

ALIBABA CLOUD

Alibaba Cloud

大数据计算服务
工具及下载

文档版本：20210713

 阿里云

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.查询编辑器	06
2.MaxCompute客户端 (odpscmd)	09
3.MaxCompute Studio	17
3.1. 认识MaxCompute Studio	17
3.2. 安装MaxCompute Studio	20
3.2.1. 安装IntelliJ IDEA	21
3.2.2. 安装MaxCompute Studio	23
3.2.3. 查看和更新MaxCompute Studio版本	28
3.3. 配置MaxCompute Studio	29
3.4. 管理项目连接	31
3.5. 管理数据和资源	36
3.5.1. 管理项目数据	36
3.5.2. 导入导出表数据	38
3.5.3. 可视化管理表	41
3.6. 开发SQL程序	43
3.6.1. 创建MaxCompute Script Module	43
3.6.2. 开发及提交SQL脚本	46
3.7. 开发Java程序	49
3.7.1. 概述	49
3.7.2. 创建MaxCompute Java Module	51
3.7.3. 开发UDF	52
3.7.4. 开发MapReduce	55
3.7.5. 开发Graph	59
3.7.6. 查询非结构化数据	61
3.7.7. 打包、上传和注册	64
3.8. 开发Python程序	66

3.8.1. 配置Python开发环境	66
3.8.2. 开发Python UDF	69
3.8.3. 开发PyODPS脚本	72
3.9. 管理MaxCompute作业	73
3.9.1. 作业浏览	73
3.9.2. 作业实例详情	75
3.10. 工具集成	78
3.10.1. 集成MaxCompute客户端	78
4.相关下载	81

1. 查询编辑器

MaxCompute控制台提供查询编辑器，方便您快速执行SQL语句并进行数据分析操作。本文为您介绍如何通过查询编辑器使用MaxCompute服务。

概述

MaxCompute的查询编辑器集成在DataWorks的数据分析工具上，实现编辑MaxCompute SQL、查询数据、分析数据（电子表格）、在线分享数据及下载数据等功能。您可以通过查询编辑器快捷使用MaxCompute服务的相关功能：

- 支持编辑及运行SQL命令、授权命令（例如ACL授权）。
- 支持查询加速，对于中、小数据量查询作业执行时间从分钟级缩减至秒级。查询加速详情请参见[查询加速](#)。
- 默认开放MaxCompute公开数据集，您可以通过查询编辑器基于公开数据集体验及测试MaxCompute。
- 支持在线使用电子表格Web Excel对数据查询的结果进行分析、下载或者分享给其他成员。

使用场景

查询编辑器的使用场景如下：

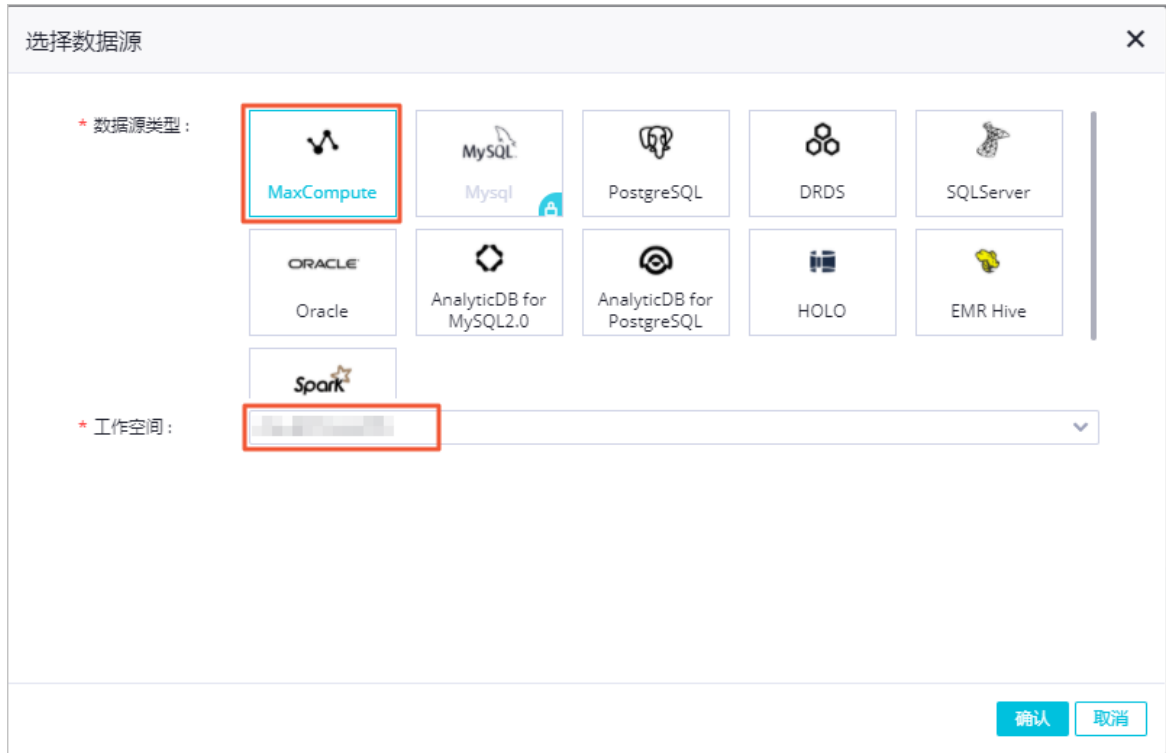
- 初次体验及测试MaxCompute的使用者：可以通过查询编辑器，使用公开数据集快速体验MaxCompute的核心功能。
- 数据分析师：您可以通过查询编辑器查询数据，并通过分析模式的Web Excel对查询结果进行分析。您也可以下载查询结果至本地，减少数据的流动，更好的保障数据安全。
- 安全管理员：MaxCompute项目右侧的[项目权限管理](#)提供了角色权限管理功能，但是正在试用过程中，很多场景需要通过命令行进行权限管理。安全管理员可通过查询编辑器快速执行大部分安全命令操作。

进入查询编辑器

1. 登录MaxCompute控制台，在左上角选择区域，单击[查询编辑](#)，即可进入查询编辑器界面。



2. 在选择数据源对话框，选择数据源类型为MaxCompute，工作空间为已创建好的项目空间。



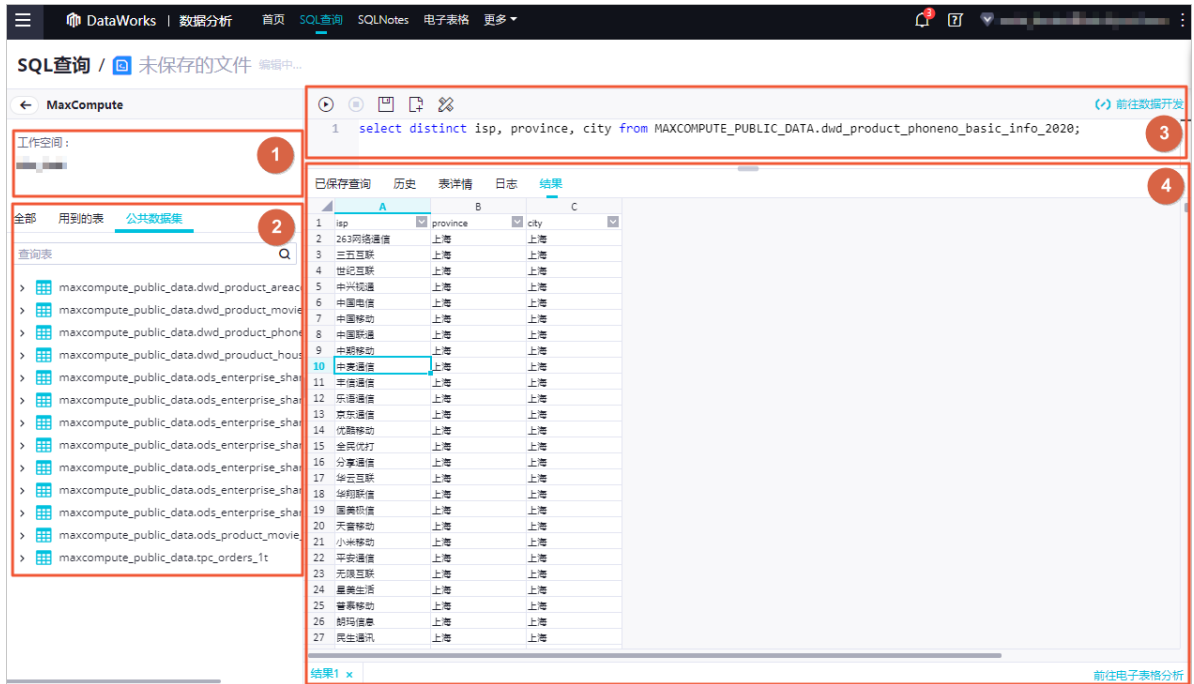
说明 如果您选择的工作空间模式为标准模式，在查询编辑器中提交作业实际是在开发项目（带dev标识）中提交。

3. 单击**确认**，即可进入查询编辑器界面。

功能介绍

查询编辑器界面分为查询模式和分析模式两种。该功能集成在DataWorks的数据分析工具上，详情请参见[数据分析](#)。基本界面功能如下：

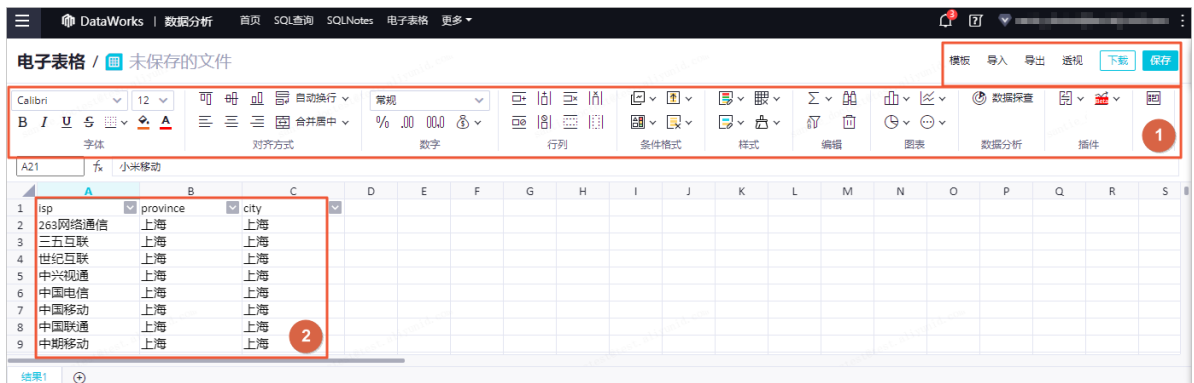
- 查询模式



序号	描述
①	工作空间名称。
②	工作空间下的所有表、使用过的表及公开数据集信息。
③	编辑器。支持执行SQL、ACL授权命令等。您进入查询编辑器界面时，可直接编辑脚本，无需先创建文件，同时也可以保存文件。
④	您可以查看已保存的查询文件、查询历史、查询日志及查询结果信息。

● 分析模式

当您在查询模式下成功运行SQL代码后，可以在查询模式界面右下角，单击前往电子表格分析进入分析模式。



序号	描述
①	电子表格编辑及管理工具栏。
②	查询结果。

2.MaxCompute客户端 (odpscmd)

您可以通过MaxCompute客户端访问MaxCompute项目并运行命令。本文为您介绍如何安装、配置和运行客户端并提供客户端相关使用说明信息。

前提条件

在使用MaxCompute客户端前，请您确认已满足如下条件：

- 待安装MaxCompute客户端的设备已安装Java 8或以上版本。
- 已创建MaxCompute项目。
更多创建MaxCompute项目操作，请参见[创建MaxCompute项目](#)。
- 使用MaxCompute客户端的RAM用户已被添加至MaxCompute项目所属的DataWorks工作空间。
更多添加工作空间成员操作，请参见[添加工作空间成员并设置角色](#)。

背景信息

安装并配置MaxCompute客户端后，您可以通过安装路径下的脚本文件或在系统的命令行窗口启动MaxCompute客户端，并执行命令。MaxCompute客户端支持如下操作：

- 获取当前登录用户信息，请参见[获取当前登录用户](#)。
- 退出MaxCompute客户端，请参见[退出MaxCompute客户端](#)。
- 获取全部命令的帮助信息，请参见[获取命令帮助](#)。
- 如果您通过系统的命令行窗口启动MaxCompute客户端，需要指定参数来执行命令，更多参数信息，请参见[启动参数](#)。


使用限制

MaxCompute客户端从v0.28.0版开始支持JDK 1.9，v0.28.0以下版本只支持JDK 1.8。您可以启动MaxCompute客户端后，在命令行界面查看客户端版本号。更多启动客户端操作，请参见[运行MaxCompute客户端](#)。

注意事项


- MaxCompute客户端的输出格式不承诺向前兼容，不同版本间的客户端命令格式及行为有差异，请勿依赖客户端的输出格式执行任何解析工作。更多客户端版本，请参见[aliyun-odps-console](#)。
- MaxCompute客户端的命令行使用两个减号（--）作为注释。

安装并配置MaxCompute客户端

 **说明** 客户端从v0.27.0版本开始支持MaxCompute 2.0新数据类型，推荐使用新数据类型。支持的数据类型列表，请参见[2.0数据类型版本](#)。

安装并配置MaxCompute客户端的操作流程如下：

1. 下载[MaxCompute客户端安装包 \(Github\)](#)。

 **说明** 如果上方链接无法下载，您可以尝试单击此处的[MaxCompute客户端安装包 \(OSS\)](#)进行下载。更多关于Github链接访问失败的问题，推荐您直接在搜索引擎中查找相关解决方案。

2. 解压下载的安装包文件，得到bin、conf、lib和plugins文件夹。

3. 进入conf文件夹，配置odps_config.ini文件。

odps_config.ini文件内容如下。

```
project_name=
access_id=
access_key=
end_point=
log_view_host=
https_check=
# confirm threshold for query input size(unit: GB)
data_size_confirm=
# this url is for odpscmd update
update_url=
# download sql results by instance tunnel
use_instance_tunnel=
# the max records when download sql results by instance tunnel
instance_tunnel_max_record=
# IMPORTANT:
# If leaving tunnel_endpoint untouched, console will try to automatically get one from odps service, which might charge networking fees in some cases.
# Please refer to Endpoint
# tunnel_endpoint=
# use set.<key>=
# e.g. set.odps.sql.select.output.format=
```

odps_config.ini文件中使用井号（#）作为注释。参数详细信息如下。

参数	是否必填	描述	示例
project_name	是	访问的目标MaxCompute项目名称。 如果您创建了标准模式的工作空间，在配置project_name时，请注意区分生产环境与开发环境（_dev）的项目名称，请参见 简单模式和标准模式的区别 。 您可以登录 MaxCompute控制台 ，在 项目管理 页签获取MaxCompute项目名称。	doc_test_dev
access_id	是	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。	无
access_key	是	AccessKey ID对应的AccessKey Secret。 您可以进入 AccessKey管理 页面获取AccessKey Secret。	无

参数	是否必填	描述	示例
end_point	是	<p>MaxCompute服务的连接地址。</p> <p>您需要根据创建MaxCompute项目时选择的地域以及网络连接方式配置Endpoint。各地域及网络对应的Endpoint值，请参见Endpoint。</p> <div style="border: 1px solid #ccc; padding: 5px; background-color: #e6f2ff;"> <p> 注意 如果Endpoint配置有误，会出现无法访问错误，请务必仔细确认。</p> </div>	http://service.cn-hangzhou.maxcompute.aliyun.com/api
log_view_host	否	<p>Logview地址。您可以通过该地址查看作业的详细运行信息，并为报错处理提供依据。固定取值为：http://logview.odps.aliyun.com。</p> <div style="border: 1px solid #ccc; padding: 5px; background-color: #e6f2ff;"> <p> 说明 推荐您配置该参数，如果不配置该参数，在作业报错时无法快速定位问题。</p> </div>	http://logview.odps.aliyun.com
https_check	否	<p>是否开启HTTPS访问机制，对访问MaxCompute项目的请求进行加密。取值范围如下：</p> <ul style="list-style-type: none"> ◦ True：采用HTTPS机制。 ◦ False：采用HTTP机制。 <p>默认值为False。</p>	True
data_size_confirm	否	<p>输入数据量的最大值，单位为GB。取值范围无限制。推荐设置为100GB。</p>	100
update_url	否	<p>预留参数，暂无需关注。</p>	无
use_instance_tunnel	否	<p>是否使用InstanceTunnel下载SQL执行结果。取值范围如下：</p> <ul style="list-style-type: none"> ◦ True：使用InstanceTunnel下载SQL执行结果。 ◦ False：不使用InstanceTunnel下载SQL执行结果。 <p>默认值为False。</p>	True
instance_tunnel_max_record	否	<p>客户端返回的SQL执行结果的最大记录数。如果use_instance_tunnel值为True，需要配置该参数。最大值为10000。</p>	10000

参数	是否必填	描述	示例
tunnel_endpoint	否	Tunnel服务的外网访问链接。如果您未配置Tunnel Endpoint, Tunnel会自动路由到MaxCompute服务所在网络对应的Tunnel Endpoint。如果您配置了Tunnel Endpoint, 则以配置为准, 不进行自动路由。 各地域及网络对应的Tunnel Endpoint值, 请参见Endpoint。	http://dt.cn-hangzhou.maxcompute.aliyun.com
set.<key>	否	设置MaxCompute项目的属性。 更多属性信息, 请参见属性列表。	set.odps.sql.decimal.odps2=true

运行MaxCompute客户端

MaxCompute客户端可通过如下方式启动, 您可以任选其中一种:

- 方式一: 在MaxCompute客户端安装路径下的bin文件夹中, 双击 odpscmd.bat 文件, 即可启动MaxCompute客户端。返回如下信息, 表明已成功连接MaxCompute项目。

```

odpscmd
Aliyun ODPS Command Line Tool
Version 0.35.1-public
©Copyright 2020 Alibaba Cloud Computing Co., Ltd. All rights reserved.
Connecting to http://service. aliyun.com/api, project: _dev
Project timezone: Asia/Shanghai
Connected!
odps@ _dev>
    
```

- 方式二: 在系统的命令行执行窗口, 进入MaxCompute客户端安装路径下的bin目录, 执行 odpscmd 命令, 即可启动MaxCompute客户端。返回如下信息, 表明已成功连接MaxCompute项目。

```

D:\ \bin>odpscmd
odpscmd
Aliyun ODPS Command Line Tool
Version 0.35.1-public
©Copyright 2020 Alibaba Cloud Computing Co., Ltd. All rights reserved.
Connecting to http://service. aliyun.com/api, project: _dev
Project timezone: Asia/Shanghai
Connected!
odps@ _dev>
    
```


您可以在上述图中的光标位置处输入命令，按下回车即可运行。命令示例及返回结果如下。

```
odps@_dev>
>select abs(-1);

ID = '...'
Log view:
http://logview.odps.aliyun.com/logview/?h=http://service.cn-hangzhou.maxcompute.aliyun.com/...
A3WGdCR2
lbnQiOlt
Y3M6b2Rw
jV3eHJzYSJdfV0sI1ZlcnNpb24iOiIxIn0=
Job Queueing...
Summary:
resource cost: cpu 0.00 Core * Min, memory 0.00 GB * Min
inputs:
outputs:
Job run time: 0.000
Job run mode: service job 2.0
Job run engine: execution engine
M1:
  bubble: 0
  instance count: 1
  run time: 0.000
  instance time:
    min: 0.000, max: 0.000, avg: 0.000
  input records:
  output records:
    AdhocSink1: 1 (min: 1, max: 1, avg: 1)
  metrics_output_count:
    AdhocSink1: 1 (min: 1, max: 1, avg: 1)
    Project1: 1 (min: 1, max: 1, avg: 1)
    Values1: 1 (min: 1, max: 1, avg: 1)
  metrics_inner_time_ms:
    AdhocSink1: 49 (min: 49, max: 49, avg: 49)    MaxInstance: 0
    GlobalInit: 41 (min: 41, max: 41, avg: 41)    MaxInstance: 0
    Project1: 0 (min: 0, max: 0, avg: 0)    MaxInstance: 0
    Values1: 0 (min: 0, max: 0, avg: 0)    MaxInstance: 0

+-----+
|  _c0  |
+-----+
|  1    |
+-----+

1 records (at most 10000 supported) fetched by instance tunnel.
```

 **说明** MaxCompute客户端支持的详细命令语法信息，请参见[常用命令列表](#)或[SQL命令及函数](#)。

获取当前登录用户

您可以在上述运行窗口执行如下命令获取当前登录用户的信息。

```
odps@project_name>whoami;
```

返回结果如下。

```
odps@_dev>whoami;
Name: ALIYUN$... aliyunid.com
Source IP: ...
End_Point: http://service....aliyun.com/api
Project: _dev
```

- Name: 当前登录的账号信息。

- Source IP：MaxCompute客户端所在设备的IP地址。
- End_Point：MaxCompute服务的连接地址。
- Project：MaxCompute项目的名称。

退出MaxCompute客户端

您可以在上述运行窗口执行如下命令退出MaxCompute客户端。

```
odps@project_name>quit;  
--等价于如下命令。  
odps@project_name>q;
```

获取命令帮助

您可以通过如下方式快速获取MaxCompute客户端的命令帮助，您可以任选其中一种：

- 方式一：在MaxCompute客户端查看命令帮助信息。
 - 查看全部命令的帮助。

```
odps@project_name>help;  
--等价于如下命令。  
odps@project_name>h;
```

- 通过指定关键字查看相关命令帮助信息。

例如获取与表操作相关的命令帮助。

```
odps@project_name>help table;  
--返回结果如下。  
Usage: alter table <tablename> merge smallfiles  
Usage: export table <tablename>  
Usage: show tables [in <project_name>] [like '<prefix>']  
list|ls tables [-p,-project <project_name>]  
Usage: describe|desc [<projectname>.<tablename> [partition(<spec>)]]  
Usage: read [<project_name>.<table_name> [(<col_name>[,...])] [PARTITION (<partition_spec>)] [line_num]
```

- 方式二：在系统的命令行执行窗口，切换到MaxCompute客户端安装路径下的bin目录，执行如下命令查看全部命令的帮助信息。在命令行执行窗口启动MaxCompute客户端时，您可以指定一系列参数，更多参数信息，请参见[启动参数](#)。

```
..\odpscmd\bin>odpscmd -h
```

启动参数

在系统的命令行执行窗口，您可以通过指定一系列参数快速执行命令，命令使用方法如下。

```
Usage: odpscmd [OPTION]...
where options include:
--help          (-h)for help
--config=<config_file>    specify another config file
--project=<prj_name>      use project
--endpoint=<http://host:port>    set endpoint
-k <n>           will skip begining queries and start from specified position
-r <n>          set retry times
-f <"file_path;">      execute command in file
-e <"command;[command;]...">    execute command, include sql command
```

启动参数说明如下。

命令	说明	命令示例
<code>--help</code> 或 <code>-h</code>	获取MaxCompute客户端的全部命令帮助信息。	<code>odpscmd --help</code>
<code>--config</code>	指定配置文件odps_config.ini的路径。默认配置文件的路径是 <code>odpscmd_public/conf/odps_config.ini</code> 。	<code>odpscmd --config=D:/odpscmd/conf/odps_config.ini</code>
<code>--project</code>	指定访问的MaxCompute项目名称。	<code>odpscmd --project=doc_test</code>
<code>--endpoint</code>	指定连接MaxCompute服务的Endpoint信息。更多Endpoint信息，请参见Endpoint。	<code>odpscmd --endpoint=http://service.cn-shanghai.maxcompute.aliyun.com/api</code>
<code>-k</code>	运行从指定位置开始的语句。当 <code>n ≤ 0</code> 时，从第一条语句开始执行。每条语句以分号 (;) 分隔。	忽略前两条语句，直接从第三条语句开始执行。 <code>odpscmd -k 3 -e "drop table table_name:create table table_name (dummy string):insert overwrite table table_name select count(*) from table_name;"</code>
<code>-r</code>	设置作业运行失败时的重试次数。	<code>odpscmd -r 2 -e "select * from sale_detail;select * from table_test;"</code>

命令	说明	命令示例
<code>-f</code>	指定读取的文件。	<p>1. 准备本地脚本文件 <code>script.txt</code>，假设存放在D盘，文件内容如下。</p> <pre>drop table if exists test_table_mj; create table test_table_mj (id string, name string); drop table test_table_mj;</pre> <p>2. 在系统命令行窗口，换到MaxCompute客户端安装路径下的bin目录，运行如下命令。</p> <pre>..\odpscmd\bin>odpscmd -f D:/script.txt;</pre>
<code>-e</code>	指定执行的命令。	<pre>odpscmd -e "select * from sale_detail;"</pre>

在Shell（或Windows命令行）执行窗口，用户可能需要使用 `odpscmd -e SQL语句` 执行得到的动态返回值，Shell的变量会获取这个动态返回值，然后在Shell中执行后续作业。此场景需要返回值不包含运行信息、表头等额外信息。为便于Shell调用，您可以通过执行 `set odps.sql.select.output.format={"needHeader":false,"fieldDelim":""}`；命令，关闭表头显示。

例如表noheader中有一列三行数据，为1、2、3。执行如下命令，将计算结果重定向stdout到目标句柄，计算结果仅包含具体数据，不包含其他额外信息：

```
odpscmd -e "set odps.sql.select.output.format={"needHeader":false,"fieldDelim":""};select * from noheader
;" >/temp/test.txt
--返回结果如下。
1
2
3
```

版本更新记录

MaxCompute客户端近期版本的更新说明如下，详细信息请单击对应版本链接获取。

版本	变更类型	描述
v0.37.4-public	功能增强	<ul style="list-style-type: none"> 优化命令对应的帮助信息。 <code>desc extended partition</code> 命令支持返回更多分区属性信息。
v0.36.0-public	新功能	支持创建External Project连接DLF（Data Lake Formation），实现湖仓一体功能。
	缺陷修复	修复下载TIMESTAMP类型数据时，纳秒（nano）部分处理不正确的问题。

3.MaxCompute Studio

3.1. 认识MaxCompute Studio

MaxCompute IDEA

MaxCompute Studio是阿里云MaxCompute平台提供的安装在开发者客户端的大数据集成开发环境工具，是一套基于流行的集成开发平台IntelliJ IDEA的开发插件，帮助您便捷、快速地进行数据开发。本文将为您介绍MaxCompute Studio的功能界面和常用的应用场景。

基本用户界面

MaxCompute Studio是IntelliJ IDEA平台上的一套插件，共享了IntelliJ IDEA的[基本开发界面](#)。

MaxCompute Studio在IntelliJ IDEA的基础上提供以下功能：

- **SQL编辑器 (SQL Editor)**：提供SQL语法高亮、代码补全、实时错误提示、本地编译、作业提交等功能。
编译器视图 (Compiler View)：显示本地编译的提示信息和错误信息，在编辑器中定位代码。
- **项目空间浏览器 (Project Explorer)**：[连接MaxCompute项目空间](#)，浏览项目空间表结构、自定义函数、资源文件。
表详情视图 (Table Details View)：提供表、视图等资源的详情显示和示例数据 (Sample Data)。
- **作业浏览器 (Job Explorer)**：浏览、搜索MaxCompute的历史作业信息。
 - 作业详情视图 (Job Details View)：显示作业的运行详细信息，包括执行计划和每个执行任务的详细信息，[Logview工具](#)能够显示的全部信息。
 - 作业输出视图 (Job Output View)：显示正在运行的作业的输出信息。
 - 作业结果视图 (Job Result View)：显示SELECT作业的输出结果。
- **MaxCompute控制台 (MaxCompute Console)**：集成了[MaxCompute客户端](#)，可以输入和执行MaxCompute客户端命令。

连接MaxCompute项目空间

使用Studio的大部分功能前需要您先[创建项目空间连接](#)。建立项目空间连接后，即可在[项目空间浏览器](#)中查看相关的数据结构和资源信息。Studio会自动为每一个项目空间连接建立一个本地的元数据备份，以提高对MaxCompute元数据的访问频率和降低延时。

创建MaxCompute项目空间连接后，才可以通过Studio进行编辑SQL脚本、提交作业、查看Job信息、打开MaxCompute控制台等操作。MaxCompute项目空间的更多详情请参见[项目空间](#)。在Studio中管理项目空间的更多详情请参见[项目空间连接](#)。

管理数据

您可以通过Studio的[项目空间浏览器](#)快速浏览项目空间的表结构、自定义函数、资源文件。通过树形控件，可以列出所有项目空间连接下的数据表、列、分区列、虚拟视图、自定义函数名称、函数签名、资源文件及类型等，并支持快速定位。

双击某个数据表，即可打开[表详情视图](#)，查看数据表的元信息、表结构和示例数据。如果您没有项目空间的相应权限，Studio会提示对应的错误信息。

Studio集成了[MaxCompute Tunnel](#)工具，可以支持本地数据的上传和下载，更多详情请参见[导入并导出数据](#)。

编写SQL脚本

您可以在Studio中编写MaxCompute SQL脚本，非常方便。

1. 打开Studio，导航至File > New > Project 或者File > New > Module...
2. 创建一个MaxCompute Studio类型的项目或者模块。
3. 导航至File > New > MaxCompute SQL 脚本 或者在项目或模块名称处，单击右键，选择New > MaxCompute SQL 脚本，即可创建一个MaxCompute SQL脚本文件。

说明 创建MaxCompute SQL脚本时，Studio会提示您选择一个关联的MaxCompute项目空间，您可以通过SQL编辑器上的工具条最右侧的项目空间选取器进行更改，编辑器会根据SQL脚本关联的项目空间对SQL语句自动进行元数据（比如表结构等）的检查并汇报错误，提交运行时也会发送到关联的项目空间执行。更多详情请参见[编写SQL脚本](#)。

SQL代码智能提示

Studio提供的SQL编辑器可以根据您写入的代码，智能提示SQL语句的语法错误、类型匹配错误或者警告等，实时地标注在代码上。

```

1
2 select a.key, b.value + c.value value
3 from src a join srcp b, src c
4 where a.key = b.key and a.key = c.key;
    table meta.srcp cannot be resolved
    
```

通过代码补全功能，Studio可以根据代码上下文，提示您项目空间名称、表、字段、函数、类型、代码关键词等，并根据您的选择，自动补全代码。

```

1 select * from meta
    meta
    meta_audit_asids
    meta_audit_java_sandbox_events
    meta_audit_odps_authentication
    meta_audit_odps_authentication
    meta_audit_odps_authorization_m
    meta_audit_odps_authorization
    
```

编译和提交作业

- 编译作业

单击SQL编辑器工具条上的



图标，可以对SQL脚本执行本地编译，如果有语法或者语义错误，编译器窗口会报告错误。

```

7
8 -- select clause in the front
9 select * from table_test;
10
11 -- from clause in the front
12 from table_test table_alias select *;
13
14 -- table name with project prefix
    
```

MaxCompute Compiler

- Information: Parsing ...
- Information: Type checking ...
- Information: Latency.compiler_parse_error : 44170
- Information: Build failed(2)
- Error:(9, 15) table meta.table_test cannot be resolved
- Error:(12, 6) table meta.table_test cannot be resolved

- 提交作业

单击SQL编辑器工具条上的



图标，会在本地编译之后，把SQL脚本提交到MaxCompute指定的项目空间排队执行。

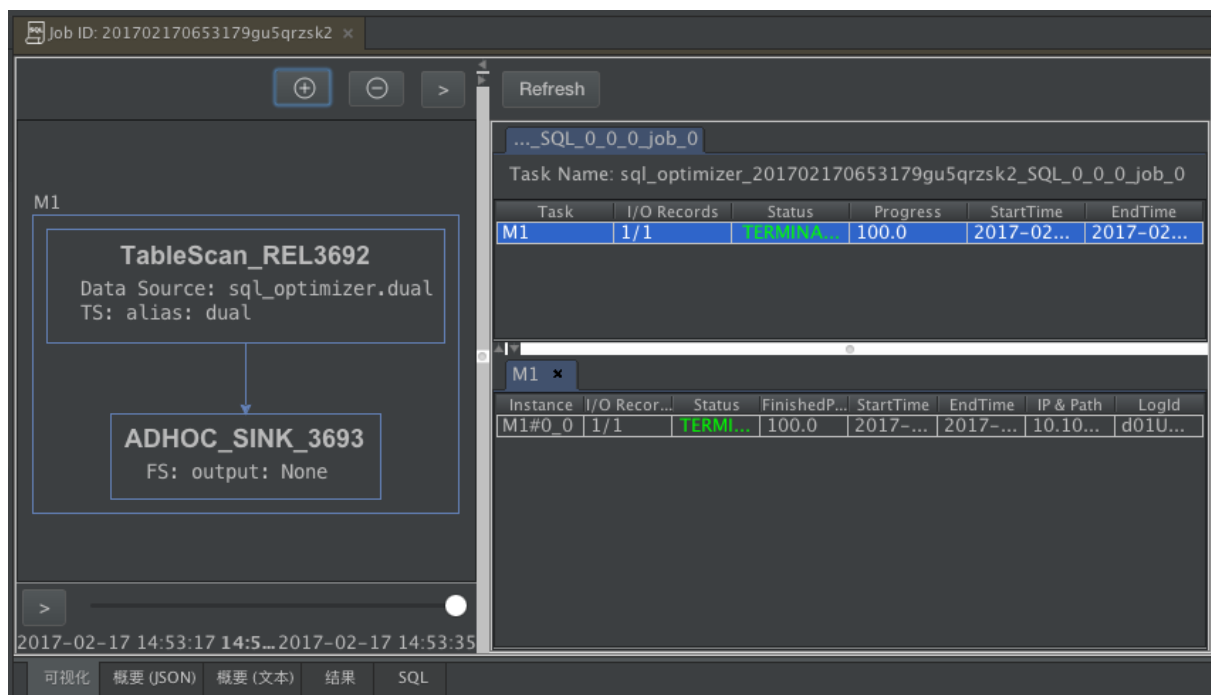
查看历史作业

打开作业浏览器，您即可查看指定项目空间上近期执行的作业。

说明 这个列表只能显示以当前连接使用的用户ID提交的作业。

Instanceld	Status	Owner	StartTime	EndTime
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	SUCCESS	ODPS...	2017-...	2017-...
20170...	FAILED	ODPS...	2017-...	2017-...

双击其中一个作业，便可查看作业的详情信息。



如果知道一个任务的Logview URL，可以导航至MaxCompute > Open Logview，打开该任务的详情页面。

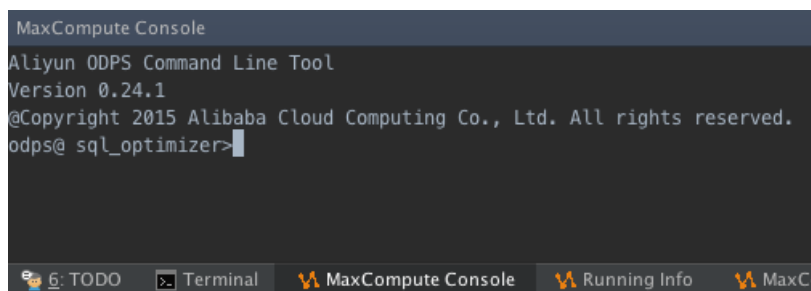
开发MapReduce和UDF

Studio还支持MapReduce和Java UDF的开发。

连接MaxCompute客户端

Studio集成了最新版本的MaxCompute客户端，您也可以在Studio的配置页面中指定本地已经安装好的MaxCompute客户端路径。

您在项目空间浏览器中选定一个项目空间，右键单击菜单选择Open in Console即可打开MaxCompute控制台窗口。



后续步骤

现在，您已经学习了MaxCompute Studio的功能界面和常用的应用场景，您可以继续学习下一个教程。在该教程中您将学习如何安装MaxCompute Studio。详情请参见[工具安装与版本信息](#)。

3.2. 安装MaxCompute Studio

3.2.1. 安装IntelliJ IDEA

MaxCompute Studio是运行在IntelliJ IDEA上的插件，本文将为您介绍如何安装MaxCompute Studio的基础平台IntelliJ IDEA。

前提条件

IntelliJ IDEA支持在Windows、macOS或者Linux操作系统上安装，硬件及系统环境要求请参见[Requirements for IntelliJ IDEA](#)。

操作步骤

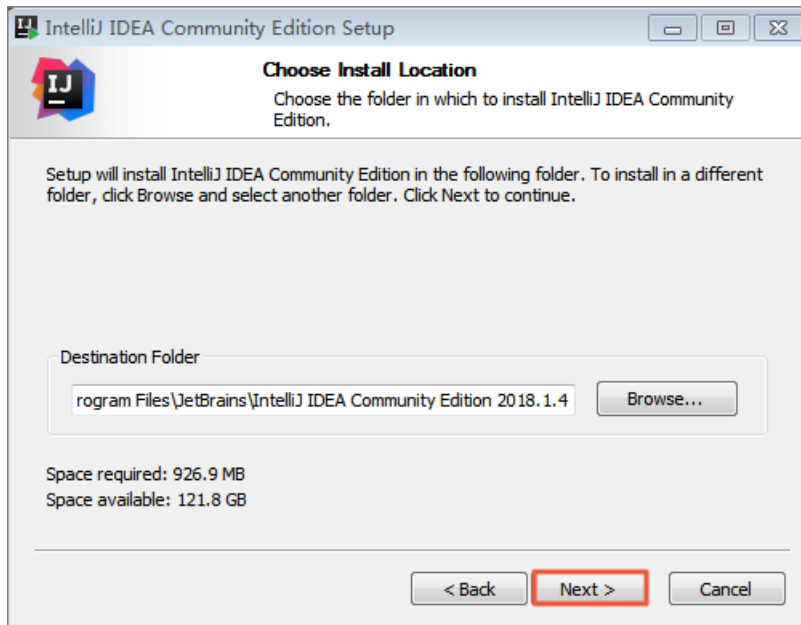
1. 下载[IntelliJ IDEA](#)。请根据操作系统（Windows、macOS、Linux）下载对应IntelliJ IDEA版本。本文以Windows操作系统为例，支持IntelliJ IDEA 14.1.4以上版本。

 说明 IntelliJ IDEA支持PyCharm、Ultimate版本或免费的Community版本。

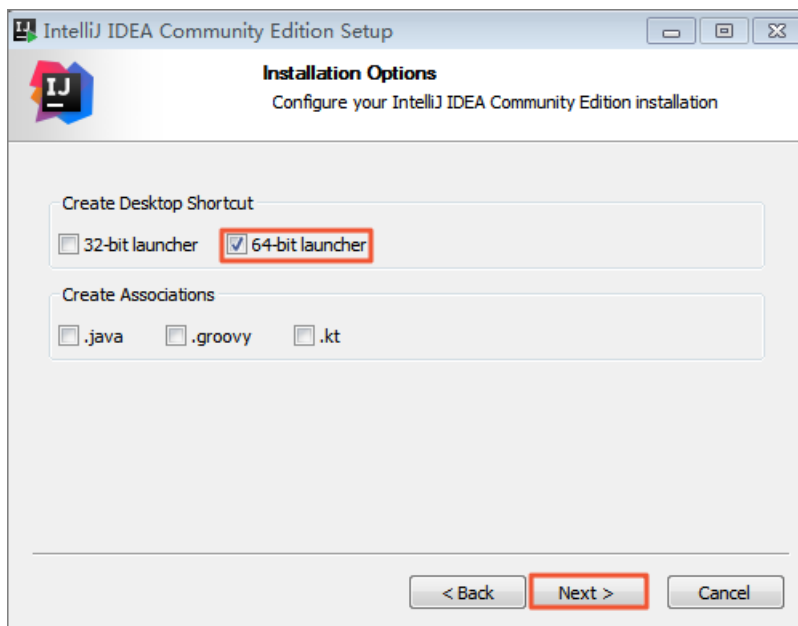
2. 下载完成后，双击安装程序，进入安装界面，单击Next。



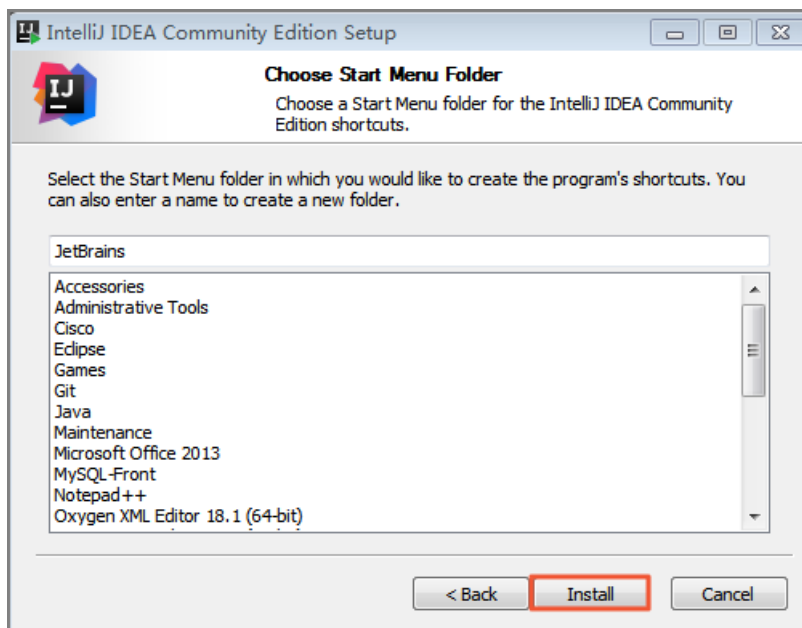
3. 指定安装目录后，单击Next。



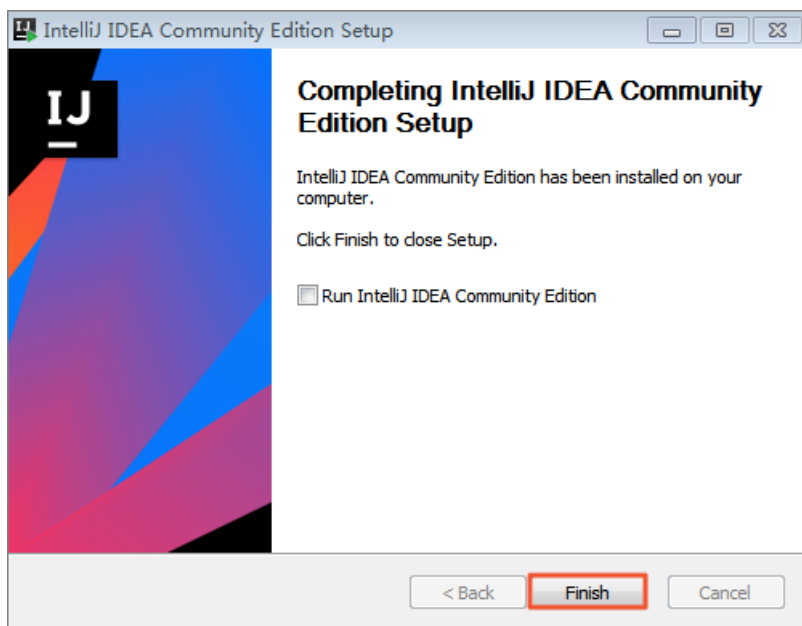
4. 选择相应的系统类型，单击**Next**。如下图所示。



5. 单击**Install**，开始安装。



6. 安装完成后，单击Finish。



后续步骤

安装MaxCompute Studio插件，详情请参见[安装MaxCompute Studio](#)。

3.2.2. 安装MaxCompute Studio


本文为您介绍如何安装MaxCompute Studio，即在IntelliJ IDEA上安装MaxCompute Studio插件。

前提条件

MaxCompute Studio安装环境需要满足以下条件：

- Window、macOS或Linux系统客户端。
- 安装IntelliJ IDEA的18.2.4以上版本。支持Ultimate版本、PyCharm版本和免费的Community版本。

- 已安装JRE 1.8。最新的IntelliJ IDEA版本捆绑了JRE 1.8，无需单独安装。
- 已安装JDK 1.8。（此项仅在需要开发和调试Java UDF时安装。）

 **说明** MaxCompute Studio从0.28.0版本起开始支持JDK 1.9，之前的版本只支持JDK 1.8。

背景信息

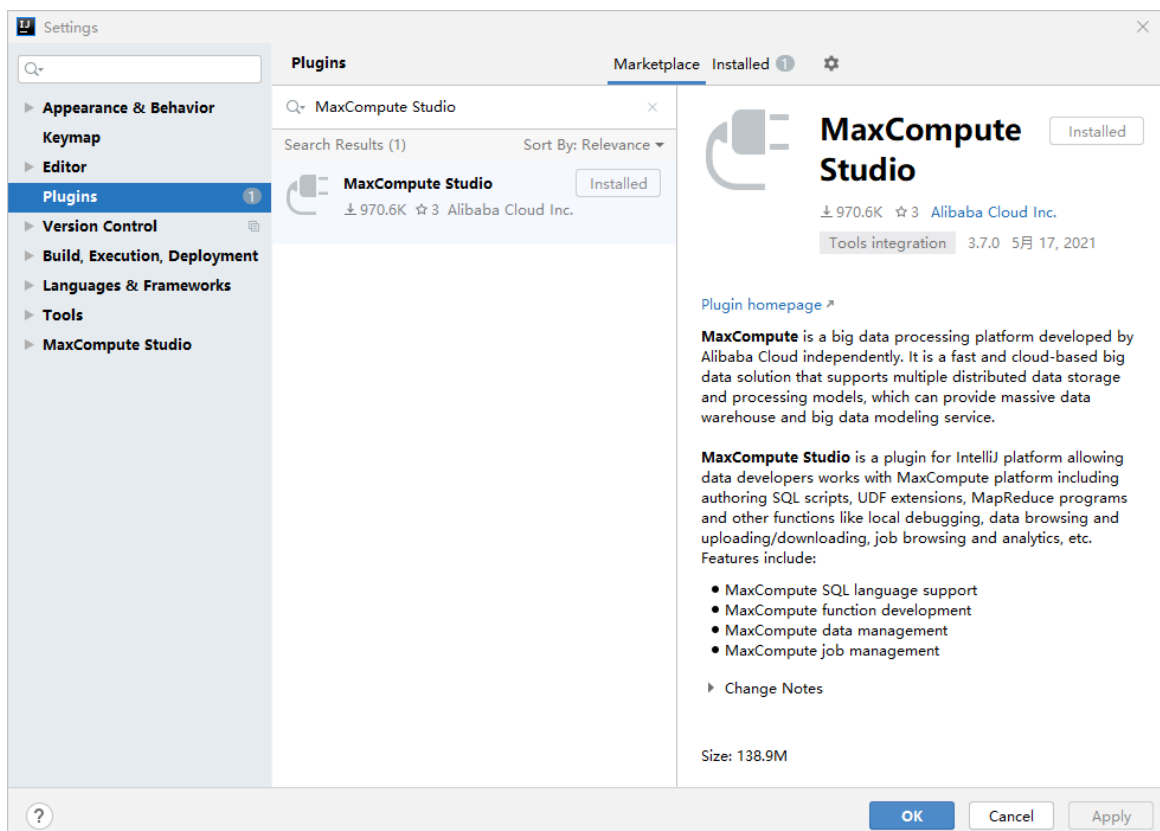
MaxCompute Studio是IntelliJ IDEA的插件，有以下两种安装方式：

- 通过插件库在线安装（推荐）。
- 通过本地文件安装。

在线安装（推荐）

MaxCompute Studio插件已对全部公网用户开放，您可以通过IntelliJ IDEA官方插件库安装。

1. 在顶部菜单栏，单击File > Settings...
2. 在Settings对话框的左侧导航栏上，单击Plugins。
3. 在Plugins对话框的搜索框中搜索MaxCompute Studio。
4. 找到MaxCompute Studio插件，单击Install进行安装。



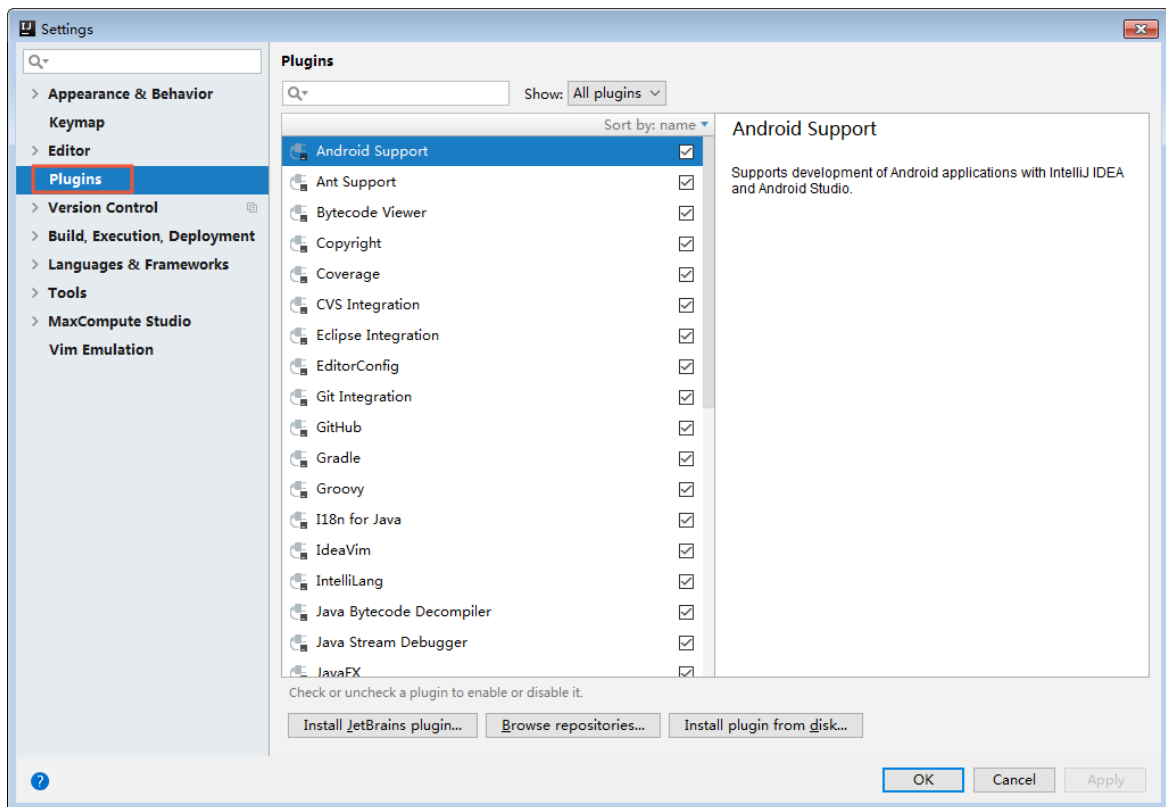
5. 确认安装后，重新启动IntelliJ IDEA，完成插件安装。

本地安装

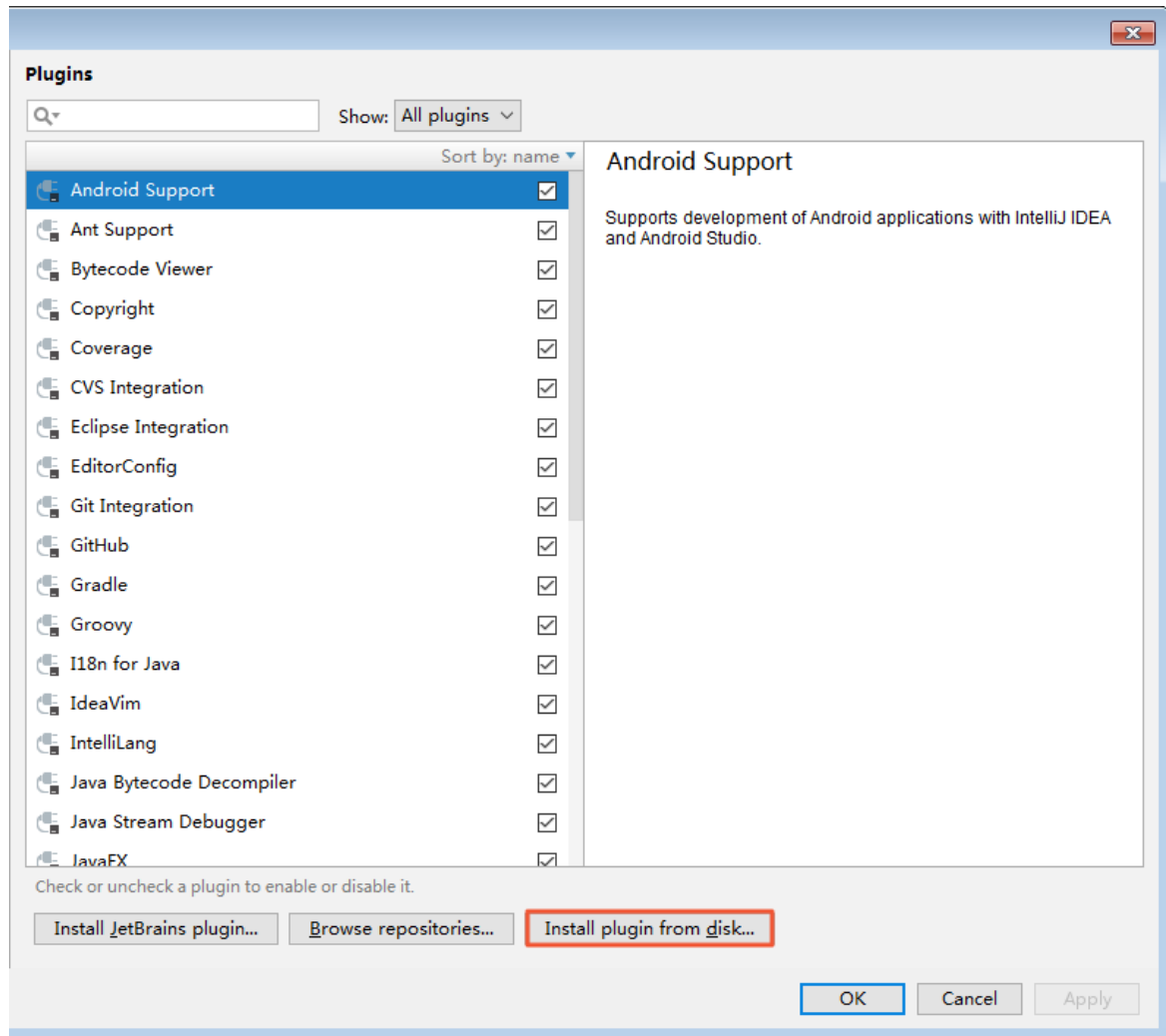
您可以将MaxCompute Studio下载至本地环境中，然后在本地环境中进行安装。

1. 下载MaxCompute Studio插件包。

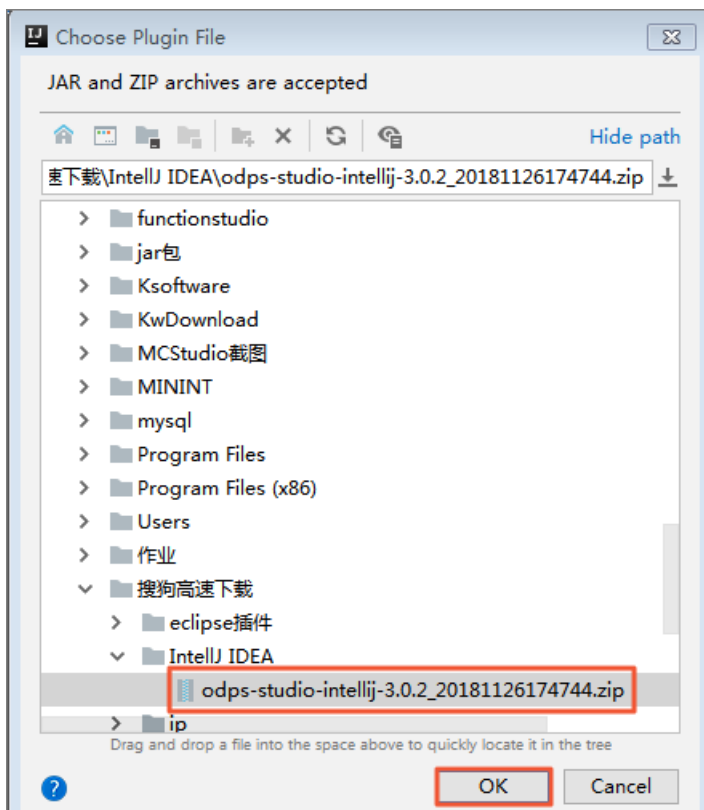
2. 在IntelliJ IDEA界面顶部菜单栏，单击File > Settings > Plugins。



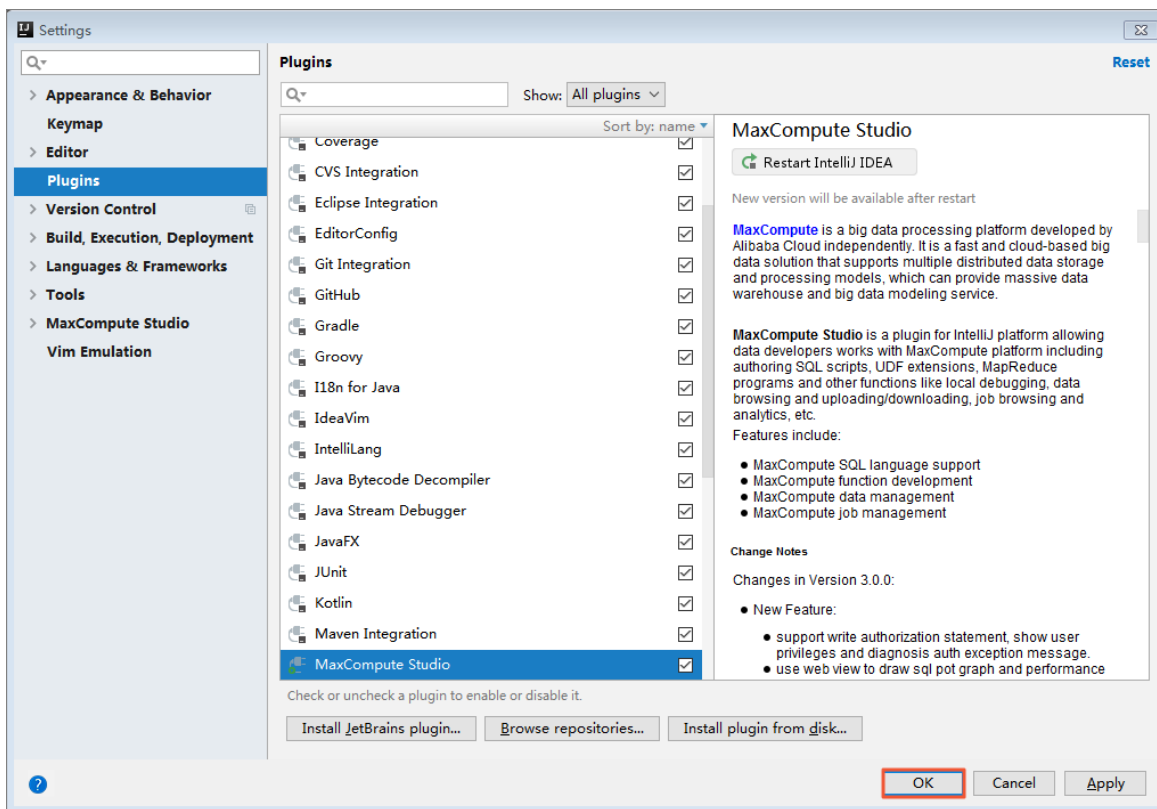
3. 在Plugins对话框，单击Install plugin from disk...。



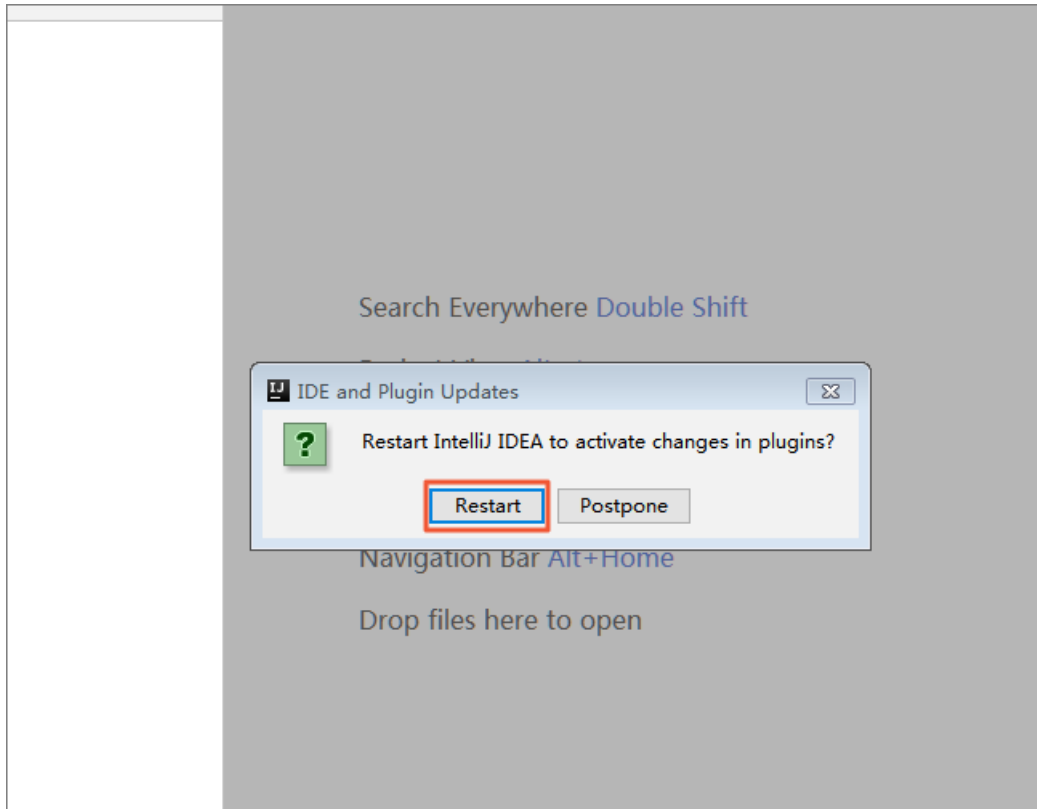
4. 在Choose Plugin File对话框中，找到已下载至本地的插件文件并选中，单击OK。



5. 返回插件首页后，单击OK，开始安装本地插件。



6. 安装完成后，弹出重新启动的提示窗口，单击Restart，重新启动IntelliJ IDEA。



后续步骤

安装MaxCompute Studio插件后，需要配置MaxCompute项目连接才可以使用数据和资源，详情请参见[管理项目连接](#)。

3.2.3. 查看和更新MaxCompute Studio版本

本文为您介绍如何查看和更新MaxCompute Studio版本。

查看MaxCompute Studio版本信息

1. 在顶部菜单栏，单击File > Settings...
2. 在Settings对话框的左侧导航栏，单击MaxCompute Studio。
3. 在MaxCompute Studio对话框的Updates Checking区域可以看到当前MaxCompute Studio的版本号，以及最近发布的版本信息。

检查新版本

- 默认情况下，MaxCompute Studio会自动检测新版本，当有新的可用版本时，会自动通知您。



收到更新提示后，您可以选择：

- 安装：在更新提示中，单击安装，将会自动下载并安装此新版本，安装完成后重启IntelliJ IDEA即可。
 - 配置：在更新提示中，单击配置，您可以配置是否自动检查新版本。
- 如果关闭了自动更新功能，您可以通过以下步骤检查MaxCompute Studio的版本并选择安装：

- i. 在顶部菜单栏，单击File > Settings...
- ii. 在Settings对话框的左侧导航栏，单击MaxCompute Studio。
- iii. 在MaxCompute Studio对话框的Updates Checking区域可以看到当前MaxCompute Studio的版本号，以及最近发布的版本信息。
- iv. 单击Check new versions检测最新可用版本并安装。

 说明 您可以选中Automatically checks for new version打开自动检测新版本功能。

3.3. 配置MaxCompute Studio

本文为您介绍如何配置MaxCompute Studio以及各项配置项信息。

安装MaxCompute Studio插件后，在顶部菜单栏，单击File > Settings...，即可进入MaxCompute Studio配置页面。

MaxCompute Studio配置选项页


在Settings页面左侧导航栏上，单击MaxCompute Studio。MaxCompute Studio配置选项页提供以下配置项：

- 基本参数
 - Local meta store location：本地元数据仓库存储路径。指定本地存储MaxCompute项目空间元数据的路径。MaxCompute Studio的默认设置是本地用户目录下的`.odps.studio\meta`目录。
 - Table preview rows：表的最大预览行数。
 - 本地作业保存目录：MaxCompute Studio作业的本地保存路径。
 - Python path to resolve UDF：Python的本地安装路径。
- 版本更新选项
 - Automatically checks for new version：控制MaxCompute Studio是否自动检查更新版本。默认情况下是选中状态，支持自动更新。
 - Check new versions：用于手动检查新版本。单击此按钮后，如果有新版本可以更新，将显示Install new version按钮。单击Install new version安装最新版本软件，安装完成后需要重启IntelliJ IDEA。

SDK & Console配置选项页

在Settings页面左侧导航栏上，单击MaxCompute Studio > SDK & Console，进入SDK & Console配置选项页。

SDK & Console配置选项页面提供了Installed Location配置项，用以指定本地安装MaxCompute客户端的安装路径。MaxCompute Studio会自动检测路径中安装的MaxCompute客户端的版本，如果检测失败，会返回错误信息。

 说明 MaxCompute Studio 2.6.1之后的版本自带了最新的MaxCompute客户端，不需要您特别指定。如果您希望使用自己特定版本的MaxCompute客户端，可以在此处指定路径。

MaxCompute SQL配置选项页

在Settings页面左侧导航栏上，单击MaxCompute Studio > MaxCompute SQL，进入MaxCompute SQL配置选项页。

MaxCompute SQL配置选项页面提供以下配置项：

- 语法高亮
 - 选中Enable syntax coloring，启动语法高亮功能。
- 代码自动补全
 - 选中Enable code completion，启动代码自动补全功能。
 - 选中Invoke code completion when you enter a space，启动输入空格时自动结束代码补全功能。
- 代码格式化
 - 选中Enable code formatting，启动代码格式化功能。
- 脚本提交选项
 - 选中Show job detail when script submitted，脚本提交时会显示作业详情。
 - 选中Record sql history when script submitted，脚本提交时记录SQL历史记录。
 - 选中Show sql cost confirm dialog when script submitted，提交脚本时显示SQL成本确认对话框。
 - 选中Pin sql execution result tab by default，默认情况下锁定SQL执行结果选项卡。
 - 作业名称job name：
 - use script file name as default：默认使用脚本的名称作为工作名称。
 - input job name when first submit：第一次提交时输入作业名称。
 - input job name for every submit：每一次提交时输入作业名称。
- 编译器选项


此处设置的选项为全局默认的编译器选项。以下选项还可以在SQL编辑器的工具栏中为每个文件单独设置。

 - 编译器模式（Compiler Mode）
 - 单句模式（Statement Mode）：在该模式下，编译器将SQL文件中的单条语句作为单元进行编译、提交。
 - 脚本模式（Script Mode）：在该模式下，编译器将整个SQL文件作为单元进行编译、提交。脚本模式有利于编译器和优化器最大程度地优化执行计划，提高整体执行效率。
 - 类型系统
 - 旧有类型系统（Legacy TypeSystem）：原有MaxCompute的类型系统。
 - MaxCompute 类型系统（MaxCompute TypeSystem）：MaxCompute 2.0引入的新类型系统。
 - Hive 类型系统（Hive Compatible TypeSystem）：MaxCompute 2.0引入的Hive兼容模式下的类型系统。
 - 编译器版本
 - 默认编译器（Default Version）：默认版本的编译器。
 - 实验性编译器（Flighting Version）：实验性的版本的编译器，包含正在测试中的编译器的新特性。

Accounts配置选项页

在Settings页面左侧导航栏上，单击MaxCompute Studio > Accounts，进入Accounts配置选项页。

Accounts配置选项页面用于管理访问MaxCompute的所用账户，关于账户更多信息请参见 [用户认证](#)。

 **说明** MaxCompute Studio需要通过用户指定的账号访问MaxCompute的项目空间和执行提交作业等操作，目前MaxCompute Studio支持的账号类型为阿里云账号（AccessKey）。

- 添加账户
 - i. 在右侧导航栏上，单击+ > **Aliyun Account By Accesskey**。
 - ii. 在**Add MaxCompute Account** 页面配置参数。
 - **Account Name**：该账户在MaxCompute Studio中的标识名称。
 - **Using properties file**：从配置文件中读取AccessKey ID和AccessKey Secret。如果您选择了此种方式，需要上传在**用户认证**中的配置文件`conf/odps_config.ini`。
 - **Using properties**：手动填入AccessKey ID和AccessKey Secret。此选项与**Using properties file**选项二选一即可。
 - **Access Id**：填入阿里云账号的AccessKey ID。
 - **Access Key**：填入阿里云账号的AccessKey Secret。
 - iii. 单击OK。添加完成后账号会出现在**Accounts**配置选项页面的列表中。
- 删除账户

该操作仅在MaxCompute Studio配置中删除账户配置，对您账户本身不产生影响：

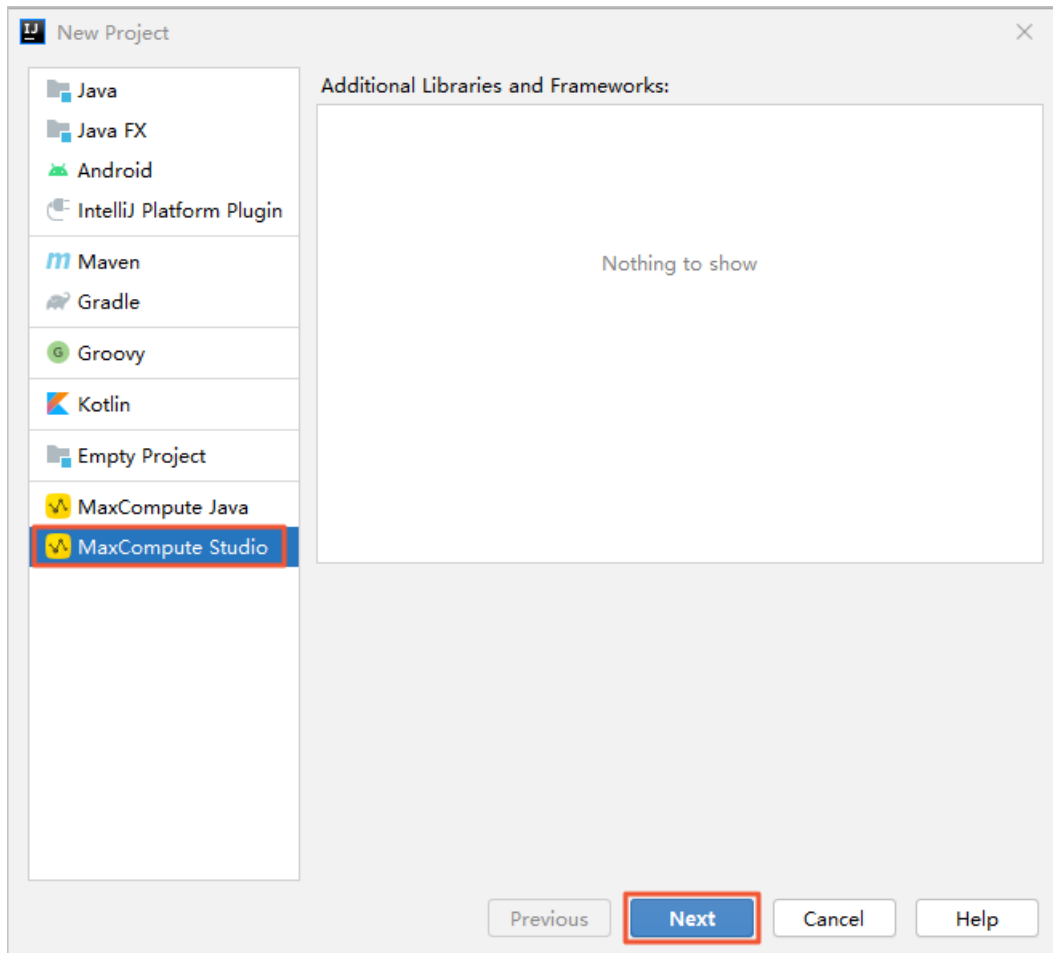
 - i. 在Account列表中选择要删除的账户名称。
 - ii. 在右侧导航栏上，单击-。
 - iii. 在弹出的确认对话框中，选择OK完成删除。
- 修改账户信息
 - i. 在Account列表中选择要修改的账户名称。
 - ii. 在右侧导航栏上，单击。
 - iii. 在弹出的**Edit MaxCompute Account**窗口中编辑Account配置信息如下：
 - **Access Id**：填入阿里云账号的AccessKey ID。
 - **Access Key**：填入阿里云账号的AccessKey Secret。
 - iv. 单击OK完成修改。

3.4. 管理项目连接

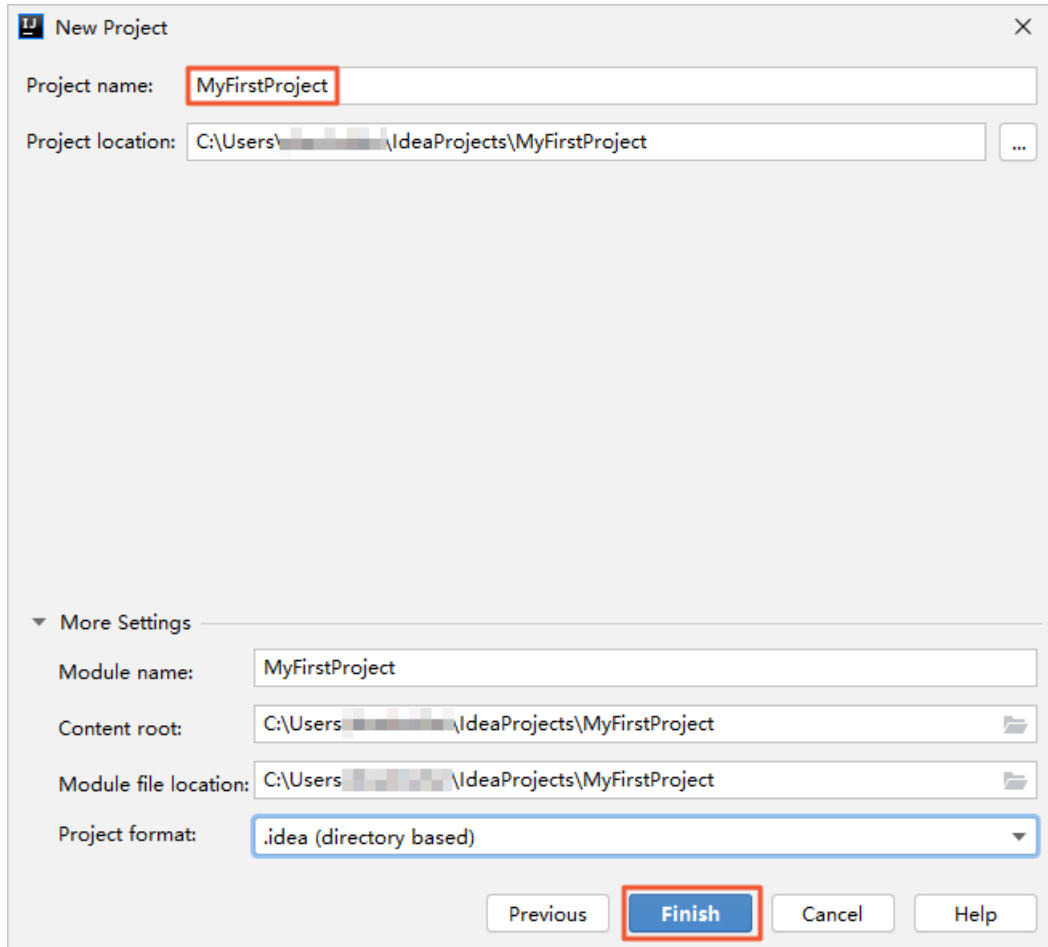
您必须通过MaxCompute Studio连接MaxCompute项目后，才可以在MaxCompute Studio上查看MaxCompute项目的信息，例如表、视图、自定义函数（UDF）或资源。本文为您介绍如何创建或修改MaxCompute项目连接。

步骤一：创建MaxCompute Studio项目

1. 启动IntelliJ IDEA，在顶部菜单栏，单击File > New > Project。
2. 在New Project对话框的左侧导航栏，选择MaxCompute Studio，单击Next。



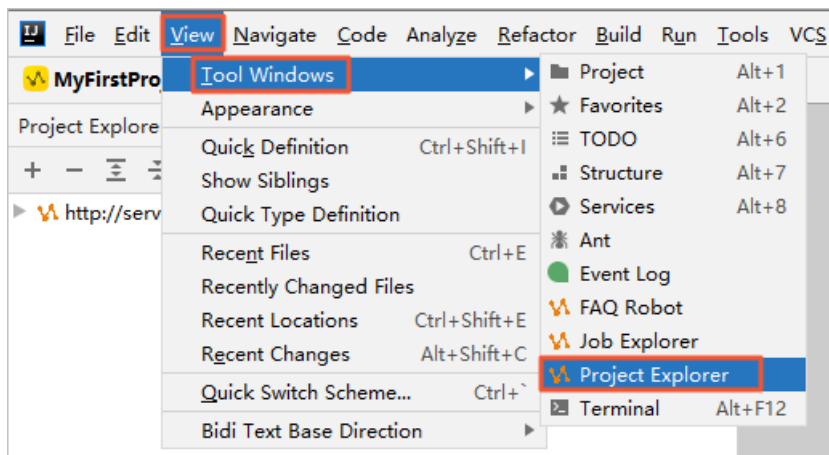
3. 填写Project name，单击Finish，完成项目创建。



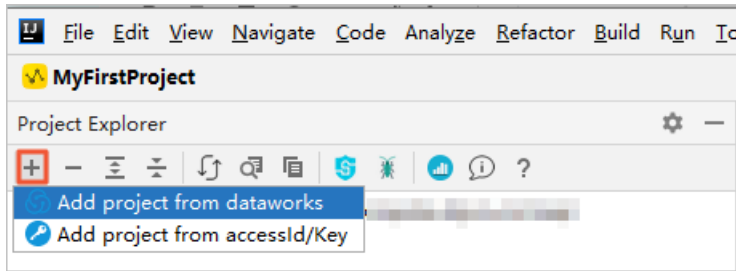
说明 如果有已经打开的Project，将会提示您是否在当前窗口中打开，即关掉之前的Project，选择This Window。

步骤二：创建MaxCompute项目连接

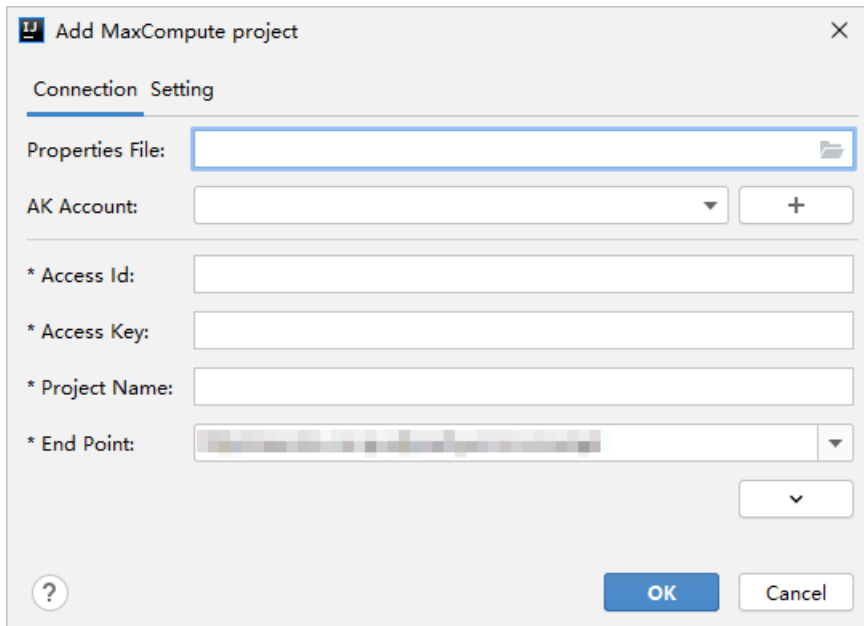
1. 在顶部菜单栏，单击View > Tool Windows > Project Explorer。




2. 单击左上角的+，选择Add project from accessId/Key。



3. 在Add MaxCompute project对话框，配置Connection页签信息。



说明

- 单击对话框左下角的即可查看在线文档。
- 如果出现超时错误，单击对话框中的Setting页签，修改数据同步相关参数：
 - sync one table timeout(s)：同步表超时参数。默认为5s。
 - sync one function timeout(s)：同步函数超时参数。默认为30s。

通过配置文件自动配置参数。

参数	说明
Properties File	上传MaxCompute项目客户端(odpscmd)的配置文件odps_config.ini, 详情请参见 安装并配置客户端 。用于初始化Access Id、Access Key、Project Name和End Point配置项。

参数	说明
AK Account	连接MaxCompute项目使用的阿里云账号。单击右侧+，在Accounts页面，选择已存在的账号。 <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> ? 说明 如果没有账号信息，您需要在Accounts页面，单击+ > Aliyun Account By AccessKey，通过配置文件自动识别或手动方式添加账号信息。更多配置Accounts信息，请参见Accounts配置选项页。 </div>

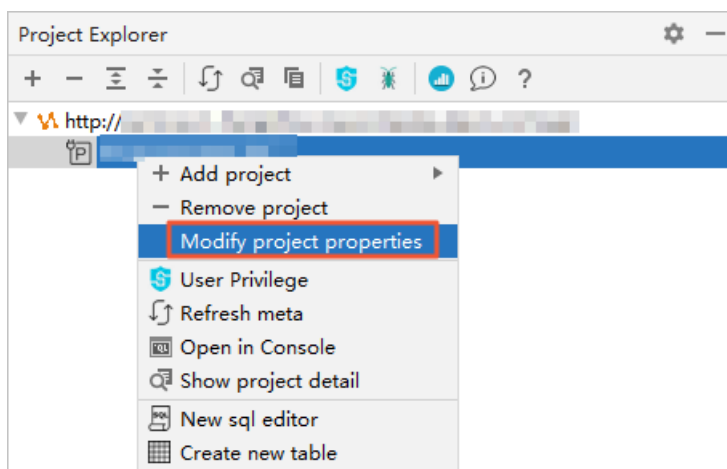
○ 手动配置参数。

参数	说明
Access Id	连接MaxCompute项目时的AccessKey ID。
Access Key	连接MaxCompute项目时的AccessKey Secret。
Project Name	MaxCompute项目的名称。 <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> ? 说明 此处为MaxCompute项目名称，非工作空间名称。登录MaxCompute控制台，在项目管理页签，您可以查看到具体的MaxCompute项目名称及所属DataWorks工作空间。 </div>
End Point	MaxCompute项目的Endpoint。

4. 配置完成后，单击OK。在左侧Project Explorer页面中会显示MaxCompute项目的信息，包括该项目中的表、视图、函数以及资源。

步骤三：修改MaxCompute项目连接

1. 在Project Explorer页面，右键单击需要修改的MaxCompute项目，选择Modify project properties。



2. 在Modify MaxCompute project对话框，修改MaxCompute项目的配置。

后续步骤

连接MaxCompute项目后，您可以管理和查看项目内的数据和资源，详情请参见[管理数据和资源](#)。

3.5. 管理数据和资源

3.5.1. 管理项目数据

本文为您介绍如何在MaxCompute Studio上查看项目空间中的表、视图、函数和资源。

前提条件

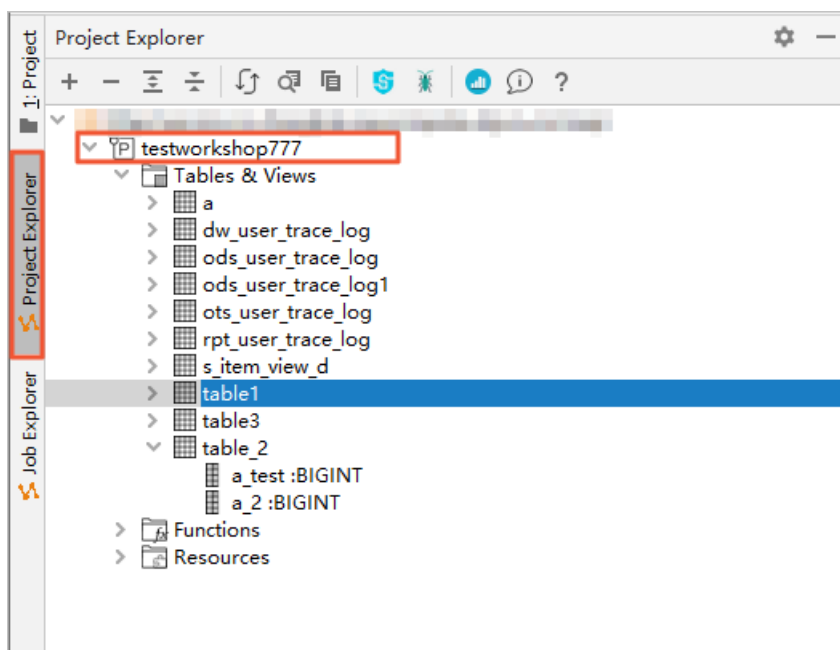
已连接MaxCompute项目，详情请参见[管理项目连接](#)。

背景信息

您可以在Project Explorer区域查看已添加连接的MaxCompute项目中的表、视图、函数和资源。


浏览和更新项目数据


1. 进入MaxCompute Studio页面，在左侧导航栏，单击Project Explorer，即可浏览已连接的MaxCompute项目。



2. 单击Tables & Views、Functions或Resources节点前的下拉箭头，可以查看该项目下的所有表、视图、函数和资源。

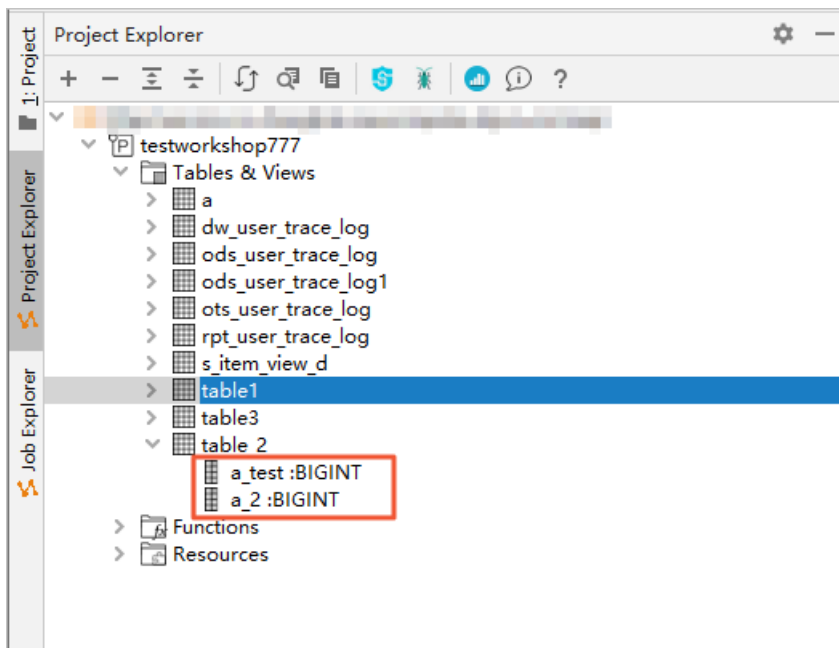
 **说明** 此操作需要您拥有该MaxCompute项目的 `DESC TABLE` 权限。

3. 在工具栏中，单击更新本地元数据。

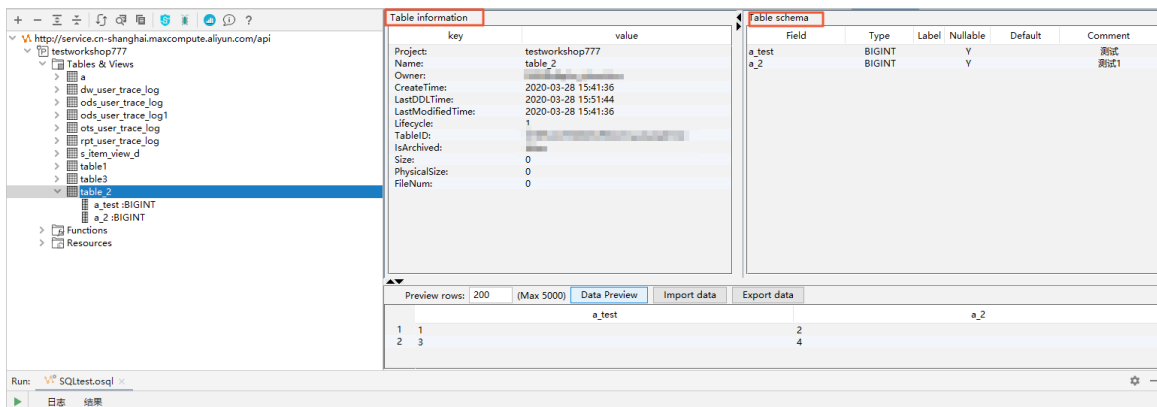
 **说明** MaxCompute Studio会将MaxCompute项目的元数据下载到本地，当MaxCompute项目的元数据有更新时，需要手动触发一次刷新，将变化的元数据重新加载到本地。

查看表或视图详细信息

1. 在Tables & Views节点树中，单击表名展开节点，可以快速查看表的字段和字段类型。



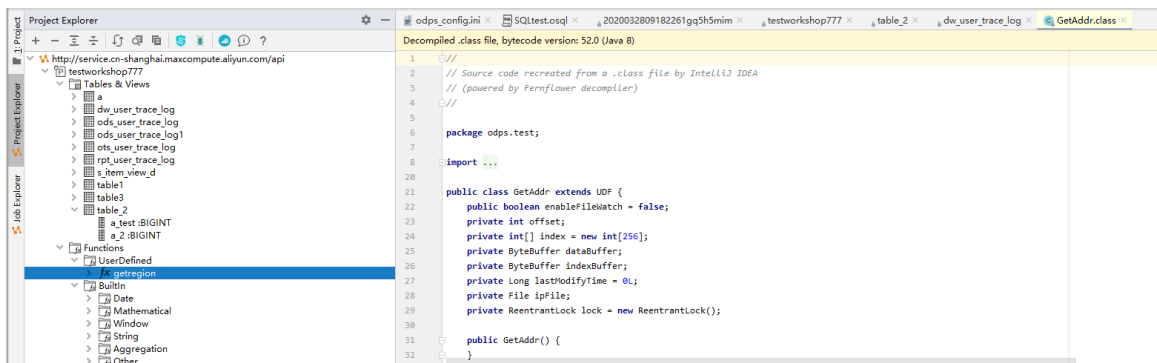
2. 在表名上单击右键，选择Show table detail，可以查看表的详细信息。例如，表的所属者、表的大小、表的列信息和Schema。



3. 在表名上单击右键，选择Find usages，可以查询表被脚本引用的情况。

查看函数详细信息

1. 在Functions节点树中，单击User Defined节点，可以查看您创建的函数。

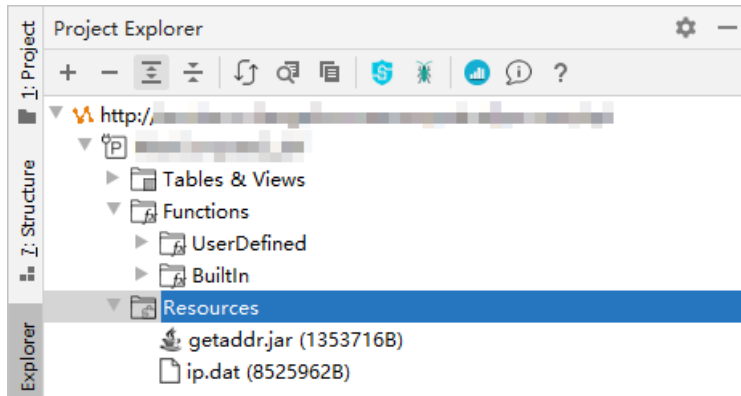


❓ 说明 Functions节点树的Built In节点下分类显示系统的内建函数，双击即可显示该函数的使用说明。内建函数详情请参见[内建函数](#)。

2. 双击创建的函数，即可打开函数对应的代码。

查看资源详细信息

您可以在Resources节点树中，查看创建的资源。



3.5.2. 导入导出表数据

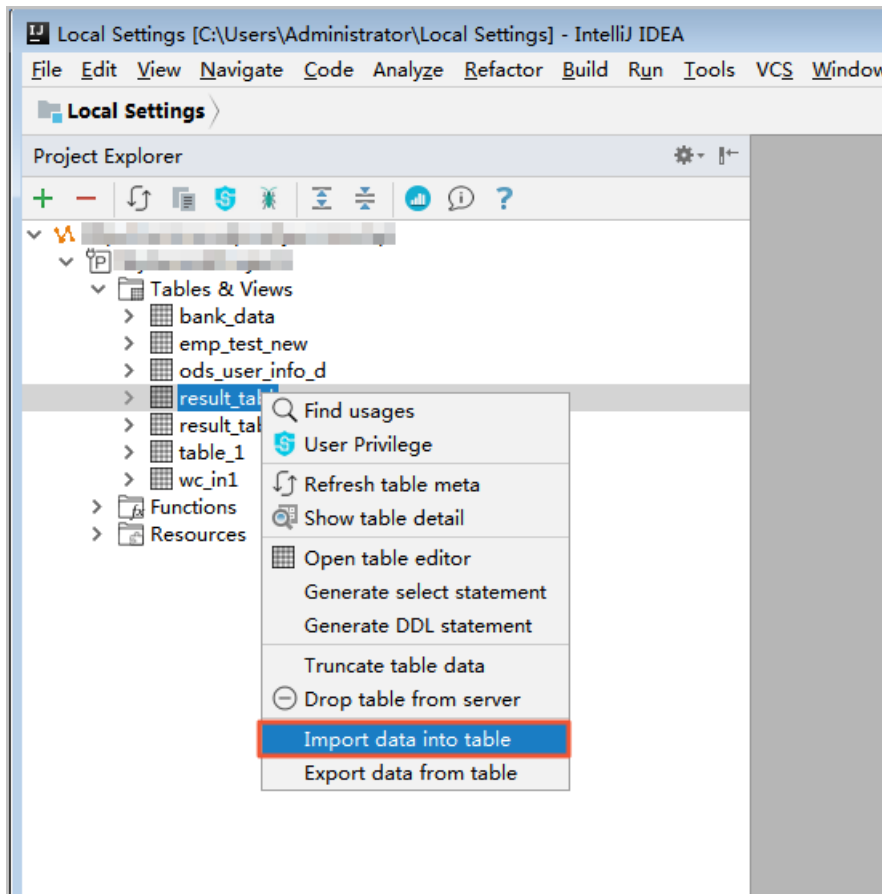
MaxCompute Studio可以将CSV、TSV等格式的本地数据文件导入至MaxCompute表中，也可将MaxCompute表中的数据导出到本地文件。MaxCompute Studio通过Tunnel导入导出数据。

前提条件

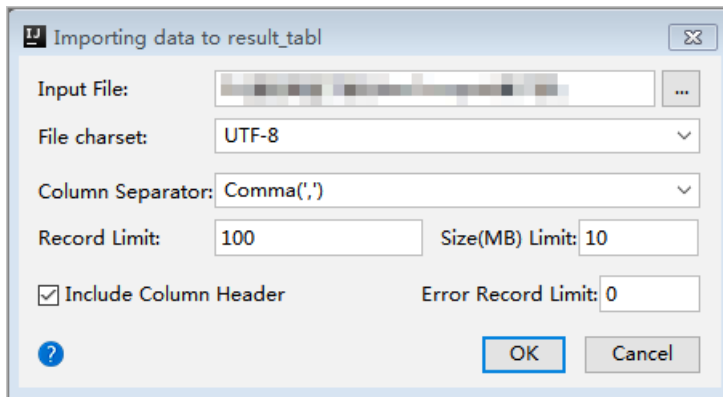
- 导入导出数据使用MaxCompute Tunnel，因此要求MaxCompute Studio中添加的MaxCompute项目必须配置了Tunnel。详情请参见[安装并配置MaxCompute客户端](#)。
- 导入导出使用的账号必须具备MaxCompute项目中表的操作权限。

导入数据

1. 在Project Explorer区域，单击MaxCompute项目的Tables & Views节点前的下拉箭头，右键单击需要导入数据的表，选择Import data into table。



2. 在Importing data to table_name对话框中，配置导入文件参数。



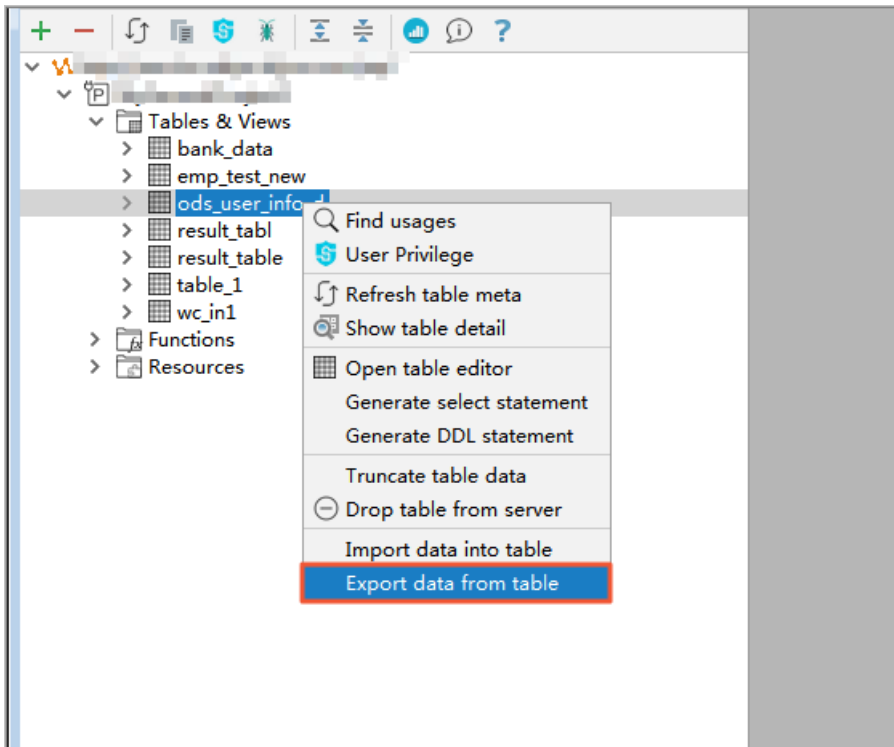
- **Input File**：导入数据文件的本地路径。
- **File charset**：导入数据文件的编码格式。编码格式包含UTF-8、UTF-16、UTF-16BE、UTF-16LE、ISO-8859-1、US-ASCII和GBK。默认为UTF-8。
- **Column Separator**：列分隔符。包含Comma(',')、Space(' ')和Tab('\t')。默认为Comma(',')。
- **Record Limit**：导入数据的最大行数。
- **Size(MB) Limit**：导入数据量最大值，单位为MB。
- **Error Record Limit**：容错行数。
- **Include Column Header**：是否导入列标题。

3. 单击OK，完成数据导入。

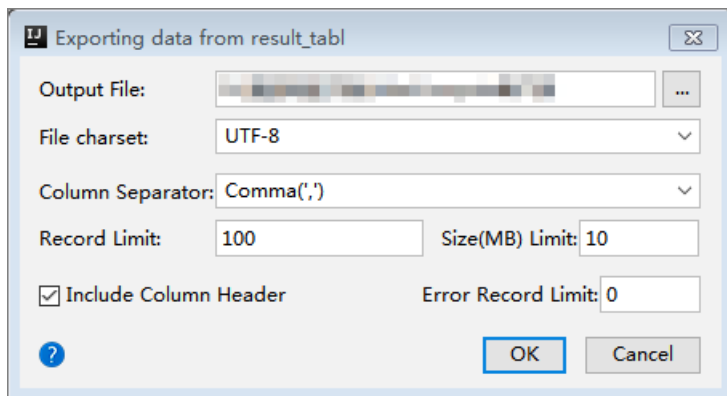
4. 提示Success，表示数据导入成功，您可以在表中查看导入的数据。

导出数据

1. 在Project Explorer区域，单击MaxCompute项目的Tables & Views节点前的下拉箭头，右键单击需要导出数据的表，选择Export data from table。



2. 在Exporting data from table_name对话框中，配置导出数据文件参数。



- **Output File**：导出数据文件的本地路径。
- **File charset**：导出数据文件的编码格式。编码格式包含UTF-8、UTF-16、UTF-16BE、UTF-16LE、ISO-8859-1、US-ASCII和GBK。默认为UTF-8。
- **Column Separator**：列分隔符。包含Comma(',')、Space(' ')和Tab('\t')。默认为Comma(',')。
- **Record Limit**：导出数据的最大行数。
- **Size(MB) Limit**：导出数据量最大值，单位为MB。
- **Error Record Limit**：容错行数。
- **Include Column Header**：是否导出列标题。

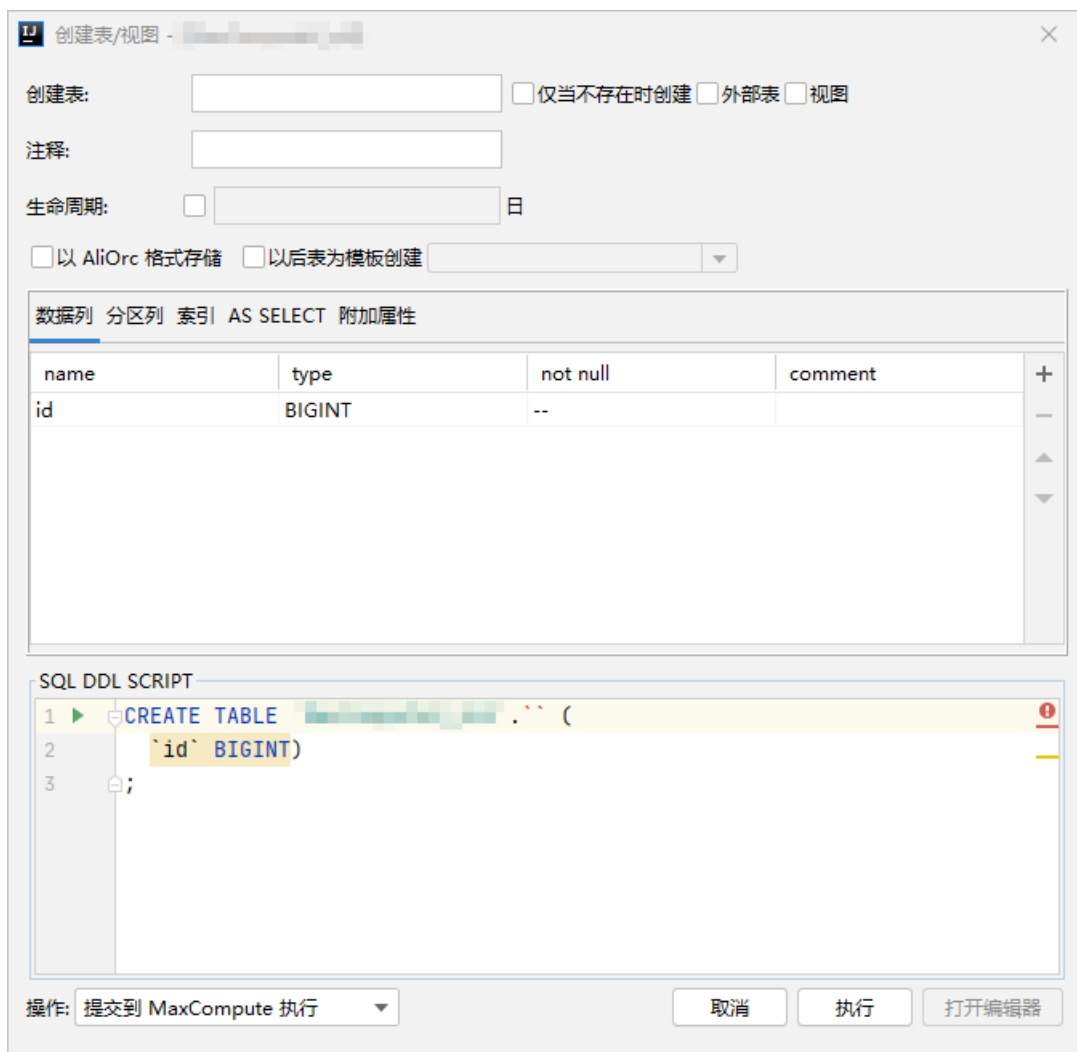
3. 提示Success，表示数据导出成功，您可以在导出文件中查看导出的数据。

3.5.3. 可视化管理表

MaxCompute Studio的Project Explorer提供了可视化表结构编辑器。本文为您介绍如何通过Project Explorer可视化创建、修改和删除表。

创建表

1. 在Project Explorer区域，右键单击MaxCompute项目下的Tables & Views，选择Create new table。
2. 在创建表/视图对话框，配置参数信息。



界面参数配置原则请遵循MaxCompute相关要求，详情请参见[表操作](#)。

说明

可视化建表无法设置Flag，默认使用以下两个Flag：

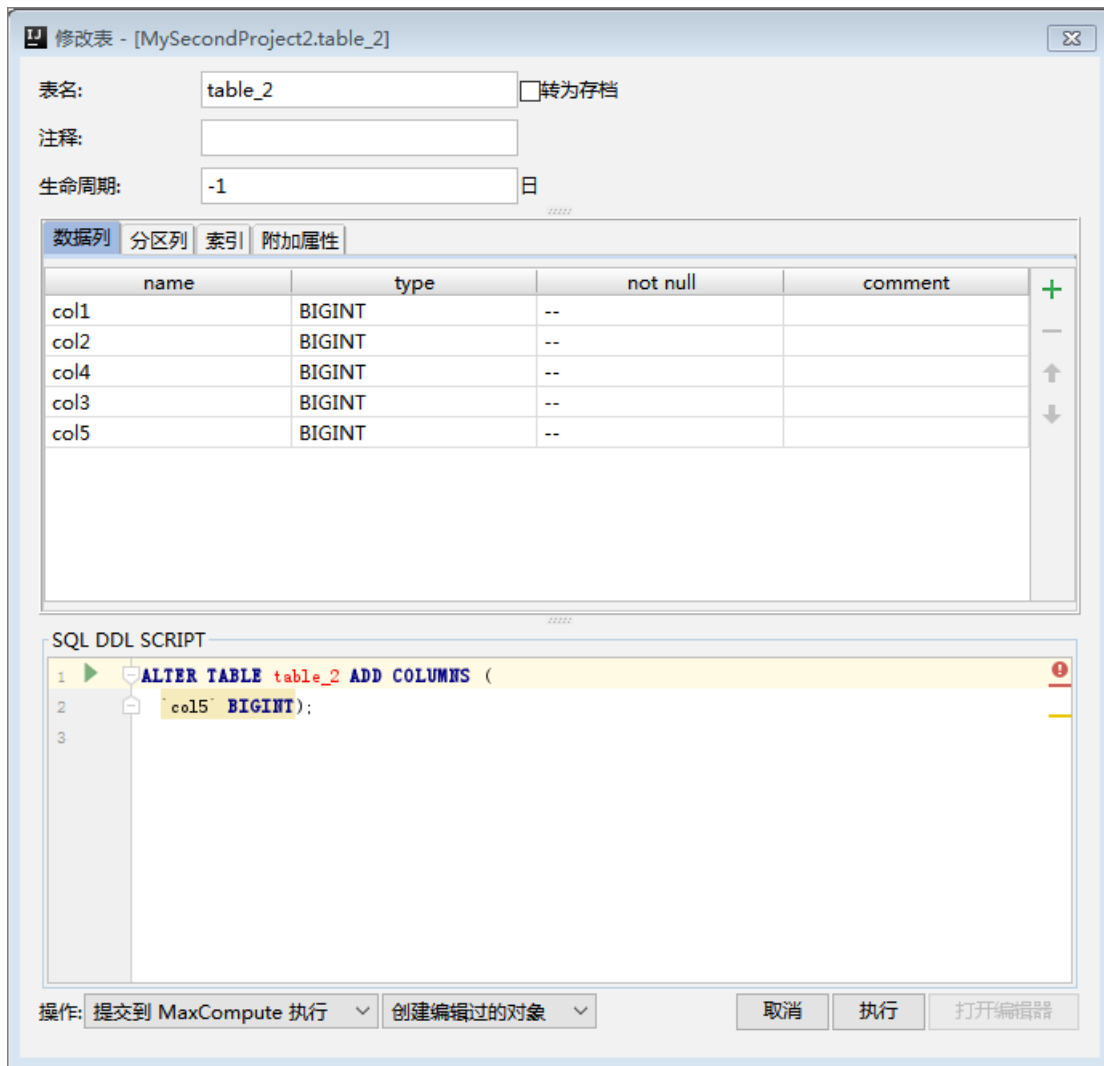
- o odps.sql.submit.mode=script
- o odps.sql.type.system.odps2=true

- 3. 单击**执行**，提示SUCCESS，完成创建。

修改表

- 1. 在Project Explorer区域，单击MaxCompute项目的Tables & Views节点前的下拉箭头，右键单击需要修改的表，选择**Open table editor**。
- 2. 在修改表对话框，对表进行编辑。

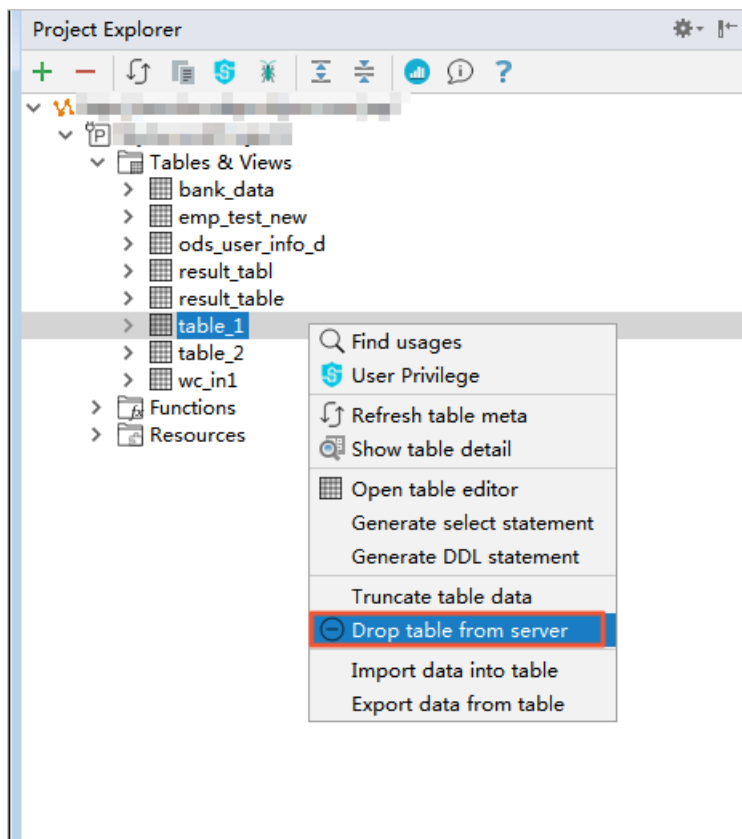
您可以新增列，修改表名称、表注释、生命周期、列名称和列注释。详情请参见[表操作](#)。



- 3. 单击**执行**，完成表修改。

删除表

- 1. 在Project Explorer区域，单击MaxCompute项目的Tables & Views节点前的下拉箭头，右键单击需要删除的表，选择**Drop table from server**。



2. 在Confirmation Required对话框，单击OK，即可将表从MaxCompute项目中删除。

3.6. 开发SQL程序

3.6.1. 创建MaxCompute Script Module

使用MaxCompute Studio开发SQL程序前，需要先创建MaxCompute Script Module。本文为您介绍如何创建MaxCompute Script Module。

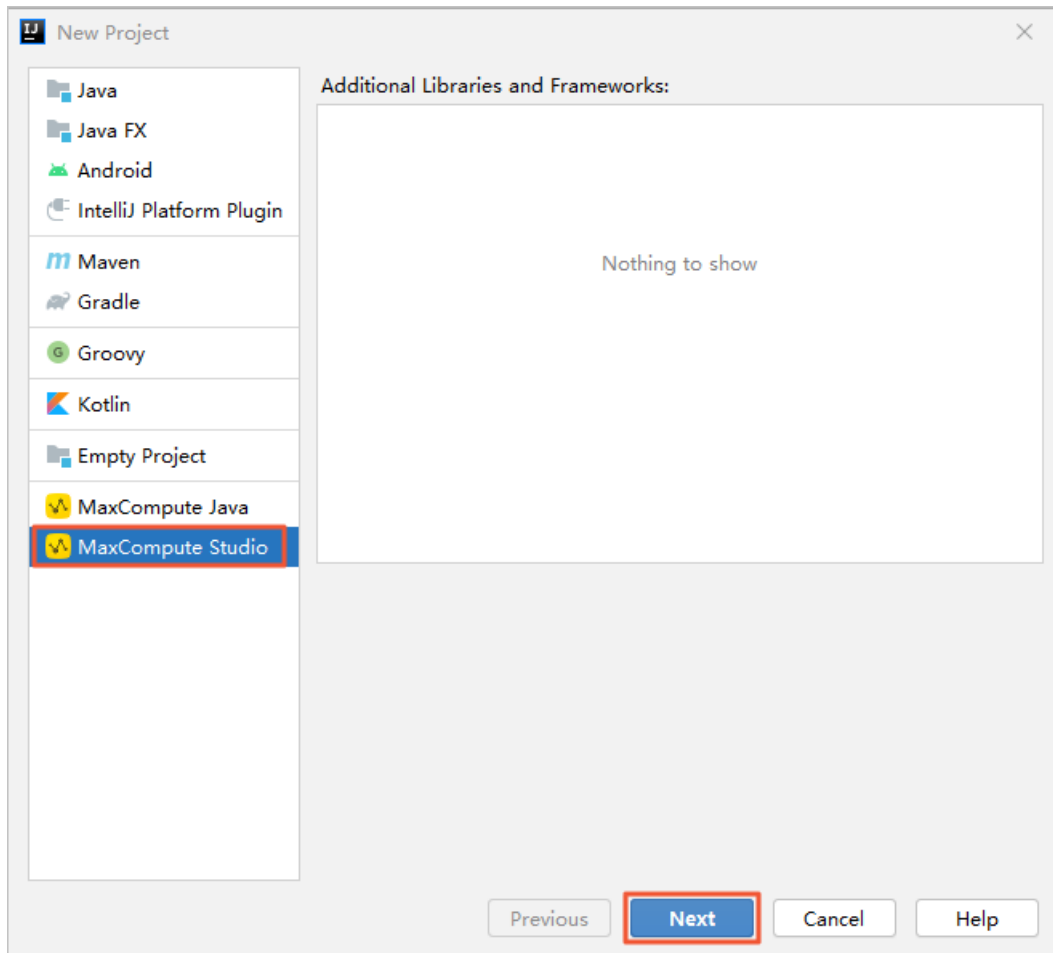
背景信息

创建MaxCompute Script Module时存在以下两种情况：

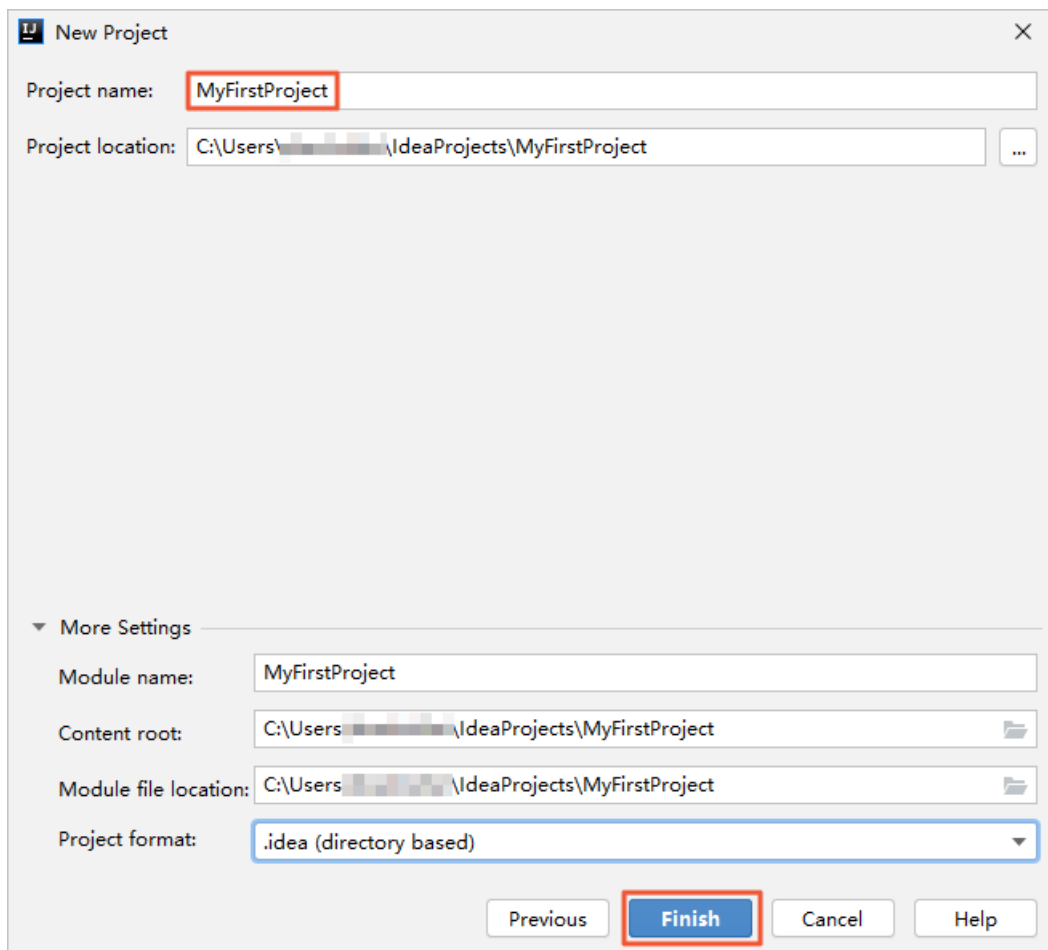
- 本地没有Script文件：需要通过IntelliJ IDEA创建一个全新的Module。
- 本地已有Script文件：假如本地某个文件夹下已经存在脚本，此时需要用MaxCompute Studio来编辑脚本，您可直接打开一个Module，无需全新创建。

本地没有Script文件时创建Module

1. 启动IntelliJ IDEA，在顶部菜单栏，单击File > New > Project。
2. 在New Project对话框的左侧导航栏，选择MaxCompute Studio，单击Next。



3. 填写Project name，单击Finish，完成项目创建。



说明 如果有已经打开的Project，将会提示您是否在当前窗口中打开，即关掉之前的Project，选择This Window。

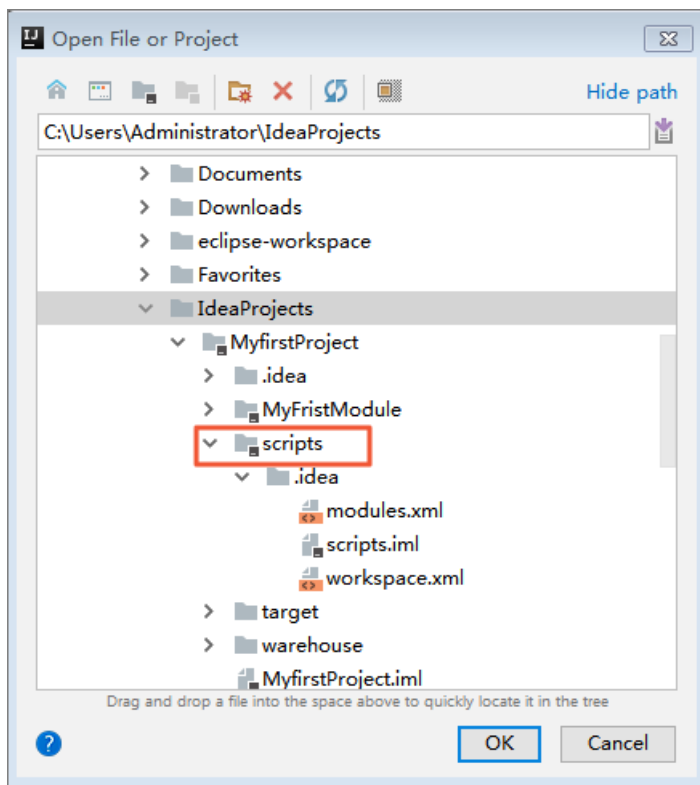
本地已有Script文件时创建Module

本地已有Script文件时无需新建Module，只需要在已有的Module目录下添加MaxCompute连接配置文件即可。

1. 在MaxCompute Studio的本地.*IdeaProjects\MaxCompute_Studio_Project_Name\scripts*文件夹下新建一个MaxCompute的连接配置文件*odps_config.ini*，文件中包含MaxCompute连接的鉴权信息，示例如下。

```
# 连接的MaxCompute项目名称。
project_name=xxxxxxx
# 云账号的AccessKey ID。
access_id=xxxxxxx
# 云账号的AccessKey Secret。
access_key=xxxxxxx
# 连接的MaxCompute服务所在区域的Endpoint信息。
end_point=xxxxxxx
```

2. 启动IntelliJ IDEA，在顶部菜单栏，单击File > Open，选择本地.*IdeaProjects\MaxCompute_Studio_Project_Name\scripts*文件夹下的*odps_config.in*文件。



说明 MaxCompute Studio会自动查找该文件夹下的odps_config.in文件，根据这个文件中的配置信息抓取MaxCompute服务端的元数据，然后编译文件夹下的所有脚本。

3.6.2. 开发及提交SQL脚本

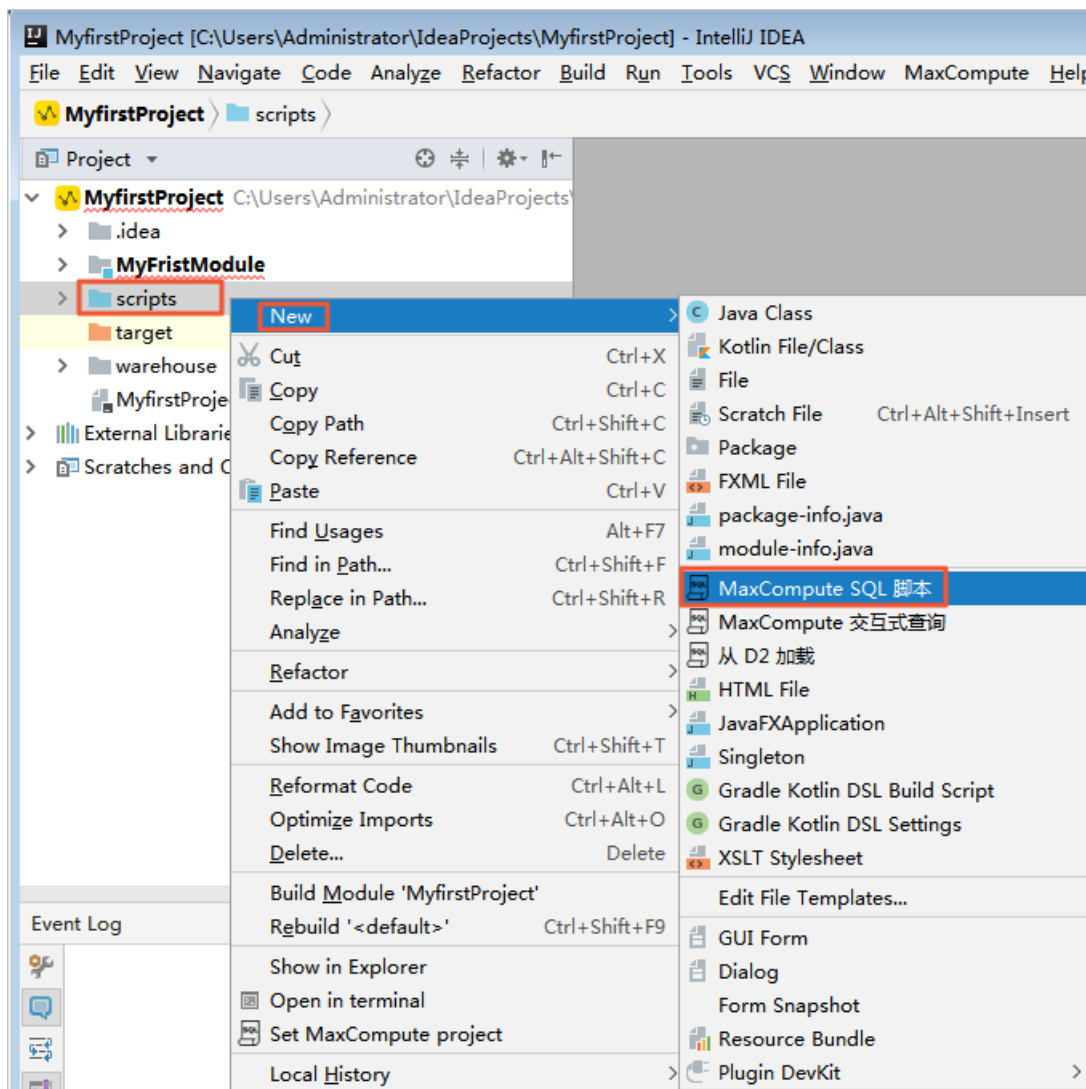
本文为您介绍如何在MaxCompute Studio上开发SQL脚本。包括编写和运行SQL脚本。

前提条件

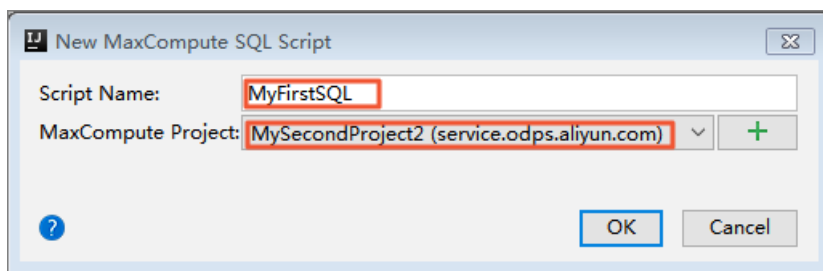
- 已连接MaxCompute项目，详情请参见[管理项目连接](#)。
- 已创建MaxCompute Script Module，详情请参见[创建MaxCompute Script Module](#)。

编写SQL脚本

1. 在Project区域下，右键单击scripts，选择New > MaxCompute SQL脚本。



2. 在New MaxCompute SQL Script对话框，配置参数信息，单击OK。



- o Script Name：脚本名称。
- o MaxCompute Project：目标MaxCompute项目。单击+即可新建一个MaxCompute项目连接，配置详情请参见[管理项目连接](#)。

3. 在脚本编辑界面中编写SQL。SQL语法详情请参见[SQL概述](#)。

说明

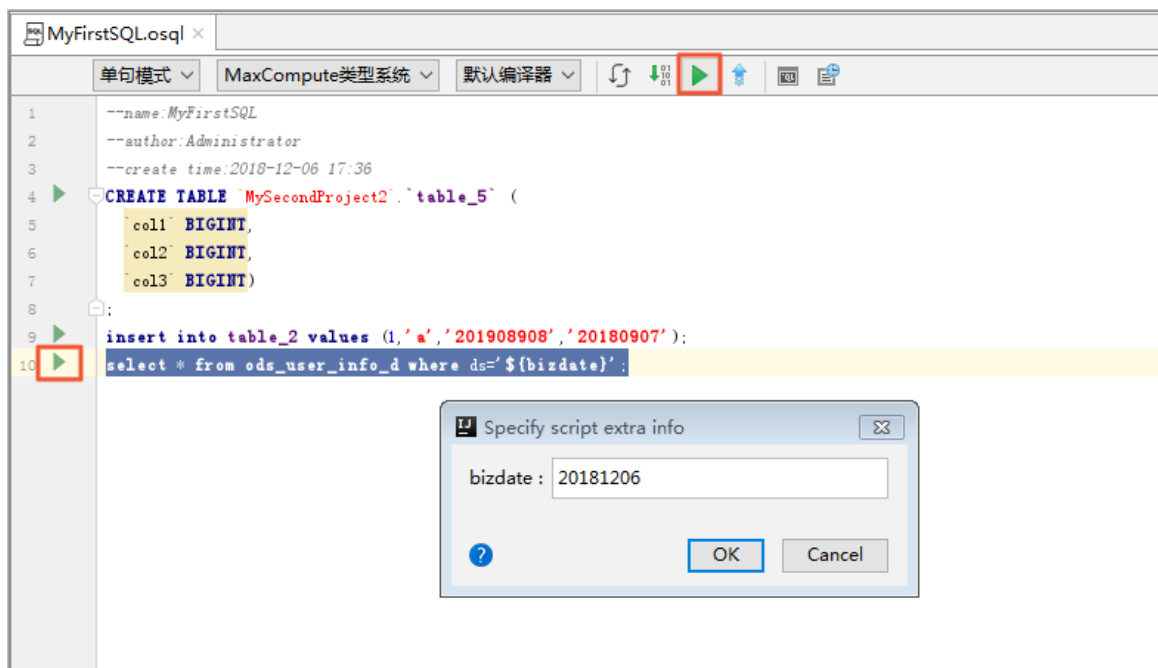
- 支持跨项目空间资源依赖。例如，脚本绑定了项目A的同时，允许访问项目B下的 table1 (ProjectB.table1)。
- MaxCompute Studio支持设置SQL脚本编辑器，详情请参见概述。

提交SQL脚本

在提交SQL脚本前您需根据自身需求进行相关设置。MaxCompute Studio提供了丰富的设置功能，您可以在编辑器页面上方的工具栏中快速设置。设置主要分为以下三种：

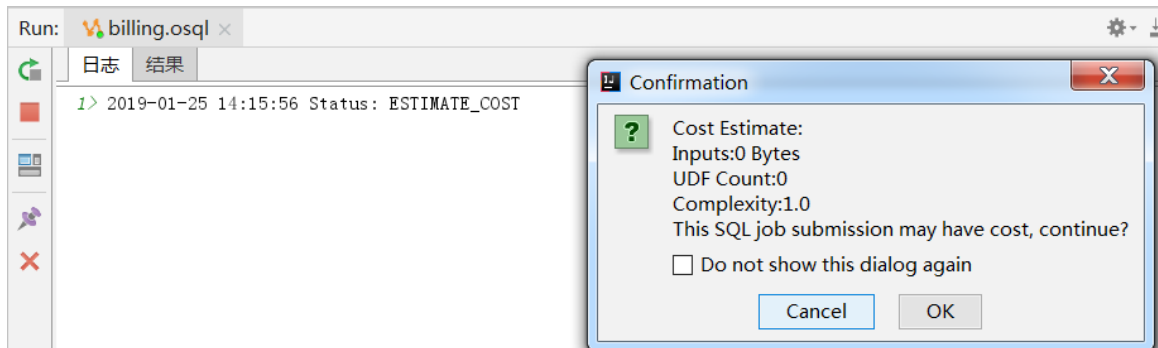
- 编辑器模式：
 - 单句模式：会将提交的脚本文件按 ; 分隔，逐条提交到MaxCompute服务端执行。
 - 脚本模式：为最新开发模式，可将整条脚本一次提交到MaxCompute服务端，由MaxCompute服务端提供整体优化，效率更高，推荐使用此模式。
- 类型系统：类型系统主要解决SQL语句的兼容性问题。分为以下三种类型：
 - 旧有类型系统：MaxCompute旧类型的系统。
 - MaxCompute类型系统：MaxCompute 2.0引入的新类型系统。
 - Hive类型系统：MaxCompute 2.0引入的Hive兼容模式下的类型系统。
- 执行模式：
 - 默认：稳定版本。
 - 查询加速：包含查询加速 (MCQA) 新特性。
 - 加速失败重跑：支持作业在查询加速失败时，重新执行作业。

1. 完成SQL脚本编写后，单击工具栏或侧边栏上的 ▶ 图标，即可将SQL脚本提交到MaxCompute服务端运行。



说明 当SQL中存在变量时（如上图中的\${bizdate}），会弹出对话框，提示您输入变量值。

- 2. 在SQL任务运行前，IntelliJ IDEA会向您提示预估的SQL费用。确认费用后，在Confirmation对话框，单击OK。

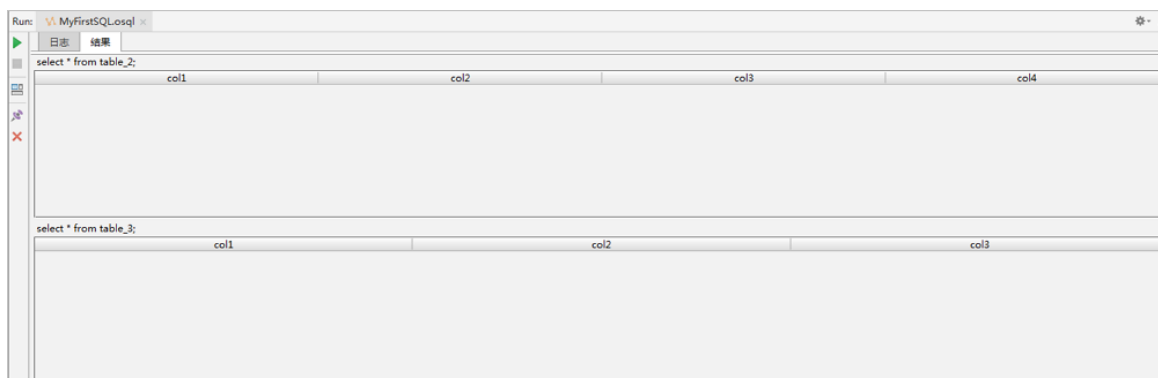


说明

- 在工具栏上，单击 图标，可以更新SQL脚本中使用的元数据，例如表、UDF。如果MaxCompute服务端存在表或函数，但MaxCompute Studio提示表和函数不存在时，请尝试使用该功能更新元数据。
- SQL依赖于您在Project Explore窗口中添加的项目元数据，系统先在本地进行编译，无编译错误后会提交到服务端执行。
- SQL执行过程中会显示运行日志。当SQL开始在MaxCompute服务端运行时，会自动打开任务详情页签，显示运行作业的基本信息。

- 3. 在控制台结果页签查看SQL运行结果。

单句模式下存在多条语句时，系统会显示每条语句的运行结果。



3.7. 开发Java程序

3.7.1. 概述

本文为您介绍使用MaxCompute Studio开发Java程序的流程以及相关目录。

开发流程

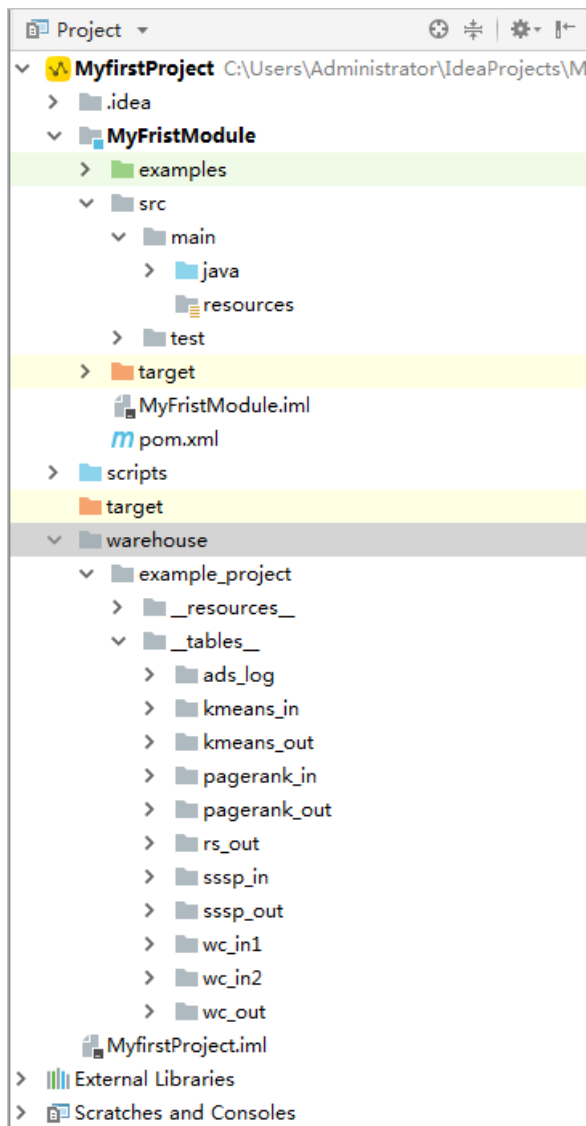
通过MaxCompute Studio开发Java程序的流程如下：

1. 创建MaxCompute Java Module。
2. 开发Java程序。您可以参考如下示例开发不同的Java程序：
 - o 开发UDF
 - o 开发MapReduce
 - o 开发Graph
 - o 查询非结构化数据
3. 打包、上传和注册。

Module目录

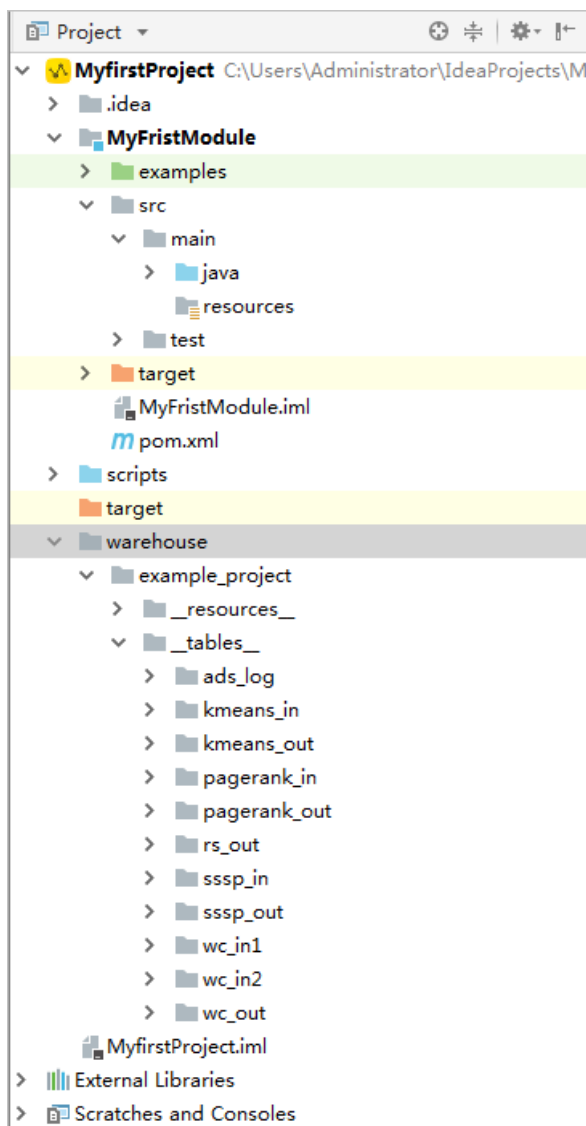
创建MaxCompute Java Module后，MaxCompute Studio会自动创建一个Module。Module目录内容如下：

- *examples*: 示例代码，包括单元测试示例。您可以参考示例开发单元测试脚本。
- *src/main/java*: 开发Java程序的源码。
- *warehouse*: 存储MaxCompute项目的表（包括Schema和数据）和资源。



warehouse目录

warehouse目录存储MaxCompute项目的表（包括Schema和数据）和资源，用于执行UDF或MapReduce。



- warehouse目录包含项目名、资源（_resources_）、表（_tables_）、表名、表结构（_schema_）和表数据（data）。
- 表结构（_schema_）文件中配置项目名、表名、列名和类型，并通过冒号（:）分隔。分区表需要配置分区列。图中wc_in1为非分区表，wc_in2为分区表。
- data文件采用标准CSV格式存储表的数据：
 - 特殊字符为逗号（,）、单个双引号（"）和换行符（\n 或 \r\n）。
 - 列分隔符为逗号（,），行分隔符为换行符（\n 或 \r\n）。
 - 如果列内容包含特殊字符，需要在该列内容前后加上双引号（"）。例如 3,No 写为 "3,No"。
 - 如果列内容包含单个双引号（"），则所有的单个双引号（"）需要转义成双引号（""）。例如 a"b"c 写为 ""a""b""c"。
 - \N 表示该列为NULL，如果该列内容为 \N（STRING类型），需要转义为 """"\N""""。
 - 文件字符编码为UTF-8。

3.7.2. 创建MaxCompute Java Module

MaxCompute Studio支持开发Java UDF、MapReduce和Graph等程序，首先您需要新建一个MaxCompute Java Module。本文为您介绍如何新建MaxCompute Java Module。

前提条件

已连接MaxCompute项目，详情请参见[管理项目连接](#)。

操作步骤

1. 启动IntelliJ IDEA，在顶部菜单栏，单击File > New > Module...
2. 在New Module对话框的左侧导航栏，单击MaxCompute Java。
3. 配置Module SDK文件位置，单击Next。
4. 填写Module name，单击Finish。

执行结果

完成上述步骤，MaxCompute Studio会自动创建一个Maven Module，同时完成下列内容：

- 引入MaxCompute相关依赖，详情请查看pom.xml文件。
- 创建examples示例代码目录。详情请参见[Module目录](#)。
- 创建warehouse目录用于存储本地调试所需的数据。详情请参见[warehouse目录](#)。

后续步骤

完成MaxCompute Java Module创建后，即可开发Java程序。详情请参见：

- [开发UDF](#)
- [开发MapReduce](#)
- [查询非结构化数据](#)
- [开发Graph](#)

3.7.3. 开发UDF

本文为您介绍如何在MaxCompute Studio上开发UDF，包括编写UDF和调试UDF。

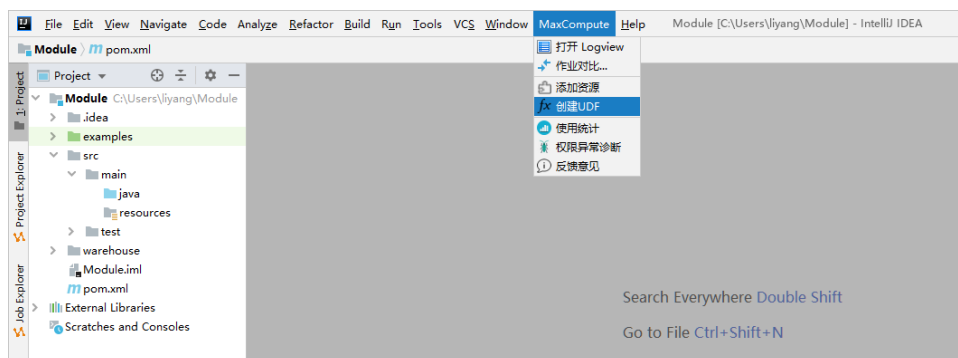
前提条件

您需要完成以下操作：

- [管理项目连接](#)
- [创建MaxCompute Java Module](#)

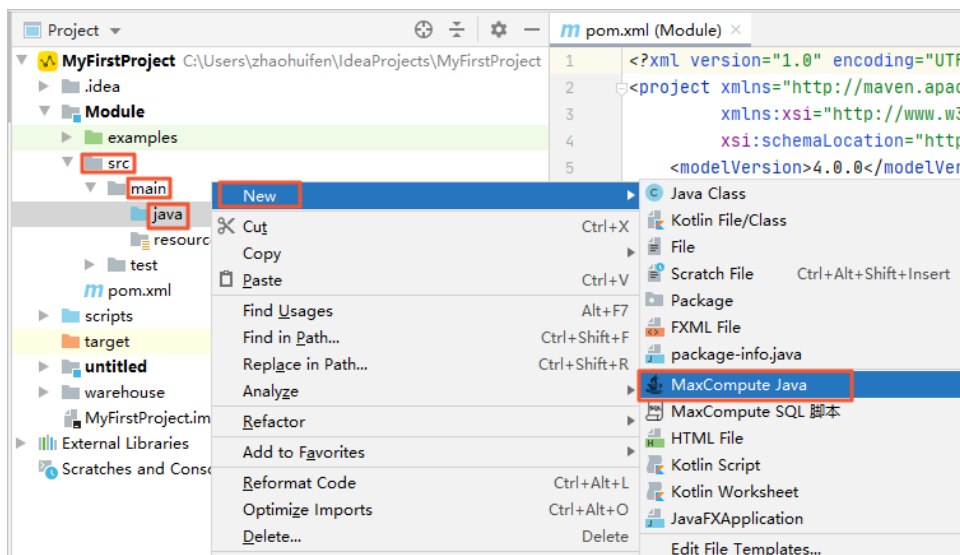
背景信息

您可以按照本文介绍自行开发UDF，也可以单击MaxCompute > 创建UDF直接创建函数。如下图所示。

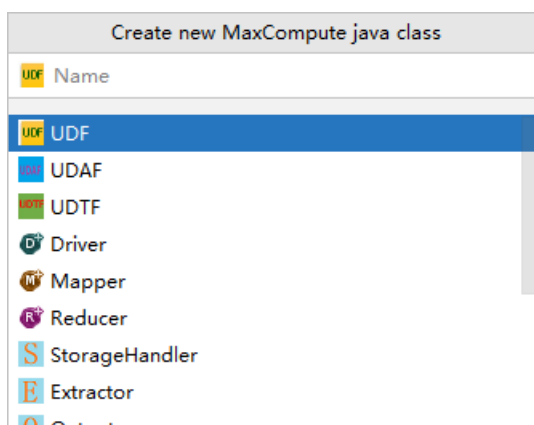


编写UDF

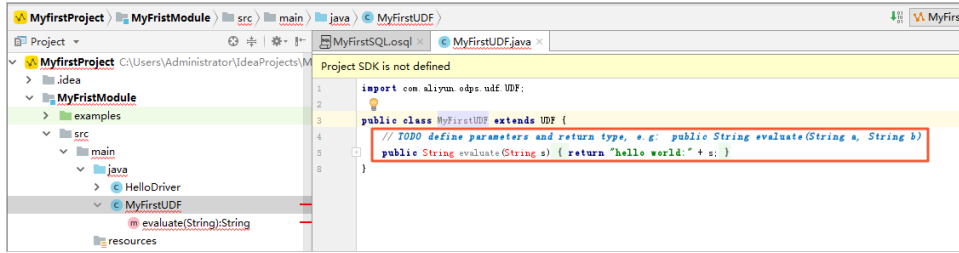
1. 在Project区域，右键单击Module的源码目录（即src > main > java），选择new > MaxCompute Java。



2. 填写Name并选择类型为UDF，按下Enter键。



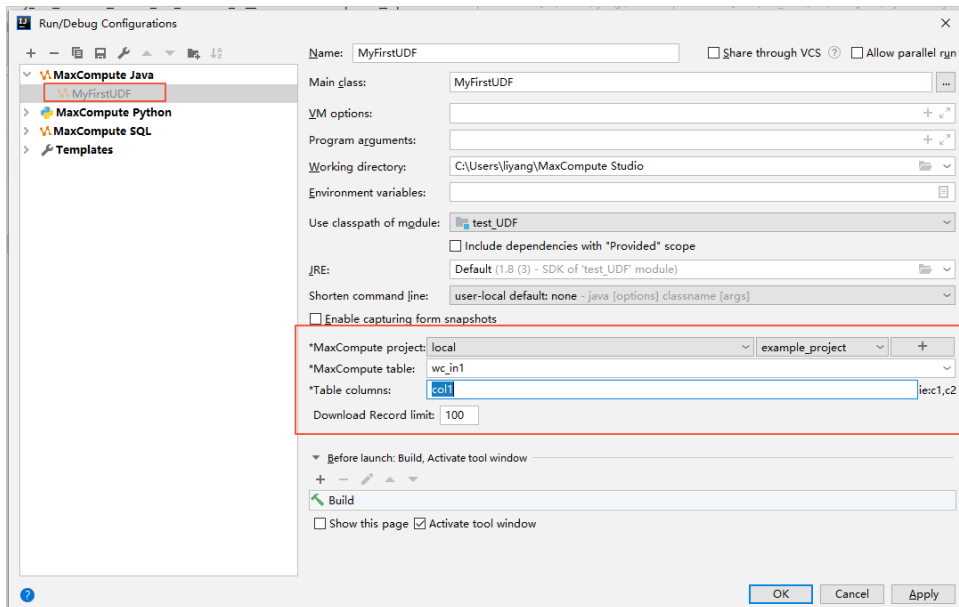
- o Name：创建的MaxCompute Java Class名称。如果还没有创建Package，在此处填写packagename.classname，会自动生成Package。
 - o 选择类型为UDF。
3. 创建成功后，在编辑界面开发Java程序。



通过本地运行调试UDF

通过本地运行方式测试，查看UDF的运行结果是否符合预期。

1. 右键单击编写完成的Java脚本，选择Run。
2. 在Run/Debug Configurations页面上配置运行参数。



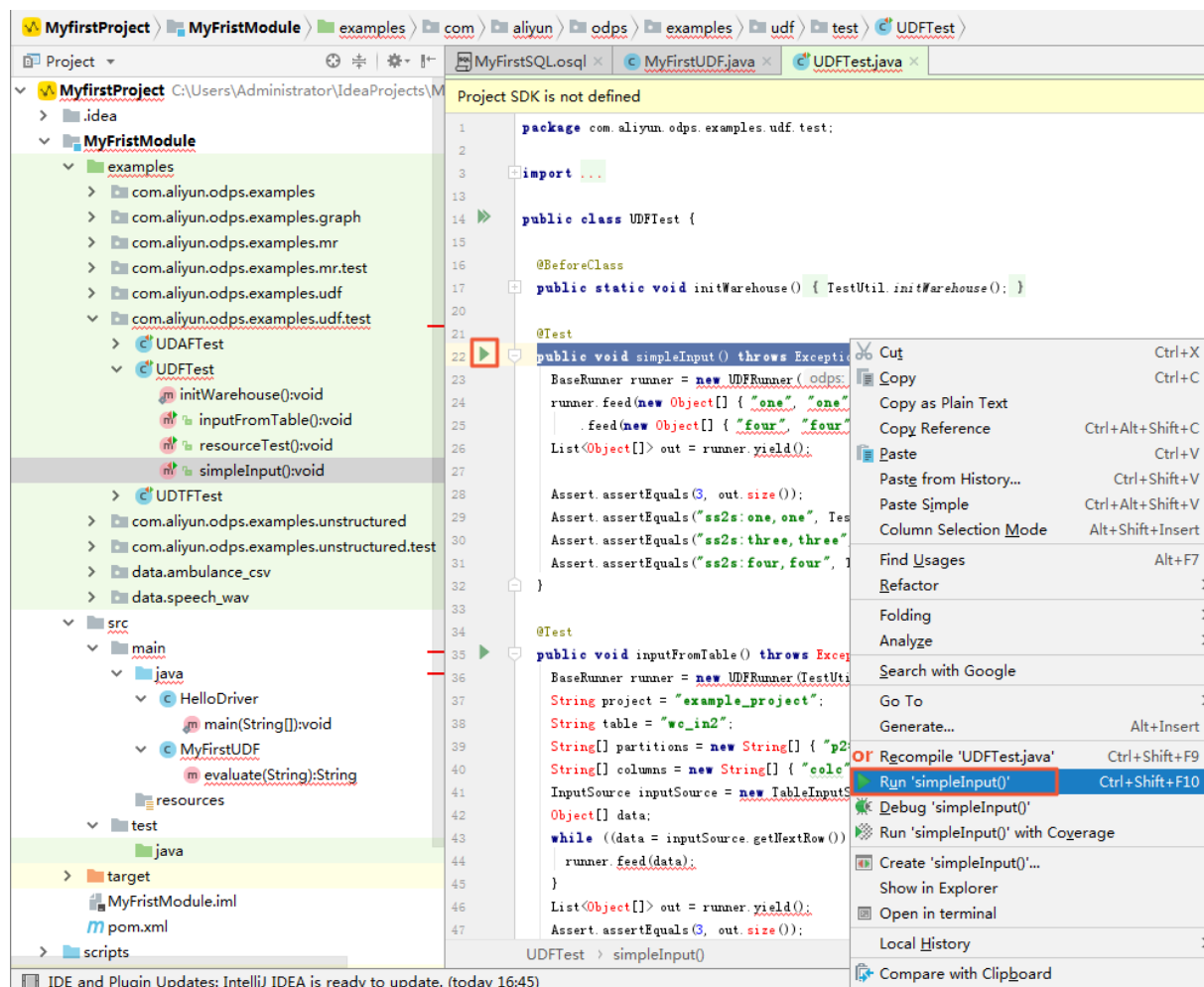
- MaxCompute project：UDF运行使用的MaxCompute空间。本地运行时选择local。
 - MaxCompute table：UDF运行时需要使用的MaxCompute表的名称。
 - Table columns：UDF运行时需要使用的MaxCompute表的列信息。
3. 单击OK，开始运行。

说明

- 本地运行会读取warehouse中指定的表数据作为输入，您可以在控制台查看日志输出。
- 如果指定的MaxCompute项目的表数据未被下载至warehouse目录中，会先下载数据；如果数据已经下载，则跳过此步骤。

通过单元测试调试UDF

您可以参照 *examples* 目录下的单元测试实例，编写自己的测试用例。



后续步骤

完成开发和调试UDF之后，需要对UDF代码打包、上传和注册。详情请参见[打包、上传和注册](#)。

3.7.4. 开发MapReduce

本文为您介绍如何在MaxCompute Studio上开发MapReduce，包括编写MapReduce、调试MapReduce、打包、上传和运行MapReduce。

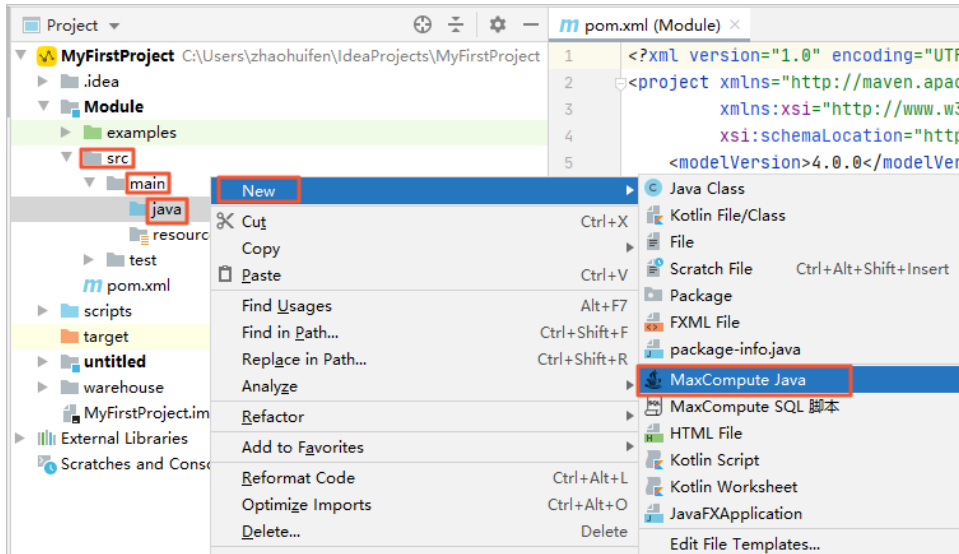
前提条件

您需要完成以下操作：

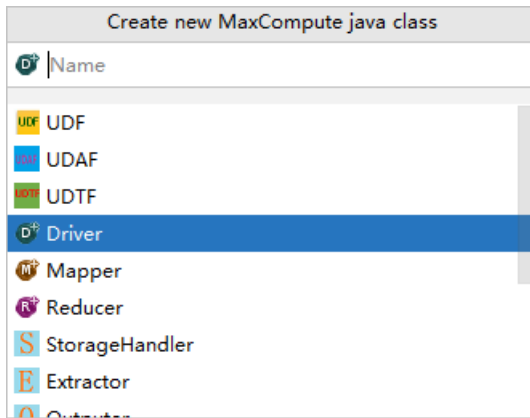
- [管理项目连接](#)
- [创建MaxCompute Java Module](#)

编写MapReduce

1. 在Project区域，右键单击Module的源码目录（即src > main > java），选择new > MaxCompute Java。



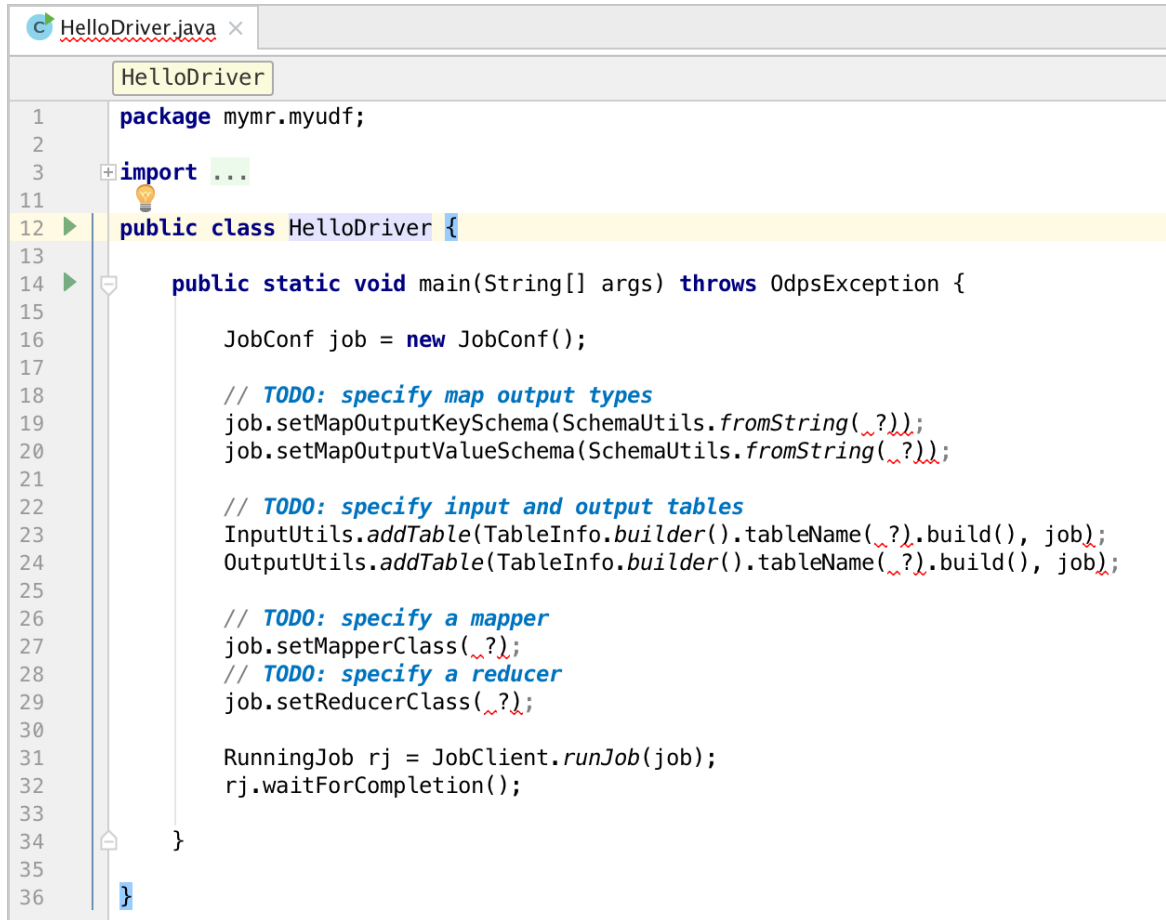
2. 填写Name并选择类型为Driver，按下Enter键。



- o Name: 创建的MaxCompute Java Class名称。如果还没有创建Package，在此处填写packagename.classname，会自动生成Package。
- o 选择类型为Driver、Mapper或Reducer。

3. 创建成功后，在编辑界面开发Java程序。

Java模板已自动填充框架代码，您只需设置输入表、输出表、Mapper和Reducer类等信息。

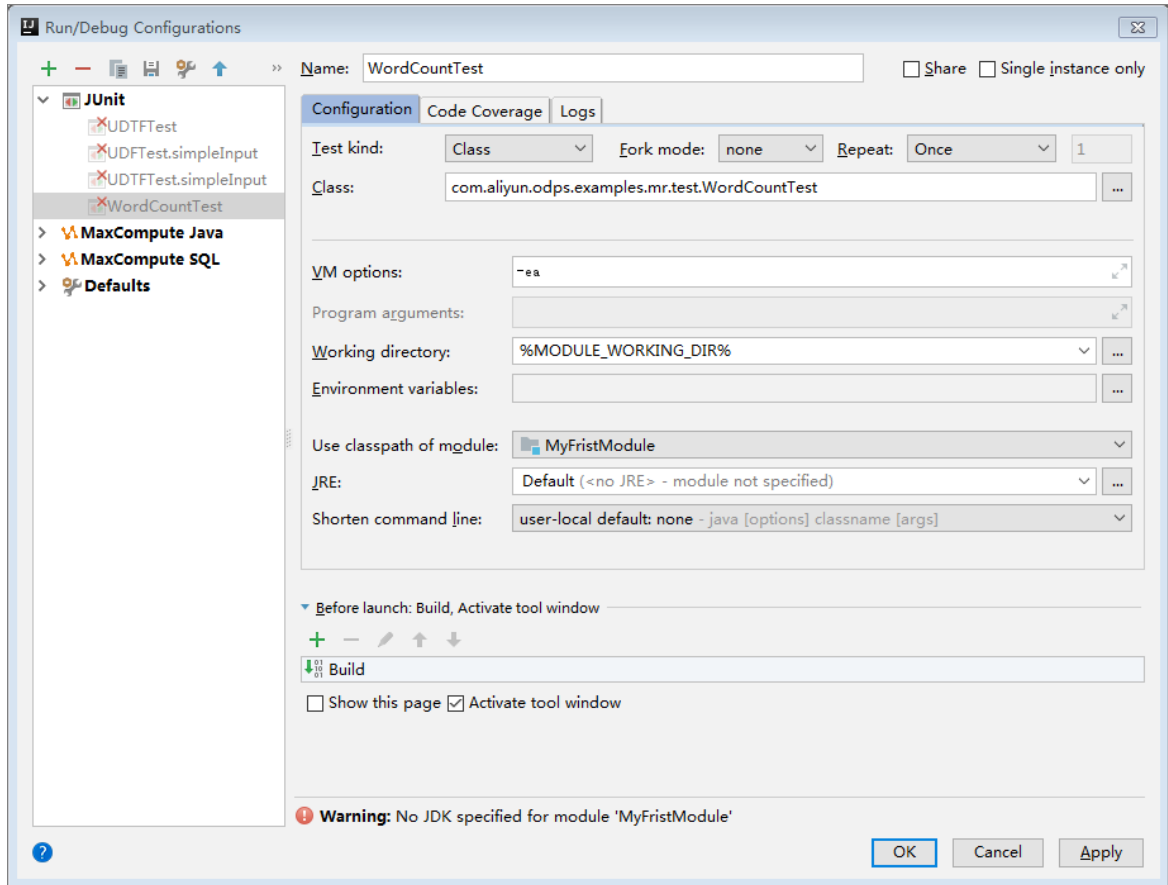


```
1 package mymr.myudf;
2
3 import ...
4
11
12 public class HelloDriver {
13
14     public static void main(String[] args) throws OdpsException {
15
16         JobConf job = new JobConf();
17
18         // TODO: specify map output types
19         job.setMapOutputKeySchema(SchemaUtils.fromString(""));
20         job.setMapOutputValueSchema(SchemaUtils.fromString(""));
21
22         // TODO: specify input and output tables
23         InputUtils.addTable(TableInfo.builder().tableName("").build(), job);
24         OutputUtils.addTable(TableInfo.builder().tableName("").build(), job);
25
26         // TODO: specify a mapper
27         job.setMapperClass("");
28         // TODO: specify a reducer
29         job.setReducerClass("");
30
31         RunningJob rj = JobClient.runJob(job);
32         rj.waitForCompletion();
33
34     }
35
36 }
```

通过本地运行调试MapReduce

通过本地运行方式测试，查看MapReduce的运行结果是否符合预期。

1. 右键单击编写完成的Java脚本，选择Run。
2. 在Run/Debug Configurations页面上选择此次运行的MaxCompute项目名称。



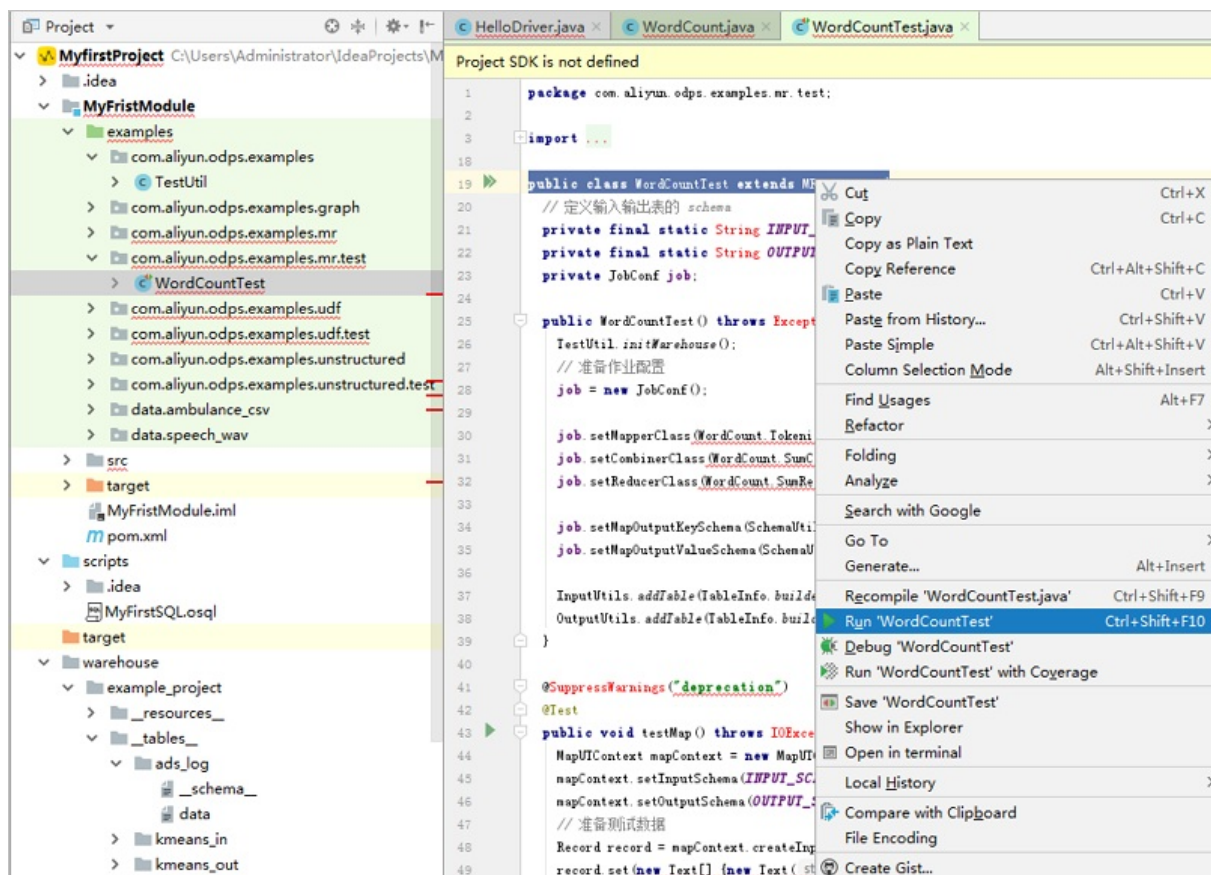
3. 单击OK，开始运行。

① 说明

- 本地运行会读取warehouse中指定的表数据作为输入，您可以在控制台查看日志输出。
- 如果指定的MaxCompute项目的表数据未被下载至warehouse目录中，会先下载数据；如果数据已经下载，则跳过此步骤。

通过单元测试调试MapReduce

您可以参照 *examples* 目录下的 *WordCount* 单元测试示例，编写测试用例。



打包上传

调试成功之后，将Java程序打成JAR包，并作为资源上传至MaxCompute服务端。详情请参见[打包、上传和注册](#)。

运行MapReduce

通过MaxCompute客户端运行MapReduce。

1. 在左侧导航栏，单击Project Explorer。
2. 右键单击项目名称，选择Open in Console。
3. 在Console区域，执行如下命令运行MapReduce。

更多命令请参见[JAR命令](#)。

```
jar-libjars wordcount.jar -classpath D:\odps\clt\wordcount.jar com.aliyun.odps.examples.mr.WordCount wc_in wc_out;
```

3.7.5. 开发Graph

本文为您介绍如何使用MaxCompute Studio开发Graph，包括编写Graph、调试Graph、打包上传和运行Graph。

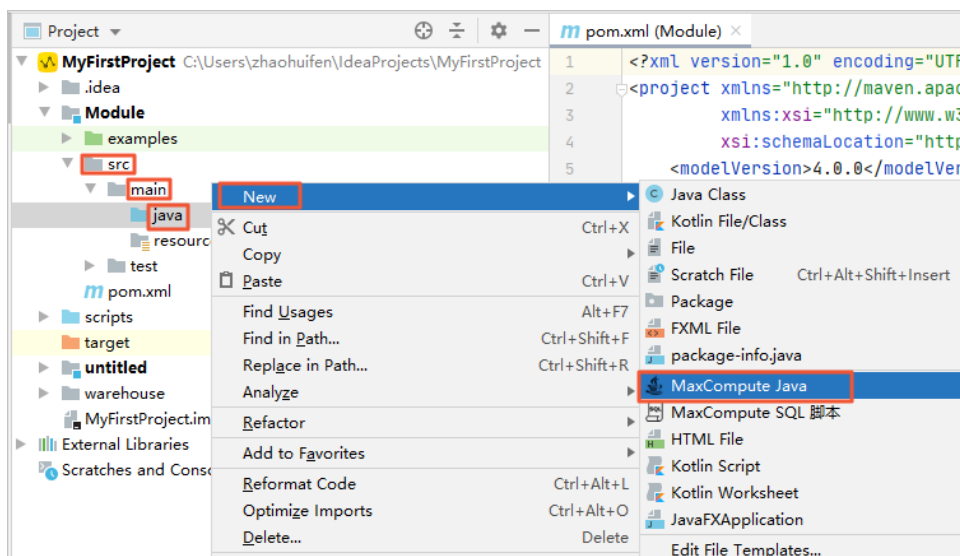
前提条件

您需要完成以下操作：

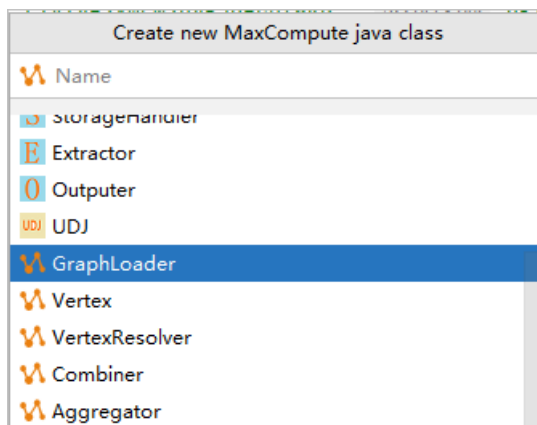
- 管理项目连接
- 创建MaxCompute Java Module

编写Graph

1. 在Project 区域，右键单击Module的源码目录（即src > main > java），选择new > MaxCompute Java。



2. 填写Name并选择类型为GraphLoader或Vertex，按下Enter键。

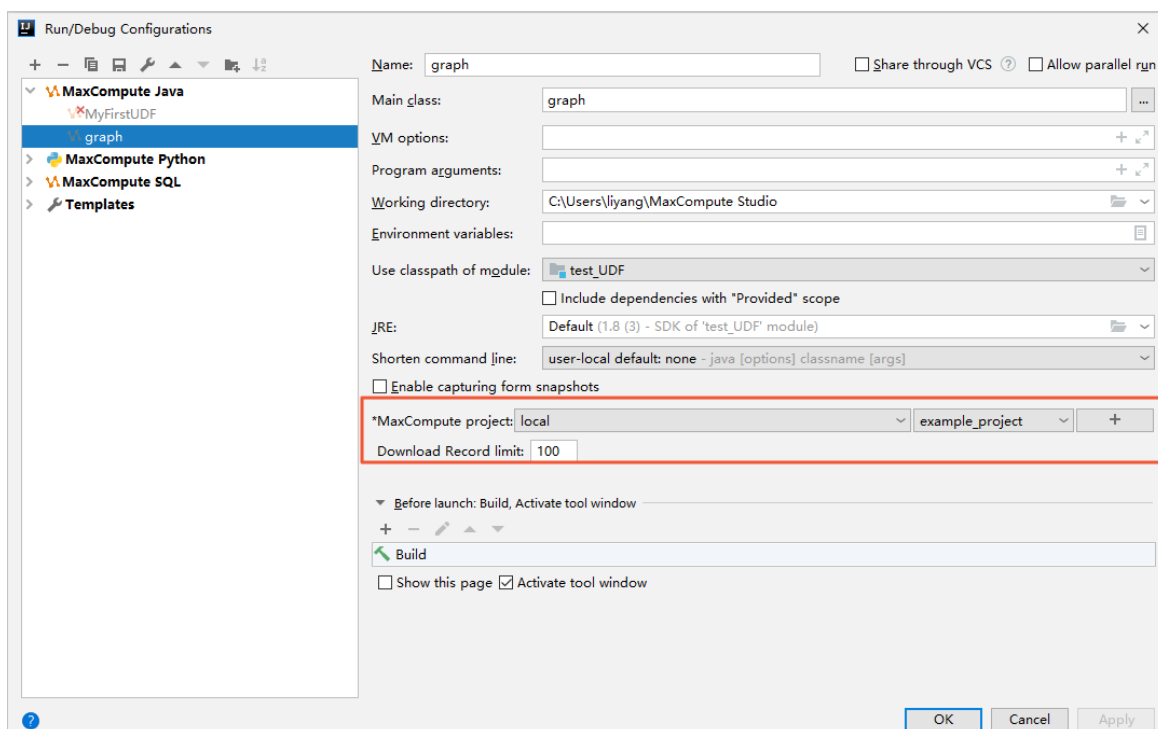


- Name：填写创建的MaxCompute Java Class名称，如果还没创建package，可以在此处填写 packagename.classname，会自动生成package。
 - 选择类型为GraphLoader或Vertex。
3. 创建成功后，在编辑界面开发Java程序。

通过本地运行调试Graph

通过本地运行方式测试，查看Graph的运行结果是否符合预期。

1. 右键单击编写完成的Java脚本，选择Run。
2. 在Run/Debug Configurations页面上配置运行参数。



- **MaxCompute project**：选择运行Graph的MaxCompute项目。
- **Download Record limit**：下载数据记录限制。默认为100条。

3. 单击OK，开始运行。

说明

- 本地运行会读取warehouse中指定的表数据作为输入，您可以在控制台查看日志输出。
- 如果指定的MaxCompute项目的表数据未被下载至warehouse目录中，会先下载数据；如果数据已经下载，则跳过此步骤。

打包上传

调试成功之后，将Java程序打成JAR包，并作为资源上传至MaxCompute服务端。详情请参见[打包、上传和注册](#)。

运行Graph

通过MaxCompute客户端运行Graph。

1. 在左侧导航栏，单击Project Explorer。
2. 右键单击项目名称，选择Open in Console。
3. 在Console区域，执行如下命令运行Graph。

更多命令，请参见[JAR命令](#)。

```
jar -libjars xxx.jar -classpath /Users/home/xxx.jar com.aliyun.odps.graph.examples.PageRank pagerank_in pagerank_out;
```

3.7.6. 查询非结构化数据

MaxCompute 2.0支持通过外部表的方式直接访问OSS、OTS等。MaxCompute Studio对此提供了一些代码模板方便您快速进行非结构化数据查询开发。本文为您介绍如何使用MaxCompute Studio查询非结构化数据。

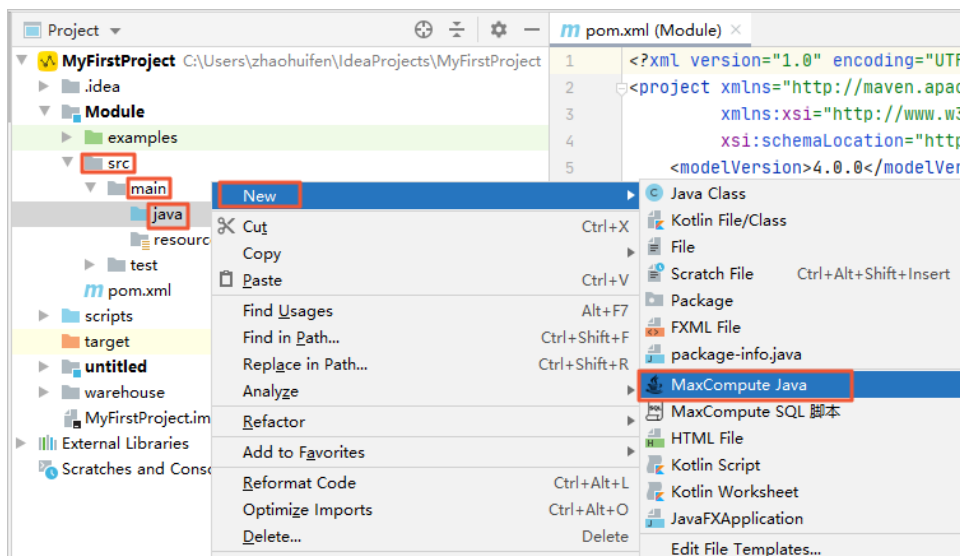
前提条件

您需要完成以下操作：

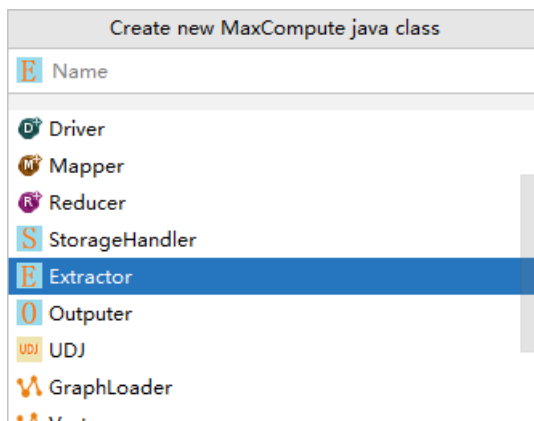
- 管理项目连接
- 创建MaxCompute Java Module

编写StorageHandler、Extractor和Outputter

1. 在Project区域，右键单击Module的源码目录（即src > main > java），选择new > MaxCompute Java。



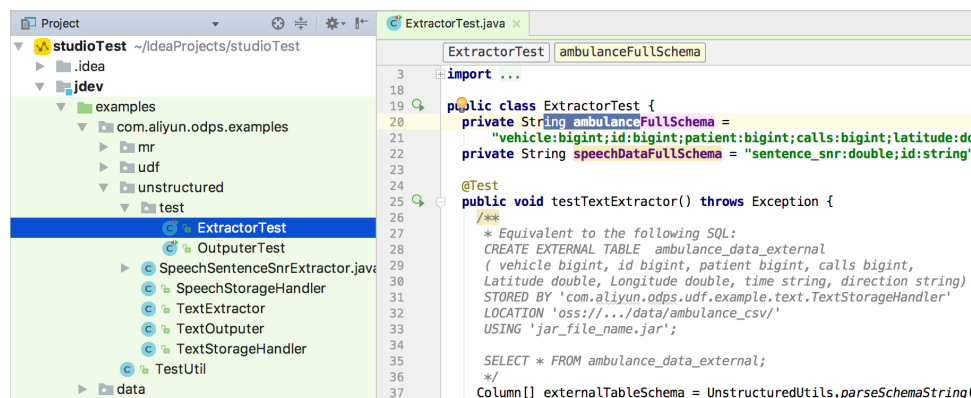
2. 填写Name并选择类型为Extractor、StorageHandler或Outputter，按下Enter键。



- o Name：创建的MaxCompute Java Class名称。如果还没有创建Package，在此处填写packagename.classname，会自动生成Package。
 - o 选择类型为Extractor、StorageHandler或Outputter。
3. 创建Extractor成功后，在编辑界面开发Java程序。代码框中模板已自动填充框架代码，只需要自行编写需要的逻辑代码即可。

通过单元测试调试Extractor和Outputter

您可以参照 *examples* 目录下的单元测试示例，编写测试用例。

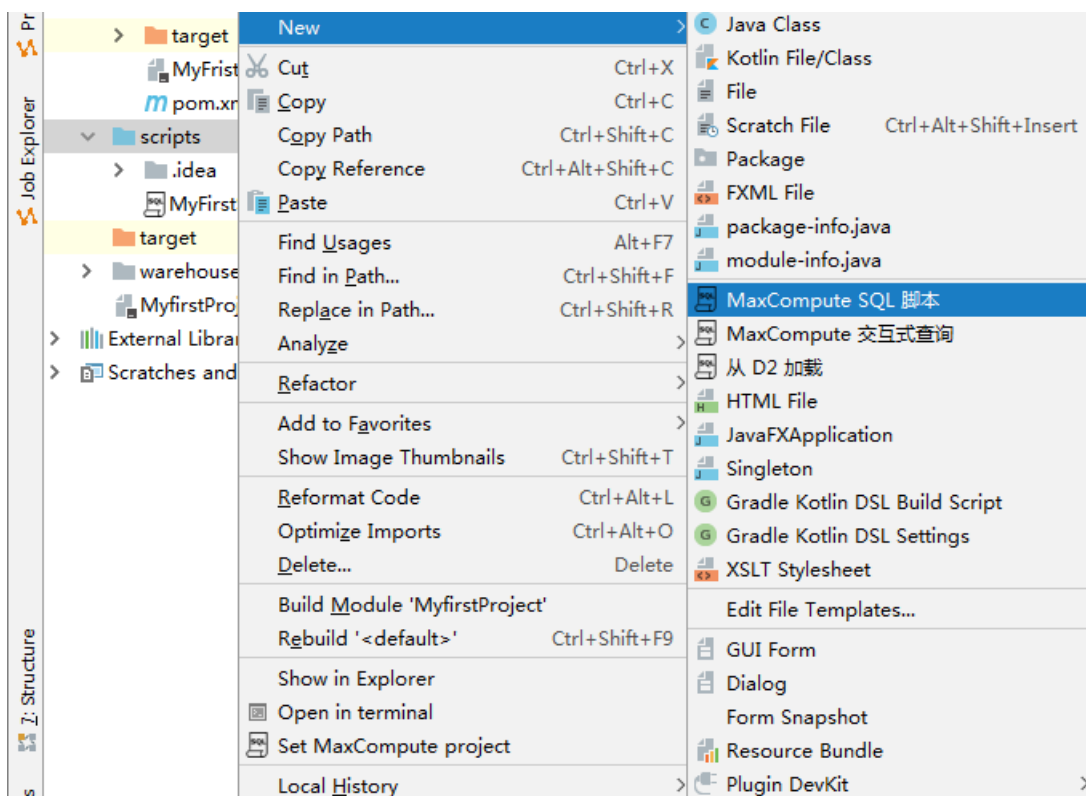


打包上传

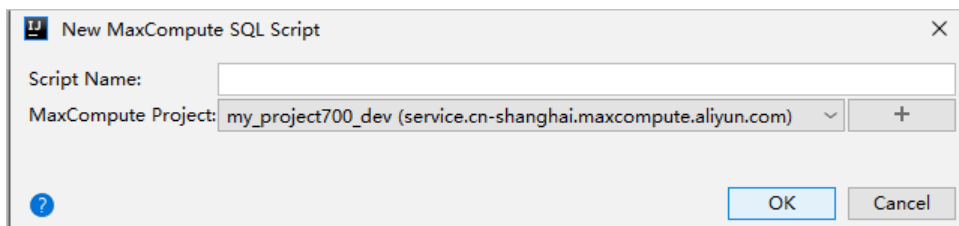
调试成功之后，将Java程序打成JAR包，并作为资源上传至MaxCompute服务端。详情请参见[打包、上传和注册](#)。

查询非结构化数据

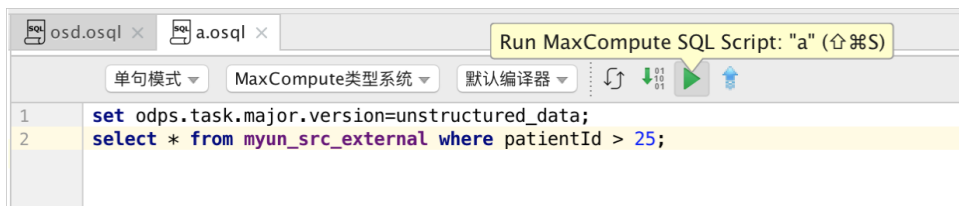
1. 在Project区域，右键单击scripts，选择new > MaxCompute SQL脚本。



2. 在Script Name后输入SQL脚本名称，MaxCompute Project中选择执行脚本的MaxCompute项目，单击OK。



3. 在编辑器中输入创建外部表的SQL语句，并单击▶图标，完成外部表创建。
4. 新建一个MaxCompute SQL脚本，输入如下查询语句，并单击▶图标，查询数据。



3.7.7. 打包、上传和注册

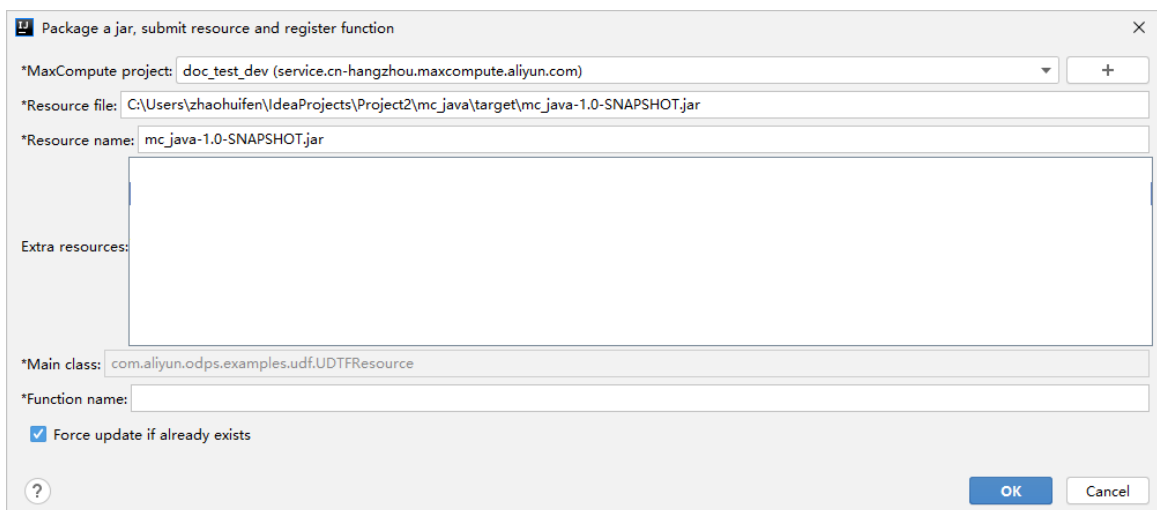
程序开发完成后，需要打包发布至MaxCompute上才可以使用。本文为您介绍如何打包、上传和注册资源。

背景信息

UDF、MapReduce和Graph等Java程序发布到服务端供生产使用前，要经历打包、上传和注册三个步骤。MaxCompute Studio提供了一键发布功能，您可以通过Deploy to server...一键式发布UDF。

打包

1. 右键单击已经编译成功的Java代码，选择Deploy to server...。
2. 在Package a jar, submit resource and register function对话框中，配置相关参数。



- **MaxCompute project**：指定目标MaxCompute项目的名称。
- **Resource file**：指定打包后的JAR包所在路径。
- **Resource name**：指定打包的JAR包名称。
- **Function name**：指定打包的函数名称，即在。
- **Force update if already exists**：选择当资源或函数已存在时是否强制更新。

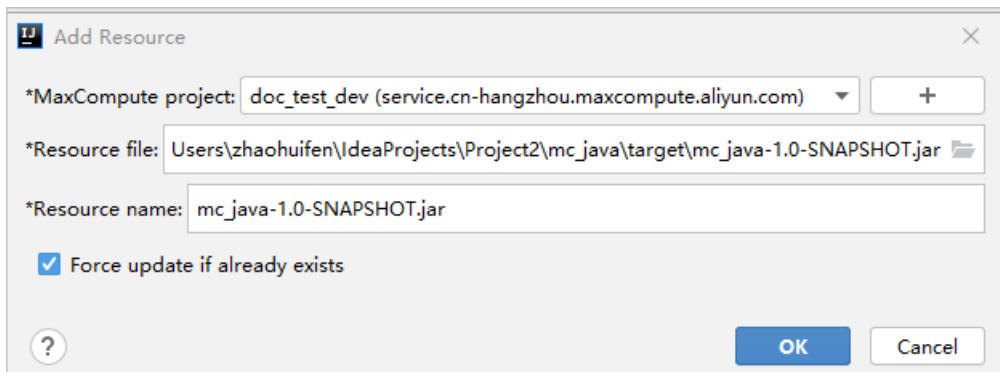
3. 单击OK，完成打包。

说明 如果您有特殊的打包需求，可以自行修改pom.xml打包相关配置。

上传JAR包

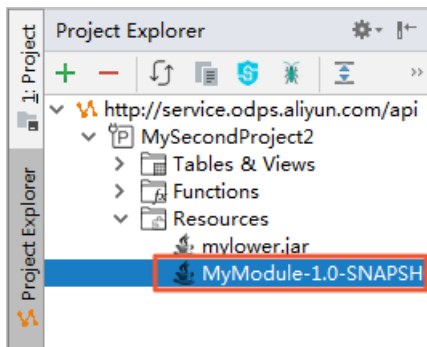
打包成功后，需要将该JAR包上传到MaxCompute服务端。

1. 在顶部菜单栏，单击MaxCompute > 添加资源。
2. 在Add Resource对话框中配置相关信息，单击OK。



- o MaxCompute project：指定目标MaxCompute项目的名称。
- o Resource file：指定打包后的JAR包所在路径。
- o Resource name：指定打包的JAR包名称。
- o Force update if already exists：选择当资源或函数已存在时是否强制更新。

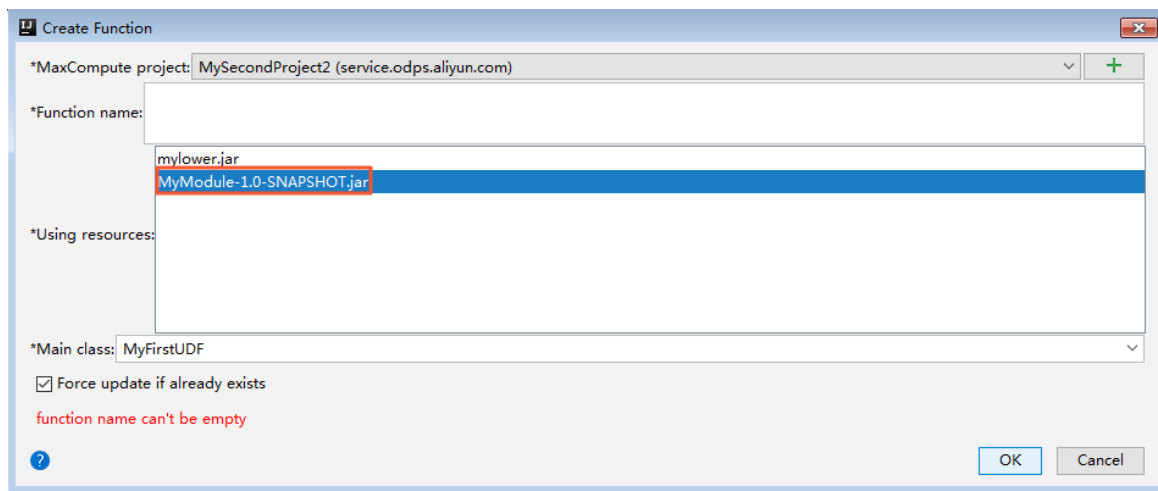
3. 在左侧导航栏，单击Project Explorer。
4. 在Project Explorer区域的Resources节点下可以看到该资源。



注册UDF

JAR包上传完成后，需要注册UDF函数后您才可以调用该函数。

1. 单击顶部菜单栏上的MaxCompute，选择创建UDF。
2. 在Create Function页面配置如下参数，然后单击OK。



- **MaxCompute project**：选择要上传的Project名称。
 - **Function name**：指定函数名称。
 - **Using resources**：指定函数依赖的JAR包名称。
 - **Main class**：指定JAR的主类。
 - **Force update if already exists**：当资源或函数已存在时是否强制更新。
3. 在左侧导航栏，单击Project Explorer。
 4. 在Project Explorer区域的Functions节点下看到该函数。



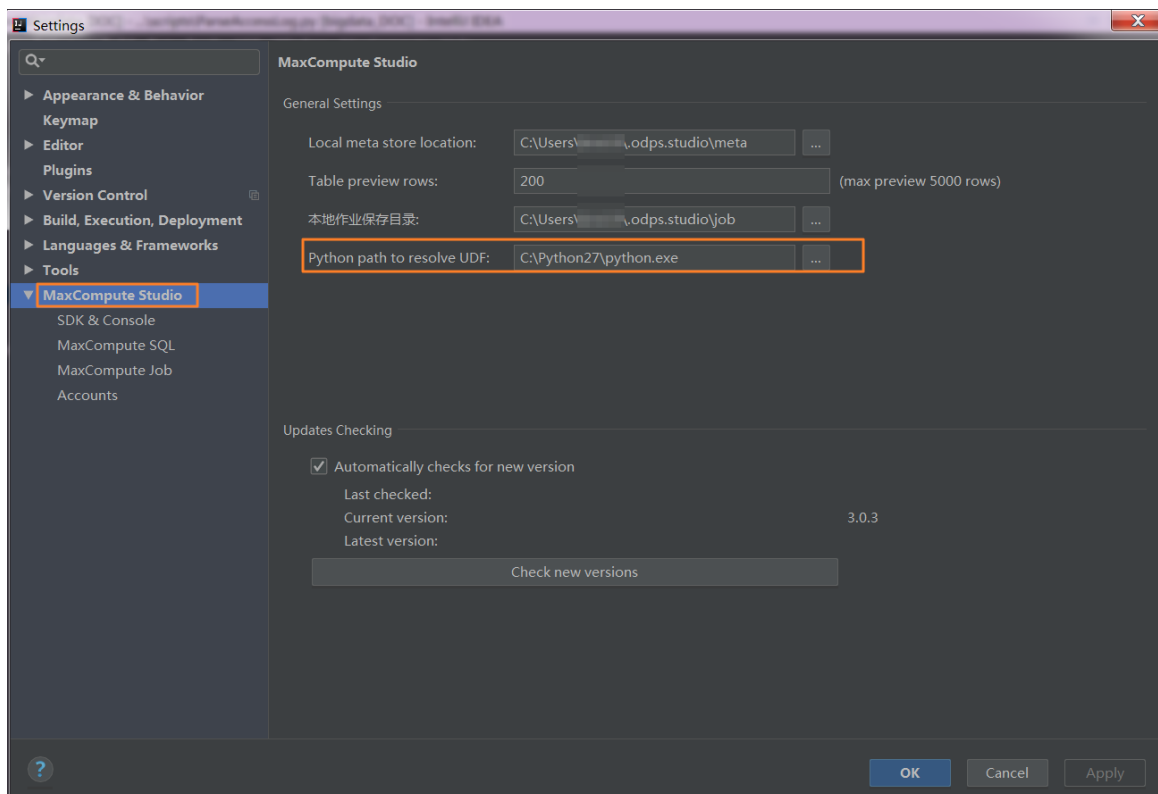
3.8. 开发Python程序

3.8.1. 配置Python开发环境

MaxCompute Studio支持您在IntelliJ IDEA中完成Python开发，例如UDF和PyODPS脚本。本文为您介绍如何配置Python开发环境。

安装PyODPS

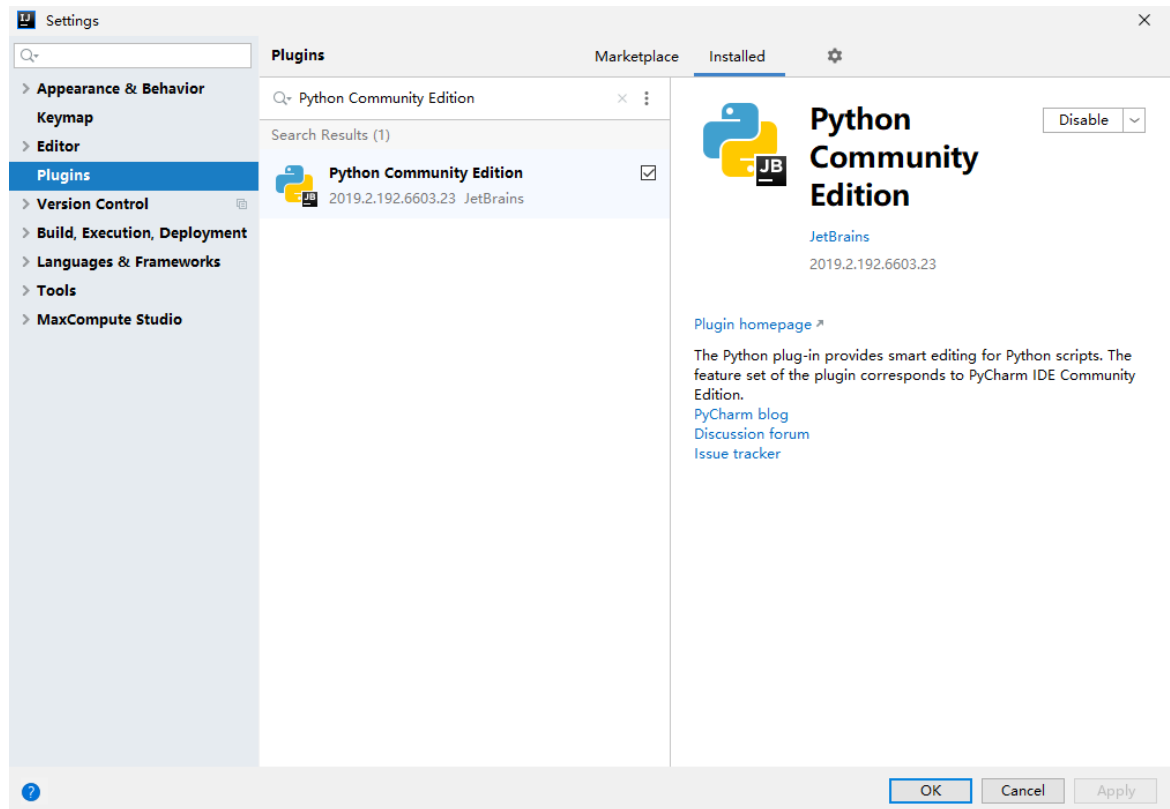
1. PyODPS是MaxCompute的PyODPS SDK，详情请参见[安装指南及使用限制](#)。
2. 运行IntelliJ IDEA，在顶部菜单栏上，单击File > Settings。
3. 在Settings页面左侧导航栏，单击MaxCompute Studio。
4. 设置Python path to resolve UDF为本地Python安装目录。



安装Python插件

在IntelliJ IDEA的插件仓库中搜索Python或者Python Community Edition插件并安装。

1. 在IntelliJ IDEA顶部菜单栏上，单击File，选择Settings。
2. 在Settings页面左侧导航栏，单击Plugins。
3. 在搜索框搜索Python Community Edition，并安装。

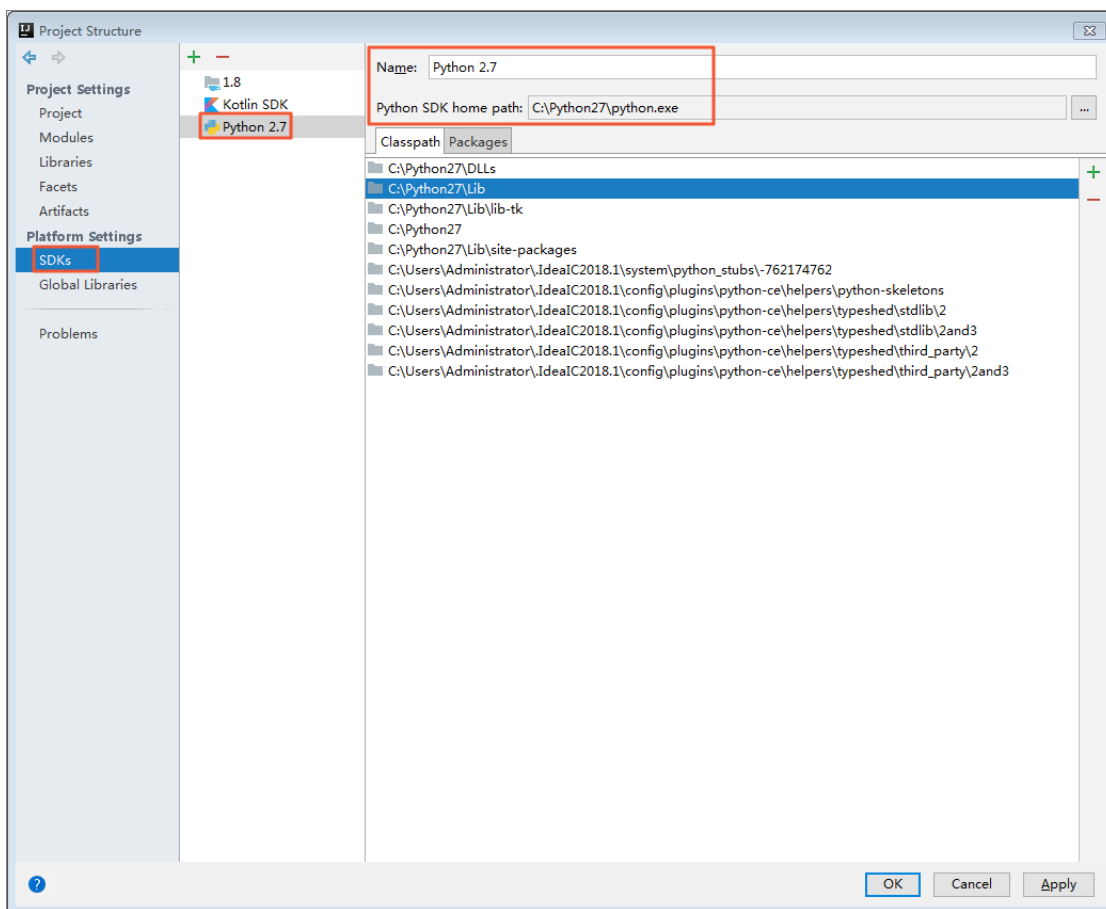


配置Python依赖

配置Studio Module对Python的依赖，即可进行MaxCompute Python的开发。

1. 在顶部菜单栏，单击File > Project Structure。
2. 增加SDK。
 - i. 在左侧导航栏上，单击SDKs。

ii. 单击上方的加号添加Python SDK。

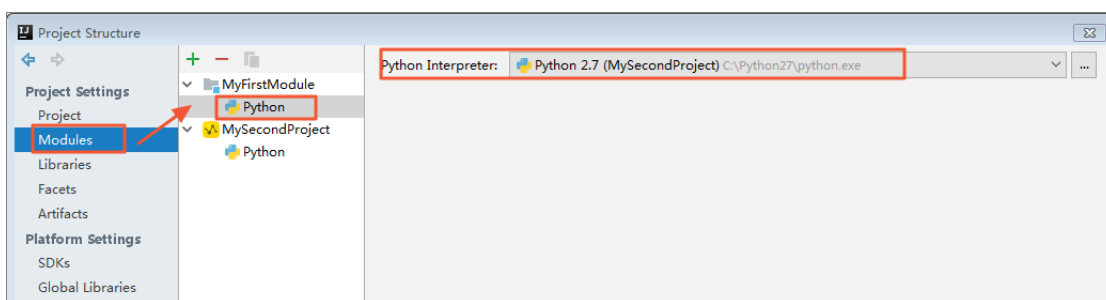


iii. 单击Apply完成添加。

3. 配置Modules。

i. 在左侧导航栏上，单击Modules。

ii. 单击+ 添加Module依赖Python Facets。



iii. 单击Apply保存配置。

3.8.2. 开发Python UDF

MaxCompute Studio支持Python UDF开发，本文为您介绍如何开发、测试和注册发布Python UDF。

前提条件

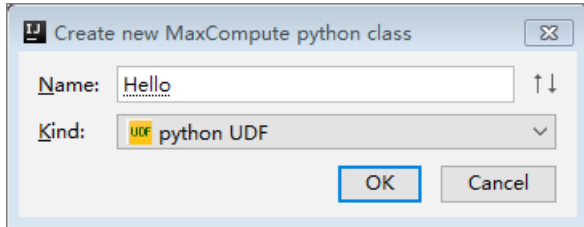
您必须完成以下操作：

- [管理项目连接](#)

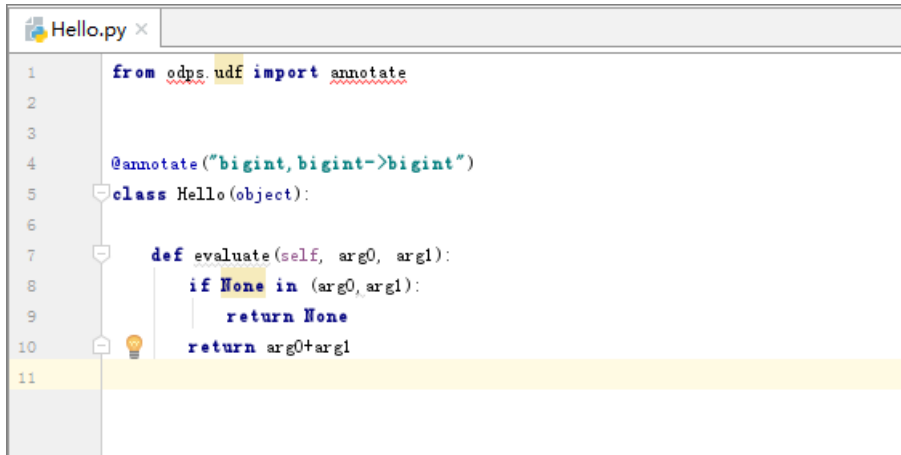
- 配置Python开发环境

开发Python UDF

- 在Project区域MaxCompute Studio目录下，右键单击scripts，选择New > MaxCompute Python。
- 在Create new MaxCompute python class对话框中输入类名Name，选择类型为Python UDF，单击OK完成。



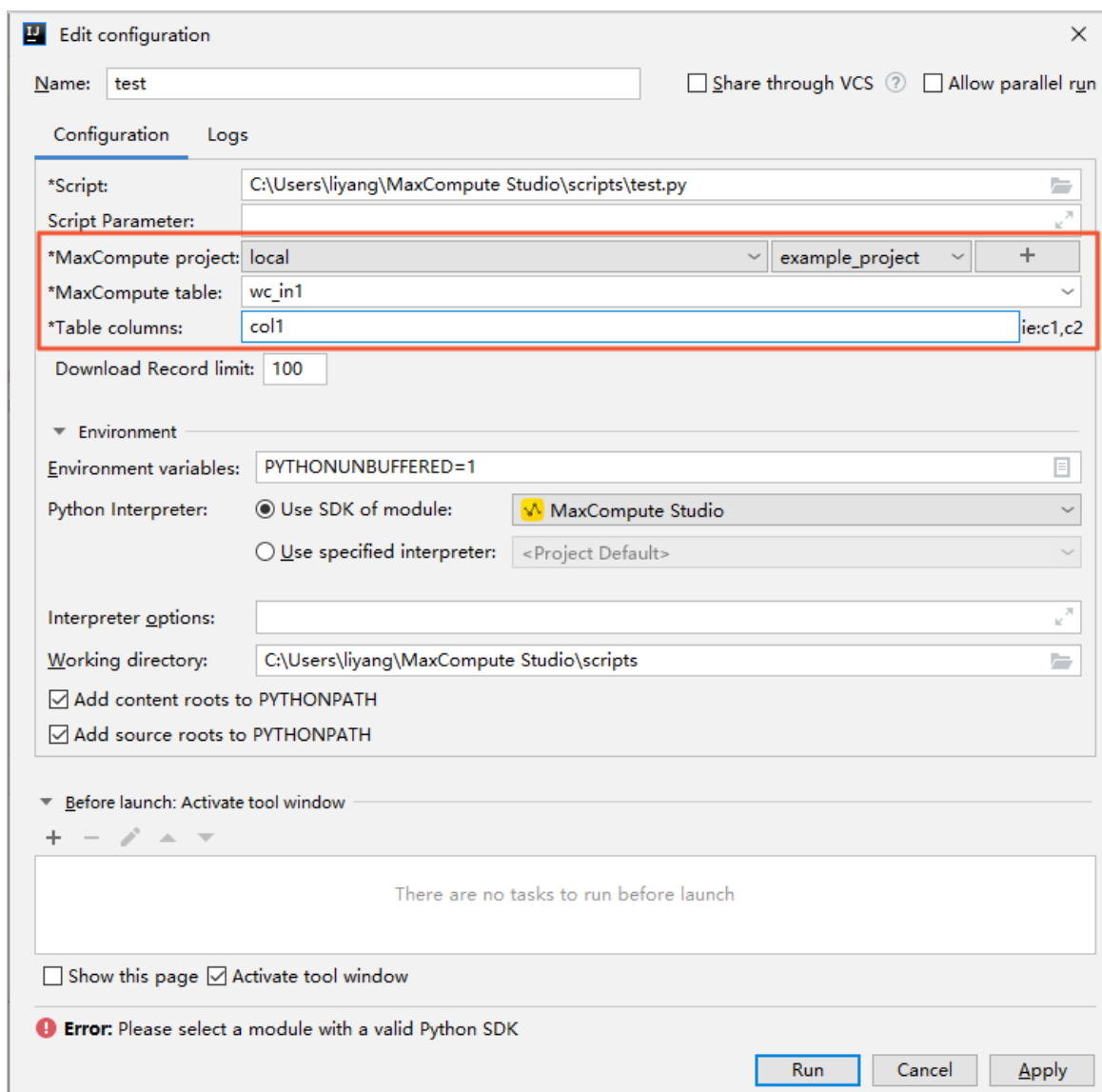
- 在编辑框中编写UDF代码。



测试UDF

UDF开发完成后，需要测试代码是否符合预期。MaxCompute Studio支持本地测试，即下载表的部分示例数据在本地运行并进行调试。

- 右键单击已经编辑完成的Python UDF脚本，选择RUN。
- 在Edit configuration页面，配置相关参数，单击OK。



- **MaxCompute project**：UDF运行使用的MaxCompute空间。本地运行时选择local。
- **MaxCompute table**：UDF运行时需要使用的MaxCompute表的名称。
- **Table columns**：UDF运行时需要使用的MaxCompute表的列信息。
- **Download Record limit**：下载数据记录限制。默认为100条。

说明

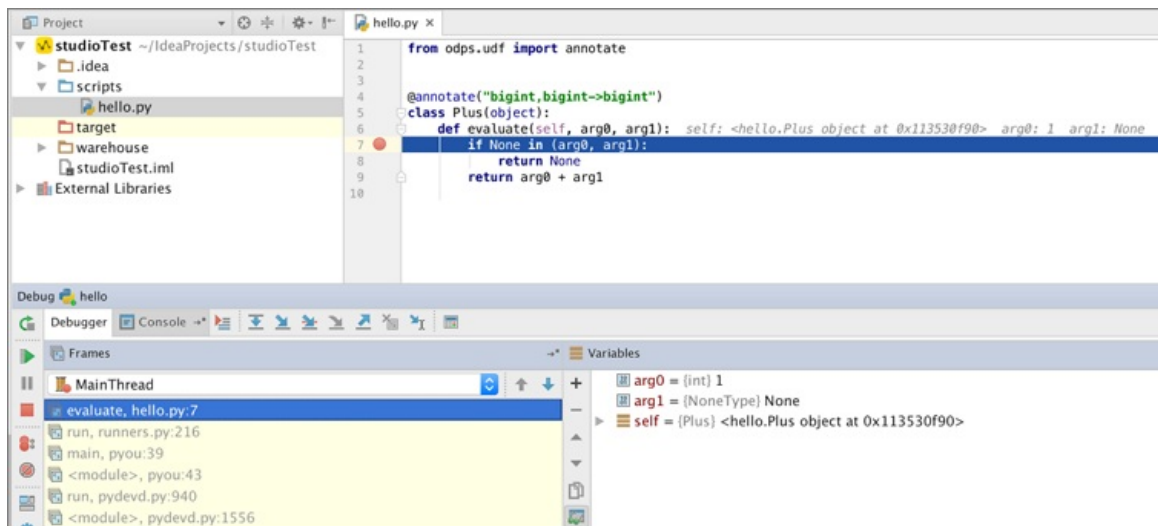
- 如果已经下载数据，则不会再次重复下载。如果需要再次下载，请在MaxCompute客户端使用Tunnel命令下载数据。
- 默认下载100条数据，如果需要更多数据测试，请在MaxCompute客户端使用Tunnel命令或者MaxCompute Studio的表下载功能下载数据。
- 下载完成后，您可以在warehouse目录下该表的data文件中看到下载的示例数据。

3. 本地运行框架会根据您指定的列，获取data文件中指定列的数据，调用UDF本地运行。

说明 本地运行是通过PyODPS的pyou脚本实现的，命令为 `pyou hello.Plus<data>`。安装完PyODPS后可以使用相应的命令检查该脚本是否存在。

- 如果您是Windows系统，请运行 `${python}/../Scripts/pyou` 命令。
- 如果您是mac OS系统，请运行 `${python}/../pyou` 命令。

4. 您可以在控制台查看打印结果。



发布Python UDF

Python UDF测试通过后，即可发布到生产环境上使用。详情请参见[打包](#)、[上传](#)和[注册](#)。

3.8.3. 开发PyODPS脚本

本文为您介绍如何开发PyODPS脚本。

前提条件

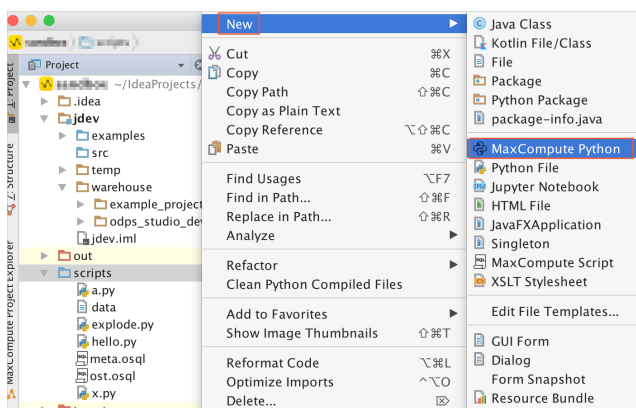
已配置Python开发环境，详情请参见[配置Python开发环境](#)。

背景信息

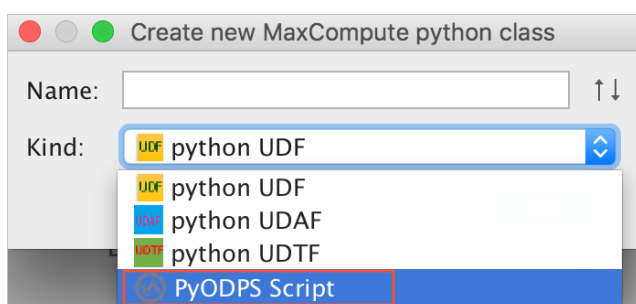
PyODPS是MaxCompute Python版本的SDK，提供对MaxCompute对象的基本操作和DataFrame框架。您可以通过PyODPS在MaxCompute上分析数据。

操作步骤

1. 在Project区域右键单击scripts，选择New > MaxCompute Python。

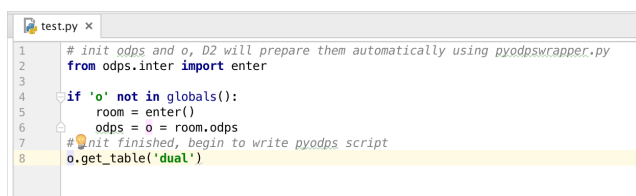


2. 在Create new MaxCompute python class对话框，填写Name，从Kind列表中选择PyODPS Script。



3. 新建MaxCompute PyODPS脚本后，PyODPS脚本模板会通过PyODPS Room自动初始化 odps 和 o 两个对象。

通过DataWorks开发PyODPS脚本时，系统会自动创建Room。通过Intellij IDEA开发PyODPS脚本时，需要创建Room，详情请参见[PyODPS文档](#)。



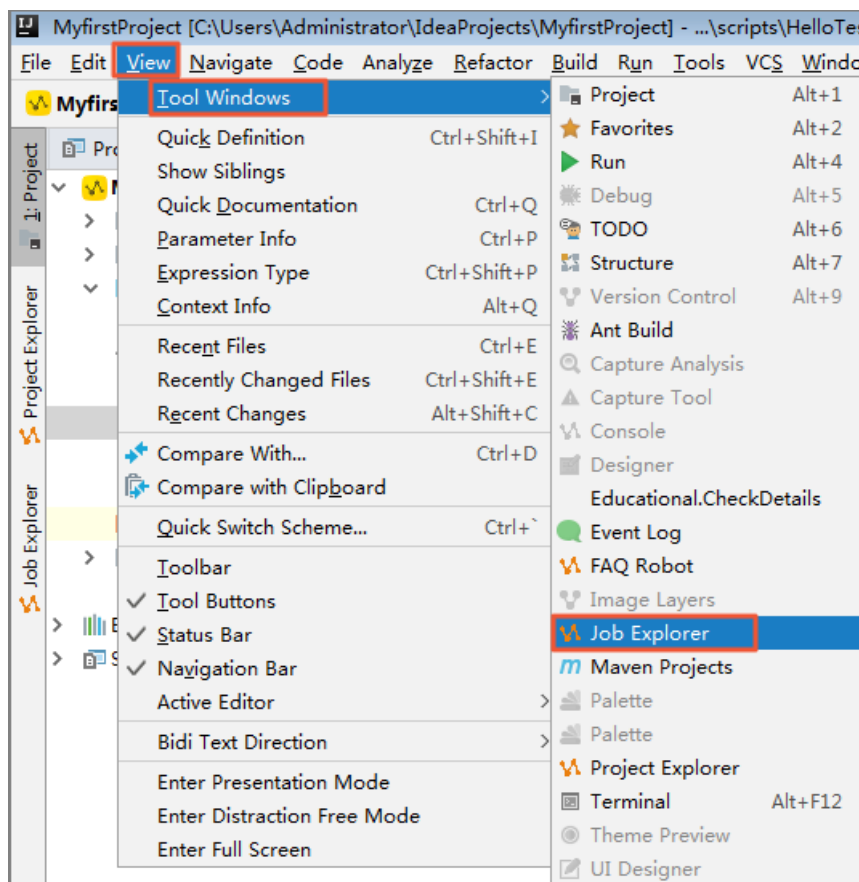
3.9. 管理MaxCompute作业

3.9.1. 作业浏览

通过MaxCompute Studio的Job Explorer可以方便查看当前用户提交的MaxCompute实例情况，包括运行状态、作业类型、起止时间等。

打开Job Explorer

在顶部菜单栏，单击View > Tool Windows > Job Explorer打开Job Explorer。



查看项目的作业实例

Job Explorer支持按照状态、提交者、日期和用时等信息查询提交的作业列表。例如：

- 按照状态查询提交的作业列表。选择状态为失败，查看过去24小时内执行失败的作业。
- 按照日期查询提交的作业列表。从日期列表框，滑动指针选择时间，获取指定时间段的作业列表。

说明

- 默认只显示符合条件的前1000条作业。如果需要超过1000条的作业信息，请考虑更新过滤条件。
- 您可以单击Id与名称、提交时间对作业进行简单的排序。

查看作业排队队列

活动状态的作业如果正在排队队列中等待调度，队列位置会展示当前排队的位置，优先级会展现作业的全局优先级。

说明 在Running Instances下的作业状态、队列位置等信息会自动更新，作业结束后会从列表中移除。

保存作业日志

目前作业的Logview日志默认保存7天。如果需要查看更长时间的Logview信息，建议您将Logview保存在本地。

1. 双击列表中的作业，在右侧打开作业的详细信息。
2. 在工具栏上，单击**保存**，将日志保存到本地。

如果需要自定义保存文件的目录，可以在MaxCompute Studio的Settings页面中进行配置。

3.9.2. 作业实例详情

本文为您介绍如何查看作业实例详情。

查看作业实例详情

MaxCompute Studio支持2种方式查看MaxCompute作业实例详情。

- 通过Logview URL或本地的Logview离线文件以只读方式打开作业详情。

使用Logview查看作业的详细信息是比较常用的方式。使用Logview还可以查看其他用户在其他项目空间中提交的任务状态。MaxCompute Studio支持通过输入一个有效的Logview URL打开任意一个作业详情。

在IntelliJ IDEA顶部菜单栏中，单击**MaxCompute > 打开Logview**，选择导出本地的Logview离线文件或者将有效Logview URL复制到**Open job detail by logview**对话框中。

- 双击作业列表下的作业，可以查看该实例的详细信息。

作业详情视图

作业详情页面包括顶部的工具栏，左半部分的基本属性栏以及右半部分详细视图页，其中详细视图页主要包含四个视图：

- **执行图**：以DAG图的方式显示作业整体信息，可查看子任务间的依赖关系以及各个子任务的详细执行计划。
- **详情**：以表视图的方式展示作业详细信息，包括子任务列表、各子任务的Worker列表、Worker处理数据量、执行时间及状态信息等。
- **脚本**：显示该作业提交时所对应的SQL语句以及提交作业的参数配置信息。
- **结果**：显示该作业运行结果。



还包含如下信息：

- **时序图**：显示作业执行时间线，可以从不同粒度查看执行的时序，并提供了多种过滤器。
- **概要 (JSON)**：以JSON格式显示作业运行详细信息。
- **分析**：提供作业执行散点图、长尾柱状图及数据倾斜图。

工具栏

- **<、>**：页面左右折叠。用来收起或完全展开左右侧视图，允许您最大化某一个视图进行查看。
- **⊗**：停止作业。用来中断正在执行的作业，需要具有响应权限才能停止作业（项目所有者或管理员）。
- **🔄**：刷新详情。对于运行中作业的基本信息，例如状态、Quota等会自动刷新。但右侧各个详情视图不会自动刷新，需要您手动刷新。
- **📄**：拷贝Logview。
- **🌐**：在浏览器打开作业详情。即生成Logview URL，并通过浏览器打开。
- **📁**：将作业详情信息保存为本地文件。
- **☑ Enable auto refresh**：是否开启自动刷新。对于运行中作业，开启自动刷新后，MaxCompute Studio会对作业执行全量定时刷新。

基本信息页面

基本信息页面展示了作业的基本信息，包括ID、创建人、状态、起止时间、计算资源用量、输入项（作业的输入表）、输出项等（作业的输出表）。运行中作业的基本信息会自动定时刷新。

双击表名即可查看对应表的基本信息页面。

执行图

执行图作为日常主要使用工具，以可视化的方式展示Fuxi Job、Fuxi Task以及Operation的依赖关系，同时提供一系列辅助工具，例如作业回放、进度图和热度图等，是排查问题的好帮手。

上图中各序号对应说明如下：

1. 可单击跳转其他层次。
2. 缩放辅助工具。
3. 依赖表。
4. Fuxi Task节点。
5. 鹰眼。
6. 展示。
7. 默认打开Fuxi Task层依赖。

执行图可展示三个维度的作业依赖关系：Fuxi Job层、Fuxi Task层、Operation层。可单击向上箭头进行维度切换，默认会展示Fuxi Task层依赖关系。

- Fuxi Job层

单击**MaxCompute Job**打开Fuxi Job层。Fuxi Job层节点内包含Fuxi Task名称、起止时间等。

- Fuxi Task层

双击任一Fuxi Job节点即可进入Fuxi Task层。

当有多个Fuxi Job时，默认打开最后一个Fuxi Job的Fuxi Task层。该层可展示Fuxi Task的依赖关系，输入输出表及分区等信息。当作业结束后单击工具栏中的**进度图**，可以选择**热度图**、**输出热度图**、**Task时间热度图**和**Instance热度图**等。进度图表示节点的完成进度，热度图通过颜色区分节点热度。

Fuxi Task节点内容如下：

- Instance Count：表示为 $a/b/c$ ，指某一时刻正在运行的子任务实例个数为a，已结束任务实例个数b，总任务实例个数c。
- I/O Records：同理为某一时刻的输入记录量和输出记录量。
- 百分比与橙色进度条：表示该任务运行情况，该比例根据子任务运行实例分析得出。
- 子任务间连线：显示输出的记录数量。箭头表示数据流动方向。

- Operation层

双击任一Fuxi Task即可打开Operation层。

Operation层揭示了Fuxi Task内在的运行方式，单击任一节点即可显示Operation完整信息。

 **说明** 非SQL类型作业，仅能展示Fuxi Job和Fuxi Task层作业，不展示Operation层。

详情页

主要针对SQL DML类作业，展示作业在计算集群上的Fuxi Task列表、计算节点列表等。通常一个作业对应一个或多个Fuxi Job，每个Fuxi Job拆分成多个Fuxi Task（阶段），每个Fuxi Task包含多个Fuxi Instance（Worker）。右键单击Fuxi Instance可以查看作业运行的标准输出、标准错误和Debug Info。

对应的序号说明如下：

1. Fuxi Job Tab。
2. Fuxi Task列表。
3. 每个Fuxi Task详细信息及计算节点列表。

作业回放

MaxCompute Studio支持作业回放功能，作业回放就像播放媒体文件一样，可在12s内回顾该作业执行的历史轨迹。该功能主要用于帮助用户了解MaxCompute实例在不同时刻运行状态，快速判断子任务级运行顺序及消耗时间，掌握作业执行关键路径，从而针对运行较慢的子任务进行优化。

单击 > 按钮即可开始播放，再次单击则暂停。您也可以手动拖动进度条。

说明 回放功能仅通过时间估算某一个时刻IO数据量，从而确定完成进度，并不能代表该时刻真实IO数据量。Running状态作业不支持回放功能。

时序图

以甘特图的方式展示作业分布式执行的详细数据，可以调整展示粒度，将每一个计算节点都在甘特图中展示。可以通过甘特图直观的看出作业运行的时间瓶颈和长尾节点等。同时提供多种过滤器，能够直接筛选出作业执行的关键路径、最大数据节点和最长时间节点等。

分析页

分析页展示作业的长尾节点（Worker）、数据倾斜节点（Worker）。展示节点散点图、柱状图和辅助作业执行瓶颈诊断。散点图和柱状图支持从图中节点的准详情页查看Fuxi Instance详情。

结果页

结果页会根据作业类型及提交作业时的参数设置展示不同页面。

- SELECT 语句并且设置 `odps.sql.select.output.format = HumanReadable`，结果以文本方式展示。
- SELECT 语句并且未设置output format参数，结果以TABLE方式展示。
- 对于数据输出到表的脚本，展示输出表名及表详情页的链接。
- 对于异常作业，结果页显示异常详情。

3.10. 工具集成

3.10.1. 集成MaxCompute客户端

MaxCompute Studio集成了MaxCompute客户端，您可以在MaxCompute Studio中直接运行MaxCompute客户端。

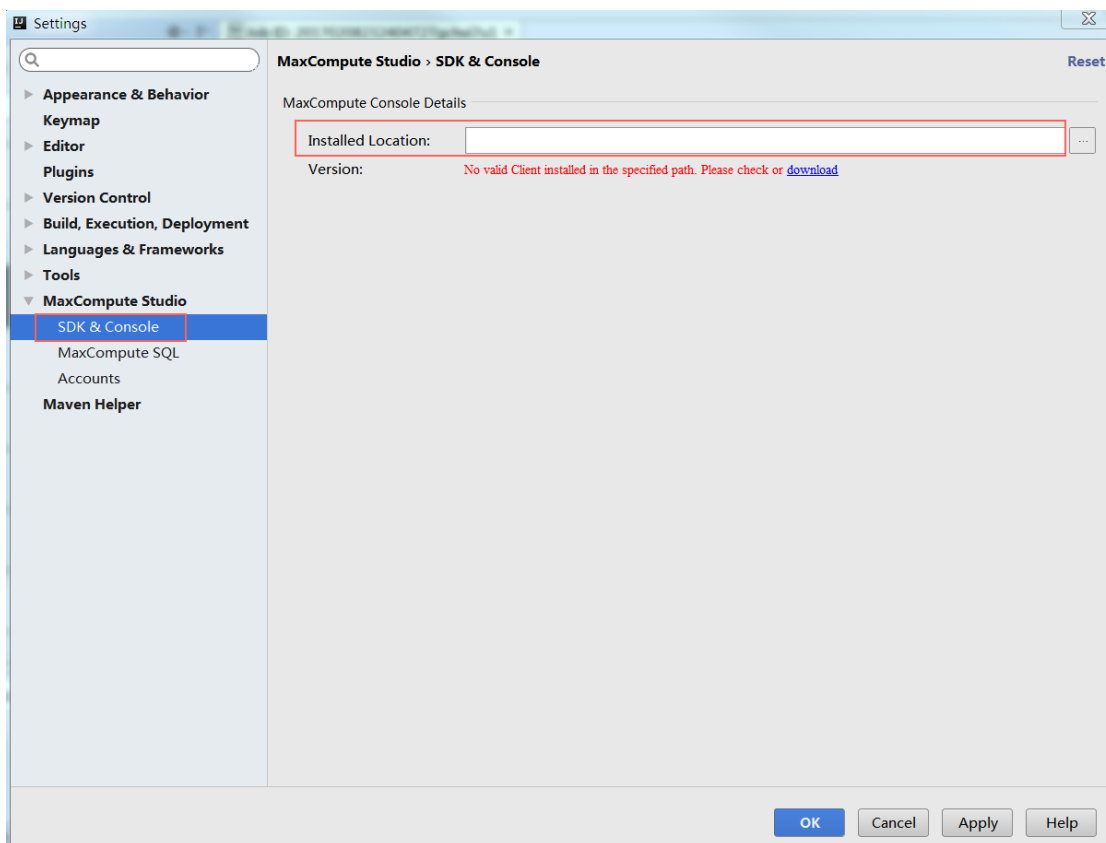
背景信息

MaxCompute Studio中已包含最新版MaxCompute客户端程序，并指定为默认客户端。您也可自行指定其他版本客户端程序。

操作步骤

1. 配置客户端安装路径。
 - i. 在IntelliJ IDEA顶部菜单栏，单击File > Settings...
 - ii. 在Settings页面左侧导航栏，选择MaxCompute Studio > SDK & Console。

iii. 在Installed Location后选择MaxCompute客户端的本地安装路径。

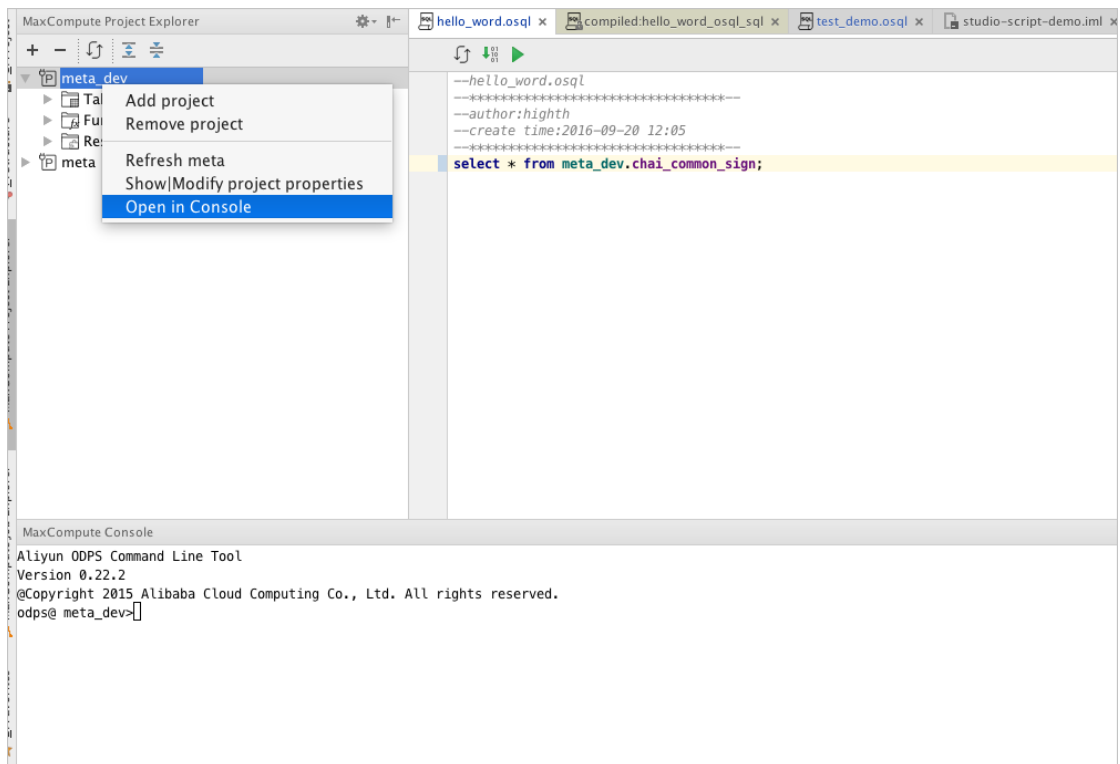


iv. 单击OK，完成客户端安装路径配置。

2. 在MaxCompute Studio中打开MaxCompute客户端。

i. 在IntelliJ IDEA左侧导航栏，单击Project Explorer。

- ii. 在选中的项目节点上，单击右键选择**Open in Console**，打开MaxCompute客户端。



- iii. 在**Console**区域，运行客户端命令。详情请参见[常用命令列表](#)。

4.相关下载

本文将为您提供在使用MaxCompute过程中，可能用到的相关工具及插件的下载地址。

- SDK下载信息：如果您使用Maven，可以从[Maven库](#)中搜索odps-sdk，获取不同版本的Java SDK。
- 客户端：进入[客户端下载页面](#)，下载所需版本客户端。
- IntelliJ IDEA开发插件：下载[IDEA工具](#)，[Studio插件](#)即可下载所需的IntelliJ IDEA开发插件。
- JDBC：MaxCompute提供开源JDBC，您可以在Git Hub[下载JDBC](#)。
- PHP SDK：您可以在Git Hub[下载PHP SDK](#)。