

Alibaba Cloud MaxCompute

產品簡介

檔案版本：20190415

目錄

1 什麼是MaxCompute.....	1
2 通告.....	4

1 什麼是MaxCompute

大資料計算服務（MaxCompute，原名ODPS）是一種快速、完全託管的GB/TB/PB級資料倉儲解決方案。MaxCompute為您提供了完善的資料匯入方案以及多種經典的分散式運算模型，能夠更快速的解決海量資料計算問題，有效降低企業成本，並保障資料安全。

同時，DataWorks和MaxCompute關係緊密，DataWorks為MaxCompute提供了一站式的資料同步、任務開發、資料工作流程開發、資料管理和資料營運等功能，詳情請參見[DataWorks](#)。

MaxCompute主要服務於批量結構化資料的儲存和計算，可以提供海量資料倉儲的解決方案以及針對大資料的分析建模服務。隨著社交資料收集手段的不斷豐富及完善，越來越多的行業資料被積累下來。資料規模已經增長到了傳統軟體行業無法承載的海量資料（百GB、TB乃至PB）層級。

在分析海量資料情境下，由於單台伺服器的處理能力限制，資料分析者通常採用分散式運算模式。但分布式的計算模型對資料分析人員提出了較高的要求，且不易維護。使用分布式模型，資料分析人員不僅需要瞭解業務需求，同時還需要熟悉底層計算模型。MaxCompute的目的是為您提供一種便捷的分析處理海量資料的手段，您可以不必關心分散式運算細節，便可達到分析大資料的目的。



說明：

MaxCompute已經在阿里巴巴集團內部得到大規模應用，例如大型互連網企業的資料倉儲和BI分析、網站的日誌分析、電子商務網站的交易分析、使用者特徵和興趣挖掘等。

MaxCompute學習路徑

您可以通過[MaxCompute####](#)快速瞭解MaxCompute的相關概念、基礎操作、進階操作等。

產品優勢

- 大規模計算儲存

MaxCompute適用於100GB以上規模的儲存及計算需求，最大可達EB層級。

- 多種計算模型

MaxCompute支援SQL、MapReduce、Graph等計算類型及MPI迭代類演算法。

- 強資料安全

MaxCompute已穩定支撐阿里全部離線分析業務7年以上，提供多層沙箱防護及監控。

- 低成本

與企業自建專有雲相比，MaxCompute的計算儲存更高效，可以降低20%-30%的採購成本。

功能概述

· 資料通道

- 支援批量、歷史資料通道

TUNNEL是MaxCompute為您提供的Data Transmission Service，提供高並發的離線資料上傳下載服務。支援每天TB/PB層級的資料匯入匯出，特別適合於全量資料或歷史資料的大量匯入。Tunnel 為您提供Java編程介面，並且在MaxCompute的用戶端工具中，有對應的命令實現本地檔案與服務資料的互連。

- 即時、增量資料通道

針對即時資料上傳的情境，MaxCompute提供了延遲低、使用方便的DataHub服務，特別適用於增量資料的匯入。DataHub還支援多種資料轉送外掛程式，例如Logstash、Flume、Fluentd、Sqoop等，同時支援Log Service Log Service中的投遞日誌到MaxCompute，進而使用DataWorks進行日誌分析和挖掘。

· 計算及分析任務

MaxCompute支援多種計算模型，詳情如下。

- **SQL**：MaxCompute只能以表的形式儲存資料，並對外提供了SQL查詢功能。您可以將MaxCompute作為傳統的資料庫軟體操作，但其卻能處理TB、PB層級的海量資料。



说明:

- MaxCompute SQL不支援事務、索引及Update/Delete等操作。
- MaxCompute的SQL文法與Oracle、MySQL有一定差別，您無法將其他資料庫中的SQL語句無縫遷移到MaxCompute上來。
- 在使用方式上，MaxCompute SQL最快可以在分鐘、乃至秒層級完成查詢，無法在毫秒層級返回結果。
- MaxCompute SQL的優點是學習成本低，您不需要瞭解複雜的分散式運算概念。如果您具備資料庫操作經驗，便可快速熟悉MaxCompute SQL的使用。

- **UDF**：即使用者自訂函數。

MaxCompute提供了很多####來滿足您的計算需求，同時您還可以通過建立自訂函數來滿足不同的計算需求。

- **MapReduce**：MaxCompute MapReduce是MaxCompute提供的Java MapReduce編程模型，它可以簡化開發流程，更為高效。您若使用MaxCompute MapReduce，需要對分

散式運算概念有基本瞭解，並有相對應的編程經驗。MaxCompute MapReduce為您提供Java編程介面。

- **Graph**: MaxCompute提供的Graph功能是一套面向迭代的圖計算處理架構。圖計算作業使用圖進行建模，圖由點（Vertex）和邊（Edge）組成，點和邊包含權值（Value）。通過迭代對圖進行編輯、演化，最終求解出結果，典型應用：[PageRank](#)、[#####](#)、[K-#####](#)等。

- SDK

SDK是MaxCompute提供給開發人員的工具包，詳情請參見[SDK##](#)。

- 安全

MaxCompute提供了功能強大的安全服務，為您的資料安全提供保護，詳情請參見[####](#)。

後續步驟

現在，您已經學習了MaxCompute的產品優勢、功能特性等相關簡介，您可以繼續學習下一個教程。在該教程中您將瞭解MaxCompute的相關收費情況，詳情請參見[####](#)。

2 通告

2018年9月26日-元資料倉儲組件最佳化

為了提升您的MaxCompute使用效能，系統即將對元資料倉儲組件進行最佳化，屆時您會提前收到以下資訊：

尊敬的阿里雲使用者您好，大資料計算服務MaxCompute為向您提供更好的產品效能和穩定性，將於北京時間2018年9月26日16:00 - 20:00對底層元資料倉儲組件進行最佳化，您的應用可能出現1分鐘左右任務無法提交、運行中任務失敗的情況。極端情況下，應用不可用時間將延長至30分鐘。請您盡量避免在遷移視窗提交作業。感謝您的理解和支援！謝謝。有任何問題，可隨時通過企業釘釘群或工單聯絡反饋。

2018年-MaxCompute例行最佳化

為了提升您的MaxCompute使用效能，系統每個月將會對檔案數佔用較多的project進行小檔案合并，屆時您將會提前收到以下資訊：

尊敬的客戶，您好：為了提升您的MaxCompute使用效能，我們計劃於XX年XX月XX日對檔案數佔用較多的project（只會涉及到5天內未被訪問分區）進行小檔案合并，此操作是後台自動做，使用者無需參與。Merge執行過程中，作業讀取正在被Merge的表會有小機率失敗，業務上重試即可成功，屆時請關注相關業務，如有異常隨時通過企業釘釘群進行反饋。



说明：

如果您收到報錯：ODPS-0010000:System internal error - PanguOpen4Read4R，說明作業在讀取正在被Merge的表，請在業務上重試。

2018年9月17日支援setproject 2.0新類型和相容hive模式

為了提升MaxCompute 2.0的使用體驗，支援對project層級進行新類型和相容hive模式開啟，即對原來session層級的set odps.sql.type.system.odps2=true|false和 set odps.sql.hive.compatible=true|false 兩個功能，升級為支援setproject層級。project的Owner可根據需要對project進行設定，命令為：

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

對setproject的詳細說明請參見：[####](#)。



说明：

開啟project層級的新類型的主要影響：

- 某些隱式類型轉換會被禁用，包括string -> bigint, string -> datetime, double->bigint, decimal -> double, decimal -> bigint都是有精度損失或者可能報錯的。使用者依舊可以用cast來做強制轉換。
- 常量類型會變化。單獨一個整形常量，如123，在舊類型下是bigint類型，在新類型下是int類型。
- udf resolve結果可能變化，比如udf包含bigint和int兩個重載。舊類型下一定走bigint的重載，而新類型下可能會被解析到int的重載。
- session層級會覆蓋project層級設定。

2018年5月底—NewSQL開啟計費

MaxCompute New SQL採用的計費標準為一次 SQL 計算費用=計算輸入資料量* SQL 複雜度* SQL 價格。當天的所有計量資訊在第二天做一次性匯總收費，並直接體現在您的賬戶賬單中。

2018年1月16日流量切換公告

1月16日14~17點對MaxCompute前端、tunnel服務進行流量切換，如果出現服務連線逾時或訪問失敗，可提交#####。

2018年1月15日開放label、Package安全模型

2018年1月15日，MaxCompute的安全模型開發label和Package，即#####和#####。

2018年1月10日禁止Full scan功能開放

2018年1月10日 20點新建立的項目（project）預設禁止分區表全表掃描，即對該項目的分區表執行全表掃描語句時會返回失敗，必須指定分區條件，若實在需要，可以在對分區表全表掃描的sql語句前加一個set語句 `set odps . sql . allow . fullscan = true ;`，執行的時候，set語句和SQL語句一起提交執行。

2017年11月28-29日內部維護升級公告

由於MaxCompute內部維護升級出現問題，2017-11-28 22:00到2017-11-29 18:40期間的計算計量和儲存計量未及時出賬，相應賬單會延後推送，但不影響最終計費結果，由此給您造成的不便深感抱歉。

2017年11月1日—華北2（北京）Region開服

2017年11月1日，阿里雲數加>MaxCompute 華北2（北京）資料中心將正式開服售賣。

售價與華東2、華南1一致，詳情請參見#####。

開通時，選擇什麼Region您需要考慮的最主要因素是MaxCompute與其他阿里雲產品之間的關係。

不同地區資料是不能互連，僅針對MaxCompute而言，假如您既購買華東2又購買華北2的服務，那麼MaxCompute之間的資料不互連。

如果您的Elastic Compute Service不在北京Region，若想訪問串連北京的MaxCompute，需跨Region進行訪問。跨Region訪問的各種服務串連請參見#####。

2017年10月11日—預付費支援自動續約

2017年10月11日，MaxCompute的預付費執行個體開始支援自動續約，可在#####中心操作。

MaxCompute續約的操作介紹請參見####。

2017年9月19日—香港Region開服

2017年9月19日，阿里雲數加>MaxCompute香港（香港）資料中心將正式開服售賣。

售價與華東2、華南1一致，詳情請參見#####。

開通時，選擇什麼Region您需要考慮的最主要因素是MaxCompute與其他阿里雲產品之間的關係。

不同地區資料是不能互連，僅針對MaxCompute而言，假如您既購買華東2又購買香港的服務，那麼MaxCompute之間的資料不互連。

如果您的Elastic Compute Service不在香港，若想訪問串連香港的MaxCompute，需跨Region進行訪問。跨Region訪問的各種服務串連請參見#####。

2017年9月7日—華南1（深圳）Region開服

2017年9月7日，阿里雲數加>MaxCompute華南1（深圳）資料中心將正式開服售賣，這是數加>MaxCompute在國內開服的第二個地區。

關於售價

華南1地區價格與華東2一致，主要分三部分進行收費：儲存、計算和下載，其中計算（指SQL和MR計算任務）分預付費、按量後付費兩種模式，儲存和下載都是按量後付費。做預算的具體的售價資訊請參見#####。

關於開通

請您確保雲帳號是實名認證的帳號，在開通購買頁面進列區域選擇時，注意選擇華南1。那麼什麼情境適合選擇地區是華南1呢？

選擇地區時，您需要考慮的最主要因素是MaxCompute與其他阿里雲產品之間的關係，樣本如下：

- ECS在華南1地區。ECS訪問、下載華南1的MaxCompute資料都可以通過內網，快速又避免跨Region走公網下載，從而產生費用。
- 資料在華南1地區。如RDS、OSS、TableStore等在華南1區，那麼在和華南1的MaxCompute進行資料轉送時可以走內網，避免跨Region配置阿里雲內網或VPC網路時不保證連通性的情況。

關於不同地區資料是否互連

僅針對MaxCompute而言，假如您既購買華東2又購買華南1的服務，那麼MaxCompute之間的資料不互連。

如果您的Elastic Compute Service不在華南1，若想訪問串連華南1的MaxCompute，需跨Region進行訪問。跨Region訪問的各種服務串連請參見#####。

關於資料跨Region遷移

若您原來已經開通購買華東2地區，目前想使用華南1地區，所以希望將華東2的所有資料移轉到華南1。建議您通過以下方式進行遷移：

通過配置Data Integration任務進行資料同步，如在華南1地區建立好的項目中，配置Data Integration任務，來讀取華東2地區的表，寫入到華南1地區的表中。如果遇到問題可以提工單進行諮詢。



说明：

如果您對新服購買有疑惑，可通過工單諮詢或加入MaxCompute購買釘釘群（群號：11782920）進行諮詢。

2017年8月16日-MR開啟計費

MaxCompute MapReduce任務開啟隨用隨付。計費標準如下：

MR 任務當日計算費用=當日總計算時 * 0.46 元（人民幣）

一個MR任務一次執行成功的計算時=任務已耗用時間（小時）*任務調用的core數量。

如一個MR任務一次執行成功是調用了100core並花費0.5小時，那麼本次執行計算時為：0.5 小時 × 100core= 50個計算時。

詳情請參見#####。

2017年7月—開始分批升級MaxCompute2.0



說明:

2017年9月底已經全面升級開放MaxCompute2.0, MaxCompute2.0 NewSQL及非結構化（外表）處理能力公測階段，不收取費用，具體收費時間等待官方通知。

MaxCompute2.0已經正式開始，升級後的MaxCompute2.0完全擁抱開源生態，支援更多的語言功能，帶來更快的運行速度，同時新版本會執行更嚴格的文法檢測。升級的具體時間會在分批升級時另行通知。

升級後需要對部分SQL文法做出調整，詳情請參見[#####SQL##](#)進行修改。

本次MaxCompute2.0產品升級以下功能：

- 全新的SQL引擎

更快的SQL執行引擎：降低企業大資料分析成本，SQL執行效率更高。

- 生態相容

MaxCompute顯著提升了針對Hive等開源產品的相容性，直接相容Hive的UDF/UDAF/UDTF，為Hive開發的大部分UDF/UDAF/UDTF可不經修改，直接在MaxCompute中運行。

詳情請參見[INSERT##](#)、[SELECT##](#)、[####](#)、[####](#)、[####](#)、[####](#)、其他函數和[JAVA UDF](#)等相關文檔。

- 非結構化處理能力

通過MaxCompute SQL直接處理OSS/TableStore（OTS）資料，從而處理音頻、視頻、映像、氣象等非結構化資料以及K-V類型的資料。

詳情請參見[##](#)、[##OTS#####](#)和[##OSS#####](#)。

- MaxCompute Studio

MaxCompute編譯器基於MaxCompute2.0全新自主研發的SQL引擎，尤其配合使用MaxCompute Studio，提供了豐富的錯誤提示、警告的功能及作業排隊展示。詳情請參見[##MaxCompute#####](#)。

2017年06月01日

尊敬的阿里雲客戶：

MaxCompute目前只針對SQL作業進行收費，MapReduce作業近期將開啟收費。

MaxCompute近期將進行2.0版本升級，升級過程中會逐批進行，本次版本變更將帶來一些新特性：

- SQL執行引擎

降低大資料分析成本，SQL執行效率更高。

- 非結構化處理能力

通過MaxCompute SQL直接處理OSS/TableStore（OTS）資料，從而處理音頻、視頻、映像、氣象等非結構化資料以及K-V類型資料。

- 生態相容

MaxCompute MapReduce SDK高度相容了開源Hadoop MapReduce的使用情境。

- 基於Intelij的本地開發外掛程式

本地對MaxCompute作業進行開發，調試。

- 即時資料上傳

新增DataHub模組，協助企業資料即時上傳。

- 新型演算法平台

基於GPU的深度學習演算法架構，線上預測服務等。

如果您在系統升級過程中遇到問題，請及時提交工單聯絡我們。

2016年09月16日

尊敬的阿里雲客戶：

MaxCompute在11月1日進行了計費變更。本次變更升級儲存及計算兩個方面：

- 升級後，修複部分SQL作業無法計費的問題。
- 升級後，解決部分情境下MaxCompute儲存費用存在的問題.升級後，儲存將按照實際情況收費。

以上兩處之前均不存在向您多收取費用的情況，對於歷史作業不做費用補收。您可以通過計量計費說明中擷取賬單詳情的相關指引，從阿里雲官網擷取每天詳細的計量計費資訊。

目前MaxCompute只有SQL作業收費（不包括UDF），UDF/MapReduce/Graph/PAI等作業僅是公測狀態。具體收費計劃請關注官網公告。

感謝您長期以來對MaxCompute產品的支援。