

ALIBABA CLOUD

阿里云

交互式分析Hologres

产品简介

文档版本：20200923

 阿里云

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
 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。
<code>Courier</code> 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。
<i>斜体</i>	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] 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

1.什么是Hologres	05
2.产品优势	06
3.应用场景	07
4.基本概念	09

1.什么是Hologres

Hologres是阿里巴巴自主研发的一款交互式分析产品，兼容PostgreSQL 11协议，与大数据生态无缝连接，支持高并发和低延时地分析处理PB级数据。

随着收集数据的方式不断丰富，企业信息化程度越来越高，企业掌握的数据量呈TB、PB或EB级别增长。同时，数据中台的快速推进，使数据应用主要为数据支撑、用户画像、实时圈人及广告精准投放等核心业务服务。高可靠和低延时地数据服务成为企业数字化转型的关键。

Hologres致力于低成本和高性能地大规模计算型存储和强大的查询能力，为您提供海量数据的实时数据仓库解决方案和实时交互式查询服务。

功能概述

- 加速查询MaxCompute数据

Hologres与MaxCompute在底层无缝连接，您无须移动数据，即可使用标准的PostgreSQL语句查询分析MaxCompute中的海量数据，快速获取查询结果。

- 快速搭建实时数据仓库

Hologres深度集成实时计算Flink版，支持高并发地实时写入和实时查询数据，帮助您快速搭建企业实时数据仓库。

- 数据仓库服务化

Hologres数据仓库的服务化体现如下：

- 支持行存储和列存储两种存储模式。
- 优化了高并发和复杂的查询场景，为PB级数据的查询提供秒级查询响应。
- 支持交互式和服务式的多种查询场景。

- 无缝对接主流BI工具

Hologres兼容PostgreSQL协议，提供JDBC或ODBC接口。您可以将查询到的海量数据轻松对接各种BI工具，实现探索式多维分析，无需迁移数据即可支持更丰富的应用场景。

2. 产品优势

Hologres兼容PostgreSQL生态，支持快速查询分析MaxCompute的数据、实时查询实时写入的数据、联邦分析实时数据与离线数据，帮助您快速搭建企业实时数据仓库。

Hologres具有如下优势：

- 极速响应
 - 支持以亚秒级的响应速度查询PB级数据，支持向量化计算以及列存储智能索引，性能大幅领先开源系统，帮助您实时多维分析业务。
- 快速查询分析MaxCompute的数据
 - 与MaxCompute无缝连接，您无需导入导出即可查询MaxCompute表的数据，快速获取查询结果。
 - 您可以单独查询MaxCompute表的数据，也可以与实时数据进行联合计算，低成本和高时效地构建企业实时数据仓库。
- 高并发地实时写入和查询
 - 支持高并发地实时写入和查询实时数据，写入速度可以达到数亿TPS。
- 计算存储分离架构
 - 采用计算与存储分离的云原生架构，存储资源和计算资源分离部署并独立扩展。
 - 支持动态升降配，您可以根据业务需求灵活地扩容或缩容Hologres资源。资源越多，任务的并发数量越大。
 - 支持行存储和列存储两种存储模式。您可以进行简单的点查询、复杂的多维查询以及即席查询（Ad Hoc）等多样化的查询分析。
- 简单易用
 - 兼容PostgreSQL生态，与大数据计算引擎及智能云研发平台DataWorks无缝连接。您可以使用标准的PostgreSQL语句进行数据开发，无需付出额外的学习成本，即可快速上手。

3.应用场景

Hologres是兼容PostgreSQL生态的实时交互式分析产品，与大数据生态无缝连接，支持实时分析海量实时和离线数据，对接第三方BI工具，实现可视化分析业务。本文为您介绍基于Hologres核心功能的典型应用场景。

实时数据仓库

实时写入业务数据至实时计算，使用ETL (Extract Transformation Load) 方式清洗、转换及整理数据。通过Hologres实时查询并输出数据至第三方分析工具，实时分析数据。典型应用场景如下：

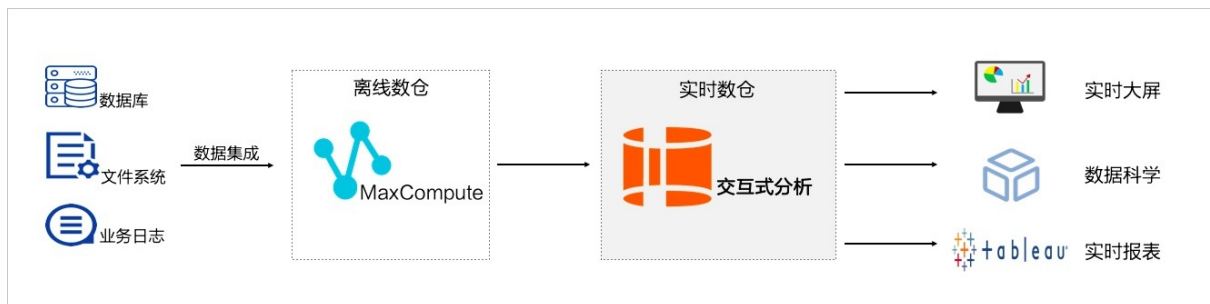
- 数据部门搭建实时数据仓库、展示实时大屏和分析实时Reporting报表。
- 运维和数据应用部门执行实时监控、实时异常检测预警与实时调试。
- 业务部门进行实时风控、实时推荐、实时效果分析和实时训练。



加速查询MaxCompute离线数据

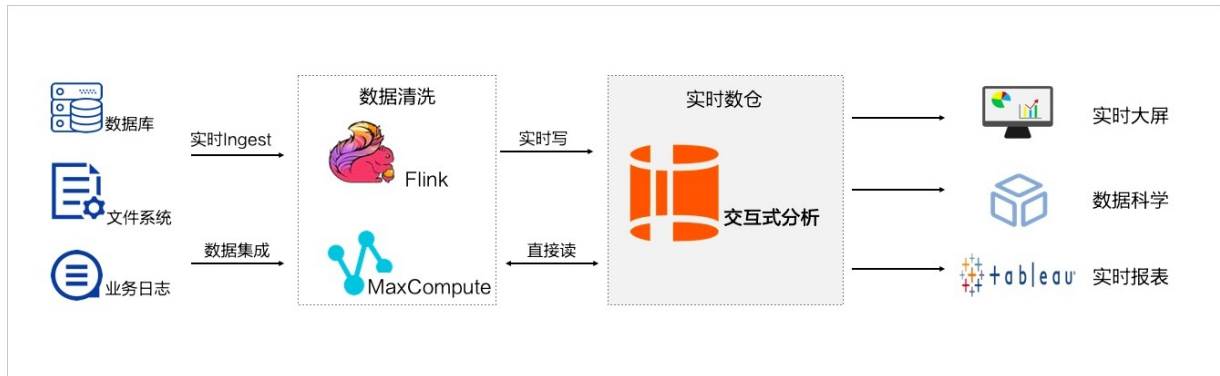
写入业务数据至离线数据仓库MaxCompute，通过Hologres直接加速查询或导入数据至Hologres查询，并对接BI分析工具，实现实时分析离线数据。典型应用场景如下：

- 实时查询MaxCompute离线数据。
- 分析MaxCompute离线数据报表。
- 输出MaxCompute离线数据的在线应用，例如RESTful API的使用。



联邦分析实时数据和离线数据

业务数据分为冷数据和热数据，冷数据存储在离线数据仓库MaxCompute中，热数据存储在Hologres中。Hologres可以联邦分析实时数据和离线数据，对接BI分析工具，快速响应简单查询与复杂查询的业务需求。



4. 基本概念

本文为您介绍Hologres中，账号、用户角色、实例、数据库、Schema、内部表、外部表和分区表的基本概念。

账号

- 阿里云账号

阿里云账号即主账号，用于创建和管理Hologres实例，可以授权给其他用户管理员权限。

主账号支持的表达格式如下表所示。

账号格式	描述	示例
ALIYUN\$<AccountName>@aliyun.com	<AccountName>为主账号的登录名称。	ALIYUN\$company@aliyun.com
<AccountName>@aliyun.com		company@aliyun.com
<Account ID>	主账号的ID。	1344445678xxx

- RAM子账号

RAM子账号即子账号。阿里云账号授予RAM子账号一定权限后，RAM子账号可以在权限范围内创建和管理实例中的数据库和表。

子账号支持的表达格式如下表所示。

账号格式	描述	示例
p4_<Account ID>	<ul style="list-style-type: none"> ◦ p使用小写字母。 ◦ <Account ID>为主账号的ID。 	p4_12345678xxx
<subUserName>@<Account Name>.onaliyun.com	<ul style="list-style-type: none"> ◦ <subUserName>为子账号的登录名称。 ◦ <AccountName>为主账号的登录名称。 ◦ <AccountID>为主账号的ID。 	holouser@company.onaliyun.com
<subUserName>@<Account Name>		
<subUserName>@<Account ID>.onaliyun.com		
RAM\$<subUserName>		
RAM\$<AccountName>:<subUserName>		
RAM\$<Account ID>:<subUserName>		
<subUserName>@<Account ID>		

用户角色

- 超级管理员

系统默认设置购买实例的主账号为超级管理员Superuser。Superuser拥有整个实例的所有权限，例如创建数据库、删除数据库、创建角色以及为角色授权。

- 普通用户

普通用户需要被Superuser授权后才能访问实例，并在权限范围内进行数据开发。普通用户也可以被授权为Superuser。

实例

实例（Instance）是您使用和管理数据库存储服务的实体。您对数据库的所有操作都需要在实例中完成。

您需要购买才能获取实例，购买流程请参见[购买Hologres](#)。

数据库

数据库是按照数据结构来组织、存储和管理数据的仓库。您可以在数据库中创建、更新或删除表，以及执行与函数相关的操作。

购买Hologres实例后，系统自动创建postgres数据库。该数据库分配到的资源较少，仅用于管理，开发实际业务需要新建数据库。新建数据库请参见[CREATE DATABASE](#)。

Schema

Schema是数据库对象的集合，相当于文件系统中的目录。实例中的表和函数等对象存放于各个Schema中。

成功创建数据库后，系统默认创建命名为public的Schema，您对实例对象的所有操作默认都在该Schema中执行。您也可以新建Schema，详情请参见[CREATE SCHEMA](#)。

说明

- Schema不能嵌套Schema。
- 同一个Schema中的对象名称不能重复。
- 不同Schema中的对象名称可以重复。

表

表是由行和列组成的数据存储单元。列的数量和顺序是固定的，每一列拥有一个名称。行的数量是变化的，表示在给定的时刻，表中存储的数据量。

表的每一列都有一个数据类型，表示该列的取值类型，详情请参见[数据类型](#)。

Hologres中的表包括以下两种类型：

- 内部表

指数据存储在Hologres中的表。表中的数据类型可以是Hologres支持的任何一种类型。

- 外部表

指在Hologres中不存储数据，只进行字段映射的表。外部表的数据都是只读的，您在外部表中不能执行DML语句或创建索引。

您可以在Hologres中创建外部表，加速查询外部数据源的数据，例如查询MaxCompute的数据。

Hologres中创建外部表请参见[CREATE FOREIGN TABLE](#)。

分区表

分区表是根据键值属性划分父表为许多小的子集，这些子集称为分区。不同类型的数据存放于不同分区，Hologres当前仅支持对单个键值进行列表分区。查询数据时，您可以通过分区键值过滤查询语句，排除无关的分区，从而提升查询速度。

Hologres中创建或删除分区表请参见[DROP PARTITION TABLE](#)。