更多PyODPS的内容介绍请参考相关[文档](#)。

2018年9月29日-SQL语言升级

MaxCompute Sprint29中发布的新功能如下。

- UDJ

为了满足用户很多跨表操作的需求，MaxCompute S29版本在UDF框架中引进了新的扩展机制：**UDJ(User Defined Join)**。UDJ提供了让用户自定义Join实现逻辑的接口。如何定义和使用UDJ，详情请参考文档[UDJ](#)。

- UDT

为了提升SQL语言编译过程的易用性与语言的表达能力，比如使用Select Transform可以不创建Function甚至不上传资源的情况下执行其他语言的脚本，MaxCompute2.0基于新一代的SQL引擎增加了**UDT (User Defined Type)** 新功能。

UDT功能允许用户在SQL中直接引用第三方语言的类或者对象，获取其数据内容或者调用其方法。如在SQL中直接写java的表达式，并可以引用JDK中的类。如何使用UDT，详情请参考文档[UDT](#)。



说明：

- 目前UDT只支持Java语言。
- 所有的运算符都是MaxCompute SQL的语义，不是UDT的语义。
- 目前UDT不能用作shuffle key：包括join，group by，distribute by，sort by，order by，cluster by 等结构的key。
- DDL不支持UDT，也不能将UDT对象insert到表中；屏显的最终结果也不能是UDT类型。

- 支持OUTER LATERAL VIEW

MaxCompute Sprint30扩展了Lateral View功能，支持Outer Lateral View，当table function不输出任何一行时，对应的输入行在lateral view结果中依然保留，且所有table function输出列为null。关于Lateral View介绍请参考文档DML语句的[Lateral View](#)部分。

- 支持SCALAR SUBQUERY

子查询支持SCALAR SUBQUERY，即当SUBQUERY的输出结果为单行单列的时候，可以当做标量来使用。具体使用介绍请参考文档DML语句[子查询](#)中的“SCALAR SUBQUERY”部分。

2018年9月26日-元数据仓库组件优化

为了提升您的MaxCompute使用性能，系统即将对元数据仓库组件进行优化，届时您会提前收到以下信息：

尊敬的阿里云用户您好，大数据计算服务MaxCompute为向您提供更好的产品性能和稳定性，将于北京时间2018年9月26日16:00 - 20:00对底层元数据仓库组件进行优化，您的应用可能出现1分钟左右

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。感谢您的理解和支持！谢谢。有任何问题，可随时通过企业钉钉群或工单联系反馈。

2018年-MaxCompute例行优化

为了提升您的MaxCompute使用性能，系统每个月将会对文件数占用较多的project进行小文件合并，届时您将会提前收到以下信息：

尊敬的客户，您好：为了提升您的MaxCompute使用性能，我们计划于XX年XX月XX日对文件数占用较多的project（只会涉及到5天内未被访问分区）进行小文件合并，此操作是后台自动做，用户无需参与。Merge执行过程中，作业读取正在被Merge的表会有小概率失败，业务上重试即可成功，届时请关注相关业务，如有异常随时通过企业钉钉群进行反馈。



说明：

如果您收到报错：ODPS-0010000:System internal error - PanguOpen4Read4R，说明作业在读取正在被Merge的表，请在业务上重试。

2018年9月17日支持setproject 2.0新类型和兼容hive模式

为了提升MaxCompute 2.0的使用体验，支持对project级别进行新类型和兼容hive模式打开，即对原来session级别的set odps.sql.type.system.odps2=true|false和 set odps.sql.hive.compatible=true|false 两个功能，升级为支持setproject级别。project的Owner可根据需要对project进行设置，命令为：

```
- setproject odps.sql.type.system.odps2=false|true  
- setproject odps.sql.hive.compatible=false|true
```

对setproject的详细说明请参见：[其他操作](#)。



说明：

打开project级别的新类型的主要影响：

- 某些隐式类型转换会被禁用，包括string -> bigint, string -> datetime, double->bigint, decimal -> double, decimal -> bigint都是有精度损失或者可能报错的。用户依旧可以用cast来做强制转换。
- 常量类型会变化。单独一个整形常量，如 123，在旧类型下是bigint类型，在新类型下是int类型。

- udf resolve结果可能变化，比如udf包含bigint和int两个重载。旧类型下一定走bigint的重载，而新类型下可能会被解析到int的重载。
- session级别会覆盖project级别设置。

2018年5月底—NewSQL开启计费

MaxCompute New SQL采用的计费标准为一次SQL计算费用=计算输入数据量*SQL复杂度*SQL价格。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。

2018年1月16日流量切换公告

1月16日14~17点对MaxCompute前端、tunnel服务进行流量切换，如果出现服务连接超时或访问失败，可提交[工单进行咨询](#)。

2018年1月15日开放label、Package安全模型

2018年1月15日，MaxCompute的安全模型开发label和Package，即[列级别访问控制](#)和[跨项目空间的资源分享](#)。

2018年1月10日禁止Full scan功能开放

2018年1月10日 20点新创建的项目（project）默认禁止分区表全表扫描，即对该项目的分区表执行全表扫描语句时会返回失败，必须指定分区条件，若实在需要，可以在对分区表全表扫描的sql语句前加一个set语句set odps.sql.allow.fullscan=true;，执行的时候，set语句和SQL语句一起提交执行。

2017年11月28-29日内部维护升级公告

由于MaxCompute内部维护升级出现问题，2017-11-28 22:00到2017-11-29 18:40期间的计算计量和存储计量未及时出账，相应账单会延后推送，但不影响最终计费结果，由此给您造成的不便深感抱歉。

2017年11月1日—华北2（北京）Region开服

2017年11月1日，阿里云数加>MaxCompute 华北2（北京）数据中心将正式开服售卖。

售价与华东2、华南1一致，详情请参见[计量计费文档](#)。

开通时，选择什么Region您需要考虑的最主要因素是MaxCompute与其他阿里云产品之间的关系。

不同区域数据是不能互通，仅针对MaxCompute而言，假如您既购买华东2又购买华北2的服务，那么MaxCompute之间的数据不互通。

如果您的云服务器ECS不在北京Region，若想访问连接北京的MaxCompute，需跨Region进行访问。跨Region访问的各种服务连接请参见[访问域名和数据中心](#)。

2017年10月11日—预付费支持自动续费

2017年10月11日，MaxCompute的预付费实例开始支持自动续费，可在[阿里云控制台的续费管理中心](#)操作。

MaxCompute续费的操作介绍请参见[续费管理](#)，更多自动续费的介绍请参见[自动续费功能](#)。

2017年9月19日—香港Region开服

2017年9月19日，阿里云数加>MaxCompute香港（香港）数据中心将正式开服售卖。

售价与华东2、华南1一致，详情请参见[计量计费文档](#)。

开通时，选择什么Region您需要考虑的最主要因素是MaxCompute与其他阿里云产品之间的关系。

不同区域数据是不能互通，仅针对MaxCompute而言，假如您既购买华东2又购买香港的服务，那么MaxCompute之间的数据不互通。

如果您的云服务器ECS不在香港，若想访问连接香港的MaxCompute，需跨Region进行访问。

跨Region访问的各种服务连接请参见[访问域名和数据中心](#)。

2017年9月7日—华南1（深圳）Region开服

2017年9月7日，阿里云数加>MaxCompute华南1（深圳）数据中心将正式开服售卖，这是数加>MaxCompute在国内开服的第二个区域。

关于售价

华南1区域价格与华东2一致，主要分三部分进行收费：存储、计算和下载，其中计算（指SQL和MR计算任务）分预付费、按量后付费两种模式，存储和下载都是按量后付费。做预算的具体的售价信息请参见[官网定价页](#)或[计量计费文档](#)。

关于开通

请您确保云账号是实名认证的账号，在开通购买页面进行区域选择时，注意选择华南1。那么什么场景适合选择区域是华南1呢？

选择地域时，您需要考虑的最主要因素是MaxCompute与其他阿里云产品之间的关系，示例如下：

- ECS在华南1区域。ECS访问、下载华南1的MaxCompute数据都可以通过内网，快速又避免跨Region走公网下载，从而产生费用。

- 数据在华南1区域。如RDS、OSS、TableStore等在华南1区，那么在和华南1的MaxCompute进行数据传输时可以走内网，避免跨Region配置阿里云内网或VPC网络时不保证连通性的情况。

关于不同区域数据是否互通

仅针对MaxCompute而言，假如您既购买华东2又购买华南1的服务，那么MaxCompute之间的数据不互通。

如果您的云服务器ECS不在华南1，若想访问连接华南1的MaxCompute，需跨Region进行访问。

跨Region访问的各种服务连接请参见[访问域名和数据中心](#)。

关于数据跨Region迁移

若您原来已经开通购买华东2区域，目前想使用华南1区域，所以希望将华东2的所有数据迁移到华南1。建议您通过以下方式进行迁移：

通过配置数据集成任务进行数据同步，如在华南1区域创建好的项目中，配置数据集成任务，来读取华东2区域的表，写入到华南1区域的表中。如果遇到问题可以提工单进行咨询。



说明：

如果您对新服购买有疑惑，可通过工单咨询或加入MaxCompute购买钉钉群（群号：11782920）进行咨询。

2017年8月16日-MR开启计费

MaxCompute MapReduce任务开启按量计费。计费标准如下：

MR任务当日计算费用=当日总计算时 * 0.46元（人民币）

一个MR任务一次执行成功的计算时=任务运行时间（小时）*任务调用的core数量。

如一个MR任务一次执行成功是调用了100core并花费0.5小时，那么本次执行计算时为：0.5 小时 × 100core= 50个计算时。

详情请参见[计量计费说明](#)。

2017年7月一开始分批升级MaxCompute2.0



说明：

2017年9月底已经全面升级开放MaxCompute2.0，MaxCompute2.0 NewSQL及非结构化（外表）处理能力公测阶段，不收取费用，具体收费时间等待官方通知。

MaxCompute2.0已经正式开始，升级后的MaxCompute2.0完全拥抱开源生态，支持更多的语言功能，带来更快的运行速度，同时新版本会执行更严格的语法检测。升级的具体时间会在分批升级时另行通知。

升级后需要对部分SQL语法做出调整，详情请参见[修改不兼容SQL实战](#)进行修改。

本次MaxCompute2.0产品升级以下功能：

- **全新的SQL引擎**

更快的SQL执行引擎：降低企业大数据分析成本，SQL执行效率更高。

- **生态兼容**

MaxCompute显著提升了针对Hive等开源产品的兼容性，直接兼容Hive的UDF/UDAF/UDTF，为Hive开发的大部分UDF/UDAF/UDTF可不经修改，直接在MaxCompute中运行。

详情请参见[INSERT操作](#)、[SELECT操作](#)、[数据类型](#)、[日期函数](#)、[数学函数](#)、[字符串函数](#)、其他函数和[JAVA UDF](#)等相关文档。

- **非结构化处理能力**

通过MaxCompute SQL直接处理OSS/TableStore (OTS) 数据，从而处理音频、视频、图像、气象等非结构化数据以及K-V类型的数据。

详情请参见[前言](#)、[访问OTS非结构化数据](#)和[访问OSS非结构化数据](#)。

- **MaxCompute Studio**

MaxCompute编译器基于MaxCompute2.0全新自主研发的SQL引擎，尤其配合使用MaxCompute Studio，提供了丰富的错误提示、警告的功能及作业排队展示。详情请参见[巧用MaxCompute编译器的错误和警告](#)。

2017年06月01日

尊敬的阿里云客户：

MaxCompute目前只针对SQL作业进行收费，MapReduce作业近期将开启收费。

MaxCompute近期将进行2.0版本升级，升级过程中会逐批进行，本次版本变更将带来一些新特性：

- **SQL执行引擎**

降低大数据分析成本，SQL执行效率更高。

- **非结构化处理能力**

通过MaxCompute SQL直接处理OSS/TableStore (OTS) 数据，从而处理音频、视频、图像、气象等非结构化数据以及K-V类型数据。

- 生态兼容

MaxCompute MapReduce SDK高度兼容了开源Hadoop MapReduce的使用场景。

- 基于InteliJ的本地开发插件

本地对MaxCompute作业进行开发，调试。

- 实时数据上传

新增DataHub模块，帮助企业数据实时上传。

- 新型算法平台

基于GPU的深度学习算法框架，在线预测服务等。

如果您在系统升级过程中遇到问题，请及时提交工单联系我们。

2016年09月16日

尊敬的阿里云客户：

MaxCompute在11月1日进行了计费变更。本次变更升级存储及计算两个方面：

- 升级后，修复部分SQL作业无法计费的问题。
- 升级后，解决部分场景下MaxCompute存储费用存在的问题。升级后，存储将按照实际情况收费。

以上两处之前均不存在向您多收取费用的情况，对于历史作业不做费用补收。您可以通过计量计费说明中获取账单详情的相关指引，从阿里云官网获取每天详细的计量计费信息。

目前MaxCompute只有SQL作业收费（不包括UDF），UDF/MapReduce/Graph/PAI等作业仅是公测状态。具体收费计划请关注官网公告。

感谢您长期以来对MaxCompute产品的支持。

3 用户必读

本文根据您的角色为您推荐不同的文档阅读顺序，并为您介绍MaxCompute产品各模块使用限制。

如果您是MaxCompute初学者

如果您是初学者，建议您从如下模块开始读起：

- **产品简介**：MaxCompute产品的总体介绍以及包含的主要功能。通过阅读该章节，您会对MaxCompute有一个总体的认识。
- **快速开始**：通过示例，指导您如何进行申请账号、安装客户端、创建表、授权、导入导出数据、运行SQL任务、运行UDF/Mapreduce程序等操作。
- **基本介绍**：MaxCompute的基本概念及常用命令介绍。您可以进一步熟悉如何操作MaxCompute。
- **工具**：在分析数据之前，您需要掌握MaxCompute常用工具的下载、配置以及使用方法。

我们为您提供**工具**，您可以通过此工具对MaxCompute进行操作。

- **配置Endpoint**：MaxCompute Region的开通情况和连接方式，解答您在与其他云产品（ECS、TableStore、OSS）互访场景中遇到的网络连通性和下载数据收费等问题。

建议您熟悉以上的模块后，再有针对性地对其他模块进行深入学习。

如果您是数据分析师

如果您是数据分析师，建议您熟读SQL模块的内容。

- **SQL**：您可以查询并分析存储在MaxCompute上的大规模数据。包含的主要功能如下：
 - 支持DDL语句，您可以通过Create、Drop和Alter对表和分区进行管理。
 - 您可以通过Select选择表中的某几条记录，通过Where语句查看满足条件的记录，实现过滤功能。
 - 您可以通过等值连接Join实现两张表的关联。
 - 您可以通过对某些列Group By，实现聚合操作。
 - 您可以通过Insert overwrite/into把结果记录插入到另一张表中。
 - 您可以通过内置函数和自定义函数（UDF）来实现一系列的计算。

如果您拥有一定开发经验

如果您拥有一定的开发经验，了解分布式概念，并且某些数据分析可能无法用SQL来实现，此时推荐您学习MaxCompute更高级的功能模块。如下所示：

- [MapReduce](#)：MaxCompute提供的Java MapReduce编程模型。您可以使用MapReduce提供的接口（Java API）编写MapReduce程序，处理MaxCompute中的数据。
- [Graph](#)：一套面向迭代的图计算处理框架。使用图进行建模，图由点（Vertex）和边（Edge）组成，点和边包含权值（Value）。通过迭代对图进行编辑、演化，最终得出结果。
- [Eclipse Plugin](#)：方便您使用MapReduce、UDF以及Graph的Java SDK进行开发工作。
- [Tunnel](#)：您可以使用Tunnel服务向MaxCompute批量上传离线数据或者从MaxCompute下载离线数据。
- SDK：
 - [Java SDK](#)：向开发者提供Java接口。
 - [Python SDK](#)：向开发者提供Python接口。



说明：

目前[MapReduce](#)以及[Graph](#)功能仍处于公测中，若您想使用这部分功能，可以通过工单系统提交申请。申请时请注明您的项目空间名称，我们会在7个工作日内处理。

如果您是项目Owner或者管理员

如果您是一个项目空间的Owner或者管理员，您需要熟知以下模块：

- [安全指南](#)：您可以通过阅读该章节，了解如何进行给用户授权、跨项目空间的资源共享、设置项目空间的数据保护功能、policy授权等操作。
- [MaxCompute收费指南](#)：介绍MaxCompute的收费模式。
- 部分只有项目空间Owner才能使用的命令，例如常用命令中[其他操作](#)的SetProject操作。

使用限制汇总

- [SQL开发限制](#)
- [数据上传下载限制](#)
- [操作命令限制](#)
- [MapReduce限制](#)
- [图模型限制](#)

- 安全配置限制
- *Lightning*限制
- *PyODPS*限制

4 与其它阿里云服务的集成使用

MaxCompute (原ODPS) 是一种大数据计算服务, 能提供快速、完全托管的PB级数据仓库解决方案, 已经与阿里云部分产品集成, 可以快速实现很多业务场景。

MaxCompute与DataWorks

[DataWorks](#)是基于MaxCompute计算和存储, 提供 workflow 可视化开发、调度运维托管的一站式海量数据离线加工分析平台。在数加中, DataWorks控制台即为MaxCompute控制台。

通过DataWorks, 您既可直接编写并运行MaxCompute SQL, 又能可视化配置 workflow 并定时调度运行MaxCompute SQL、MR等任务。更多使用说明请参见[DataWorks帮助文档](#)。



说明：

您可以将DataWorks理解成MaxCompute的web客户端。

MaxCompute与数据集成

MaxCompute可以通过数据集成加载不同数据源数据, 同样也可以通过数据集成把MaxCompute的数据导出到各种业务数据库。

数据集成已经集成到DataWorks作为数据同步任务进行配置、运行。您可直接在DataWorks上配置[MaxCompute数据源](#), 再配置[读取MaxCompute表](#)或者[写入MaxCompute表](#)任务, 整个过程只需在一个平台上进行操作。

MaxCompute与机器学习

[机器学习](#)是基于MaxCompute的一款机器学习算法平台。数加上创建好MaxCompute项目, 开通好机器学习, 即可通过机器学习平台的算法组件对MaxCompute数据进行模型训练等操作。详情请参见[机器学习操作文档](#)。

MaxCompute与QuickBI

数据在MaxCompute进行加工处理后, 将Project添加为[QuickBI数据源](#), 即可在QuickBI页面[对MaxCompute表数据进行报表制作](#), 实现数据可视化分析。

MaxCompute与AnalyticDB

[AnalyticDB](#)是海量数据实时高并发在线分析 (Realtime OLAP) 的云计算服务, 与MaxCompute双剑合璧实现大数据驱动业务系统的场景。通过MaxCompute离线计算挖掘, 产出高质量数据后, 导入分析型数据库, 供业务系统调用分析。

将MaxCompute数据导入到AnalyticDB，有以下两种方式：

- 通过DMS for AnalyticDB的[导入导出](#)功能进行配置。
- 通过DataWorks配置数据同步任务，[读MaxCompute](#)和[写AnalyticDB](#)。

MaxCompute与推荐引擎

[推荐引擎](#)是在阿里云计算环境下建立的一套推荐服务框架，推荐服务通常由三部分组成：日志采集、推荐计算和产品对接，而推荐计算的离线计算输入和输出都是MaxCompute（原ODPS）表。

在推荐引擎控制台的资源管理页面，通过[添加云计算资源](#)的方式，将MaxCompute项目添加为推荐引擎的计算资源。

MaxCompute与表格存储

[表格存储#Table Store#](#)是构建在阿里云飞天分布式系统之上的分布式NoSQL数据存储服务，MaxCompute2.0支持直接通过外部表方式访问表格存储中的表数据并进行处理，详情请参见[访问OTS非结构化数据](#)。

MaxCompute与OSS

[对象存储OSS](#)是海量、安全、低成本、高可靠的云存储服务，MaxCompute2.0支持直接通过外部表方式访问表格存储中的表数据并进行处理，详情请参见[访问OSS非结构化数据](#)。

MaxCompute与OpenSearch

阿里云[开放搜索OpenSearch](#)是一款阿里巴巴自主研发的大规模分布式搜索引擎平台。数据通过MaxCompute进行计算处理后，可以在OpenSearch平台上通过[添加数据源](#)的方式将MaxCompute数据接入。

MaxCompute与移动数据分析

[移动数据分析#Mobile Analytics#](#)是阿里云推出的一款移动APP数据统计分析产品，为开发者提供一站式数据化运营服务。当移动数据分析自带的基础的分析报表不能满足APP开发者的个性化需求时，可以将数据[一键同步至Maxcompute](#)，结合自己的业务需求来进一步加工、分析自己的数据。

MaxCompute与日志服务

[日志服务](#)能快速完成数据采集、消费、投递以及查询分析等功能。日志数据采集后，需要更多的个性化分析、挖掘，您可以在日志服务上[投递日志到MaxCompute](#)，通过MaxCompute对日志数据进行个性化、深层次的数据分析、挖掘。

MaxCompute与RAM

RAM是阿里云为客户提供的用户身份管理与资源访问控制服务。MaxCompute与RAM的集成使用主要有两个场景：

- 通过**DataWorks**使用**MaxCompute**时，子账户的身份管理

主帐号开通并创建项目后，若需要通过DataWorks使用MaxCompute且多个账户协同开发，必须由主帐号到RAM服务中创建子账户，将RAM子账户添加为项目成员从而进行协同开发，详情请参见[准备RAM子账号](#)、[添加工作空间成员和角色](#)。



说明：

此时RAM只起到用户身份管理功能，相关的权限管理不在RAM上控制。

- **MaxCompute**处理非结构化数据时，通过**RAM**对非结构化数据进行授权

目前MaxCompute支持直接处理非结构化数据（包含OSS和Table Store），前提条件之一就是需要在RAM中授予MaxCompute访问OSS或Table Store的权限，详情请参见[访问OSS非结构化数据](#)、[访问OTS非结构化数据](#)。

5 文档最新动态

本文为您介绍MaxCompute2.0版本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，为您开发项目提高效率。

时间	特性	类别	描述	适用客户	产品文档
2018.12.04	大数据计算服务MaxCompute欧洲西部1（伦敦）节点正式开服售卖	新地域/可用区	大数据计算服务MaxCompute欧洲西部1（伦敦）节点开服售卖。可在国际站、中文站、日本站购买节点资源，在控制台按需求开通project。	所有	配置Endpoint
2018.11.28	新增在MaxCompute上分析IP来源的方法文档	体验优化	介绍如何在MaxCompute上分析IP来源的方法，包括下载、上传淘宝IP地址库数据及编写UDF函数、编写SQL四个步骤。	所有	使用MaxCompute分析IP来源最佳实践
2018.11.26	新增从MongoDB提取JSON字段到MaxCompute最佳实践	体验优化	介绍如何利用DataWorks数据集成直接从MongoDB提取JSON字段到MaxCompute。	所有	JSON数据从MongoDB迁移到MaxCompute最佳实践
2018.11.15	处理非结构化数据目录更新为外部表	体验优化	处理非结构化数据目录更新为外部表	所有	外部表
2018.11.13	新增JSON数据从OSS迁移到MaxCompute最佳实践	体验优化	介绍如何利用DataWorks数据集成将JSON数据从OSS迁移到MaxCompute，并使用MaxCompute内置字符串函数GET_JSON_OBJECT提取JSON信息。	所有	JSON数据从OSS迁移到MaxCompute最佳实践
2018.11.08	新增与MySQL、Oracle内建函数对照表	体验优化	帮助您根据MySQL内建函数查找对应的MaxCompute内建函数。	所有	与MySQL、Oracle内建函数对照表
2018.11.08	标题“导读”更新为“用户必读”，添加MaxCompute使用限制汇总	体验优化	标题“导读”更新为“用户必读”，添加MaxCompute SQL、数据上传下载、MapReduce	所有	用户必读

			、安全配置、Lighting、PyODPS等使用限制汇总。		
2018.11.05	MaxCompute SQL新窗口函数 cume_dist	新功能	该函数用于取得累计分布，相当于求分组中值小于等于当前值的行数占分组总行数的比例。	所有	窗口函数
2018.11.03	将Java SDK、Python SDK和PyODPS DataFrame自定义函数中使用第三方包三篇文档移动到SDK参考目录	体验优化	将用户指南中的Java SDK、Python SDK和PyODPS DataFrame自定义函数中使用第三方包三篇文档统一移动到SDK参考目录	所有	Java SDK 、 Python SDK 、 PyODPS DataFrame自定义函数中使用第三方包
2018.11.02	新增RDS迁移到MaxCompute实现动态分区最佳实践文档	体验优化	介绍如何将RDS数据迁移到MaxCompute表实现动态分区	所有	RDS迁移到MaxCompute实现动态分区
2018.10.31	MaxCompute SQL外部表功能计费	新规格	从2018年10月31日开始，MaxCompute SQL外表功能开始计费。MaxCompute下创建和OTS表存储关联的表格，使用SQL查询时，无论通过外网还是经典/VPC网络访问OTS（除非使用STS方式访问），OTS都会收取费用。当您需要内外表混合作业的时候，会分别计费。	所有	计量计费说明
2018.10.16	新增数据迁移文档	体验优化	整理将数据移迁移到MaxCompute的最佳实践相关文档。	所有	数据迁移
2018.9.29	MaxCompute SQL引擎新功能User Defined Type（简称UDT）	新功能	允许在SQL中直接引用第三方语言的类或者对象，获取其数据内容或者调用其方法。	所有	UDT
2018.9.29	MaxCompute UDF框架中新近引入的新扩展机制	新功能	满足了MaxCompute内置Join操作无法实现的灵活的跨表、多表自定义操作，并且减少	所有	UDJ

			了对分布式系统底层细节的操作，提高了执行效率。		
2018.9.27	PyODPS已全面对外开放	新规格	PyODPS是MaxCompute Python版本的SDK，提供了DataFrame框架，现已全面对外开放。DataWorks也推出了PYODPS任务类型，可在DataWorks的PYODPS节点上直接编辑python代码操作Maxcompute。	所有	PyODPS
2018.5.25	Maxcompute支持OSS压缩格式gzip	新功能	MaxCompute目前只支持通过内置extractor读取OSS上gzip压缩的CSV/TSV数据。与读取非压缩文件相比，主要区别是SERDEPROPERTIES指定的属性。	所有	访问OSS非结构化数据
2018.5.25	Maxcompute支持OSS上的Hive文件格式	新规格	存储在OSS上的各种流行的开源数据格式（ORC、PARQUET、SEQUENCEFILE、RCFILE、AVRO和TEXTFILE）可通过非结构化框架在MaxCompute进行处理。	所有	处理OSS的开源格式数据
2018.5.25	Maxcompute SQL新语法Select Transform	新功能	Select Transform功能明显简化了对脚本代码对引用，支持Java、Python、Shell、Perl等语言，且编写过程简单，适合adhoc功能的实现。	所有	Select Transform语法
2018.5.25	Maxcompute SQL新语法SEMI JOIN	新功能	MaxCompute支持SEMI JOIN（半连接）。SEMI JOIN中，右表只用来过滤左表的数据而不出现在结果集中。支持LEFT SEMI JOIN和LEFT ANTI JOIN两种语法。	所有	SEMI JOIN
2018.5.25	Maxcompute SQL支持CASE WHEN的隐式类型转换	新规格	Maxcompute SQL支持CASE WHEN的隐式类型转换	所有	CASE WHEN作用下的隐式转换

2018.5.16	MaxCompute提供全表扫描的设置操作	新功能	MaxCompute提供全表扫描的设置操作	所有	其他操作
2018.4.27	Dataworks支持MaxCompute PYODPS任务	新规格	DataWorks推出了PYODPS任务类型，集成了Maxcompute的Python SDK，可在DataWorks的PYODPS节点上直接编辑Python代码操作Maxcompute，也可以设置调度任务来处理数据，提高数据开发效率。	开发者	PYODPS节点
2018.4.14	MaxCompute自动续费支持过期实例续费	新规格	原先预付费用户实例过期，在续费中心找不到过期实例的续费方式，耽误项目正常使用，新功能可以帮助用户通过续费中心很方便的完成手动续费，保障项目的正常使用，操作步骤为登陆后选择左侧大数据—>数加控制台 -> MaxCompute 续费管理。	所有	
2018.4.10	大数据计算服务MaxCompute 美国东部1（弗吉尼亚）节点正式开服售卖	新地域/可用区	大数据计算服务 MaxCompute 美国东部1（弗吉尼亚）节点开服售卖。可在国际站、中文站、日本站购买美东节点资源，在控制台按需求开通project。	所有	
2018.4.2	MaxCompute Studio 2.9.0 新版本发布	新功能	支持代码检查和快速修复 2.支持graph开发及调试 3.更多改进及Bug Fix，请更新后查看Release Notes： MaxCompute Studio 2.9.0	开发者、数据分析师	认识Studio
2018.3.28	MaxCompute CU管家支持华东1杭州Region	新规格	本次MaxCompute为华东1预付费用户带来了计算资源管理功能；之前，MaxCompute 华东1预付费用户购买了CU计算资源，但并不清楚资源消耗的情况，现在，CU管家不仅可以帮助企业客户了解计算资源的消耗情况，还可以帮助企业客户	所有	

			排查计算任务排队情况，并根据任务重要程度，分组管理计算任务。		
2018.3.23	MaxCompute Studio 支持查询结果自定义 Tunnel endpoint	修复问题	修复功能：使用 Instance Tunnel 下载 SQL 结果时，用户可以指定 Tunnel endpoint。	所有	
2018.3.22	大数据计算服务 MaxCompute 华东1（杭州）节点正式开服售卖	新地域/可用区	大数据计算服务 MaxCompute 华东1（杭州）节点开服售卖。可在中文站、国际站、日本站购买杭州节点资源，在控制台按需求开通project。	所有	配置Endpoint
2018.3.22	增强SQL 编辑器功能	体验优化	DataWorks 通过在前端实现 MaxCompute SQL和编辑器参数等扩展语法的AST解析，来提升脚本开发效率和开发体验，就像本地编辑器一样，提供智能代码提示，语法高亮，语法检查，风险及错误提示，类型检查等一系列功能。	所有	DataWorks 增强SQL编辑器功能
2018.2.9	PyODPS 0.7.15版本发布	新版本/新规格	PyODPS 0.7.15较与之前版本新增、优化和修复了相关功能，并提升了开发者的体验与效率：1、在 Python 3.6版本上，开发者可以直接使用 DataFrame 自定义函数了；2、修复 Python 3.5 版本上，DataFrame 自定义函数可能会导致错误的问题；3、支持Xflow 运行时可以设置运行参数；4、从 pandas DataFrame 创建 PyODPS DataFrame 时，as_type 参数可以指定组合类型（如 list 或者 dict 类型）；5、修复其他BUG。	Python 开发者、数据分析师	Python SDK

2018.2.9	大数据计算服务 MaxCompute 印尼开服，亚太东南5（雅加达）节点正式售卖	新地域/可用区	大数据计算服务 MaxCompute 亚太东南5（雅加达）节点正式开服售卖。可在国际站、中文站、日本站购买雅加达节点资源，在控制台上按需求开通 project。	所有	配置Endpoint
2018.2.8	MaxCompute服务等级协议正式发布	体验优化	MaxCompute服务已对接服务等级监控（SLA）平台，用户可以直接在官网控制台可用性中心查看每月可用性数据。服务承诺可实现99.9%的正常运行时间和可用性，如果未能实现承诺，客户可以提出赔偿申请。	所有	MaxCompute服务等级协议
2018.2.5	新增函数“TRANS_ARRAY”说明	新功能	该函数用于将一行数据转为多行的UDTF，将列中存储的以固定分隔符格式分隔的数组转为多行。	所有	其他函数
2018.2.3	MaxCompute Studio 2.8.2.3版本发布	体验优化	1.MaxCompute Studio Editor的侧边栏针对完整sql语句增加执行功能。原本当Editor工作区中有多条sql语句而仅需要执行其中某条sql语句时，需要先选中对应语句，再点击工具栏的执行按钮，才会只执行该条语句。本次优化后，可直接在左侧栏对对应的sql语句操作执行，符合开发者习惯，减少操作步骤，提升使用体验。2.MaxCompute Studio Project Explorer中按类型划分内置函数。原本的函数列表不分类型，一个大列表显示所有函数，查找不方便，同时也无法区分名称相同类型和使用方式不同的函数。本次优化后内置函数列表按类型划分，与官方文档对应，方便开发者查找并使用对应的函数。	所有	

2018.2.3	添加 top instance 命令说明	新功能	添加 top instance 命令说明	所有	常用命令-实例操作
2018.2.3	添加如何查看外部表的"external info"	新功能	添加如何查看外部表的"external info" (包括 LOCATION、StorageHandler 等) 信息的介绍。	所有	常用命令-表操作、访问 OSS 非结构化数据
2018.2.2	添加 MaxCompute 管家介绍入口	体验优化	MaxCompute 管家就是解决这个问题。目前 MaxCompute 管家主要提供三个功能，系统状态、资源组分配、任务监控。添加 MaxCompute 管家 介绍入口	所有	MaxCompute 管家
2018.2.1	MaxCompute Console 支持 top instance 命令	新规格	MaxCompute Console 0.29.0 版本推出 top instance 命令，支持开发者通过该命令查询当前项目空间内正在执行或排队的作业具体信息，方便开发者、管理员查看与管理当前项目作业情况。	MaxCompute 开发者、DataWorks 开发者、MaxCompute Studio 开发者	实例操作
2018.1.31	MaxCompute CU 管家	新功能	MaxCompute 为预付费用户带来了计算资源管理功能；之前，MaxCompute 预付费用户购买了 CU 计算资源，但并不清楚资源消耗的情况，现在，CU 管家不仅可以帮助企业客户了解计算资源的消耗情况，还可以帮助企业客户排查计算任务排队情况，并根据任务重要程度，分组管理计算任务。	MaxCompute 预付费用户	MaxCompute 预付费资源监控工具-CU 管家
2018.1.22	DataHub 新增四大 Region	新地域/可用区	DataHub 新增开通华东 2 (上海)、华北 2 (北京)、华南 1 (深圳) 以及亚太东南 1 (新加坡)，可以在管控台上根据业务需求开通服务。	所有	

2018.1.18	MaxCompute Studio 2.8.2.2版本发布	新版本/新规格	MaxCompute Studio 2.8.2.2 主要发布的功能有四点：功能一：支持python udf开发。功能二：支持非结构化unit test编写。功能三：支持UDF local run时对复杂数据类型进行操作。功能四：支持自定义sql语句格式化规则。	开发者	认识Studio
2018.1.18	大数据计算服务 MaxCompute 印度开服，亚太南部1（孟买）节点正式售卖	新地域/可用区	大数据计算服务 MaxCompute 亚太南部1（孟买）节点正式开服售卖。可在国际站、中文站、日本站购买孟买节点资源，在控制台上按需求开通project。	所有	配置Endpoint
2018.1.15	于Package的跨项目空间的资源分享功能发布	新功能	Package是一种跨项目空间共享机制，主要用于解决跨项目空间授权问题	数据开发者和项目管理员	基于Package的跨项目空间的资源分享
2018.1.15	基于LabelSecurity的列级别安全访问控制功能发布	新功能	MaxCompute推出基于表字段级别的安全访问控制，更细粒度的管控	数据开发者和项目管理员	列级别访问控制
2018.1.10	MaxCompute禁止 Full Scan功能开放	新功能	禁止全表扫描分区表，避免资源浪费，同时防止全表扫描带来的不必要费用。	大数据的管理者和开发者	其他操作
2018.01	新增项目空间的数据保护文档	体验优化	介绍项目空间的数据保护背景动机，介绍保护机制以及在开启保护机制后如何让数据流出的方法	所有	项目空间的数据保护
2018.01	新增“Hive UDF兼容示例”示例	体验优化	以一个示例介绍Hive的UDF如何在MaxCompute上使用。	所有	JAVA UDF
2018.01	添加TableTunnel接口说明以及底层实现和使用限制	体验优化	添加TableTunnel接口说明以及底层实现和使用限制。	所有	批量数据通道 TableTunnel

2018.01	添加公开数据集介绍入口，给试用产品的用户提供测试数据	体验优化	添加公开数据集介绍入口，给试用产品的用户提供测试数据。	所有	公开数据集介绍
2018.01	新增Full scan功能开放的通告	新功能	由于Full Scan功能发布，在这些文档中添加相关说明。	所有	通告 、 常用命令 、 其他操作 、 SQL常见错误
2017.12	添加set flag的说明	体验优化	在会使用到2.0新扩展的数据类型的函数说明中，都加上需要set flag的说明	所有	日期函数 、 数学函数 、 字符串函数 、 其他函数
2017.12	新增对应Studio2.8.2.*的新功能说明	新功能/体验优化	新增对应Studio2.8.2.*的新功能说明，包括可视化建表、历史提交sql、保存/查看logview、通过AK账户创建项目链接、自定义脚本模板等等	所有	MaxCompute Studio模块： 可视化创建/修改表 、 作业实例详情 、 项目空间连接管理 、 提交SQL脚本 、 作业浏览
2017.12	添加3个MaxCompute Studio短视频	体验优化	添加3个短视频：Studio 安装介绍、通过Studio管理数据、Studio SQL Scripting	所有	MaxCompute Studio模块： Studio 视频介绍
2017.12	新增函数SUBSTRING说明	体验优化	新增函数SUBSTRING说明	所有	内置函数 字符串函数
2017.12	添加写入特殊类型数据的说明案例	体验优化	values语法中，添加写入特殊类型数据的说明案例，包括DATETIME、TIMESTAMP类型和ARRAY复杂类型	所有	SQL语句 更新表中的数据 、 #INSERT OVERWRITE and INTO#
2017.12	新增MaxCompute服务等级协议相关说明	体验优化	新增MaxCompute服务等级协议相关说明	所有	MaxCompute 服务等级协议

2017.11	添加MaxCompute和OSS、OTS的Owner为不同账号时的授权方式，修改访问OSS非结构化数据和访问OTS非结构化数据文档内容结构	新功能/体验优化	添加MaxCompute和OSS、OTS的Owner为不同账号时的授权方式，同时修改两个文档的内容结构，方便查找各个步骤的操作内容。	所有	访问OSS非结构化数据 、 访问OTS非结构化数据
2017.11	添加TunnelUrl配置	新功能	Tunnel SDK所有示例代码中添加TunnelUrl配置，由于Tunnel Endpoint没有自动路由，若代码中不指定Tunnel的Endpoint，默认都是走公网。	所有	批量数据通道SDK示例
2017.11	添加跨项目空间Table、Resource、Function分享示例	体验优化	添加跨项目空间Table、Resource、Function分享示例	所有	安全指南的授权
2017.11	添加OSS源数据文件的外部表分区表的使用方法	新功能/体验优化	添加OSS源数据文件的外部表分区表的使用方法	所有	用户指南的访问OSS非结构化数据
2017.11	添加Studio4个操作短视频	体验优化	添加Studio4个操作短视频，对快速了解工具非常	所有	工具及下载的Studio部分Studio视频介绍
2017.10	新增详细的MaxCompute预付费实例的续费管理操作介绍	体验优化	新增详细的MaxCompute预付费实例的续费管理操作介绍，包括手动续费、自动续费、不续费；新增MaxCompute预付费实例的升级/降配操作介绍。	所有	续费管理、升级/降配
2017.10	新增分区剪裁合理性评估最佳实践文档	体验优化	新增最佳实践文档，主要介绍如查询分区表时如何合理评估分区裁剪，以便提高查询任务效率，降低计算量。	所有	分区剪裁合理性评估
2017.10	更新用户及角色管理说明	体验优化	将原来准备工作中的添加用户、角色等说明下掉，统一由这个文档做链接引导到对应的安全指南文档。	所有	准备工作用户及角色管理

2017.10	重新整理准备工作和安全指南重复文档和去掉不支持的授权	体验优化	这些文档都将原来准备工作和安全指南中的快速开始、用户管理及授权3个地方重复的文档进行统一整理。授权中去掉了一些目前不支持的授权客体和action。	所有	安全指南的 用户管理 、 角色管理 、 授权 、 查看权限 、 项目空间的安全配置
2017.9	DML相关文档中添加MaxCompute2.0版本扩展支持功能	新规格	DML相关文档中添加MaxCompute2.0版本扩展支持的新功能，包括values、CTE、SEMIJOIN等。	所有	更新表中的数据#INSERT OVERWRITE and INTO# 、 SELECT 操作
2017.9	相关函数文档添加新支持的函数说明	新规格	相关函数文档添加新支持的函数说明。	所有	日期函数 、 数学函数 、 字符串函数 、 其他函数
2017.9	数据类型文档添加新支持的数据类型说明，JAVA UDF文档添加新数据类型使用方法、返回类型、复杂数据类型示例	新规格/体验优化	数据类型文档添加新支持的数据类型说明包括基本数据类型和复杂数据类型。JAVA UDF文档添加新数据类型使用方法、返回类型、复杂数据类型示例。	所有	数据类型 、 JAVA UDF
2017.9	添加新数据类型隐式转换对照表	体验优化	从SQL概要中拆分出来，同时添加新数据类型隐式转换对照表。	所有	类型转换
2017.9	添加华南1Tunnel Endpoint；添加香港、亚太东南1Region相关endpoint说明	新规格	添加华南1Tunnel Endpoint；添加香港、亚太东南1Region相关endpoint说明。	所有	访问域名和数据中心
2017.9	新增最佳实践：长周期指标的计算优化方案	体验优化	新增最佳实践文档。	所有	长周期指标的计算优化方案

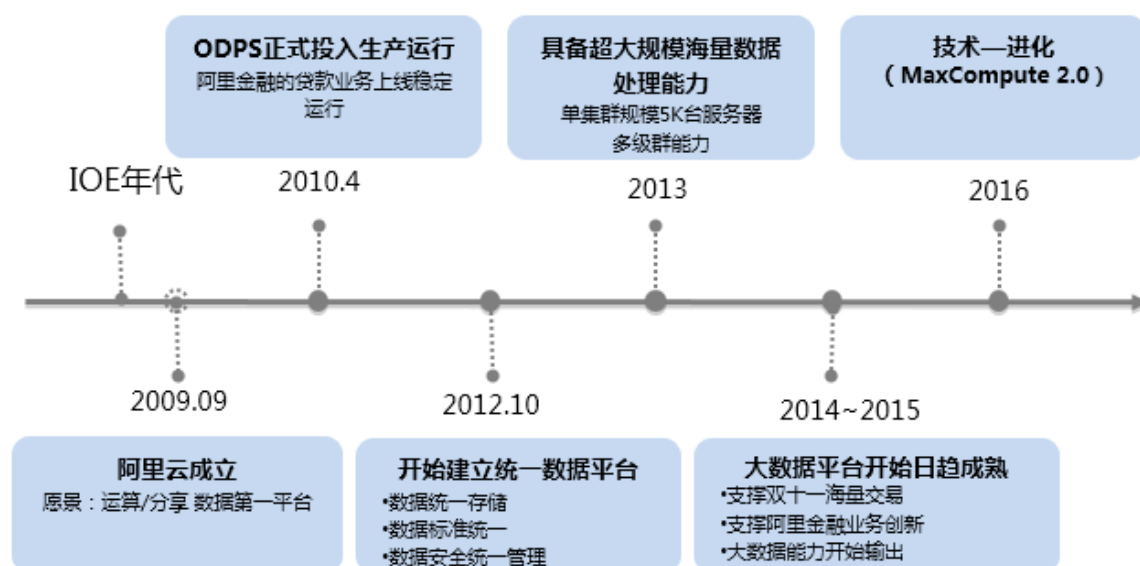
2017.9	Logview和巧用MaxCompute 编译器的错误和警告两篇文章移动	体验优化	将这两篇所在目录《JOB运行信息查看》从“工具及下载”移动到“用户指南”一级目录下。	所有	Logview 、 巧用MaxCompute编译器的错误和警告
2017.9	处理非结构化数据文档移动到“用户指南”一级目录下，OSS和OTS自定义授权内容添加AliyunODPS RolePolicy的策略内容	体验优化	将《处理非结构化数据》相关内容移动到“用户指南”一级目录下，OSS和OTS自定义授权内容添加AliyunODPS RolePolicy的策略内容。	所有	处理非结构化数据
2017.9	产品简介目录下文档内容格式规范化	体验优化	产品简介目录下文档内容格式规范化。	所有	什么是MaxCompute

6 发展历程

本文为您介绍MaxCompute从诞生到成熟的发展历程。了解更多MaxCompute的最新动态请参见[文档最新动态](#)。

从2009年9月阿里云成立，愿景就是做运算/分享数据的第一平台。2010年4月，伴随阿里金融的贷款业务上线，ODPS正式投入生产运行。2012年建立统一数据平台，2013年具备超大规模海量数据处理能力，2014~2015年大数据平台开始日趋成熟，2016年MaxCompute2.0诞生，成立之初的愿景正在逐步实现。

关键性里程碑



时间	动态
2010.04	ODPS正式投入生产运行，阿里金融的贷款业务上线稳定运行。
2013.05	ODPS公测。
2013.07	ODPS正式提供商业化服务，单集群规模5K台服务器多级群能力。
2016.09	ODPS正式更名为MaxCompute，并推出MaxCompute2.0，实现高性能，新功能，富生态。

7 通告记录

本文为您提供关于MaxCompute使用功能的各项更新记录。

2018年11月9日- Hash Clustering功能发布

Hash Clustering通过允许在建表时设置表的Shuffle和Sort属性，使得MaxCompute可以根据数据已有的存储特性来优化执行计划，提高效率，节省资源消耗。如果在初始表数据生成时，按照Hash Shuffle和Sort的方式存储，则后续查询中将避免对数据的再次Shuffle和Sort。具体如何在建表时设置、查看、管理Shuffle和Sort属性请参考[DDL语句](#)。



说明：

目前Hash Clustering还存在限制：

- 不支持Insert Into，只能通过Insert Overwrite来添加数据。
- 不支持Tunnel直接Upload到Range Cluster表，因为Tunnel上传数据是无序的。

2018年11月3日-底层服务软件缺陷修复

2018年11月3日10:00到19:00将对MaxCompute底层服务进行软件缺陷修复。正常情况下，您不会感知到升级过程。极端情况下，可能会产生约20分钟的作业运行失败或作业提交失败。一旦出现作业失败，请您重新提交失败任务或提交工单反馈。

2018年10月31日-SQL外部表功能开启计费

从2018年10月31日开始，MaxCompute SQL[外部表](#)功能开始计费。采用的计费标准为一次SQL计算费用=计算输入数据量*SQL复杂度*SQL价格。SQL价格是0.03元/GB/复杂度，复杂度系数为1。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。

2018年10月16日-Python UDF全面开放

从2018年10月16日开始，MaxCompute公共云环境下的Python UDF使用权限将全面开放，新增的project默认可用Python UDF，具体的Python UDF使用方式请参考文档[Python UDF](#)。更多PyODPS的内容介绍请参考相关[文档](#)。

2018年9月29日-SQL语言升级

MaxCompute Sprint29中发布的新功能如下。

- UDJ

为了满足用户很多跨表操作的需求，MaxCompute S29版本在UDF框架中引进了新的扩展机制：**UDJ(User Defined Join)**。UDJ提供了让用户自定义Join实现逻辑的接口。如何定义和使用UDJ，详情请参考文档[UDJ](#)。

- UDT

为了提升SQL语言编译过程的易用性与语言的表达能力，比如使用Select Transform可以不创建Function甚至不上传资源的情况下执行其他语言的脚本，MaxCompute2.0基于新一代的SQL引擎增加了**UDT (User Defined Type)** 新功能。

UDT功能允许用户在SQL中直接引用第三方语言的类或者对象，获取其数据内容或者调用其方法。如在SQL中直接写java的表达式，并可以引用JDK中的类。如何使用UDT，详情请参考文档[UDT](#)。



说明：

- 目前UDT只支持Java语言。
- 所有的运算符都是MaxCompute SQL的语义，不是UDT的语义。
- 目前UDT不能用作shuffle key：包括join，group by，distribute by，sort by，order by，cluster by 等结构的key。
- DDL不支持UDT，也不能将UDT对象insert到表中；屏显的最终结果也不能是UDT类型。

- 支持OUTER LATERAL VIEW

MaxCompute Sprint30扩展了Lateral View功能，支持Outer Lateral View，当table function不输出任何一行时，对应的输入行在lateral view结果中依然保留，且所有table function输出列为null。关于Lateral View介绍请参考文档DML语句的[Lateral View](#)部分。

- 窗口函数支持WINDOW_CLAUSE

为了避免多次调用window函数窗口时，SQL里反复写上OVER语句会显得冗余，MaxCompute在该版本中支持WINDOW语句预定义窗口。关于窗口函数介绍请参考文档[窗口函数](#)。

- 支持SCALAR SUBQUERY

子查询支持SCALAR SUBQUERY，即当SUBQUERY的输出结果为单行单列的时候，可以当做标量来使用。具体使用介绍请参考文档DML语句[子查询](#)中的“SCALAR SUBQUERY”部分。

2018年9月26日-元数据仓库组件优化

为了提升您的MaxCompute使用性能，系统即将对元数据仓库组件进行优化，届时您会提前收到以下信息：

尊敬的阿里云用户您好，大数据计算服务MaxCompute为向您提供更好的产品性能和稳定性，将于北京时间2018年9月26日16:00 - 20:00对底层元数据仓库组件进行优化，您的应用可能出现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。感谢您的理解和支持！谢谢。有任何问题，可随时通过企业钉钉群或工单联系反馈。

2018年-MaxCompute例行优化

为了提升您的MaxCompute使用性能，系统每个月将会对文件数占用较多的project进行小文件合并，届时您将会提前收到以下信息：

尊敬的客户，您好：为了提升您的MaxCompute使用性能，我们计划于XX年XX月XX日对文件数占用较多的project（只会涉及到5天内未被访问分区）进行小文件合并，此操作是后台自动做，用户无需参与。Merge执行过程中，作业读取正在被Merge的表会有小概率失败，业务上重试即可成功，届时请关注相关业务，如有异常随时通过企业钉钉群进行反馈。



说明：

如果您收到报错：ODPS-0010000:System internal error - PanguOpen4Read4R，说明作业在读取正在被Merge的表，请在业务上重试。

2018年9月17日支持setproject 2.0新类型和兼容hive模式

为了提升MaxCompute 2.0的使用体验，支持对project级别进行新类型和兼容hive模式打开，即对原来session级别的set odps.sql.type.system.odps2=true|false和 set odps.sql.hive.compatible=true|false 两个功能，升级为支持setproject级别。project的Owner可根据需要对project进行设置，命令为：

```
- setproject odps.sql.type.system.odps2=false|true
- setproject odps.sql.hive.compatible=false|true
```

对setproject的详细说明请参见：[其他操作](#)。



说明：

打开project级别的新类型的主要影响：

- 某些隐式类型转换会被禁用，包括string -> bigint, string -> datetime, double->bigint, decimal -> double, decimal -> bigint都是有精度损失或者可能报错的。用户依旧可以用cast来做强制转换。
- 常量类型会变化。单独一个 整形常量，如 123，在旧类型下是bigint类型，在新类型下是int类型。
- udf resolve结果可能变化，比如udf包含bigint和int两个重载。旧类型下一定走bigint的重载，而新类型下可能会被解析到int的重载。
- session级别会覆盖project级别设置。

2018年5月底—NewSQL开启计费

MaxCompute New SQL采用的计费标准为一次SQL计算费用=计算输入数据量*SQL复杂度*SQL价格。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。

2018年1月16日流量切换公告

1月16日14~17点对MaxCompute前端、tunnel服务进行流量切换，如果出现服务连接超时或访问失败，可提交[工单进行咨询](#)。

2018年1月15日开放label、Package安全模型

2018年1月15日，MaxCompute的安全模型开发label和Package，即[列级别访问控制](#)和[跨项目空间的资源分享](#)。

2018年1月10日禁止Full scan功能开放

2018年1月10日 20点新创建的项目（project）默认禁止分区表全表扫描，即对该项目的分区表执行全表扫描语句时会返回失败，必须指定分区条件，若实在需要，可以在对分区表全表扫描的sql语句前加一个set语句set odps.sql.allow.fullscan=true;，执行的时候，set语句和SQL语句一起提交执行。

2017年11月28-29日内部维护升级公告

由于MaxCompute内部维护升级出现问题，2017-11-28 22:00到2017-11-29 18:40期间的计算计量和存储计量未及时出账，相应账单会延后推送，但不影响最终计费结果，由此给您造成的不便深感抱歉。

2017年11月1日—华北2（北京）Region开服

2017年11月1日，阿里云数加>MaxCompute 华北2（北京）数据中心将正式开服售卖。

售价与华东2、华南1一致，详情请参见[计量计费文档](#)。

开通时，选择什么Region您需要考虑的最主要因素是MaxCompute与其他阿里云产品之间的关系。

不同区域数据是不能互通，仅针对MaxCompute而言，假如您既购买华东2又购买华北2的服务，那么MaxCompute之间的数据不互通。

如果您的云服务器ECS不在北京Region，若想访问连接北京的MaxCompute，需跨Region进行访问。跨Region访问的各种服务连接请参见[访问域名和数据中心](#)。

2017年10月11日—预付费支持自动续费

2017年10月11日，MaxCompute的预付费实例开始支持自动续费，可在[阿里云控制台的续费管理](#)中心操作。

MaxCompute续费的操作介绍请参见[续费管理](#)，更多自动续费的介绍请参见[自动续费功能](#)。

2017年9月19日—香港Region开服

2017年9月19日，阿里云数加>MaxCompute香港（香港）数据中心将正式开服售卖。

售价与华东2、华南1一致，详情请参见[计量计费文档](#)。

开通时，选择什么Region您需要考虑的最主要因素是MaxCompute与其他阿里云产品之间的关系。

不同区域数据是不能互通，仅针对MaxCompute而言，假如您既购买华东2又购买香港的服务，那么MaxCompute之间的数据不互通。

如果您的云服务器ECS不在香港，若想访问连接香港的MaxCompute，需跨Region进行访问。跨Region访问的各种服务连接请参见[访问域名和数据中心](#)。

2017年9月7日—华南1（深圳）Region开服

2017年9月7日，阿里云数加>MaxCompute华南1（深圳）数据中心将正式开服售卖，这是数加>MaxCompute在国内开服的第二个区域。

关于售价

华南1区域价格与华东2一致，主要分三部分进行收费：存储、计算和下载，其中计算（指SQL和MR计算任务）分预付费、按量后付费两种模式，存储和下载都是按量后付费。做预算的具体的售价信息请参见[官网定价页](#)或[计量计费文档](#)。

关于开通

请您确保云账号是实名认证的账号，在开通购买页面进行区域选择时，注意选择华南1。那么什么场景适合选择区域是华南1呢？

选择地域时，您需要考虑的最主要因素是MaxCompute与其他阿里云产品之间的关系，示例如下：

- ECS在华南1区域。ECS访问、下载华南1的MaxCompute数据都可以通过内网，快速又避免跨Region走公网下载，从而产生费用。
- 数据在华南1区域。如RDS、OSS、TableStore等在华南1区，那么在和华南1的MaxCompute进行数据传输时可以走内网，避免跨Region配置阿里云内网或VPC网络时不保证连通性的情况。

关于不同区域数据是否互通

仅针对MaxCompute而言，假如您既购买华东2又购买华南1的服务，那么MaxCompute之间的数据不互通。

如果您的云服务器ECS不在华南1，若想访问连接华南1的MaxCompute，需跨Region进行访问。

跨Region访问的各种服务连接请参见[访问域名和数据中心](#)。

关于数据跨Region迁移

若您原来已经开通购买华东2区域，目前想使用华南1区域，所以希望将华东2的所有数据迁移到华南1。建议您通过以下方式进行迁移：

通过配置数据集成任务进行数据同步，如在华南1区域创建好的项目中，配置数据集成任务，来读取华东2区域的表，写入到华南1区域的表中。如果遇到问题可以提工单进行咨询。



说明：

如果您对新服购买有疑惑，可通过工单咨询或加入MaxCompute购买钉钉群（群号：11782920）进行咨询。

2017年8月16日-MR开启计费

MaxCompute MapReduce任务开启按量计费。计费标准如下：

MR任务当日计算费用=当日总计算时 * 0.46元（人民币）

一个MR任务一次执行成功的计算时=任务运行时间（小时）*任务调用的core数量。

如一个MR任务一次执行成功是调用了100core并花费0.5小时，那么本次执行计算时为：0.5 小时 × 100core= 50个计算时。

详情请参见[计量计费说明](#)。

2017年7月一开始分批升级MaxCompute2.0



说明：

2017年9月底已经全面升级开放MaxCompute2.0，MaxCompute2.0 NewSQL及非结构化（外表）处理能力公测阶段，不收取费用，具体收费时间等待官方通知。

MaxCompute2.0已经正式开始，升级后的MaxCompute2.0完全拥抱开源生态，支持更多的语言功能，带来更快的运行速度，同时新版本会执行更严格的语法检测。升级的具体时间会在分批升级时另行通知。

升级后需要对部分SQL语法做出调整，详情请参见[修改不兼容SQL实战](#)进行修改。

本次MaxCompute2.0产品升级以下功能：

- 全新的SQL引擎

更快的SQL执行引擎：降低企业大数据分析成本，SQL执行效率更高。

- 生态兼容

MaxCompute显著提升了针对Hive等开源产品的兼容性，直接兼容Hive的UDF/UDAF/UDTF，为Hive开发的大部分UDF/UDAF/UDTF可不经修改，直接在MaxCompute中运行。

详情请参见[INSERT操作](#)、[SELECT操作](#)、[数据类型](#)、[日期函数](#)、[数学函数](#)、[字符串函数](#)、其他函数和[JAVA UDF](#)等相关文档。

- 非结构化处理能力

通过MaxCompute SQL直接处理OSS/TableStore（OTS）数据，从而处理音频、视频、图像、气象等非结构化数据以及K-V类型的数据。

详情请参见[前言](#)、[访问OTS非结构化数据](#)和[访问OSS非结构化数据](#)。

- MaxCompute Studio

MaxCompute编译器基于MaxCompute2.0全新自主研发的SQL引擎，尤其配合使用MaxCompute Studio，提供了丰富的错误提示、警告的功能及作业排队展示。详情请参见[巧用MaxCompute编译器的错误和警告](#)。

2017年06月01日

尊敬的阿里云客户：

MaxCompute目前只针对SQL作业进行收费，MapReduce作业近期将开启收费。

MaxCompute近期将进行2.0版本升级，升级过程中会逐批进行，本次版本变更将带来一些新特性：

- **SQL执行引擎**

降低大数据分析成本，SQL执行效率更高。

- **非结构化处理能力**

通过MaxCompute SQL直接处理OSS/TableStore (OTS) 数据，从而处理音频、视频、图像、气象等非结构化数据以及K-V类型数据。

- **生态兼容**

MaxCompute MapReduce SDK高度兼容了开源Hadoop MapReduce的使用场景。

- **基于IntelliJ的本地开发插件**

本地对MaxCompute作业进行开发，调试。

- **实时数据上传**

新增DataHub模块，帮助企业数据实时上传。

- **新型算法平台**

基于GPU的深度学习算法框架，在线预测服务等。

如果您在系统升级过程中遇到问题，请及时提交工单联系我们。

2016年09月16日

尊敬的阿里云客户：

MaxCompute在11月1日进行了计费变更。本次变更升级存储及计算两个方面：

- 升级后，修复部分SQL作业无法计费的问题。
- 升级后，解决部分场景下MaxCompute存储费用存在的问题。升级后，存储将按照实际情况收费。

以上两处之前均不存在向您多收取费用的情况，对于历史作业不做费用补收。您可以通过计量计费说明中获取账单详情的相关指引，从阿里云官网获取每天详细的计量计费信息。

目前MaxCompute只有SQL作业收费（不包括UDF），UDF/MapReduce/Graph/PAI等作业仅是公测状态。具体收费计划请关注官网公告。

感谢您长期以来对MaxCompute产品的支持。