一、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EB级数据仓库解决方案。

随着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（TB、PB、EB）级别。MaxCompute致力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的计算模型。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，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。

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

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、UDF（Java/Python）、Graph、基于DAG的处理、交互式、内存计算、机器学习等计算类型及MPI迭代类算法。简化了企业大数据平台的应用架构。

强数据安全

MaxCompute已稳定支撑阿里全部数据仓库业务9年以上，提供多层沙箱防护、细粒度权限管理及监控。

MaxCompute通过了独立的第三方审计师针对阿里云对AICPA可信服务标准中关于安全性、可用性和机密性原则符合性描述的审计。审计报告请参见[SOC 3报告](https://www.alibabacloud.com/zh/trust-center/soc)。

低成本

与企业自建专有云相比，MaxCompute的计算存储更高效，可以降低30%~50%的采购成本。

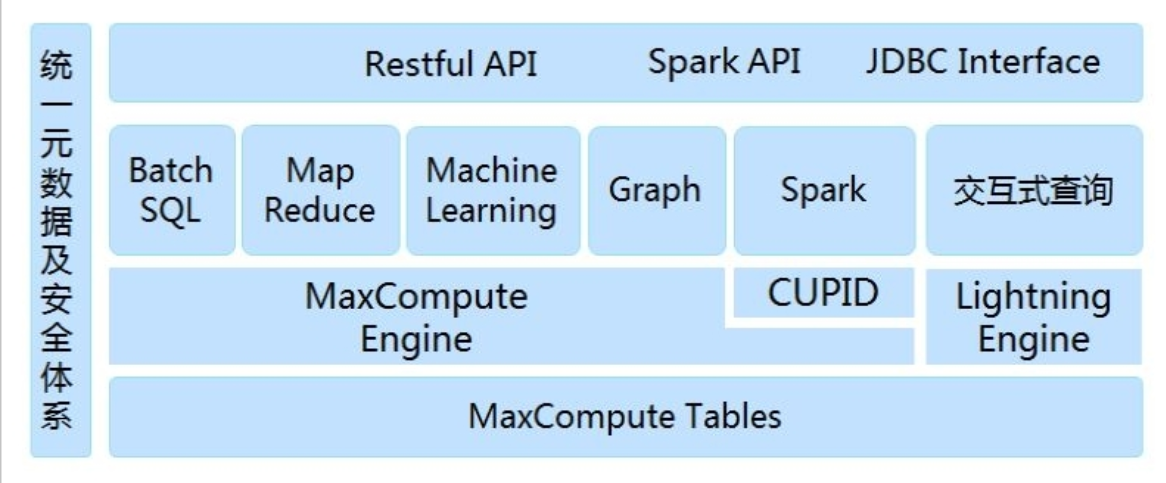
免运维

基于MaxCompute的Serverless无服务器的设计思路，用户只需关心作业和数据，而无需关心底层分布式架构及运维。

极致弹性扩展

MaxCompute提供按量付费模式下的作业级别的资源管理。用户无需受困于资源扩展难题，系统会自动扩展计算、存储、网络等资源，最大程度地节省成本。

系统架构：



产品功能：

1.数据通道

①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，提供对应的命令实现本地文件与服务数据的互通。

②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。

2. 计算及分析任务

MaxCompute支持多种计算模型，详情如下：

①SQL：MaxCompute以表的形式存储数据，支持多种数据类型版本说明，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。

（MaxCompute SQL不支持事务、索引，也不支持Update或Delete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语句无缝迁移至MaxCompute中。详情请参见与其他SQL语法的差异。

MaxCompute主要用于100GB以上规模的数据计算，因此MaxCompute SQL最快支持在分钟或秒钟级别完成查询返回结果，但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。）

②UDF：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。

③MapReduce：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。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。

④Graph：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。

⑤Spark on MaxCompute：Spark on MaxCompute是阿里云开发的大数据分析引擎，为您提供大数据处理能力。详情请参见MaxCompute Spark概述。