

Alibaba Cloud

大数据计算服务 使用教程

文档版本：20211223

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> <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.搭建互联网在线运营分析平台 ----- 05
 - 1.1. 业务场景与开发流程 ----- 05
 - 1.2. 环境准备 ----- 06
 - 1.3. 数据准备 ----- 12
 - 1.4. 数据建模与开发 ----- 17
 - 1.4.1. 新建数据表 ----- 17
 - 1.4.2. 设计工作流 ----- 22
 - 1.4.3. 节点配置 ----- 24
 - 1.4.4. 任务提交与测试 ----- 31
 - 1.5. 数据可视化展现 ----- 35
 - 1.6. 数据可视化展示 ----- 45

1.搭建互联网在线运营分析平台

1.1. 业务场景与开发流程

本教程基于大数据时代在线运营分析平台的基础需求，为开发者提供从数据高并发写入存储、便捷高效的数据加工处理到数据分析与展示的全链路解决方案。本教程帮助您了解并操作阿里云的大数据产品，完成在线运营分析平台的搭建。

业务场景

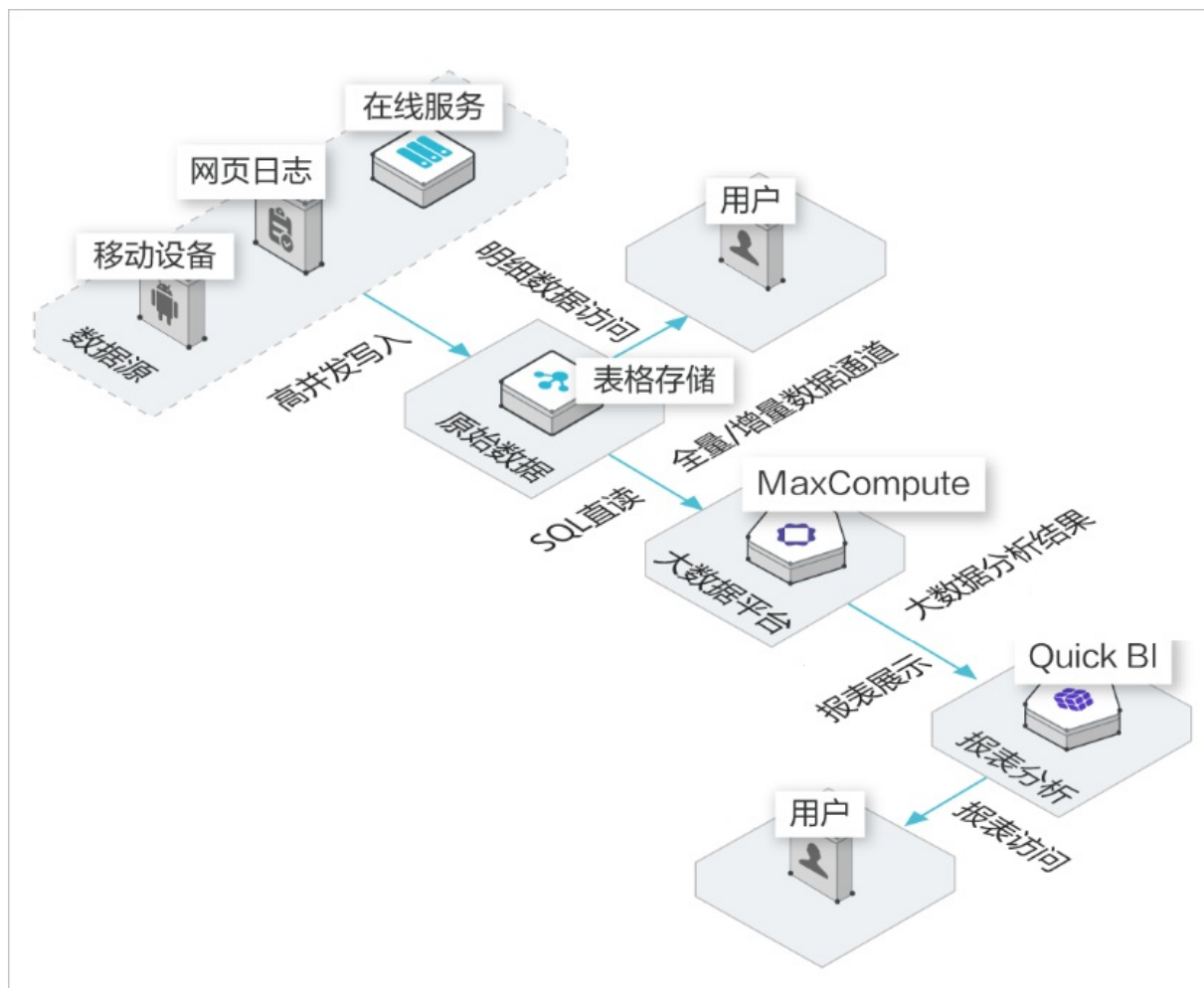
本文的示例基于真实的网站日志数据集，数据来源于某网站上的HTTP访问日志数据。基于这份网站日志，您可以实现如下分析需求：

- 统计并展现网站的PV和UV，并能够按照用户的终端类型（例如，Android、iPad、iPhone和PC等）分别统计。

② 说明 浏览次数（PV）和独立访客（UV）是衡量网站流量的两项最基本指标。用户每打开一个网站页面，记录一个PV，多次打开同一页面PV累计多次。独立访客（UV）是指一天内访问网站的不重复用户数，一天内同一访客多次访问网站只计算一次。

- 统计并展现网站的流量来源地域。

开发流程



本教程涉及的具体开发流程如下：

- 步骤一：环境准备。
- 步骤二：数据准备。
- 步骤三：新建数据表。
- 步骤四：设计工作流。
- 步骤五：节点配置。
- 步骤六：任务提交与测试。
- 步骤七：数据可视化展示。

1.2. 环境准备

本文为您介绍开始本教程前的环境准备工作，需要开通表格存储（Tablestore）、大数据计算服务（MaxCompute）、一站式大数据开发治理平台（DataWorks）和数据可视化分析平台（Quick BI）。

前提条件

- 已注册阿里云账号。如果您还没有注册阿里云账号，请进入[阿里云官网](#)，单击[免费注册](#)，即可进入阿里云账号注册页面创建新的阿里云账号。
- 已实名认证。如果您还没有实名认证，请进入[实名认证](#)页面对账号进行实名认证。

背景信息

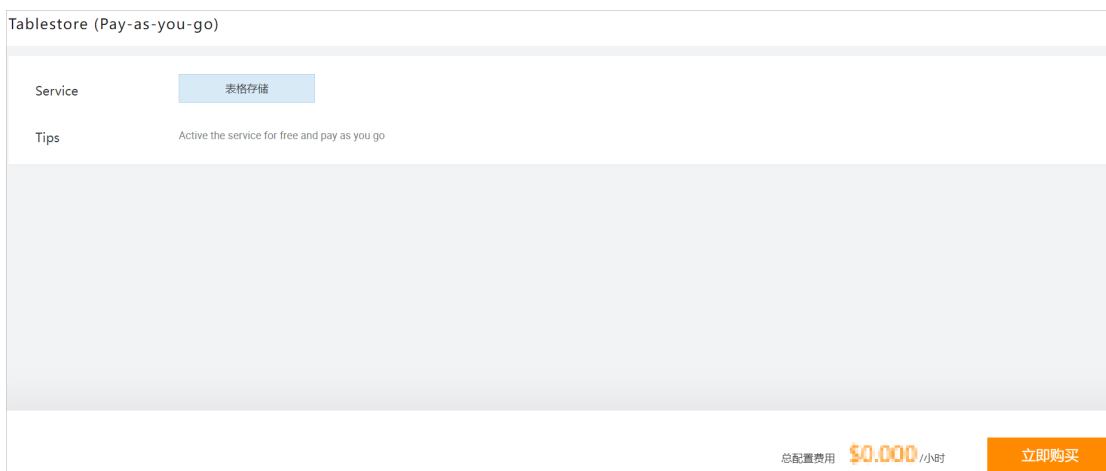
本教程涉及的阿里云产品如下：

- 表格存储Tablestore
- 大数据计算服务MaxCompute
- 数据工场DataWorks
- 智能分析套件Quick BI

操作步骤

1. 创建表格存储实例。

- 进入[表格存储Tablestore产品详情页](#)，单击[免费开通](#)。
- 在Tablestore (Pay-As-You-Go)页面，单击[立即购买](#)。



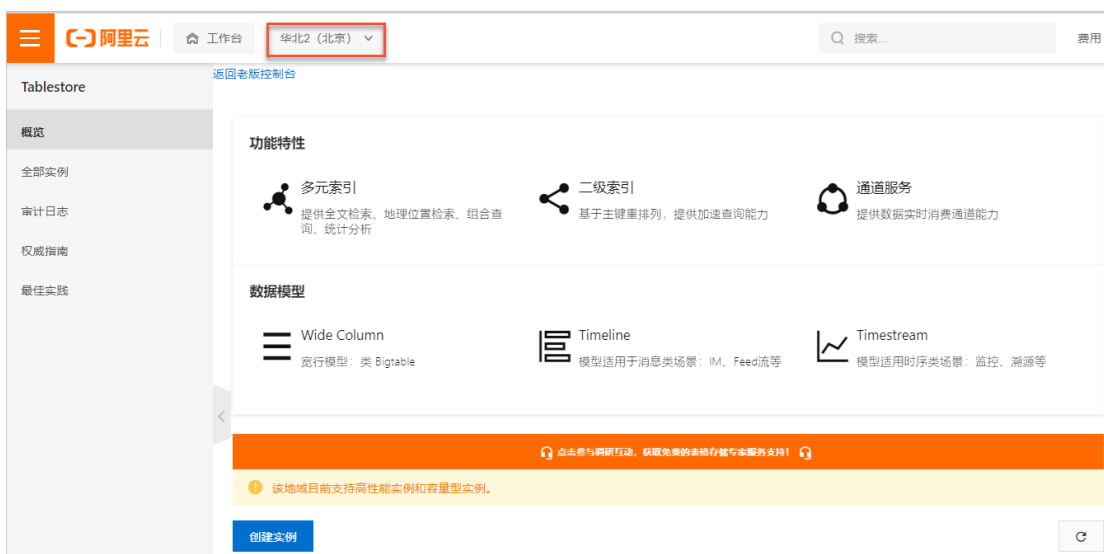
- 在[确认订单](#)页面，勾选[我已阅读并同意Tablestore \(Pay-As-You-Go\)服务协议](#)，并单击[立即开通](#)。



iv. 单击管理控制台。



v. 在管理控制台页面的概述页签，选择地域为华北2（北京），单击创建实例。



? 说明 在本教程中，表格存储服务选择华北2（北京）。您可以根据需要选择其他地域。

vi. 在购买方式对话框选择按量模式，填写实例名称，实例规格请选择容量型实例，单击确定。

购买方式

预留模式

按量模式

创建实例免费，计费按使用量收取。可以开启更多丰富数据访问管理能力。

地区：

华北2（北京）

实例名称：*

workshop-bj-mc

实例规格：

容量型实例

适合离线场景，提供更低成本的数据存储，不适合对访问延时敏感的在线场景。

实例注释：

实例注释最多256个字。

确定

取消

② 说明

- 实例名称在表格存储同一个地域内必须全局唯一，建议您选用自己可辨识且符合规则的名称。
- 实例名称在MaxCompute数据处理中也会被使用，本例中为workshop-bj-mc，关于实例的详细解释请参见[实例](#)。

vii. 完成创建后，单击左侧导航栏全部实例可以看到您刚刚创建的实例，状态为运行中。

2. 开通大数据计算服务MaxCompute。

- 进入[MaxCompute产品详情页](#)，单击立即购买。
- 选择按量计费，选择地域为华东2（上海），规格类型为默认的标准版，单击立即购买。

② 说明 MaxCompute地域与表格存储地域相同可以节省您的流量费用，因此您可以选择地域为华北2（北京）。本教程中MaxCompute地域选择为华东2（上海），以便为您展示跨地域的外部表使用过程。

3. 开通DataWorks。

- 进入[DataWorks产品详情页](#)，单击立即购买。
- 选择地域为华东2（上海），单击立即购买。

② 说明 MaxCompute地域与表格存储地域相同可以节省您的流量费用，因此您也可以选择地域为华北2（北京）。本教程中MaxCompute地域选择为华东2（上海），以便为您展示跨地域访问数据的使用过程。

4. 创建DataWorks工作空间。

- i. 进入DataWorks工作空间列表，选择地域为华东2（上海），单击创建工作空间。



- ii. 在创建工作空间面板中，填写基本配置相关内容，单击下一步。

创建工作空间

1 基本配置 ————— 2 选择引擎 ————— 3 引擎详情

当前地域

华东2（上海）

基本信息

* 工作空间名称

显示名

* 模式

描述

高级设置

* 能下载Select结果 ☒

下一步 取消

? 说明

- 工作空间名称全局唯一，建议您使用易于区分的名称。
- 为方便使用，本教程中DataWorks工作空间模式为简单模式（单环境）。在简单模式下，DataWorks工作空间与MaxCompute项目一一对应，详情请参见[简单模式和标准模式的区别](#)。

- iii. 进入选择引擎界面，选择计算引擎服务下的MaxCompute，按量付费，单击下一步。

创建工作空间

基本配置

2 选择引擎

3 引擎详情

选择DataWorks服务

☒

 数据集成、数据开发、运维中心、数据质量
您可以进行数据同步集成、工作流编排、周期任务调度和运维、对产出数据质量进行检查等。

选择计算引擎服务

☒

 MaxCompute ☐ 包年包月 ☒ 按量付费 ☐ 开发者版 [去购买](#)
开通后，您可在DataWorks里进行MaxCompute SQL、MaxCompute MR任务的开发。
[充值](#) [续费](#) [升级](#) [降配](#)

☐

 实时计算 ☐ 共享模式 ☐ 独享模式
开通后，您可在DataWorks里面进行流式计算任务开发。

☐

 E-MapReduce
开通后，您可以在DataWorks中使用E-MapReduce进行大数据处理任务的开发。

☐

 交互式分析 ☐ 包年包月 [去购买](#)
开通后，您可以在DataWorks里使用HoloStudio进行交互式分析(Interactive Analytics)的表管理、外部表管理、SQL任务的开发。

下一步

上一步

取消

iv. 进入引擎详情页面，填写选购引擎的配置，单击创建工作空间。

工作空间创建成功后，即可在工作空间列表页面查看相应内容。

> 文档版本：20211223

11

创建工作空间

基本配置

选择引擎

引擎详情

MaxCompute

* 实例显示名称

* Quota组切换

按量付费默认资源组

* MaxCompute数据类型

2.0数据类型 (推荐)

* MaxCompute项目名称

Question_01

* MaxCompute访问身份

阿里云主账号

* 是否加密:

不加密

加密

请注意: 开启MaxCompute项目存储加密后, 该项目将无法运行PAI、Hologres任务; 如需运行PAI、Hologres任务请提交工单申请关闭存储加密功能。

如当前登录执行创建MaxCompute项目的账号为RAM子账号, 为方便管理, 该子账号将被加入至MaxCompute Super_Administrator角色 (标准模式下仅开发环境项目)。

创建工作空间

上一步

取消

分类	配置	说明
MaxCompute	实例显示名称	实例名称长度需要在3-28个字符, 仅支持字母开头, 仅包含字母、下划线和数字。
	Quota组切换	Quota用来实现计算资源和磁盘配额。
	MaxCompute数据类型	MaxCompute项目的数据类型版本。
	MaxCompute项目名称	默认与DataWorks工作空间的名称一致。
	MaxCompute访问身份	MaxCompute访问身份包括阿里云主账号和任务责任人。阿里云主账号即阿里云账号
	是否加密	根据实际情况选择是否需要加密当前实例。

5. 开通Quick BI。详情请参见Quick BI购买、升级、降级、续费、欠费。

1.3. 数据准备

在数据准备阶段, 您需要通过数据Demo包生成模拟真实环境的数据, 以便后续数据开发使用。

前提条件

- 创建华北2 (北京) 区域的表格存储实例, 同时记录实例名称和实例访问地址。单击表格存储控制台中的实例名称, 即可获得实例访问地址。对于跨区域的访问, 建议您使用公网地址。详细操作请参见环境准备。
- 使用阿里云账号登录安全管理控制台, 获取并记录您的AccessKey ID和AccessKey Secret信息。

 **说明** AccessKey ID和AccessKey Secret是您访问阿里云API的密钥，具有该账户完全的权限，请您妥善保管。

操作步骤

1. 下载数据Demo包。

数据Demo包下载地址如下，本例中使用环境为Windows7 64位：

- [Mac下载地址](#)
- [Linux下载地址](#)
- [Windows7 64位下载地址](#)

2. 配置Demo环境。

完成下载后，解压下载包，编辑conf文件夹内的app.conf文件。

名称	修改日期	类型	大小
 conf	2019/6/17 10:07	文件夹	
 workshop_demo.exe	2017/12/18 16:58	应用程序	12,367 KB

app.conf文件内容示例如下。

```
endpoint = "https://workshop-bj-001.cn-beijing.ots.aliyuncs.com"
instanceName = "workshop-bj-001"
accessKeyId = "LTAIF24u7g*****"
accessKeySecret = "CcwFeF3sWTPy0wsKULMw34Px*****"
usercount = "200"
daysCount = "7"
```

其中，需要配置的参数如下：

- endpoint：表格存储实例的访问网络地址，建议您使用公网地址。您可以在Tablestore控制台，单击实例名称，在实例详情页签的实例访问地址区域获取。
- instanceName：表格存储实例的名称。您可以在Tablestore控制台的概览页面获取。
- accessKeyId和accessKeySecret：访问阿里云的密钥。

3. 启动Demo准备测试数据。

- i. 启动Windows CMD命令行工具，进入您解压缩Demo包的路径，执行如下语句查看Demo包命令用法。

```
workshop_demo.exe -h
```

该命令会列出该demo的相关命令，如下。

```
...\workshop_demo>workshop_demo.exe -h
prepare will prepare the data
raw 00005 "2017-12-19" will query user data by user id and time
new/day_active/month_active/day_pv/month_pv will query metrics data
```

- **prepare** : 准备测试数据，创建数据表，根据conf中的用户数量，为用户生成一周的行为日志数据。
- **raw** : `raw ${userid} ${date} ${Top条数}` , 查询指定用户的日志明细。
- **new/day_active/month_active/day_pv/month_pv** : 在结果表中按照如下几种类型查询报表数据：
 - 新增: new
 - 日活: day_active
 - 月活: month_active
 - 日PV: day_pv
 - 月PV: month_pv

- ii. 执行如下命令生成准备数据。

```
...\workshop_demo>workshop_demo.exe prepare
```

结果如下。

```
C:\Users\... \workshop_demo>workshop_demo.exe prepare
OTSObjectAlreadyExist Requested table already exists.
OTSObjectAlreadyExist Requested table already exists.
Prepare the metric data
Prepare the metric data
Prepare the metric data
Prepare the metric data
Prepare the metric data
Prepare User data
finished one round
finished one round
finished one round
finished one round
finished one round
total insert data count is: 41757
```

在此过程中，Demo包会自动帮助您在表格存储中创建表，结构如下：

- 原始日志数据表: user_trace_log

列名	类型	说明
md5	STRING	用户uid的md5值undefined前8位，表格存储主键。

列名	类型	说明
uid	STRING	用户uid，表格存储主键。
ts	BIGINT	用户操作时间戳，表格存储主键。
ip	STRING	IP地址。
status	BIGINT	服务器返回状态码。
bytes	BIGINT	返回给客户端的字节数。
device	STRING	终端型号。
system	STRING	系统版本：ios xxx/android xxx。
customize_event	STRING	自定义事件：登录/退出/购买/注册/点击/后台/切换用户/浏览。
use_time	BIGINT	APP单次使用时长，当事件为退出、后台、切换用户时有该项。
customize_event_content	STRING	用户关注的内容信息。

○ 分析结果表：analysis_result

列名	类型	说明
metric	STRING	报表的类型：'new'、'day_active'、'month_active'、'day_pv'、'month_pv'，表格存储主键。
ds	STRING	时间yyyy-mm-dd或yyyy-mm，表格存储主键。
num	BIGINT	对应的数据值。

4. 数据验证。

○ 用户明细查询

通过如下语句查询指定用户在某一日期指定条数的明细数据。表格数据的日期对应于您创建表格的时间。

```
raw ${userid} ${date} ${Top条数}
```

其中，\${userid}为用户ID，\${date}为指定日期，\${Top条数}为指定查询条数。例如，您创建数据时间为2019年6月15日，则可以使用 `workshop_demo.exe raw 00010 "2019-06-15" 20` 查看20条用户明细数据。

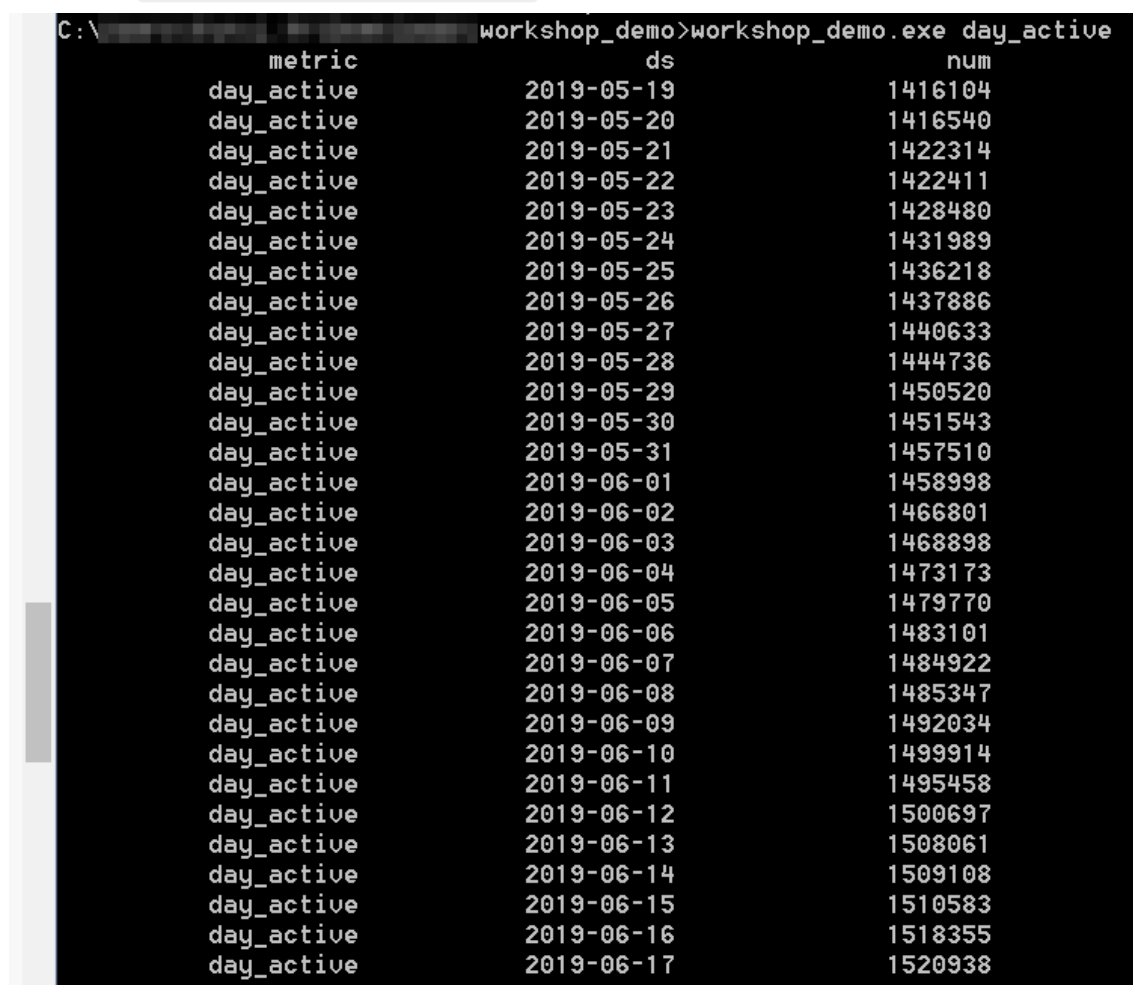
```
C:\nloads\workshop_demo>workshop_demo.exe raw 00010 "2019-06-15" 20
```

uid	device	ip	Date	status	bytes	customize_event	system
00010		2019-06-14	11:56:47	PM	759		regist
iPhone7 Plus		61.103.79.217		200		ios11	
00010		2019-06-14	11:26:34	PM	252	backstage	369
iPad min2		157.249.67.241		200		ios11	
00010		2019-06-14	11:21:30	PM	427	browse	travel
iPhone6s		222.133.108.234		200		ios10	
00010		2019-06-14	11:16:03	PM	764	switch	185
iPhone7 Plus		61.103.79.217		200		ios11	
00010		2019-06-14	11:06:03	PM	436		click
iPhone7 Plus		61.103.79.217		200		ios11	
00010		2019-06-14	10:36:54	PM	131		click
iPhone7 Plus		61.103.79.217		200		ios11	
00010		2019-06-14	10:22:26	PM	778	switch	73
iPhone6s		222.133.108.234		200		ios10	
00010		2019-06-14	10:06:29	PM	535	backstage	179
iPad min2		157.249.67.241		200		ios11	
00010		2019-06-14	09:56:11	PM	668		click
iPad min2		157.249.67.241		200		ios11	
00010		2019-06-14	09:20:45	PM	354		regist
iPhone6s		222.133.108.234		200		ios10	
00010		2019-06-14	09:15:37	PM	989		click
iPad min2		157.249.67.241		200		ios11	
00010		2019-06-14	08:51:17	PM	460	logout	462
iPhone6s		222.133.108.234		200		ios10	
00010		2019-06-14	08:26:06	PM	887	comment	funny
iPad min2		157.249.67.241		200		ios11	
00010		2019-06-14	08:10:34	PM	278	browse	finance
iPhone6s		222.133.108.234		200		ios10	
00010		2019-06-14	07:56:00	PM	480		click
iPhone7 Plus		61.103.79.217		200		ios11	
00010		2019-06-14	07:30:11	PM	68		click
iPhone6s		222.133.108.234		200		ios10	
00010		2019-06-14	07:15:09	PM	398	browse	news
iPhone7 Plus		61.103.79.217		200		ios11	
00010		2019-06-14	07:11:21	PM	21		click
iPhone6s		222.133.108.234		200		ios10	
00010		2019-06-14	06:35:07	PM	207	browse	photo
iPhone7 Plus		61.103.79.217		200		ios11	
00010		2019-06-14	06:24:43	PM	261		regist
iPhone7 Plus		61.103.79.217		200		ios11	

② 说明 由于表格存储是SchemaFree结构，表的属性列不需要预先定义。Customize_Event中不同的事件对应了不同的内容，因此Demo中将事件、内容进行对齐显示。

○ 报表结果查询

您可以使用 `workshop_demo.exe day_active` 命令查看日活数据。



```
C:\>workshop_demo.exe day_active
```

metric	ds	num
day_active	2019-05-19	1416104
day_active	2019-05-20	1416540
day_active	2019-05-21	1422314
day_active	2019-05-22	1422411
day_active	2019-05-23	1428480
day_active	2019-05-24	1431989
day_active	2019-05-25	1436218
day_active	2019-05-26	1437886
day_active	2019-05-27	1440633
day_active	2019-05-28	1444736
day_active	2019-05-29	1450520
day_active	2019-05-30	1451543
day_active	2019-05-31	1457510
day_active	2019-06-01	1458998
day_active	2019-06-02	1466801
day_active	2019-06-03	1468898
day_active	2019-06-04	1473173
day_active	2019-06-05	1479770
day_active	2019-06-06	1483101
day_active	2019-06-07	1484922
day_active	2019-06-08	1485347
day_active	2019-06-09	1492034
day_active	2019-06-10	1499914
day_active	2019-06-11	1495458
day_active	2019-06-12	1500697
day_active	2019-06-13	1508061
day_active	2019-06-14	1509108
day_active	2019-06-15	1510583
day_active	2019-06-16	1518355
day_active	2019-06-17	1520938

1.4. 数据建模与开发

1.4.1. 新建数据表

本文为您介绍如何在MaxCompute上建立数据表，用于存储原始数据及加工后的数据。

前提条件

- 已开通MaxCompute服务并创建DataWorks工作空间（本教程使用为简单模式工作空间），详情请参见[环境准备](#)。
- 已具备访问Tablestore数据的权限。当MaxCompute和Tablestore的所有者是同一个账号时，您可以[单击此处一键授权](#)。如果不是，您可以自定义授权，详情请参见[OTS外部表](#)。

操作步骤

- 进入DataWorks数据开发界面。
 - 进入[DataWorks工作空间列表](#)，选择区域为华东2（上海）。
 - 单击已创建好的工作空间后的[进入数据开发](#)，进入工作空间的数据开发界面。
- 新建业务流程。

- i. 右键单击业务流程，选择新建业务流程。



- ii. 填写业务名称和描述，单击新建。本教程中，业务流程名为Workshop。

3. 新建数据表。

- i. 创建外部表ots_user_trace_log。

 **注意** 创建外部表时，暂时不支持在向导模式下使用DDL方式创建。您可以在向导模式下手动添加外部表字段进行表配置，或创建一个MaxCompute的ODPS SQL节点，在ODPS SQL节点中使用DDL语句创建外部表。以下以使用ODPS SQL节点的DDL语句创建外部表作为操作示例。

- a. 单击新建的业务流程Workshop，右键单击MaxCompute，选择新建 > ODPS SQL，输入节点名称workshop单击提交。
- b. 双击创建的ODPS节点，进入数据开发界面后，编译如下DDL语句，用于创建名称为ots_user_trace_log的外部表。

外部表ots_user_trace_log的建表语句如下。

```
CREATE EXTERNAL TABLE ots_user_trace_log (
  md5 string COMMENT '用户uid的md5值前8位',
  uid string COMMENT '用户uid',
  ts bigint COMMENT '用户操作时间戳',
  ip string COMMENT 'ip地址',
  status bigint COMMENT '服务器返回状态码',
  bytes bigint COMMENT '返回给客户端的字节数',
  device string COMMENT '终端型号',
  system string COMMENT '系统版本ios xxx/android xxx',
  customize_event string COMMENT '自定义事件：登录/退出/购买/注册/点击/后台/切换用户/浏览/评论',
  use_time bigint COMMENT 'APP单次使用时长，当事件为退出、后台、切换用户时有该项',
  customize_event_content string COMMENT '用户关注内容信息，在customize_event为浏览和评论时，包含该列'
)
STORED BY 'com.aliyun.odps.TableStoreStorageHandler'
WITH SERDEPROPERTIES (
  'tablestore.columns.mapping'=':md5,:uid,:ts, ip,status,bytes,device,system,customize_event,use_time,customize_event_content',
  'tablestore.table.name'='user_trace_log'
)
LOCATION 'tablestore://workshop-bj-001.cn-beijing.ots.aliyuncs.com/';
```

- **STORED BY**: 必选参数, 值为 `com.aliyun.odps.TableStoreStorageHandler`, 是MaxCompute内置处理Tablestore数据的StorageHandler, 定义了MaxCompute和Tablestore的交互。
- **SERDEPROPERTIES**: 必选参数, 是提供参数选项的接口, 在使用TableStoreStorageHandler时, 以下选项必须指定:
 - **tablestore.columns.mapping**: 用于描述MaxCompute将访问的Tablestore表的列, 包括主键和属性列。

② 说明

- 以冒号 (:) 开头的参数值为Tablestore主键, 例如示例中的 `:md5`、`:uid` 和 `:ts`, 其它参数值均为属性列。
- 在指定映射时, 您必须提供指定Tablestore表的所有主键, 只需提供需要通过MaxCompute访问的属性列。提供的属性列必须是Tablestore表的列, 否则即使外表可以创建成功, 查询时也会报错。

- **tablestore.table.name**: 需要访问的Tablestore表名。如果指定的Tablestore表名错误(不存在), 则会报错, MaxCompute不会主动创建Tablestore表。
- **LOCATION**: 用来指定Tablestore的访问地址。请您根据[环境准备](#), 将自己的表格存储实例访问地址参数填写在此。

② 说明 如果您使用公网地址LOCATION 'tablestore://workshop-bj-001.cn-beijing.ots.aliyuncs.com/'报错, 显示网络不同, 可尝试更换为经典网地址LOCATION 'tablestore://workshop-bj-001.cn-beijing.ots-internal.aliyuncs.com/'。

c. 单击保存图标后, 单击运行图标。

当运行日志显示运行成功后, 外部表即已创建成功。

ii. 创建ods_user_trace_log表。

建表方法同上, 建表语句如下。ods_user_trace_log为ODS层表。

```
CREATE TABLE IF NOT EXISTS ods_user_trace_log (
  md5 STRING COMMENT '用户uid的md5值前8位',
  uid STRING COMMENT '用户uid',
  ts BIGINT COMMENT '用户操作时间戳',
  ip STRING COMMENT 'ip地址',
  status BIGINT COMMENT '服务器返回状态码',
  bytes BIGINT COMMENT '返回给客户端的字节数',
  device STRING COMMENT '终端型号',
  system STRING COMMENT '系统版本ios xxx/android xxx',
  customize_event STRING COMMENT '自定义事件: 登录/退出/购买/注册/点击/后台/切换用户/浏览/评论',
  use_time BIGINT COMMENT 'APP单次使用时长, 当事件为退出、后台、切换用户时有该项',
  customize_event_content STRING COMMENT '用户关注内容信息, 在customize_event为浏览和评论时, 包含该列'
)
PARTITIONED BY (
  dt STRING
);
```

iii. 创建dw_user_trace_log表。

建表方法同上，建表语句如下。dw_user_trace_log为DW层表。

```
CREATE TABLE IF NOT EXISTS dw_user_trace_log (
    uid STRING COMMENT '用户uid',
    region STRING COMMENT '地域，根据ip得到',
    device_brand string comment '设备品牌',
    device STRING COMMENT '终端型号',
    system_type STRING COMMENT '系统类型，Android、IOS、ipad、Windows_phone',
    customize_event STRING COMMENT '自定义事件：登录/退出/购买/注册/点击/后台/切换用户/浏览/评论',
    use_time BIGINT COMMENT 'APP单次使用时长，当事件为退出、后台、切换用户时有该项',
    customize_event_content STRING COMMENT '用户关注内容信息，在customize_event为浏览和评论时，包含该列'
)
PARTITIONED BY (
    dt STRING
);
```

iv. 创建rpt_user_trace_log表。

建表方法同上，建表语句如下。rpt_user_trace_log为ADS层表。

```
CREATE TABLE IF NOT EXISTS rpt_user_trace_log (
    country STRING COMMENT '国家',
    province STRING COMMENT '省份',
    city STRING COMMENT '城市',
    device_brand string comment '设备品牌',
    device STRING COMMENT '终端型号',
    system_type STRING COMMENT '系统类型，Android、IOS、ipad、Windows_phone',
    customize_event STRING COMMENT '自定义事件：登录/退出/购买/注册/点击/后台/切换用户/浏览/评论',
    use_time BIGINT COMMENT 'APP单次使用时长，当事件为退出、后台、切换用户时有该项',
    customize_event_content STRING COMMENT '用户关注内容信息，在customize_event为浏览和评论时，包含该列',
    pv bigint comment '浏览量',
    uv bigint comment '独立访客'
)
PARTITIONED BY (
    dt STRING
);
```

- v. (可选) 如果您使用的是标准模式, 上述步骤创建的外部表仅提交到了开发环境, 您需根据以下步骤, 依次将所有创建的外部表提交到生产环境。

- a. 单击MaxCompute下的表, 右键选择导入表。



- b. 在导入表弹窗中选择上述创建的开发环境外部表, 单击**确认**。

说明 鼠标悬浮在表名上即可在表名后看到该表是生产环境的表还是开发环境的表。如果误选了生产环境的表, 会提示您表不存在。

- c. 双击导入的外部表, 单击**提交到生产环境**。

4. 验证建表结果。

- 完成建表后, 您可以在Workshop业务流程**MaxCompute > 表**下看到新建的4张表。
- 右键单击业务流程中**MaxCompute**下的**数据开发**, 选择**新建 > ODPS SQL**。
- 在**新建节点**页面, 输入**节点名称**, 单击**提交新建ODPS SQL节点**。

iv. 在新建的ODPS SQL节点中输入如下SQL语句，单击图标。

```
DESCRIBE ots_user_trace_log;
DESCRIBE ods_user_trace_log;
DESCRIBE dw_user_trace_log;
DESCRIBE rpt_user_trace_log;
```

返回表的结构信息如下：

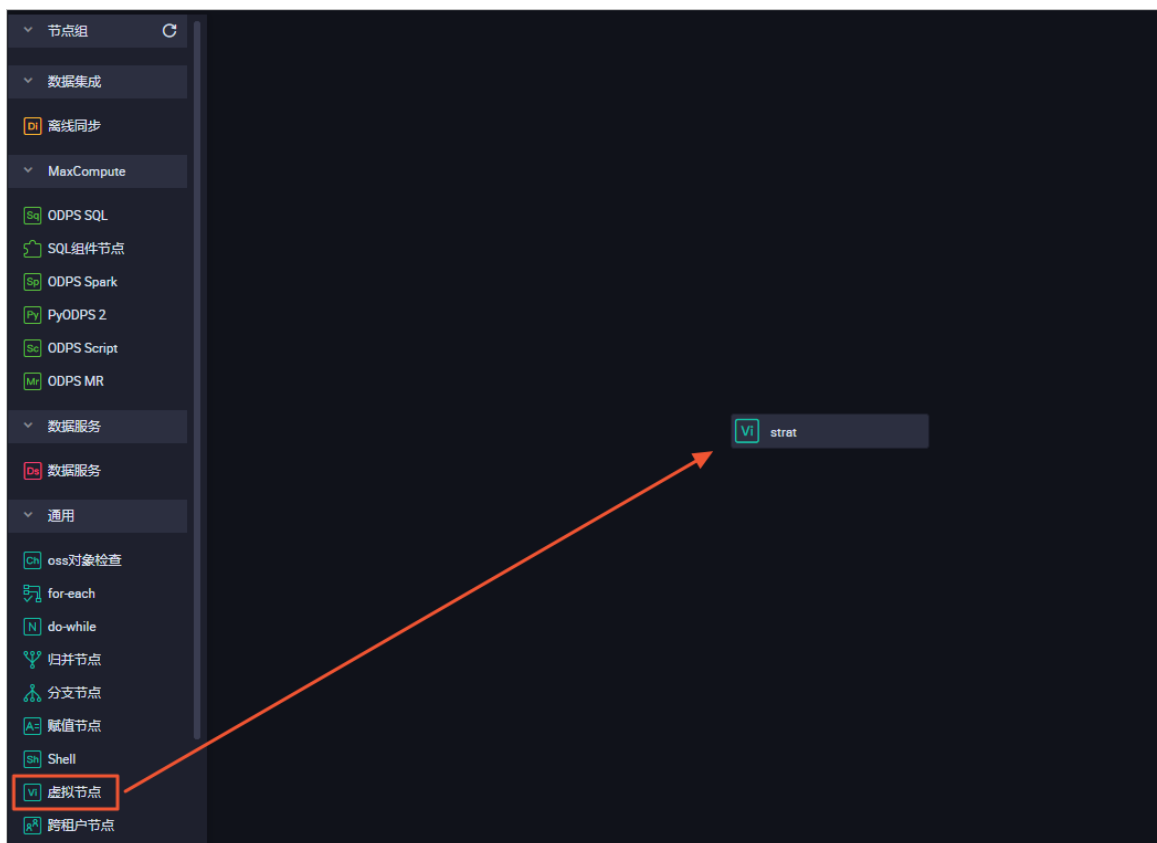
+-----+ Owner: [redacted] Project: [redacted] +-----+			
TableComment: +-----+			
CreateTime: 2020-06-16 18:56:46 +-----+			
LastDDLTime: 2020-06-16 18:56:46 +-----+			
LastModifiedTime: 2020-06-16 18:56:46 +-----+			
+-----+			
InternalTable: YES Size: 0 +-----+			
+-----+			
Native Columns: +-----+			
+-----+			
Field Type Label Comment +-----+			
+-----+			
country string 国家 +-----+			
province string 省份 +-----+			
city string 城市 +-----+			
device_brand string 设备品牌 +-----+			
device string 终端型号 +-----+			
system_type string 系统类型, Android、IOS、ipad、Windows_phone +-----+			
customize_event string 自定义事件: 登录/退出/购买/注册/点击/后台/切换用户/浏览 +-----+			
use_time bigint APP单次使用时长, 当事件为退出、后台、切换用户时有该项 +-----+			
customize_event_content string 用户关注内容信息, 在customize_event为浏览和评论时 包含该列 +-----+			
pv bigint 浏览里 +-----+			
uv bigint 独立访客 +-----+			
+-----+			
Partition Columns: +-----+			
+-----+			
dt string +-----+			
+-----+			
OK			
2020-06-16 19:56:10 INFO =====			
2020-06-16 19:56:10 INFO Exit code of the Shell command 0			
2020-06-16 19:56:10 INFO --- Invocation of Shell command completed ---			

1.4.2. 设计工作流

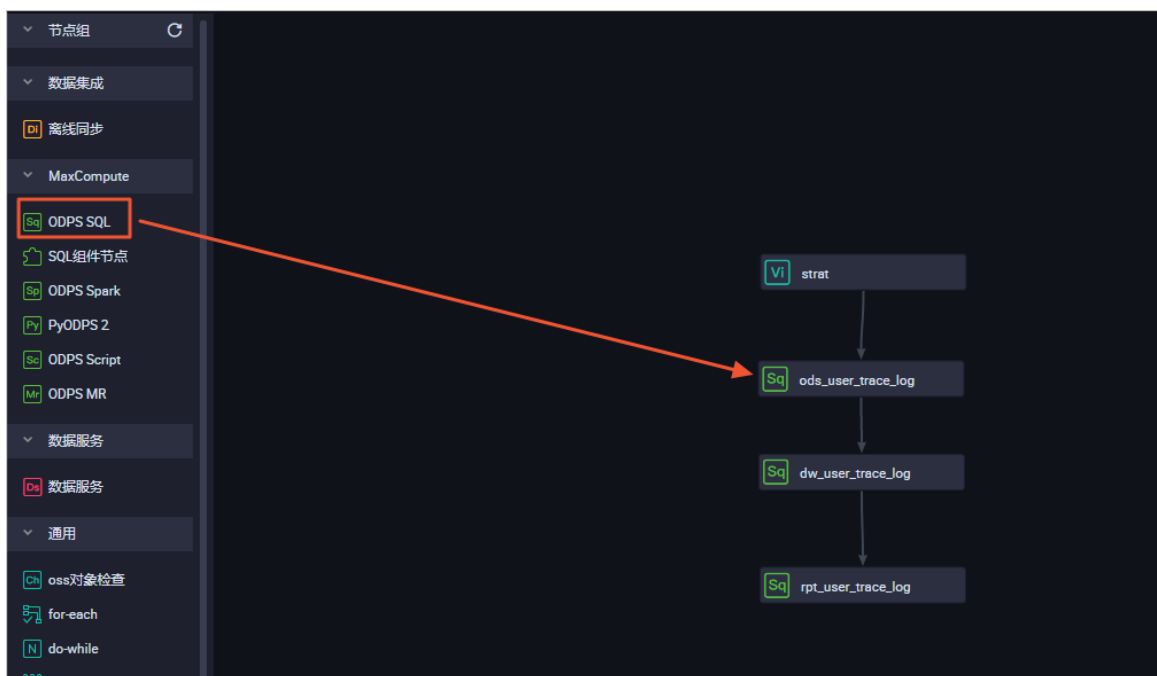
通过设计工作流，您可以明确在整体数据开发过程中各任务节点的排布。对于本教程中这种较为简单的单数据流场景，您可以选择每个数据表（数仓层次）对应一个工作流。

操作步骤

1. 进入DataWorks数据开发界面。
 - i. 进入[DataWorks工作空间列表](#)，选择区域为华东2（上海）。
 - ii. 单击已创建好的工作空间后的进入数据开发，进入工作空间的数据开发界面。
2. 在您创建的业务流程上双击，打开画布面板。
3. 向画布中拖入1个虚拟节点，命名为start。



4. 向画布中拖入3个ODPS SQL节点，依次命名为ods_user_trace_log、dw_user_trace_log、rpt_user_trace_log。通过连接不同节点，配置依赖关系如下。



❓ 说明 ods_user_trace_log、dw_user_trace_log、rpt_user_trace_log分别代表数据仓库的ODS、CDM和ADS层，详情请参见[数仓分层](#)。

1.4.3. 节点配置

完成 workflows 设计后，您需要对每个数据开发节点进行配置，填写 SQL 语句。

前提条件

本次数据开发过程中需要使用 UDF 自定义函数，您首先需要完成自定义函数的注册，详细请参见[注册自定义函数](#)。

注册自定义函数

1. 添加资源

- i. 下载用于 IP 地转换的[自定义函数 Java 包 getaddr.jar](#)以及[地址库 ip.dat](#)。
- ii. 右键单击 Workshop 业务流程下的 MaxCompute，选择新建 > 资源。需要分别新建 File 和 JAR 类型的资源。



■ File类型上传地址库ip.dat。

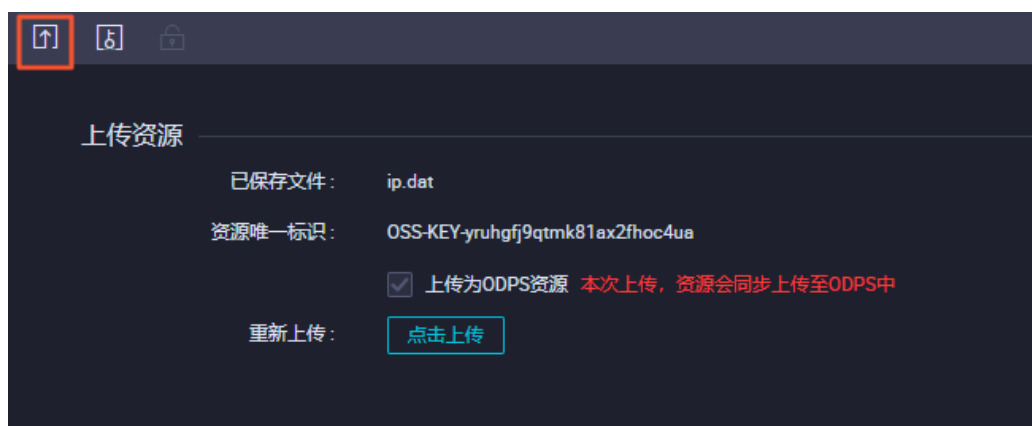
- a. 输入资源名称，选中大文件（内容超过500KB）及上传为ODPS资源，然后单击点击上传。



The screenshot shows a '新建资源' (New Resource) dialog box. It contains the following fields and options:

- * 资源名称:** An empty text input field.
- 目标文件夹:** A dropdown menu showing '业务流程/test/MaxCompute/testworkshop777/资源'.
- 资源类型:** A dropdown menu showing 'File'.
- Options:** Two checked checkboxes: '大文件 (内容超过500KB)' and '上传为ODPS资源'. A red note next to the second checkbox says '本次上传, 资源会同步上传至ODPS中'.
- 上传文件:** A button labeled '点击上传'.
- Buttons:** '确定' (Confirm) and '取消' (Cancel) buttons at the bottom right.

- b. 单击提交。



The screenshot shows an '上传资源' (Upload Resource) dialog box. It contains the following fields and options:

- 已保存文件:** A text field showing 'ip.dat'.
- 资源唯一标识:** A text field showing 'OSS-KEY-yruhgfj9qtmk81ax2fhoc4ua'.
- Options:** A checked checkbox labeled '上传为ODPS资源'. A red note next to it says '本次上传, 资源会同步上传至ODPS中'.
- 重新上传:** A button labeled '点击上传'.
- Buttons:** '确定' (Confirm) and '取消' (Cancel) buttons at the bottom right.

■ JAR类型对应Java包getaddr.jar。

a. 您需要勾选上传为ODPS资源，然后单击点击上传。



b. 上传完成后，单击提交。

② 说明 提交时，请忽略血缘不一致信息。

2. 注册函数

- 在业务流程下右键单击**MaxCompute**，选择**新建 > 函数**，将函数命名为getregion。
- 在**注册函数**页面，依次填写类名为odps.test.GetAddr，资源列表为getaddr.jar,ip.dat，命令格式为getregion(ip string)，保存后单击  提交函数注册。



配置节点

- 1. 配置虚拟节点start。
 - i. 双击start节点，进入节点配置页面。

ii. 单击右侧的调度配置，在调度依赖区域下单击使用工作空间根节点完成配置。

调度配置

依赖上一周期：☐

调度依赖

自动解析：☒ 是 ☐ 否

解析输入输出

依赖的上游节点：

请输入父节点输出名称或输出表名

+

使用工作空间根节点

父节点输出名称	父节点输出表名	节点名	父节点ID	责任人	来源	操作
my_project_simple_root	-	my_project_simple_root	700002666464	dtplus_docs	手动添加	删除

本节点的输出：

请输入节点输出名称

+

输出名称	输出表名	下游节点名称	下游节点ID	责任人	来源	操作
my_project_simple.500642587_out	-	ods_user_trace_log	700002759506	tina	系统默认添加	删除
my_project_simple.start	-	-	-	-	手动添加	删除

iii. 在时间属性区域选择重跑属性为运行成功或失败后皆可重跑。

iv. 单击  按钮，完成节点提交。

2. 配置ODPS SQL节点ods_user_trace_log

i. 双击ods_user_trace_log节点，进入节点配置界面，编写处理逻辑。SQL代码如下。

```
insert overwrite table ods_user_trace_log partition (dt=${bdp.system.bizdate})
select
    md5,
    uid ,
    ts,
    ip,
    status,
    bytes,
    device,
    system,
    customize_event,
    use_time,
    customize_event_content
from ots_user_trace_log
where to_char (FROM_UNIXTIME(ts), 'yyyymmdd')=${bdp.system.bizdate};
```

 说明 关于\${bdp.system.bizdate}释义请参见配置调度参数。

28

> 文档版本：20211223

ii. 完成代码编写后，单击右侧的调度配置，选择自动解析为否。

调度配置

暂停调度：☐

调度周期：

日

定时调度：☒

具体时间：

00:00

注：默认调度时间，从0点到0点30分随机生成

cron表达式：

00 00 00 * * ?

依赖上一周期：☐

调度依赖

自动解析：

是

☒ 否

定时输入输出

依赖的上游节点：

请输入父节点输出名称或输出表名

+

使用工作空间模板节点

自动推荐

父节点输出名称	父节点输出表名	节点名	父节点ID	责任人	来源	操作
my_project_simple.start	-	start	700002760121	tina	手动添加	删除

本节点的输出：

请输入节点输出名称

+

输出名称	输出表名	下游节点名称	下游节点ID	责任人	来源	操作
my_project_simple.500643288_out	-	-	-	-	系统默认添加	删除
my_project_simple.ods_user_trace_log	-	dws_user_trace_log	700002760123	tina	手动添加	删除

节点上下文

本节点输入参数

添加

iii. 手动删除错误的依赖关系。

调度依赖

自动解析：

是

☒ 否

定时输入输出

依赖的上游节点：

请输入父节点输出名称或输出表名

+

使用工作空间模板节点

自动推荐

父节点输出名称	父节点输出表名	节点名	父节点ID	责任人	来源	操作
my_project_simple.500643286_out	-	start	700002760121	tina	手动添加	删除

本节点的输出：

请输入节点输出名称

+

iv. 按照业务流程顺序搜索正确的上游节点，例如此处为start，并单击添加。

调度依赖

自动解析：

是

☒ 否

定时输入输出

依赖的上游节点：

1 请输入父节点输出名称或输出表名

2 +

使用工作空间模板节点

自动推荐

父节点输出名称	父节点输出表名	节点名	父节点ID	责任人	来源	操作
my_project_simple.start	-	start	700002760121	tina	手动添加	删除

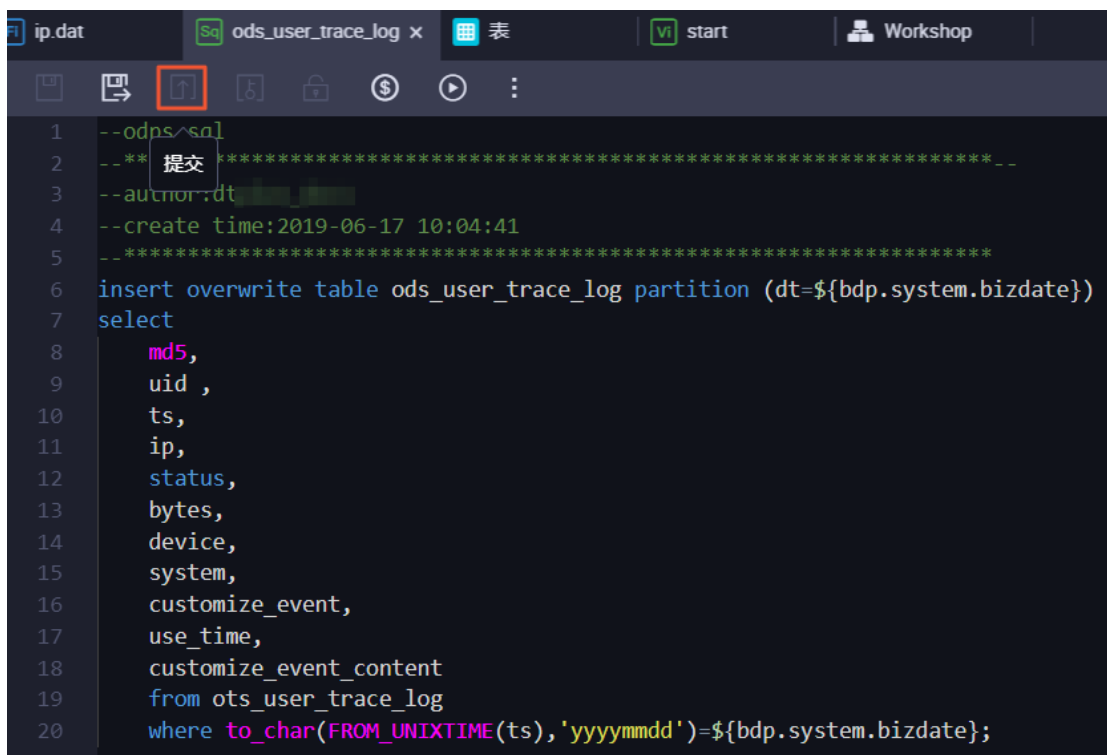
本节点的输出：

请输入节点输出名称

+

v. 在时间属性区域选择重跑属性为运行成功或失败后皆可重跑。

vi. 完成后，单击提交。



3. 配置ODPS SQL节点dw_user_trace_log

您可以使用与ods_user_trace_log节点一样的方法配置dw_user_trace_log节点，SQL代码如下。

```

INSERT OVERWRITE TABLE dw_user_trace_log PARTITION (dt=${bdp.system.bizdate})
SELECT uid, getregion(ip) AS region
, CASE
    WHEN TOLOWER(device) RLIKE 'xiaomi' THEN 'xiaomi'
    WHEN TOLOWER(device) RLIKE 'meizu' THEN 'meizu'
    WHEN TOLOWER(device) RLIKE 'huawei' THEN 'huawei'
    WHEN TOLOWER(device) RLIKE 'iphone' THEN 'iphone'
    WHEN TOLOWER(device) RLIKE 'vivo' THEN 'vivo'
    WHEN TOLOWER(device) RLIKE 'honor' THEN 'honor'
    WHEN TOLOWER(device) RLIKE 'samsung' THEN 'samsung'
    WHEN TOLOWER(device) RLIKE 'leeco' THEN 'leeco'
    WHEN TOLOWER(device) RLIKE 'ipad' THEN 'ipad'
    ELSE 'unknown'
END AS device_brand, device
, CASE
    WHEN TOLOWER(system) RLIKE 'android' THEN 'android'
    WHEN TOLOWER(system) RLIKE 'ios' THEN 'ios'
    ELSE 'unknown'
END AS system_type, customize_event, use_time, customize_event_content
FROM ods_user_trace_log
WHERE dt = ${bdp.system.bizdate};

```

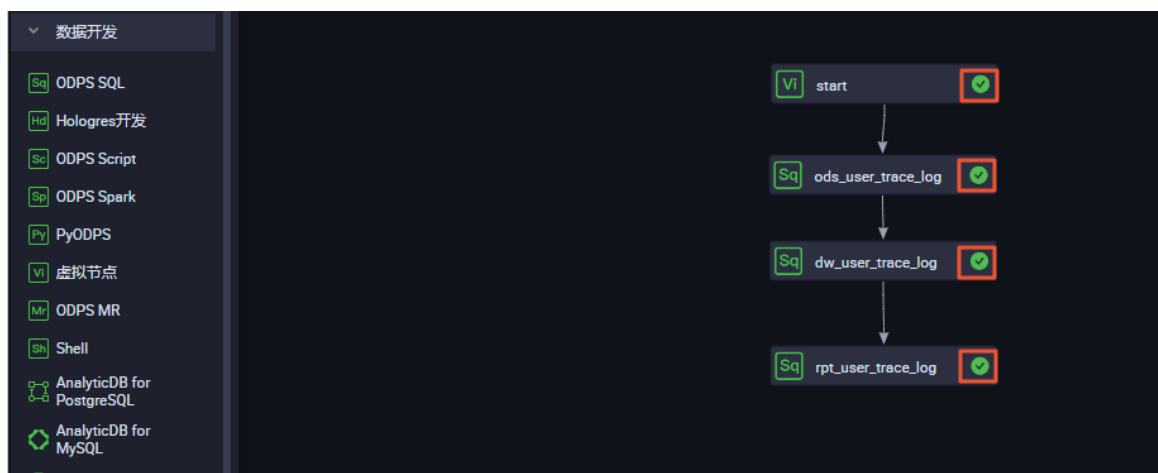
4. 配置ODPS SQL节点rpt_user_trace_log

您可以使用与ods_user_trace_log节点一样的方法配置rpt_user_trace_log节点，SQL代码如下。

```
INSERT OVERWRITE TABLE rpt_user_trace_log PARTITION (dt=${bdp.system.bizdate})
SELECT split_part(split_part(region, ',', 1), '[', 2) AS country
      , trim(split_part(region, ',', 2)) AS province
      , trim(split_part(region, ',', 3)) AS city
      , MAX(device_brand), MAX(device)
      , MAX(system_type), MAX(customize_event)
      , FLOOR(AVG(use_time / 60))
      , MAX(customize_event_content), COUNT(uid) AS pv
      , COUNT(DISTINCT uid) AS uv
FROM dw_user_trace_log
WHERE dt = ${bdp.system.bizdate}
GROUP BY uid,
        region;
```

5. 验证配置结果。

双击业务流Workshop，打开画布面板。单击  按钮。运行成功如下图所示。



如果运行状态异常，请右键单击出错节点，单击查看运行日志进行排查。

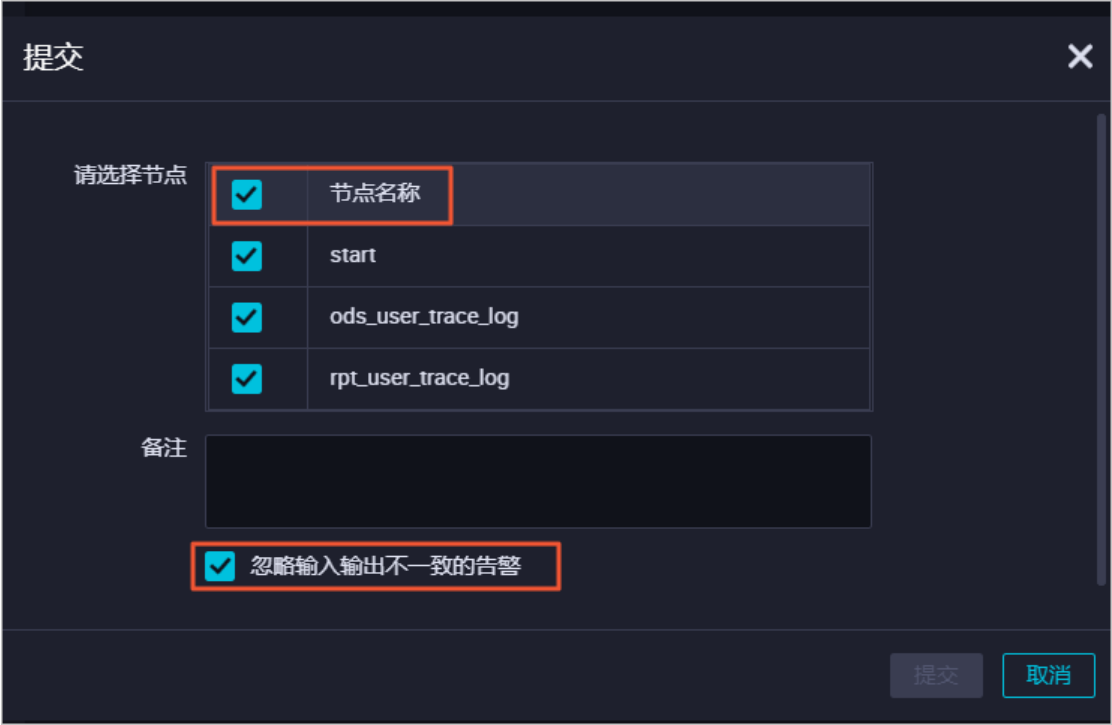


1.4.4. 任务提交与测试

您完成节点配置后，需要将任务提交到运维中心进行测试。

操作步骤

1. （可选）提交业务流程。如果您的节点在配置完成后已经提交完毕且无更新，请跳过本步骤。
 - i. 双击业务流程名称Workshop，单击图标。
 - ii. 勾选所有可提交节点及忽略输入输出不一致的告警，单击提交。



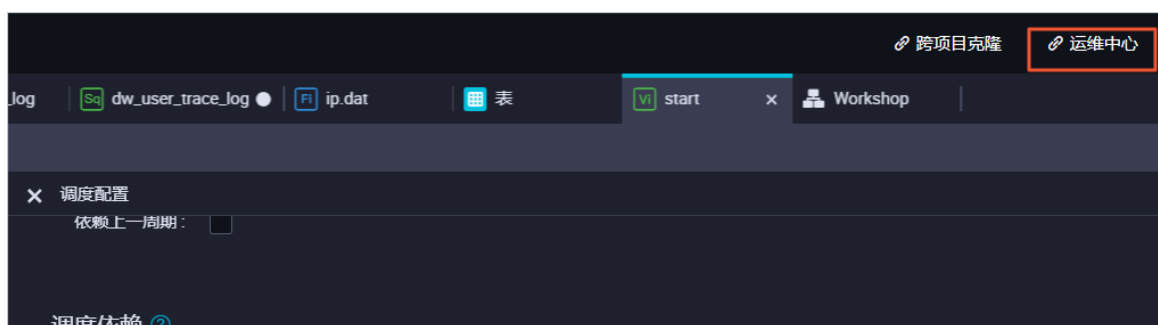
提交对话框，标题为“提交”，右上角有关闭按钮。对话框内包含以下元素：

- “请选择节点”区域：一个表格，列出了四个节点，每个节点左侧都有一个复选框，右侧是节点名称。

☒	节点名称
☒	start
☒	ods_user_trace_log
☒	rpt_user_trace_log
- “备注”区域：一个文本输入框。
- 底部：有一个复选框，勾选了“忽略输入输出不一致的告警”，以及“提交”和“取消”两个按钮。

 说明 标准空间模式下，提交通过后，需要单击发布将任务发布至生产环境。

2. 单击右上角的运维中心。



3. 在左侧导航栏，单击周期任务运维 > 周期任务，双击节点列表中的虚拟节点start。
4. 在右侧流程图上，右键单击虚拟节点start，选择补数据 > 当前节点及下游节点。



5. 在补数据页面，选中所有需要补数据的节点，选择业务日期为过去一周，单击确定。

补数据

* 补数据名称:

P_start_20190619_155104

* 选择业务日期:

2019-06-11 - 2019-06-17

* 是否并行:

不并行

* 选择需要补数据的节点:

☒

任务名称

按名称进行搜索...

Q

任务类型

☒	bigdata_DOC(1485)	
☒	start	虚节点
☒	ods_user_trace_log	ODPS_SQL
☒	dw_user_trace_log	ODPS_SQL
☒	rpt_user_trace_log	ODPS_SQL

确定

取消

说明 关于补数据实例的详情请参见[执行补数据并管理补数据实例](#)。

6. 在左侧导航栏，单击补数据实例，查看补数据实例的运行情况，并通过单击刷新查看实时状态。

> 文档版本：20211223

33

搜索700003169435

补数据名称请选择补数据名称

节点类型请选择节点类型

运行日期2020-03-17

我的节点

刷新 | 展开搜索

实例名称	状态	任务类型	责任人	定时时间	业务日期	开始时间	结束时间	REGION	操作
P_start_20200317_135530	运行中								批量终止
2020-03-16 00:00:00	运行中				2020-03-16 00:00:00...				
start	运行成功	虚节点	time	2020-03-17 00:11:00	2020-03-16 00:00:00...	2020-03-17 13:58:07	2020-03-17 13:58:07		DAG图 终止运行 重跑 更多

如果运行状态异常，右键单击出错节点，选择查看运行日志进行排查。



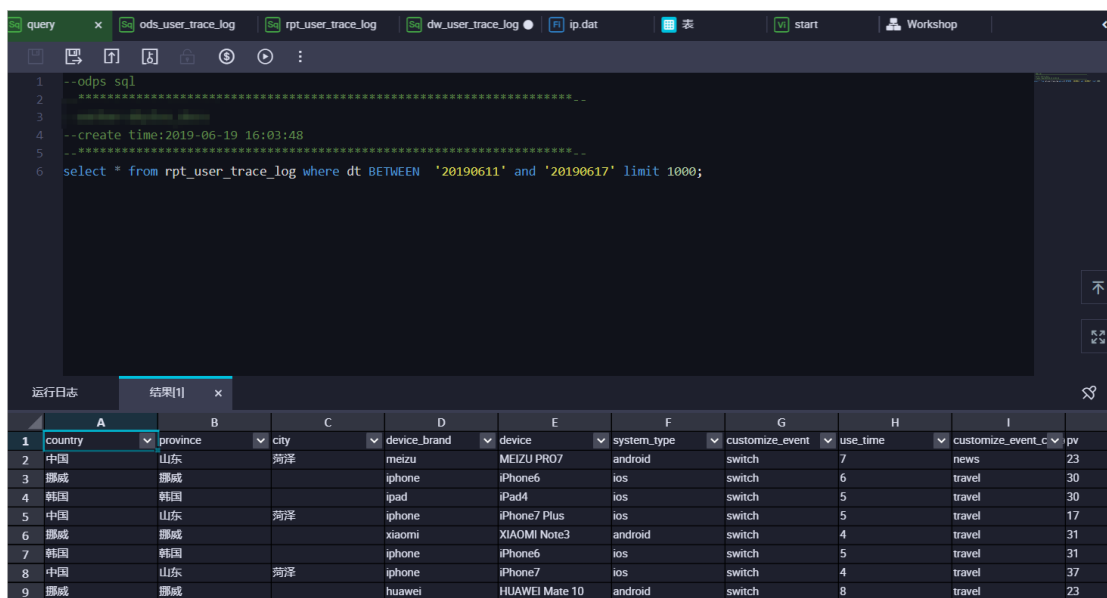
7. 补数据实例运行完成后，验证结果。
- i. 在左侧导航栏，单击业务流程Workshop > MaxCompute，右键单击数据开发，选择新建 > ODPS SQL，新建名为query的SQL节点。

- ii. 输入如下SQL语句，查询2019年6月11日到2019年6月17日之间表rpt_user_trace_log中的数据，确认数据是否成功写入rpt_user_trace_log表。

```
select * from rpt_user_trace_log where dt BETWEEN '20190611' and '20190617' limit 1000;
```

- iii. 单击图标。

查询结果如下。



The screenshot shows the ODPS SQL interface with a query executed. The results are displayed in a table with columns A through I. The data includes country, province, city, device_brand, device, system_type, customize_event, use_time, customize_event_c, and pv.

	A	B	C	D	E	F	G	H	I	
	country	province	city	device_brand	device	system_type	customize_event	use_time	customize_event_c	pv
1	中国	山东	菏泽	meizu	MEIZU PRO7	android	switch	7	news	23
2	挪威	挪威		iphone	iPhone6	ios	switch	6	travel	30
3	韩国	韩国		ipad	iPad4	ios	switch	5	travel	30
4	中国	山东	菏泽	iphone	iPhone7 Plus	ios	switch	5	travel	17
5	挪威	挪威		xiaomi	XIAOMI Note3	android	switch	4	travel	31
6	韩国	韩国		iphone	iPhone6	ios	switch	5	travel	31
7	中国	山东	菏泽	iphone	iPhone7	ios	switch	4	travel	37
8	挪威	挪威		huawei	HUAWEI Mate 10	android	switch	8	travel	23

1.5. 数据可视化展现

数据表rpt_user_trace_log加工完成后，您可以通过Quick BI创建网站用户分析画像的仪表盘，实现该数据表的可视化。

前提条件

在开始前，请确认您已经完成了[环境准备](#)、[数据准备](#)、[数据建模与开发](#)等全部步骤。

背景信息

rpt_user_trace_log表包含了country、province、city、device_brand、use_time、pv等字段信息。您可以通过仪表盘展示用户的核心指标、周期变化、用户地区分布和记录。

操作步骤

1. 使用阿里云账号登录[Quick BI控制台](#)。
2. 在Quick BI页面，单击默认空间。



3. 在默认空间中，新建MaxCompute数据源。

i. 在默认空间左侧导航栏，单击数据源。

ii. 在数据源页面，单击新建数据源。

iii. 在添加数据源界面的云数据库页签，单击MaxCompute。

iv. 在添加MaxCompute数据源对话框，配置数据源连接参数。

添加MaxCompute数据源

[查看操作指南>](#)

* 显示名称:

数据源配置列表显示名称

* 数据库地址:

http://service.odps.aliyun.com/api

* 项目名称:

请输入项目名称

* AccessKey ID:

购买此实例的accessId

* AccessKey Secret:

购买此实例的accessKey

① 温馨提示: 请添加如下白名单列表:

10.152.69.0/24,10.152.163.0/24,139.224.4.0/24

关闭

连接测试

确定

参数名	参数说明
显示名称	数据源配置列表的显示名称。名称只能由中英文、数字及下划线（_）、斜线（/）、反斜线（\）、竖线（ ）、小括号（()）、中括号（[]）组成，不超过50个字符。
数据库地址	此处有默认地址，通常无需修改。 ② 说明 数据库地址根据Region不同而变化，详细对应信息请参见Endpoint。
项目名称	MaxCompute项目名称。您可以登录MaxCompute控制台，在项目管理页查看具体的项目名称。
AccessKey ID	MaxCompute项目访问权限的AccessKey ID。您可以进入AccessKey管理进行查看。
AccessKey Secret	AccessKey ID对应的AccessKey Secret。您可以进入AccessKey管理进行查看。

- v. 配置完成数据源连接参数后，单击**连接测试**，待显示数据源连通性正常后单击**确定**。

数据源连通性正常

添加MaxCompute数据源

[查看操作指南>](#)

* 显示名称:

* 数据库地址:

* 项目名称:

* AccessKey ID:

* AccessKey Secret:

① 温馨提示: 请添加如下白名单列表:

10.152.69.0/24,10.152.163.0/24,139.224.4.0/24

关闭

连接测试

确定

4. 在数据源页面选择连接成功的MaxCompute数据源，并在rpt_user_trace_log表后单击创建数据集图标。

The screenshot shows the 'Data Sources' (数据源) management interface. On the left, a sidebar lists available data sources, with 'rpt_user_trace_log' selected and highlighted with an orange border. The main panel displays a table of data sources with columns for Name, Remarks, and Actions. The 'rpt_user_trace_log' source is highlighted in orange. A 'Create Data Set' (创建数据集) button is visible in the top right corner.

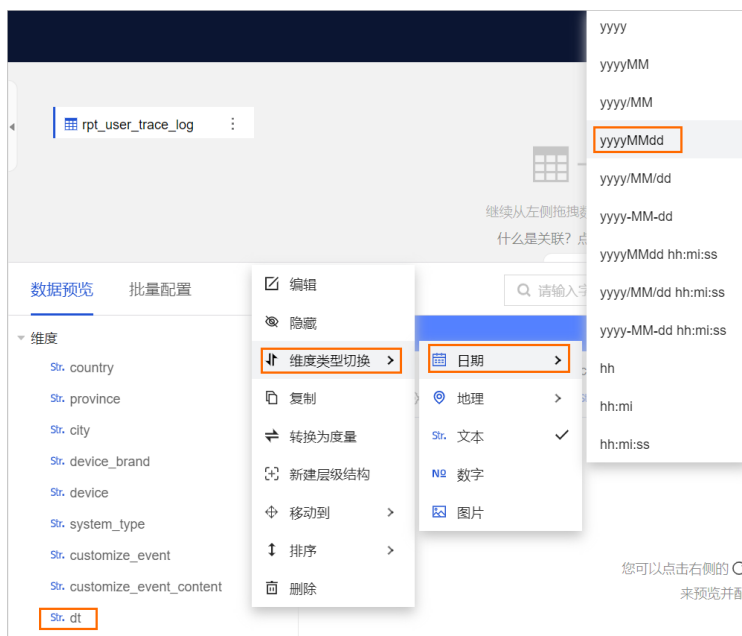
名称	备注	操作
a		
result_table1		
sale		
sale_detail		
users_phonix_mc		
rpt_user_trace_log		

5. 在数据集页面，对数据集进行编辑。

④ 说明 常见的数据集加工包括维度、度量的切换、修改维度的类型、增加计算字段、创建层次结构、修改字段的数据类型、更改度量聚合方式、制作关联模型。详情请参见概述。

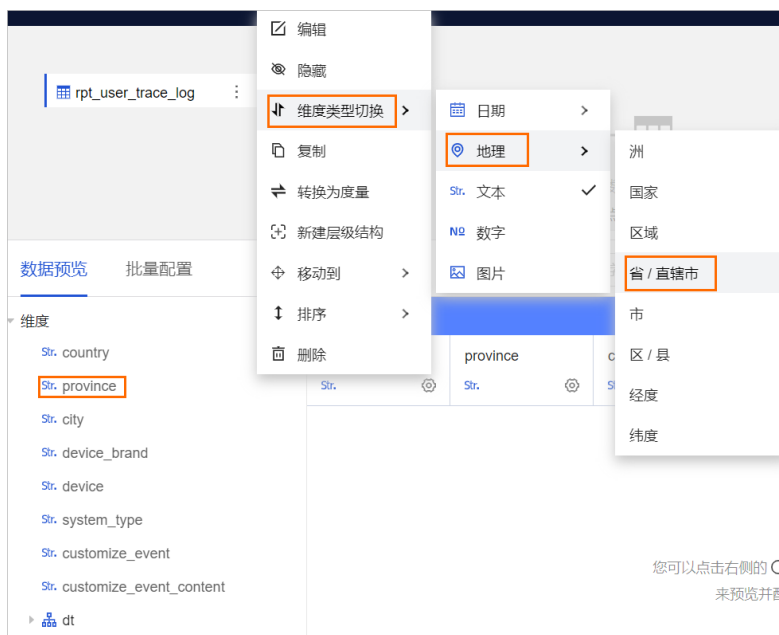
i. 切换日期字段的维度类型。

在数据预览页签，单击dt字段后的图标，选择维度类型切换 > 日期 > yyyyMMdd。



ii. 切换地理信息字段的维度类型。

a. 在数据预览页签，单击province字段后的图标，选择维度类型切换 > 地理 > 省/直辖市。



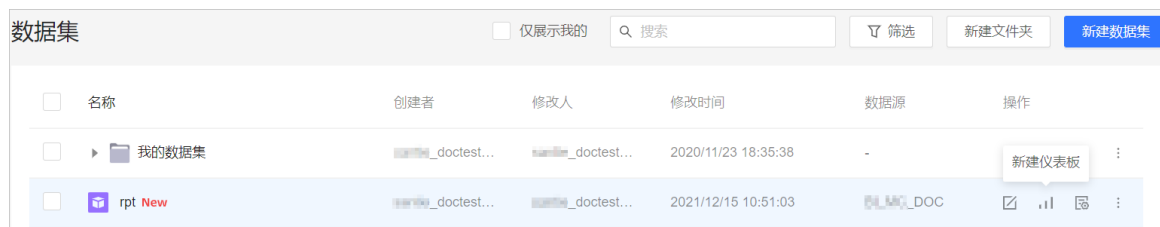
b. 在数据预览页签，单击city字段后的图标，选择维度类型切换 > 地理 > 市。

iii. 新建province层次结构。

- 在数据预览页签，单击province字段后的图标，选择维度类型切换 > 新建层级结构。
- 在新建/编辑层级结构对话框，单击确定。
- 将city字段拖拽至province层级结构下，单击保存。

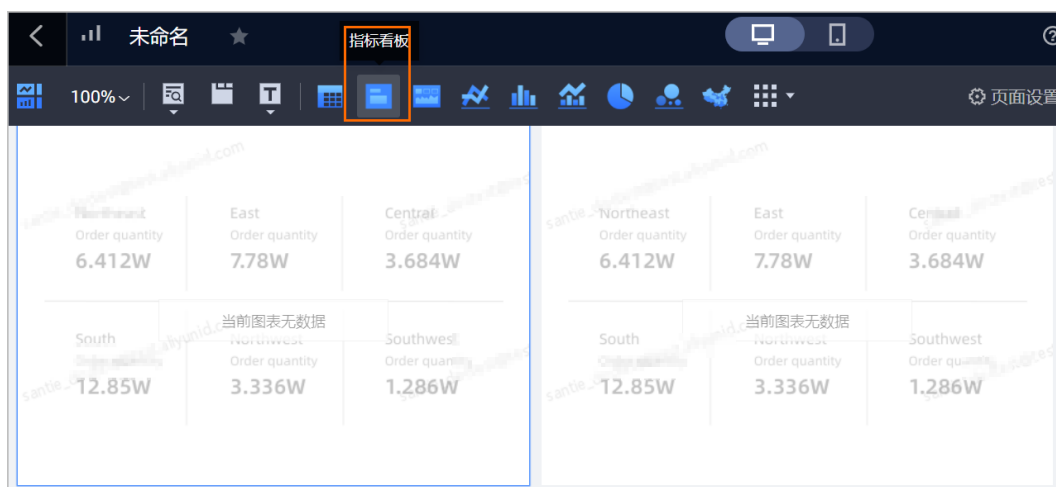


6. 在数据集页面，单击rpt数据集后的新建仪表板图标并选择常规模式。



i. 制作指标看板。

- 从仪表板空间中向空白区拖入两个指标看板，分别用于统计pv和uv。

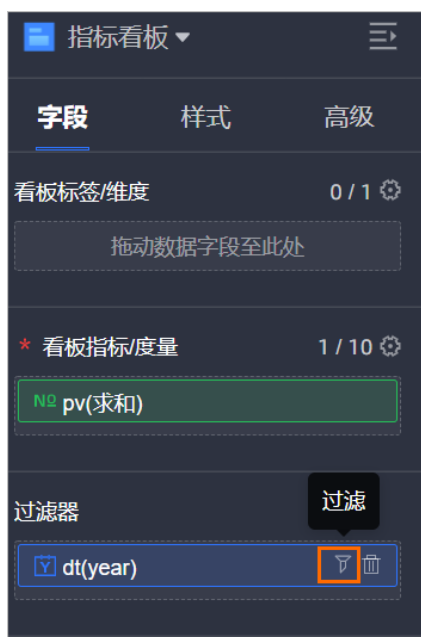


b. 选择数据来源为数据集rpt，按下图将维度字段与度量字段拖拽至指定区域。



④ 说明 pv与uv指标看板设置步骤一致，上图以统计pv为例。

c. 在字段页签，单击过滤器后的过滤图标。



- d. 在设置过滤器对话框中，设置时间区间为2019年~2019年，单击确定。

设置过滤器

数据集 rpt 选中字段 dt(year)

过滤方式 ☐ 单年 ☒ 年区间

区间类型 ☐ 开始于 ☐ 结束于 ☒ 时间区间

过滤条件 2019 2019

开始于 结束于

相对时间 相对时间

T - 2 T - 2

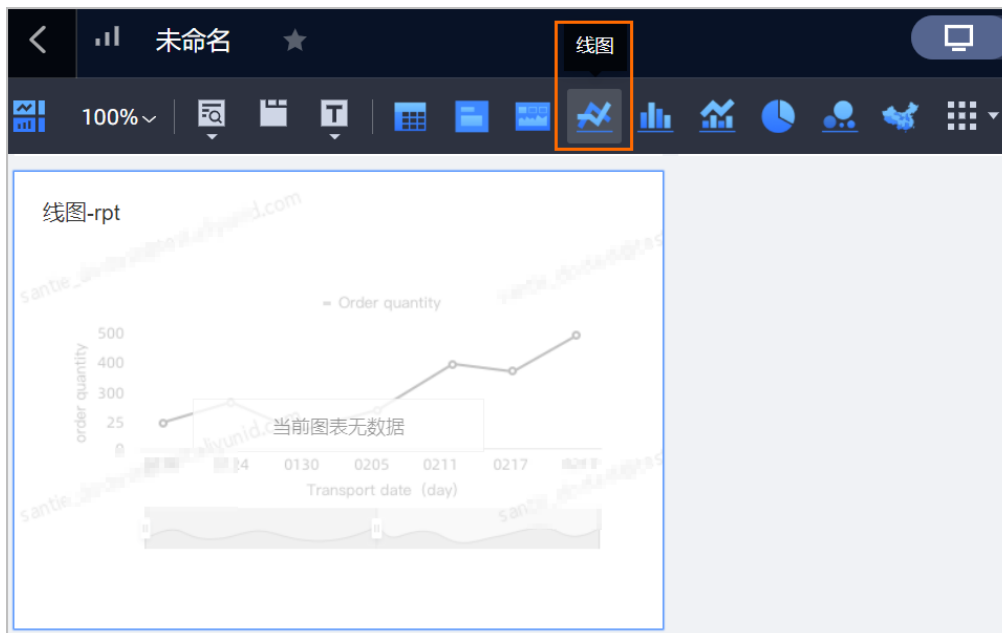
取消 确定

说明 由于数据表rpt_user_trace_log为分区表，因此必须在过滤器中设置时间区间。

- e. 在指标看板面板单击更新。

ii. 制作线图。

a. 从仪表板空间中向空白区拖入线图。



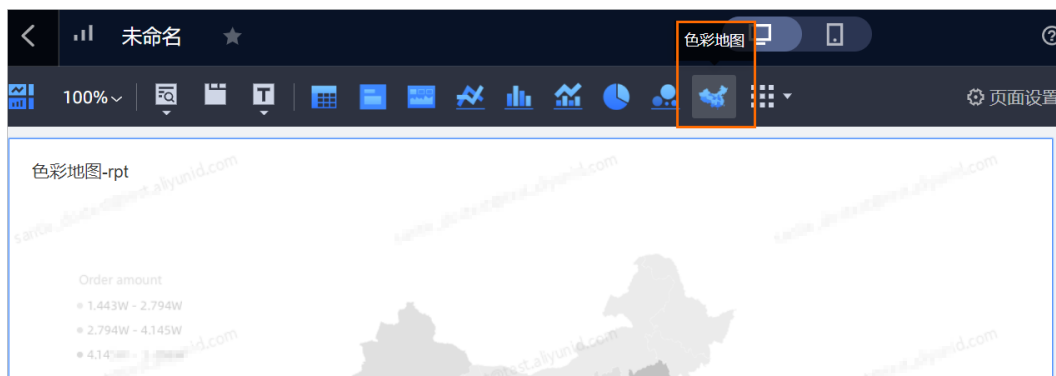
b. 选择数据源为数据集rpt，按下图将维度字段与度量字段拖拽至指定区域。



c. 设置完成后，在线图面板单击更新。

iii. 制作色彩地图。

a. 从仪表板空间中向空白区拖入色彩地图。

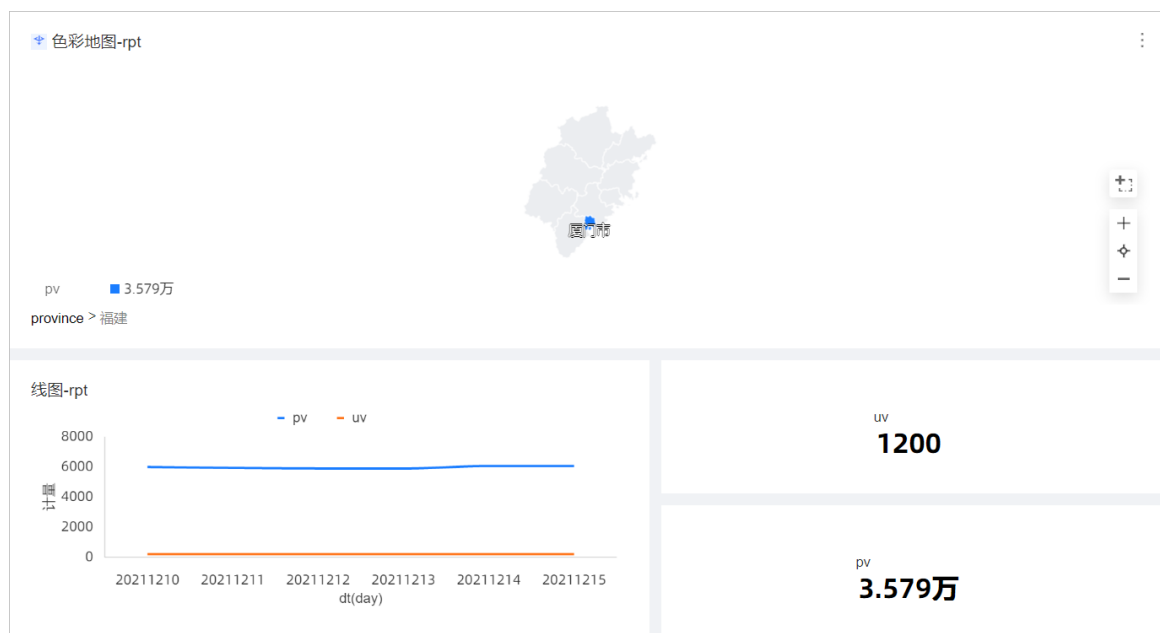


b. 选择数据源为数据集rpt，按下图将维度字段与度量字段拖拽至指定区域。



c. 设置完成后，在色彩地图面板单击更新。

7. 在仪表板页面单击保存及预览即可查看展示效果。



1.6. 数据可视化展示

数据表rpt_user_trace_log加工完成后，您可以通过Quick BI创建网站用户分析画像的仪表板，实现该数据表的可视化。

前提条件

在开始前，请确认您已经完成了[环境准备](#)、[数据准备](#)、[数据建模与开发](#)等全部步骤。

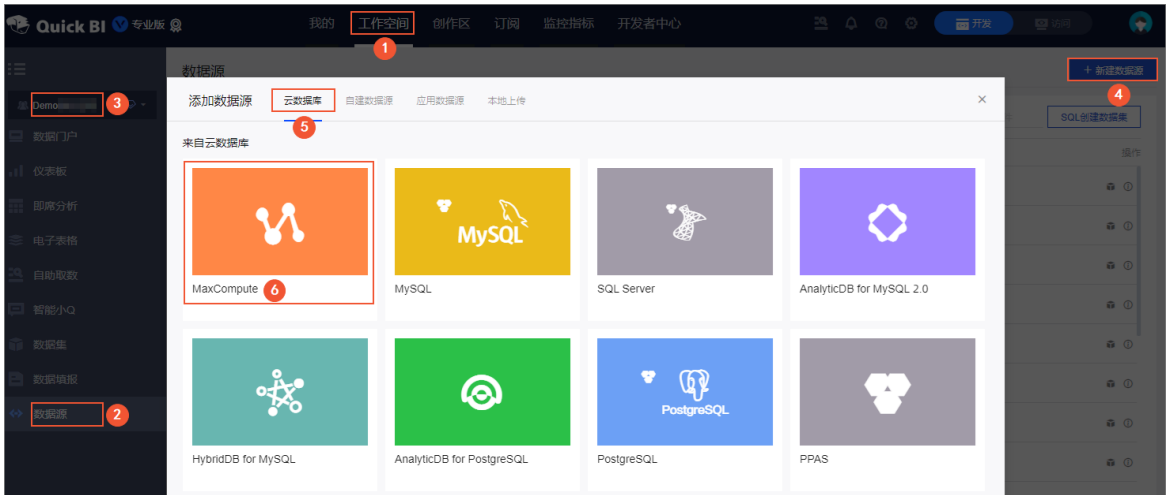
背景信息

rpt_user_trace_log表包含了country、province、city、device_brand、use_time、pv等字段信息。您可以通过以下步骤制作仪表板，用以展示用户的核心指标、周期变化、用户地区分布和记录。

- [步骤一：连接数据源。](#)
- [步骤二：创建数据集。](#)
- [步骤三：可视化展示。](#)

步骤一：连接数据源

1. 使用阿里云账号登录[Quick BI控制台](#)。
2. 在Quick BI首页，选择目标工作空间。
本例中以默认空间为例介绍。
3. 在工作台页面，按照下图指引，连接数据源。



4. 在添加MaxCompute数据源对话框，配置数据源连接参数并测试连通性。
添加成功后，您可以在数据源列表中，看到您创建的数据源。

数据源连通性正常

添加MaxCompute数据源

* 显示名称:

* 数据库地址:

* 项目名称:

* AccessKey ID:

* AccessKey Secret:

① 温馨提示: 请添加如下白名单列表:
10.152.69.0/24,10.152.163.0/24,139.224.4.0/24

关闭

连接测试

确定

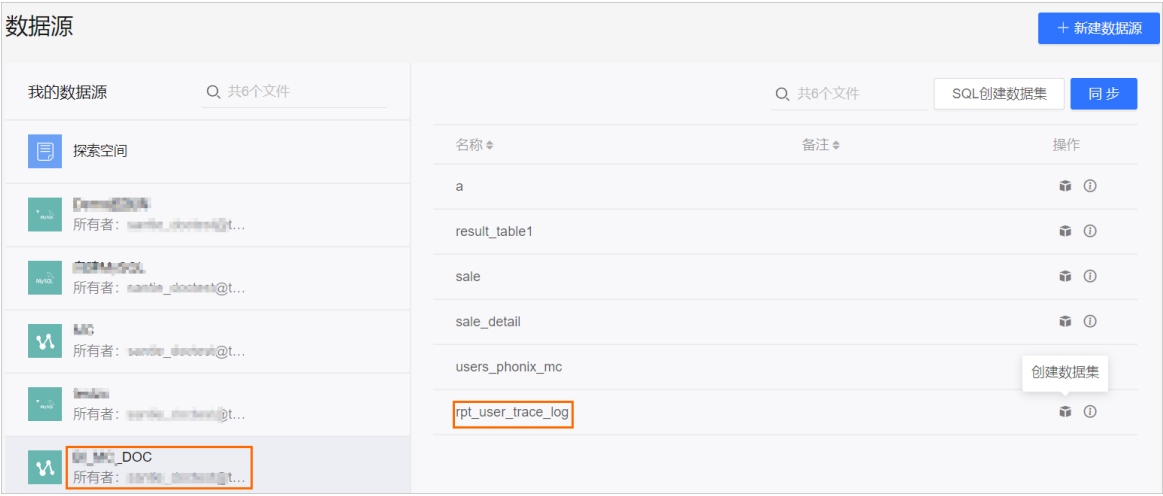
参数名	参数说明
显示名称	数据源配置列表的显示名称。名称只能由中英文、数字及下划线（_）、斜线（/）、反斜线（\）、竖线（ ）、小括号（()）、中括号（[]）组成，不超过50个字符。
数据库地址	此处有默认地址，通常无需修改。 说明 数据库地址根据Region不同而变化，详细对应信息请参见Endpoint。
项目名称	MaxCompute项目名称。您可以登录MaxCompute控制台，在项目管理页签查看具体的项目名称。

参数名	参数说明
AccessKey ID	阿里云账号或RAM用户的AccessKey ID。您可以进入 AccessKey管理 页面获取AccessKey ID。
AccessKey Secret	AccessKey ID对应的AccessKey Secret。您可以进入 AccessKey管理 页面获取AccessKey Secret。

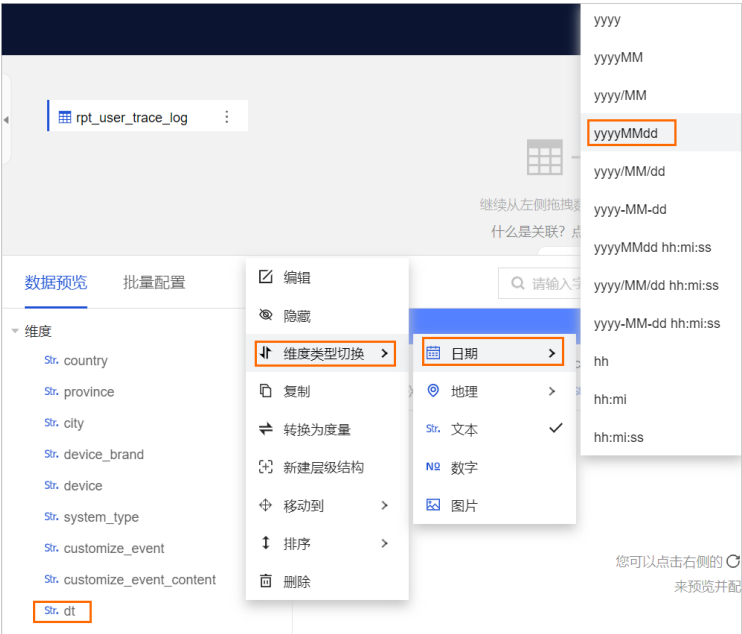
步骤二：创建数据集

Quick BI中数据集是可视化分析的基础，您可以将需要分析的数据表创建为数据集。更多操作请参见[创建并管理数据集](#)。

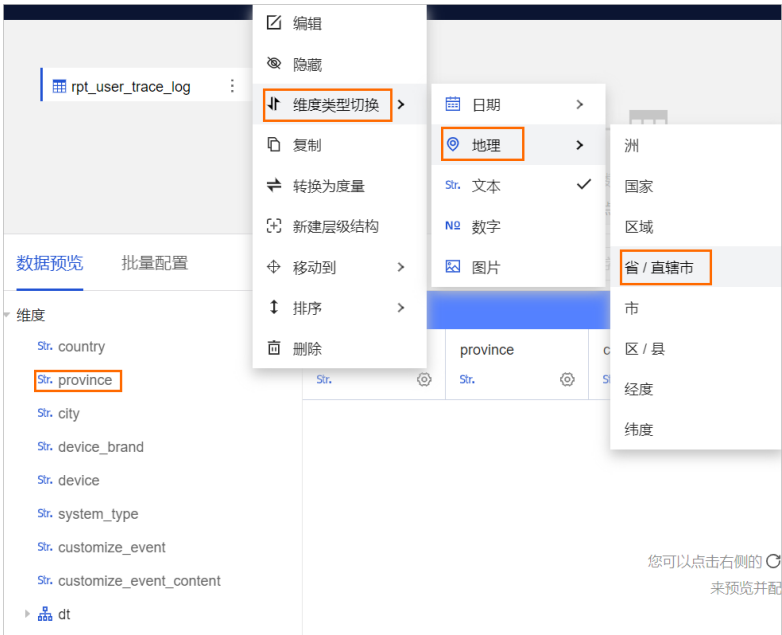
- 1. 在数据源页面选择连接成功的MaxCompute数据源，并在rpt_user_trace_log表后单击创建数据集图标。



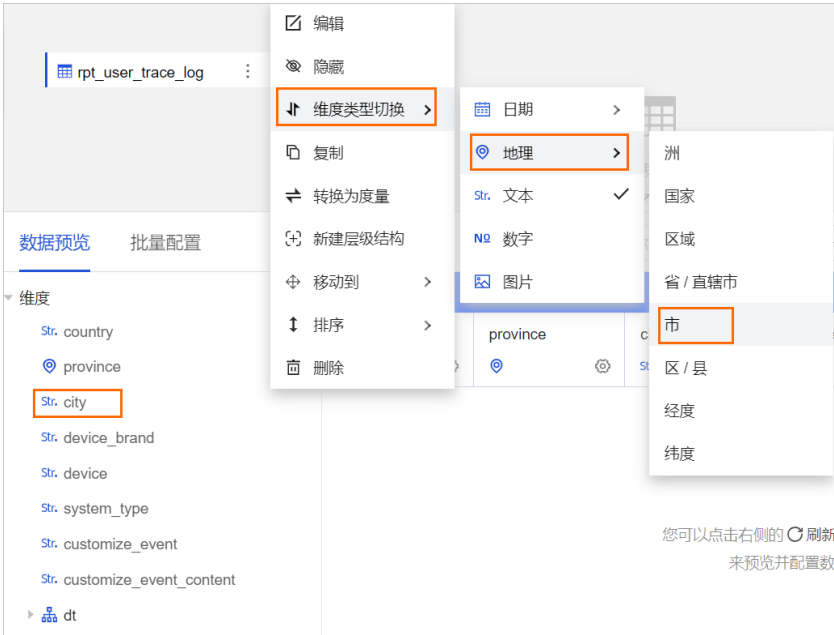
- 2. 在数据集页面，修改字段的展示格式。
 - i. 按照下图指引，修改dt字段的日期展示格式。



ii. 按照下图指引，修改province字段类型的地理信息。



iii. 按照下图指引，修改city字段类型的地理信息。



iv. 按照下图指引，新建province层级结构。



v. 将city字段拖拽至province层级结构下，钻取效果如下。



步骤三：可视化展示

可视化图表可以帮助您直观、清晰地展示数据分析结果，如需了解更多，请参见[可视化图表概述](#)。

1. 在数据集页面，单击开始分析下拉列表的创建仪表板。

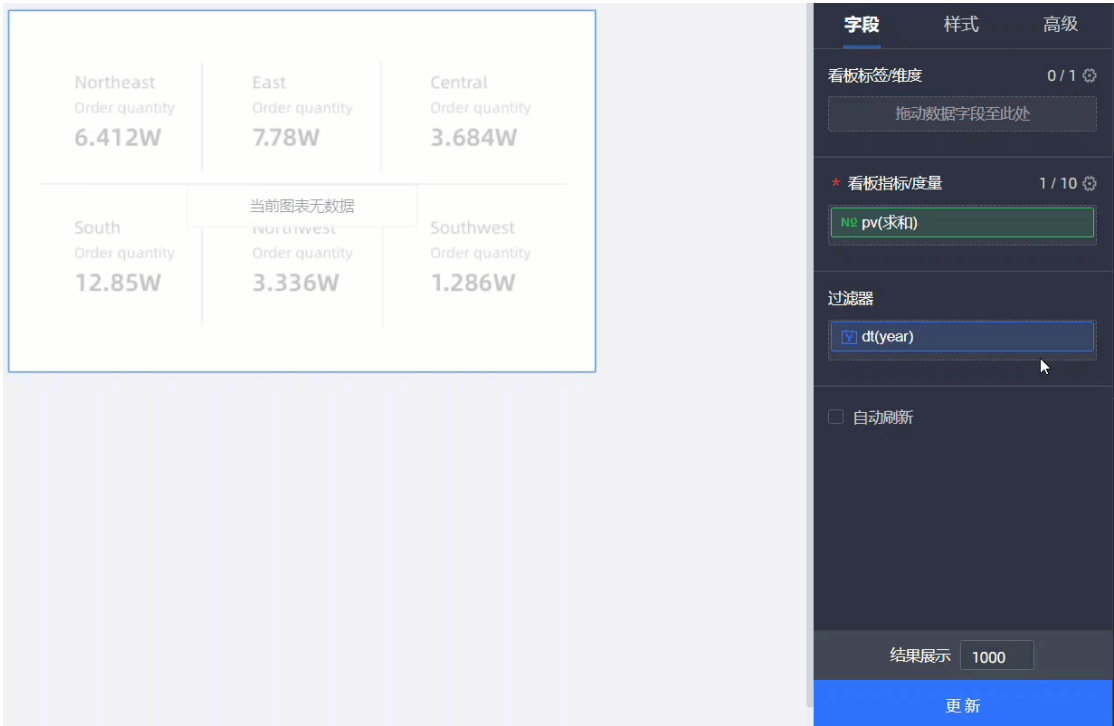


2. 制作指标看板。

i. 按照下图指引，制作pv指标看板。

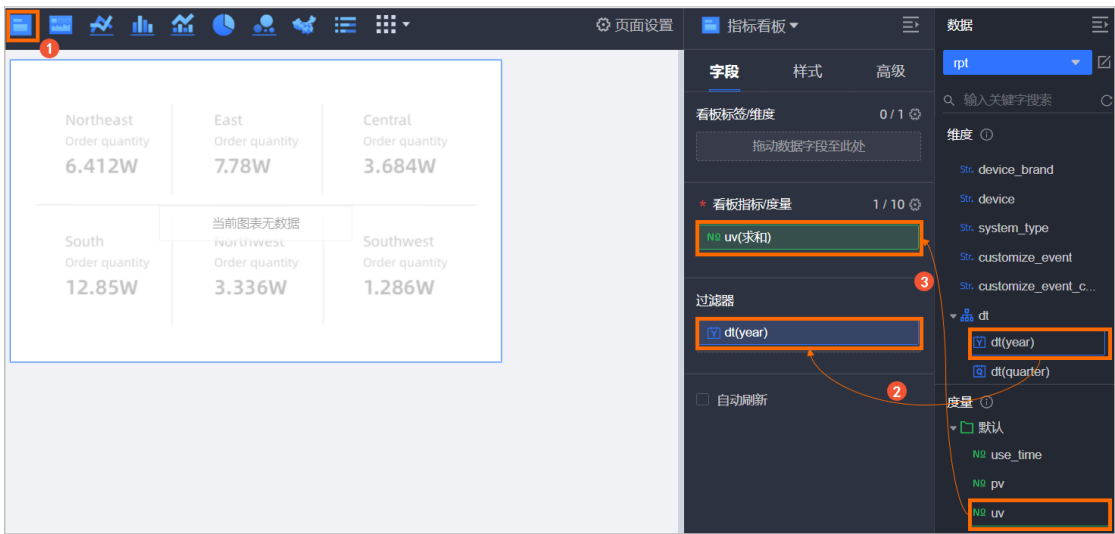


按照下图指引，设置pv指标看板过滤器。

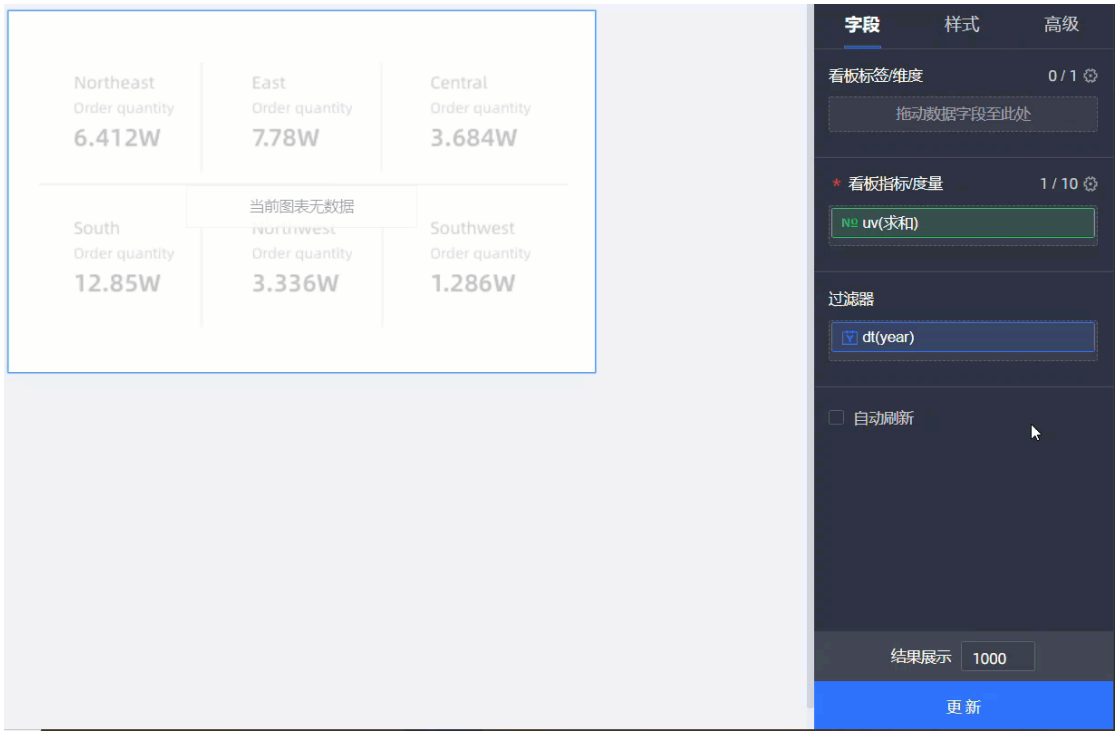


① 说明 由于数据表rpt_user_trace_log为分区表，因此必须在过滤器中设置时间区间。上图过滤器的时间区间为2021年~2021年。

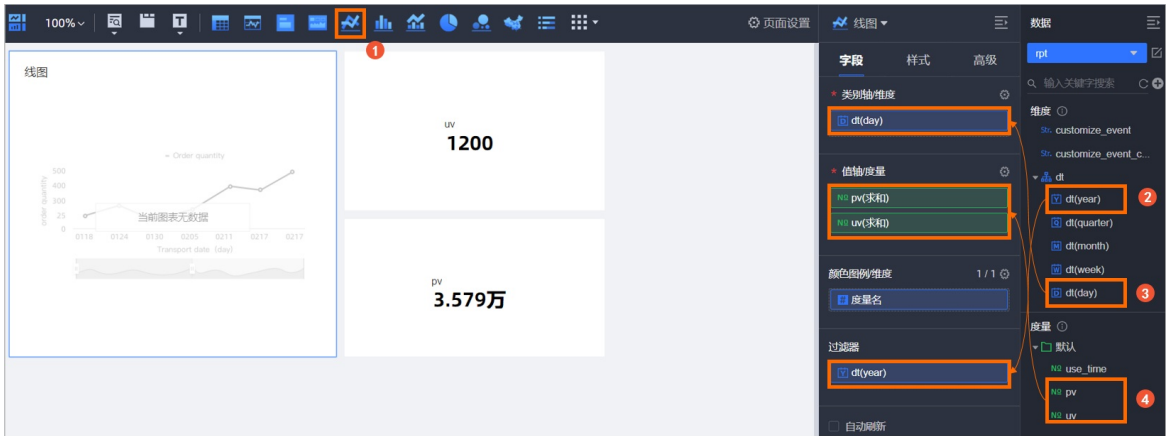
ii. 按照下图指引，制作uv指标看板。



按照下图指引，设置uv指标看板过滤器。



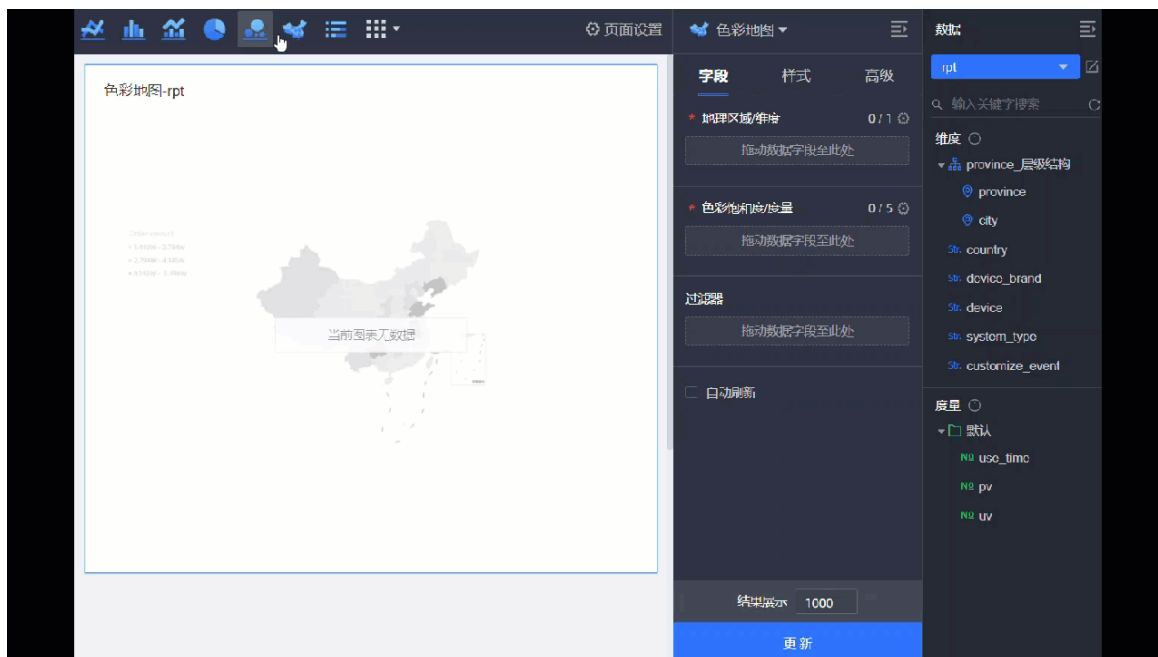
3. 按照下图指引，制作线图。



按照下图指引，设置线图过滤器。



4. 按照下图指引，制作色彩地图。



5. 设置完成后，在仪表板页面单击保存。

6. 在仪表板页面单击预览查看展示效果。

