阿里云
MaxCompute

工具及下载

文档版本：20180925

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
	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。	 禁止： 重置操作将丢失用户配置数据。
	该类警示信息可能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，恢复业务所需时间约10分钟。
	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。	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。	单击 **确定**。
`courier` 字体	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

1 客户端

本文将为您介绍如何借助客户端命令行工具使用 MaxCompute 服务的基础功能。在使用 MaxCompute 客户端前，请首先安装并配置客户端 。

说明：

- 请不要依赖客户端的输出格式来做任何的解析工作。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。
- 关于客户端的基本命令介绍，请参见 基本命令。
- 新版客户端：点击此处 即可下载新版客户端。关于不同环境下客户端的安装配置信息请参见安装并配置客户端

安装并配置好客户端后，您可借助命令行工具进行以下操作：

获取帮助

若想显示客户端的帮助信息，命令格式如下所示：

```
odps@ >./bin/odpscmd -h;
```

您也可以在交互模式下键入`h;`或`help;`（不区分大小写）。

客户端还提供了 `help [keyword];` 命令，可获取到与关键字有关的命令提示。例如：输入 `help table;` 可以得到与 table 操作相关的命令提示，如下所示：

```
odps@ odps> help table;
Usage: alter table  merge smallfiles
Usage: show tables [in ]
      list|ls tables [-p,-project ]
Usage: describe|desc [.] [partition()]
Usage: read [.] [([,..])] [PARTITION ()] [line_num]
```

启动参数

在启动时，您可指定一系列参数，如下所示：

```
Usage: odpscmd [OPTION]...
where options include:
    --help (-h)for help
    --project= use project
    --endpoint= set endpoint
    -u  -p  user name and password
    -k  will skip begining queries and start from specified position
    -r  set retry times
```

```
    -f <"file_path;"> execute command in file
    -e <"command;[command;]..."> execute command, include sql command
    -C will display job counters
```

以 -f 参数为例，操作如下：

1. 准备本地脚本文件 script.txt，假设存放在 D 盘，文件内容如下所示：

```
DROP TABLE IF EXISTS test_table_mj;
CREATE TABLE test_table_mj (id string, name string);
DROP TABLE test_table_mj;
```

2. 运行如下命令：

```
odpscmd\bin>odpscmd -f ./script.sql;
```

交互模式

直接运行客户端即可进入到交互模式，如下所示：

```
[admin: ~]$odpscmd
Aliyun ODPS Command Line Tool
Version 1.0
@Copyright 2012 Alibaba Cloud Computing Co., Ltd. All rights reserved.
odps@ odps> INSERT OVERWRITE TABLE DUAL SELECT * FROM DUAL;
```

在光标位置输入命令（以分号作为语句的结束标志），回车即可运行。

续跑

- 在用 -e 或 -f 模式运行时，如果有多条语句，想从中间某条语句开始运行，可以指定参数 -k，表示忽略前面的语句，从指定位置的语句开始运行。当指定参数 <= 0 时，从第一条语句开始执行。
- 每个以分号分隔的语句被视为一条有效语句，在运行时会打印出当前运行成功或者失败的是第几条语句。

示例如下：

假设文件/tmp/dual.sql 中有三条 SQL 语句，如下所示：

```
drop table dual;
create table dual (dummy string);
```

```
insert overwrite table dual select count(*) from dual;
```

若想忽略前两条语句，直接从第三条语句开始执行，命令格式如下所示：

```
odpscmd -k 3 -f dual.sql
```

获取当前登录用户

若想获取当前登录用户，命令格式如下所示：

```
whoami;
```

示例如下：

```
odps@ hiveut>whoami;
Name: odpstest@aliyun.com
End_Point: http://service.odps.aliyun.com/api
Project: lijunsecuritytest
```

通过以上命令，即可获取当前登录用户的云账号、使用的 End_Point 配置和项目名。

退出

若想退出客户端，命令格式如下所示：

```
odps@ > quit;
```

您也可输入如下命令退出客户端：

```
odps@ > q;
```

2 MaxCompute Studio

2.1 认识Studio

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，可以帮助您方便地进行数据开发。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。IDEA的界面详情请参见界面操作文档。

MaxCompute Studio在IntelliJ的基础上提供了以下功能界面。

- **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

Studio的大部分功能需要您首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。

> 说明：
>
> - 您需要指定作为目标的项目连接，方可通过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，因此首先创建一个MaxCompute项目空间的连接是非常必要的。
> - 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。

您双击某个数据表，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。

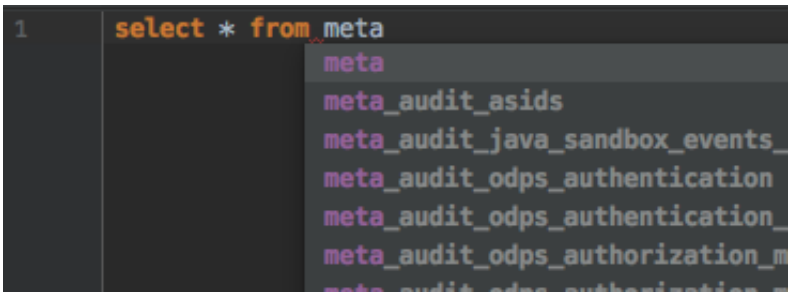
1. 打开Studio，导航至**File > New > Project**或者**File > New > Module**…。
2. 创建一个MaxCompute Studio类型的项目或者模块。
3. 导航至**File > New > MaxCompute Script** 或者右击菜单**New > MaxCompute Script**，即可创建一个MaxCompute SQL脚本文件。

   > 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，您也可以通过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。如下图所示：

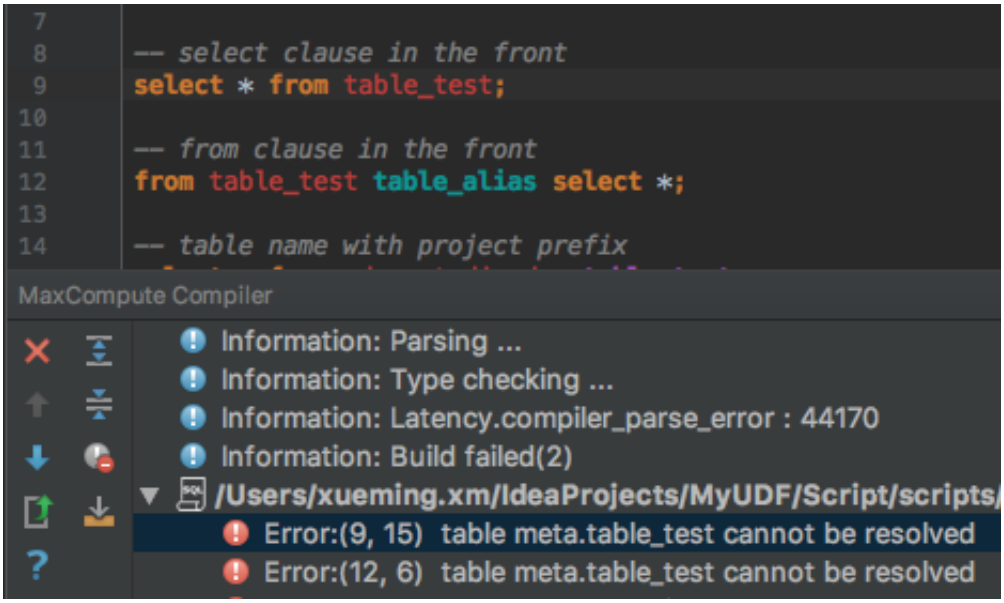
通过代码补全功能，Studio可以根据代码上下文，提示您项目空间名称、表、字段、函数、类型、代码关键词等，并根据您的选择，自动补全代码。如下图所示：



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。



- 提交作业

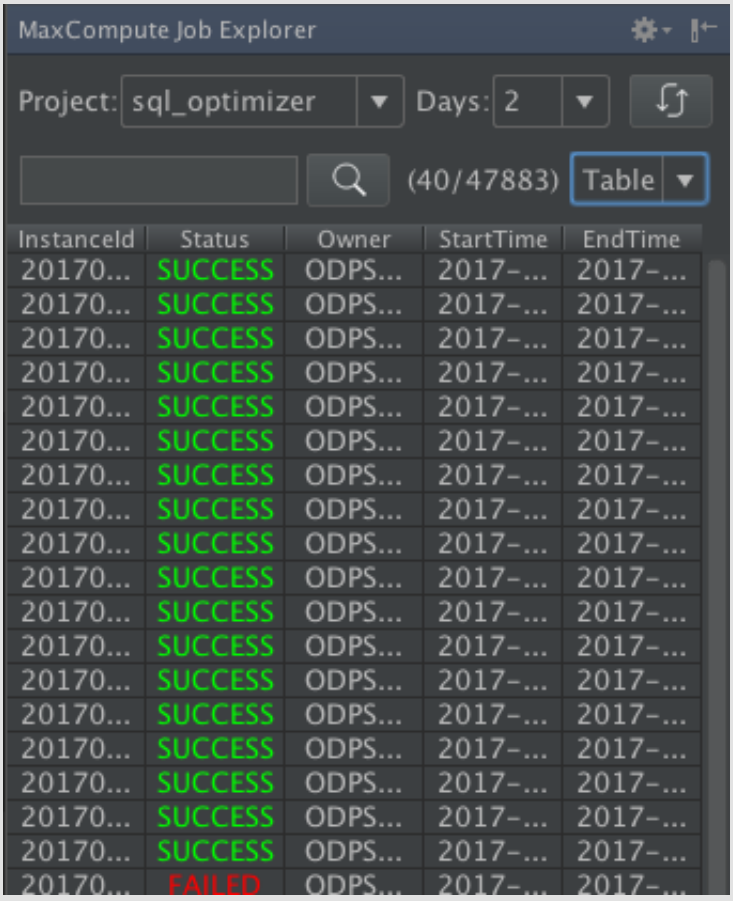
  单击SQL编辑器工具条上的 图标，会在本地编译之后，把SQL脚本提交到MaxCompute指定的项目空间排队执行。

查看历史作业

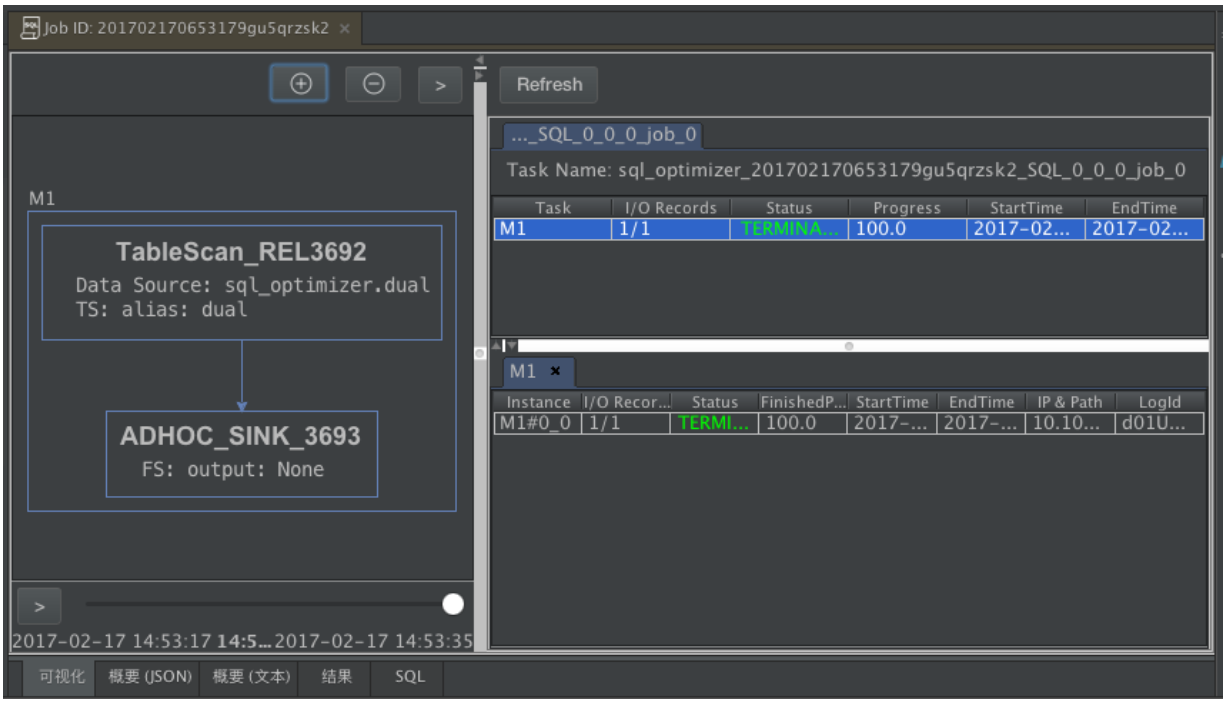
打开作业浏览器，您即可查看指定项目空间上近期执行的作业。

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。



双击其中一个作业，便可查看作业的的详情信息。如下图所示：



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。

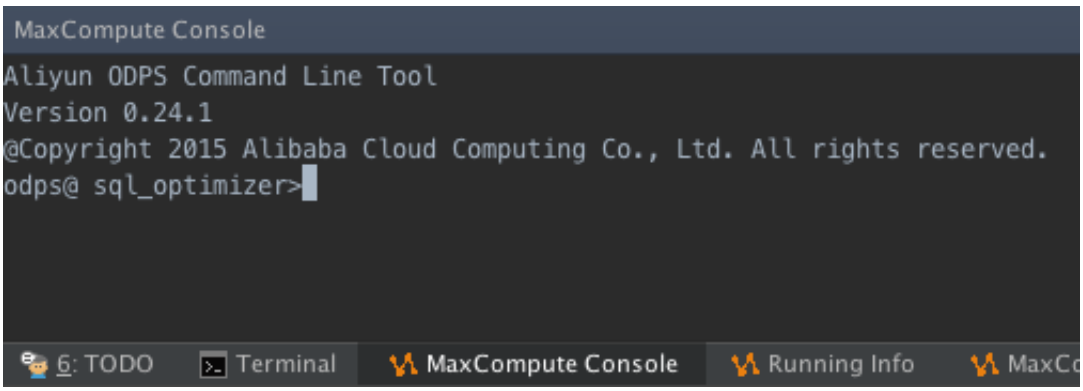
开发MapReduce和UDF

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。详情请参见安装*IntelliJ IDEA*。

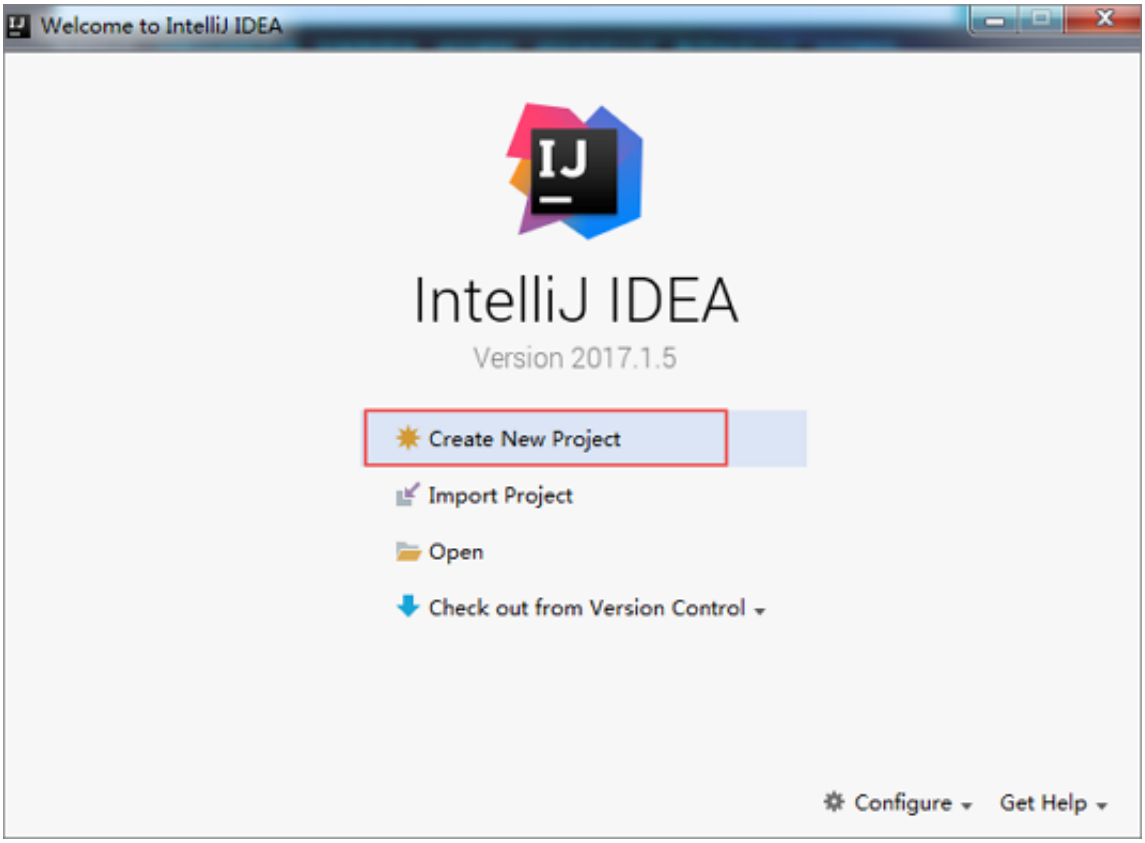
2.2 项目空间连接管理

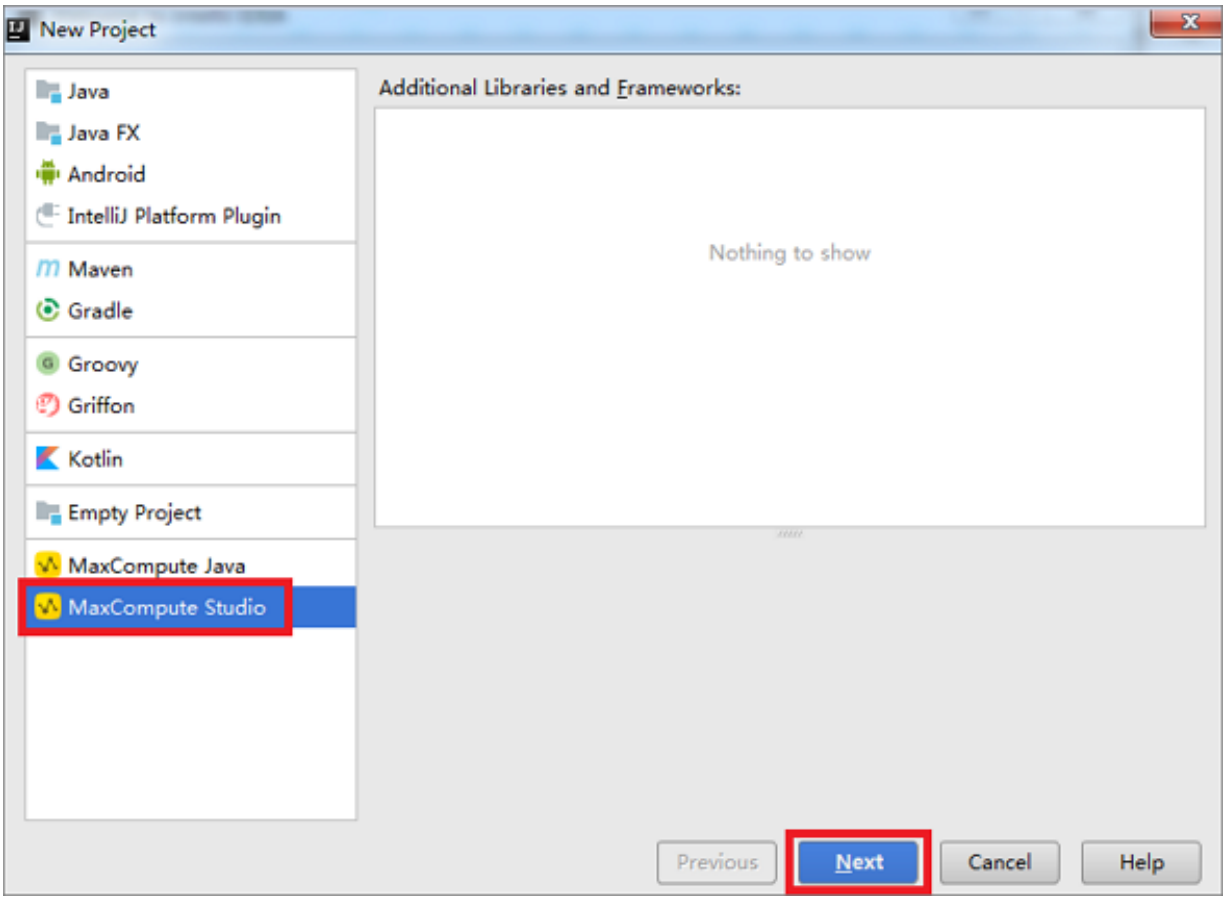
MaxCompute Studio的一大核心功能是浏览MaxCompute项目空间（Project）的资源，包括**Table**、**UDF**、**Resource**等，要想实现这一功能，首先需要创建项目连接。

前提条件

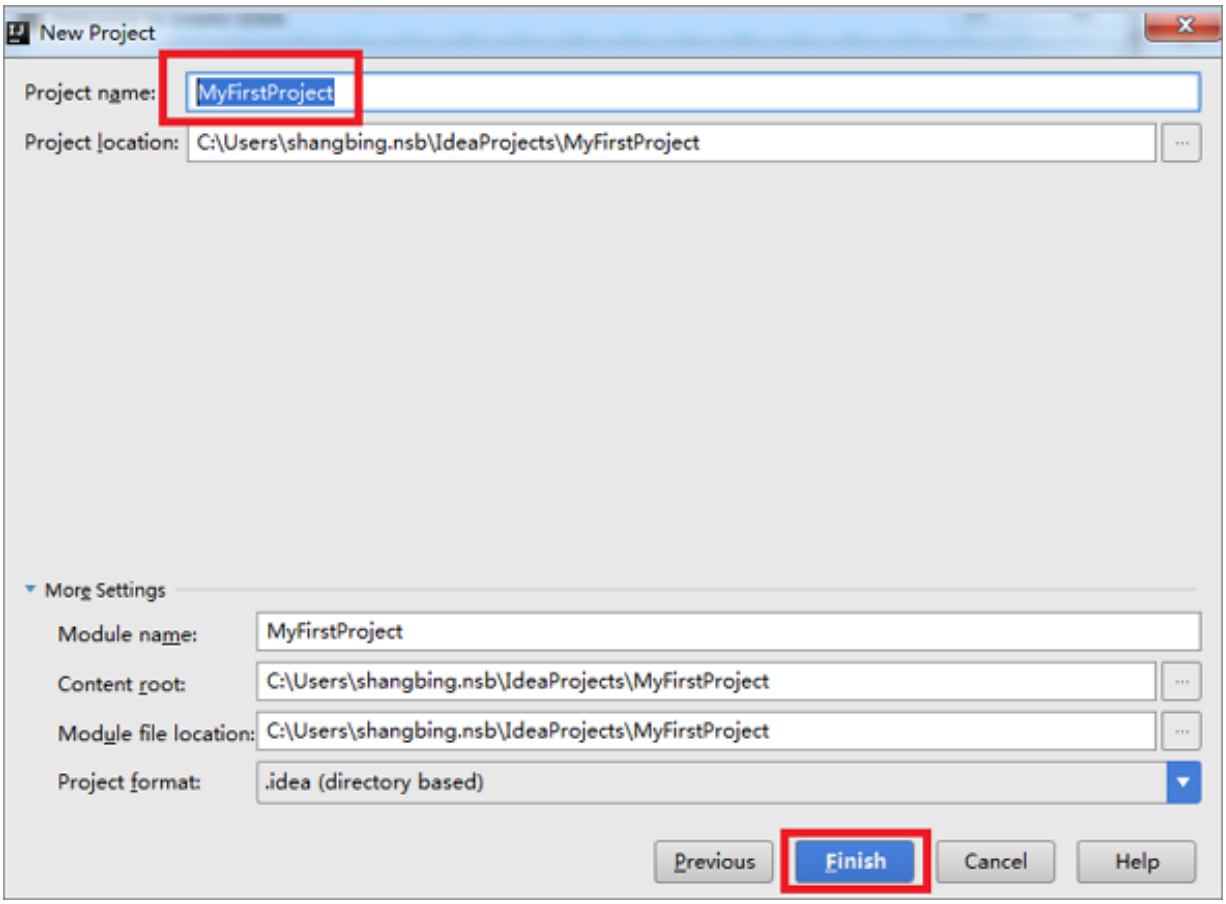
由于IntelliJ的tool window要显示必须先打开某个IntelliJ project，而配置MaxCompute Project需要进入IntelliJ界面Tool Windows中的MaxCompute Project Explorer，所以在创建MaxCompute Project链接前，先添加或者导入一个IntelliJ project。本文将以在Windows下新增project为例。

运行IntelliJ IDEA后，单击**create new project**，选择弹出页面中的**MaxCompute Studio**，单击**Next**。





填写Project name，单击**Finish**。

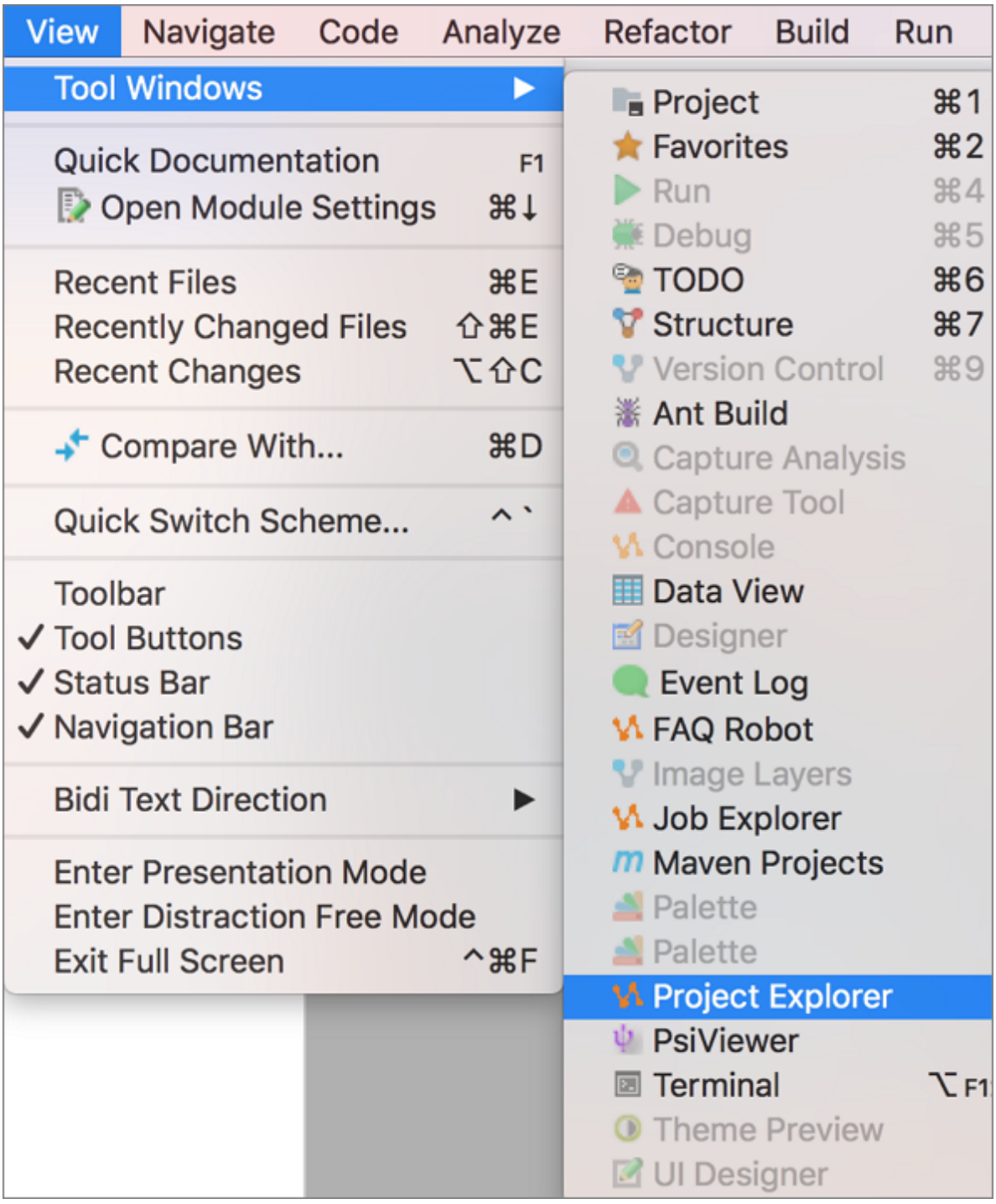


创建**MaxCompute**项目链接

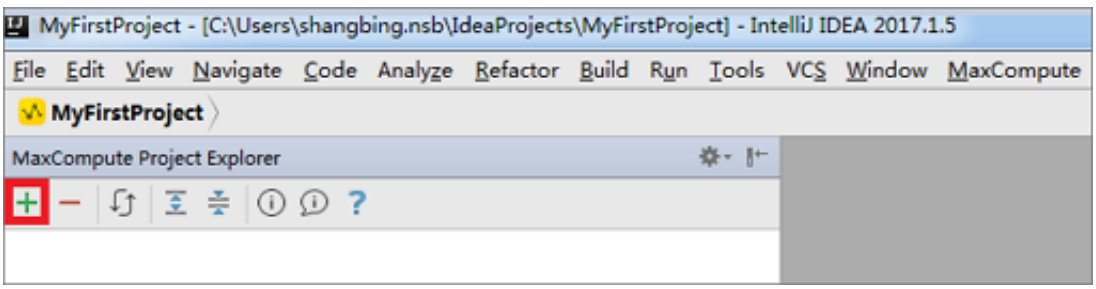
建议您根据自己的Region配置MaxCompute项目连接，否则会出现无法访问等错误。您可以参考访问域名和数据中心配置自己的endpoint和region。

操作步骤：

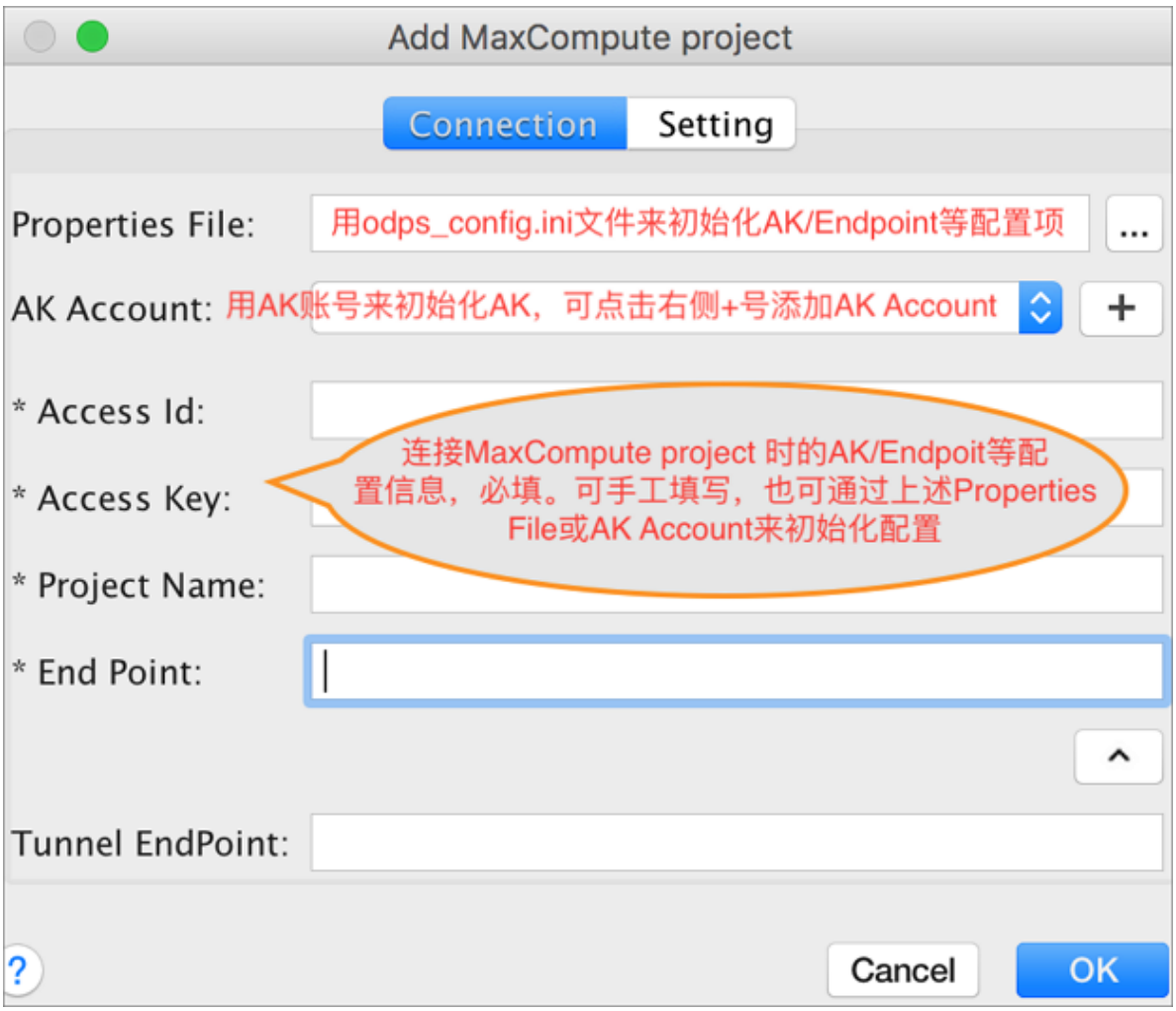
1. 单击菜单中的**view**选项，选择**Tool Windows**。
2. 单击弹出页面中的**Project Explorer**



3. 单击左上角的+，添加一个MaxCompute Project。



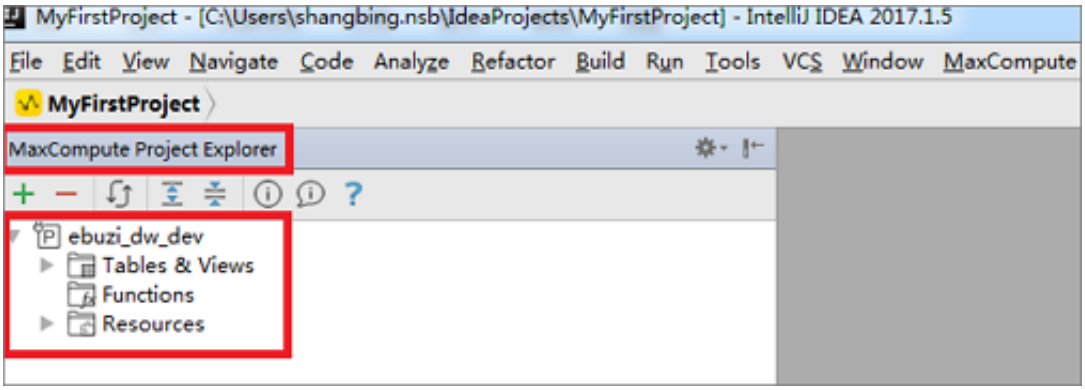
4. 在**Add MaxCompute Project**对话框中，填入相关配置选项。



说明：

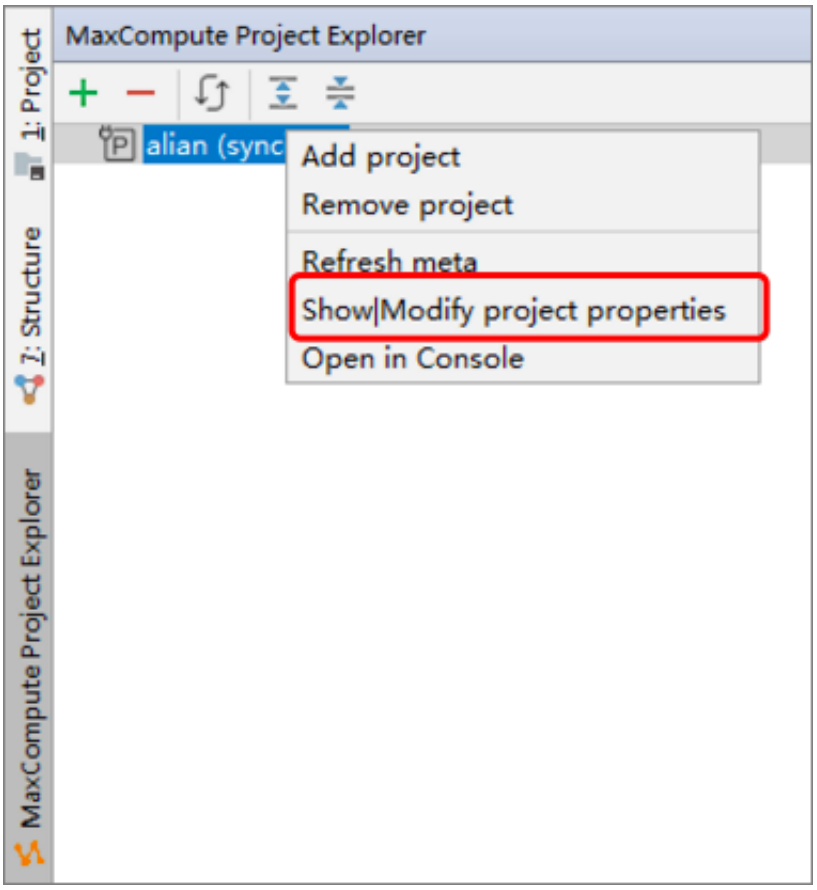
- 单击dialog左下角的?即可链接到在线文档。
- 若同步出现超时错误，可以在setting标签页酌情延长元数据同步到本地的超时时间。

5. 配置完成后，单击**OK**。左侧MaxCompute Project Explorer中会显示MaxCompute Project的信息，可以通过鼠标单击查看该project中的表、视图、函数以及资源等信息。



查看/修改MaxCompute项目链接

在Project Explorer中，对需要修改的MaxCompupte项目右键选择**Show|Modify project properties**。



在弹出框中可以查看或修改该MaxCompute project的Connection和Setting。

后续操作

现在，您已经学习了如何新建、管理项目空间连接，您可以继续学习下一个教程。在该教程中您将学习如何进行元数据查询、清理数据、上传下载数据等操作，来管理数据和资源。详情请参见 管理数据和资源。

2.3 开发Python程序

2.3.1 Python开发使用须知

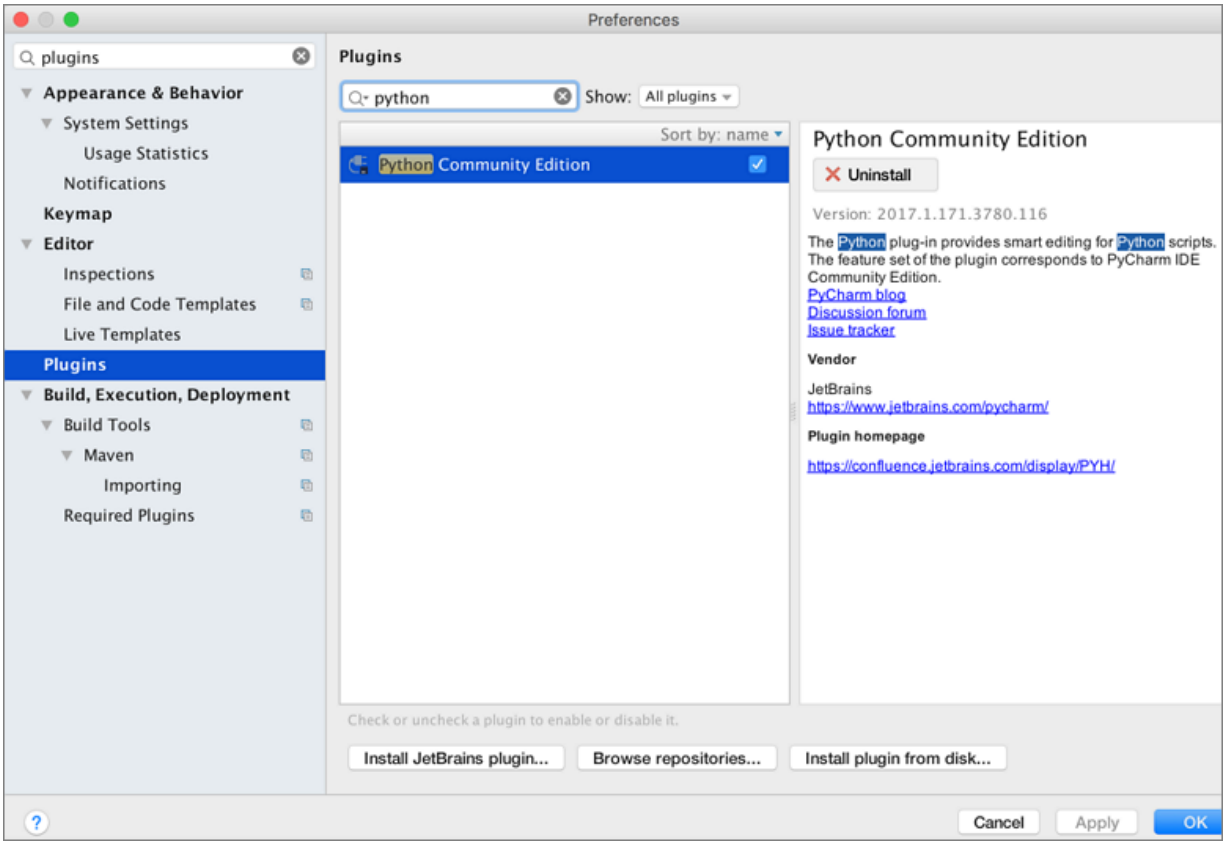
MaxCompute Studio支持您在Intellij IDEA中完成Python相关的开发，包括UDF和Pyodps脚本，但使用前必须安装**Python**、**Pyodps**和**IDEA**的**Python**插件。

安装Pyodps

Pyodps是MaxCompute的Pyodps SDK，详情请参见*Python SDK*。

安装Python插件

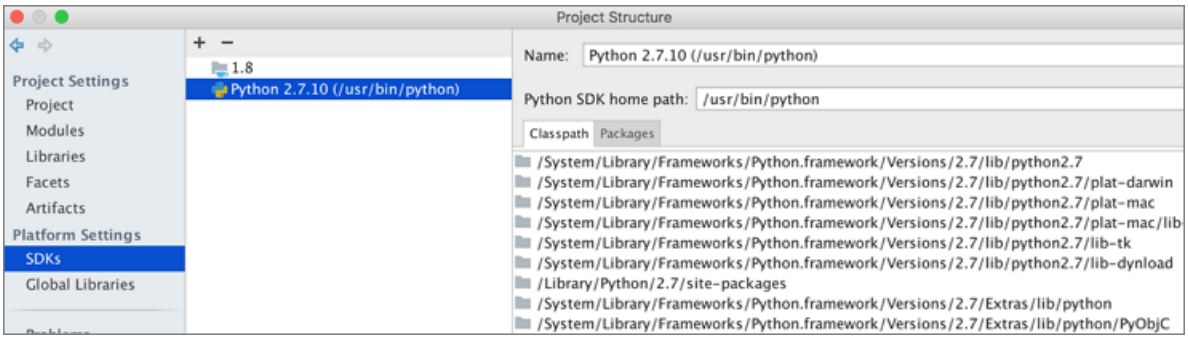
在Intellij IDEA的插件仓库中搜索**Python Community Edition**插件并安装。



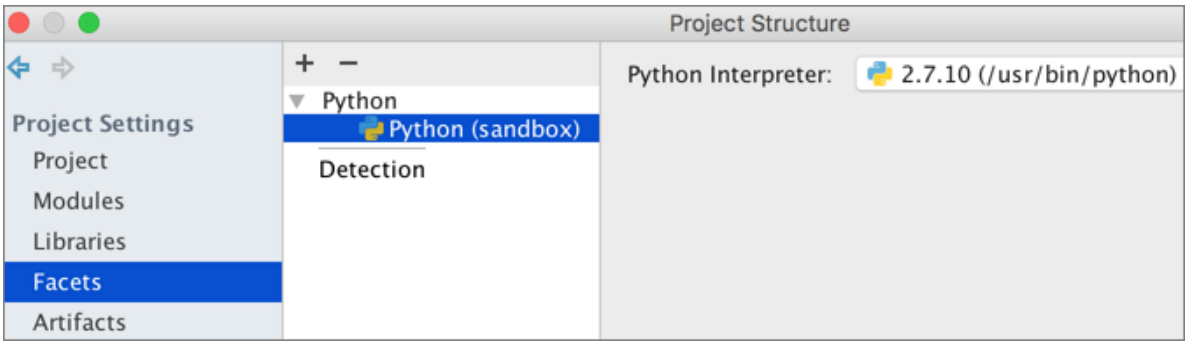
配置Python依赖

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

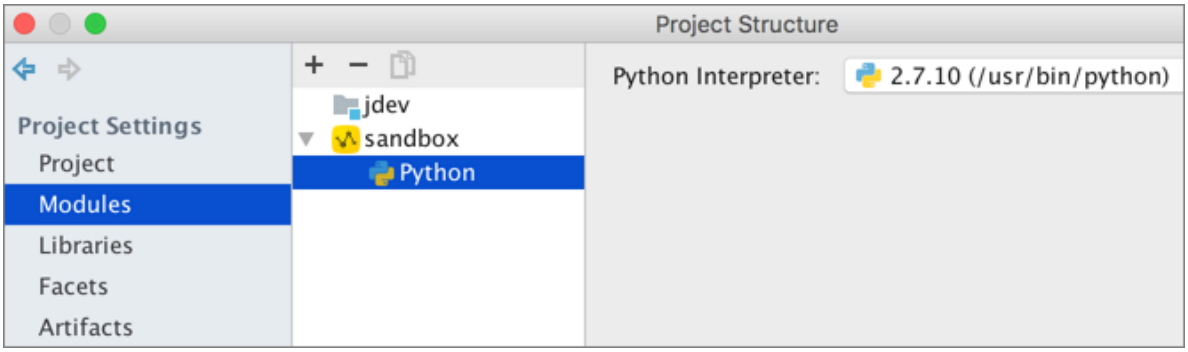
1. 导航至**File** > **Project structure**，添加Python SDK。



2. 导航至**File** > **Project structure**，添加Python Facets。



3. 导航至**File** > **Project structure**，配置Module依赖Python Facets。



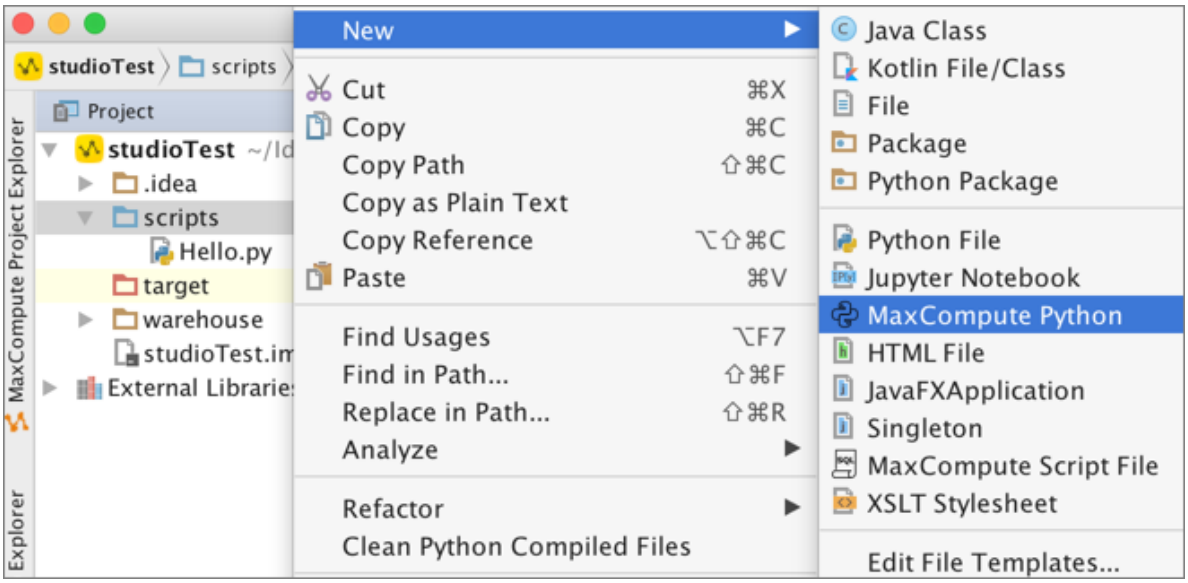
2.3.2 开发Python UDF

MaxCompute Studio支持Python UDF开发，但需要您根据*Python*开发使用须知做好准备工作。

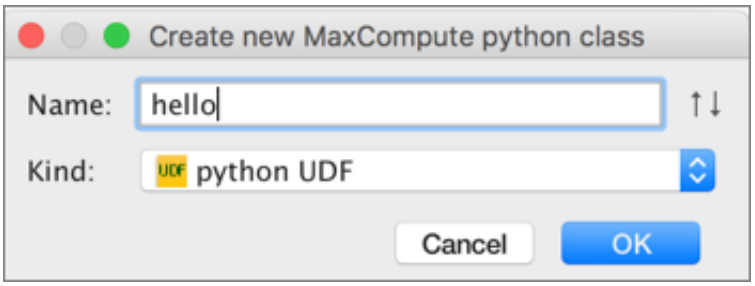
开发

1. 右键单击**New** > **MaxCompute Python**。

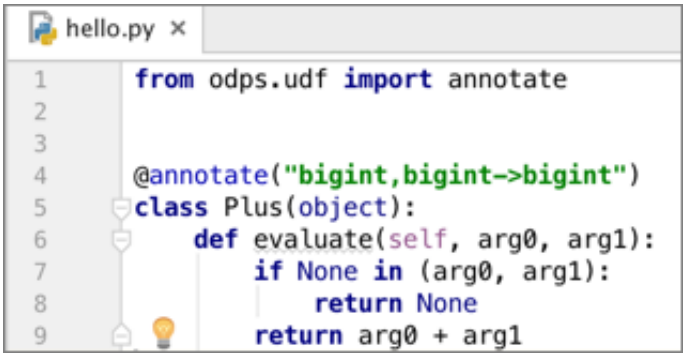
> 说明：
> 如果没有MaxCompute Python选项说明没有Python插件，请确认是否安装成功。



2. 输入类名，如hello，选择类型，此处选择**Python UDF**。填写完成后单击**OK**。



3. 模板已自动填充框架代码，您只需要编写UDF的入参出参，以及函数逻辑。



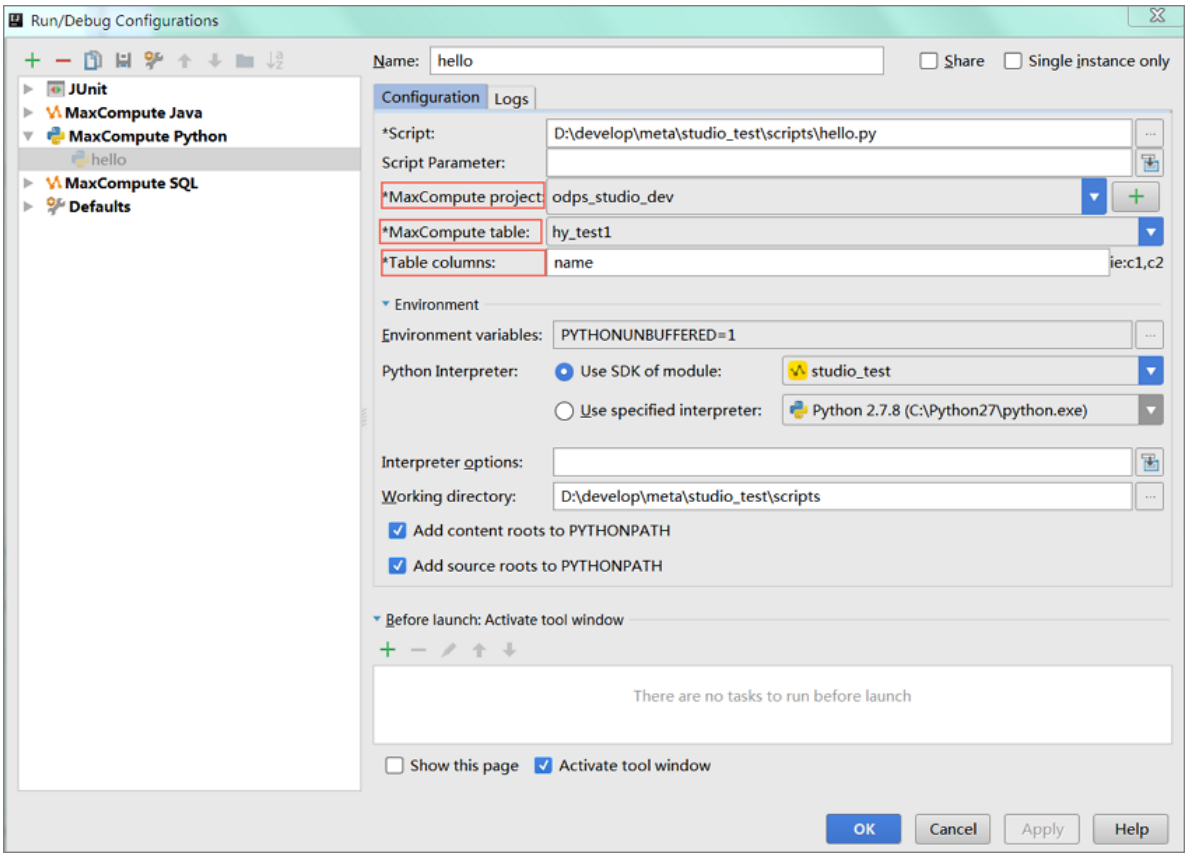
测试

UDF开发完成后，需要测试自己的代码，看是否符合预期。我们支持下载表的部分sample数据到本地运行，进行DeBug，操作如下：

1. 右键单击Editor中的UDF类，单击运行，弹出Run/Debug Configurations对话框。

> **说明：**

UDF|UDAF|UDTF一般作用于Select子句中表的某些列，此处需配置MaxCompute project、table和columns，元数据来源于**project explorer**窗口和**warehouse**下的**Mock**项目。

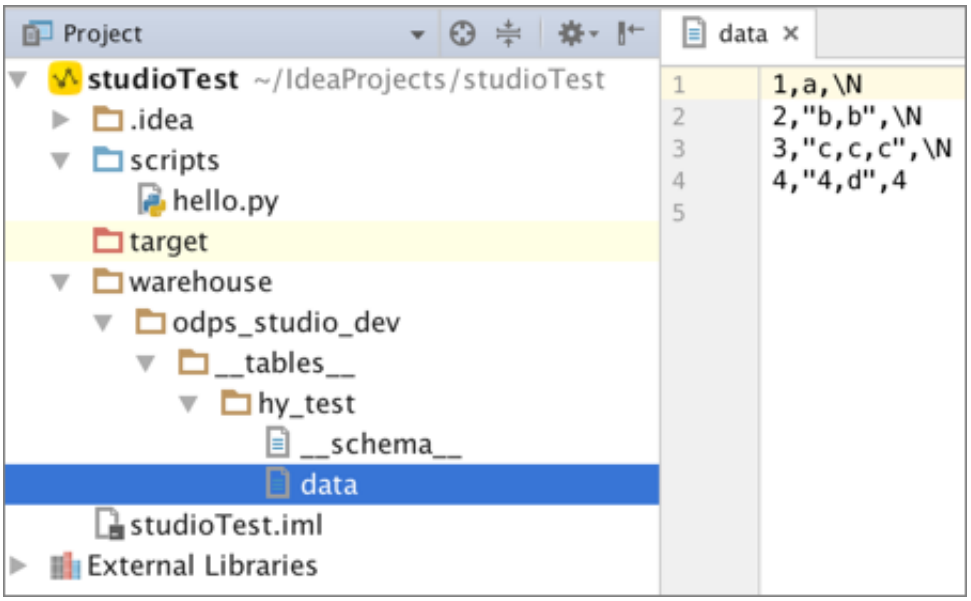


2. 单击**OK**后，通过Tunnel自动下载您指定表的sample数据到本地warehouse目录。

   说明：

   - 如果已经下载过，则不会再次重复下载，否则利用Tunnel服务下载数据。
   - 默认下载100条，如果需要更多数据测试，请自行使用console的Tunnel命令或者Studio的表下载功能。

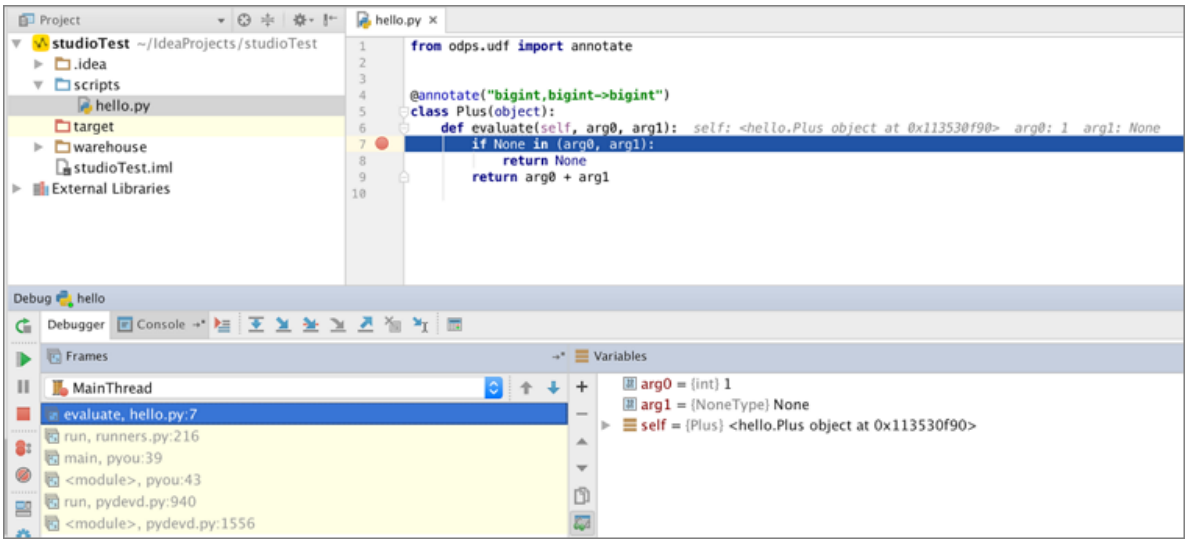
3. 下载完成后，您可以在warehouse目录看到下载的sample数据。您也可以使用Mock data（即warehouse中的数据自己mock，详情请参见开发和调试*UDF*中的本地运行的`warehouse`目录模块。



4. 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。

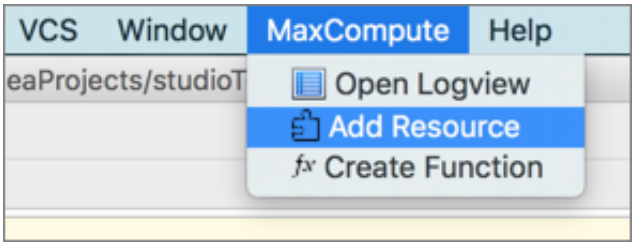
> 说明：
>
> 本地运行是通过Pyodps的pyou脚本实现的，命令如`pyou hello.Plus<data`。安装完pyodps后可以使用相应的命令检查该脚本是否存在。
>
> - 如果您是windows系统，请运行`${python}/../Scripts/pyou`。
> - 如果您是mac系统，请运行`${python}/../pyou`。

您可以在控制台看到结果打印，也可以在UDF上打断点调试。



注册发布

Python UDF测试通过后，即可注册发布到生产上进行使用。**Add Resource**后，**Create Function**即可。详情请参见打包/上传/注册。

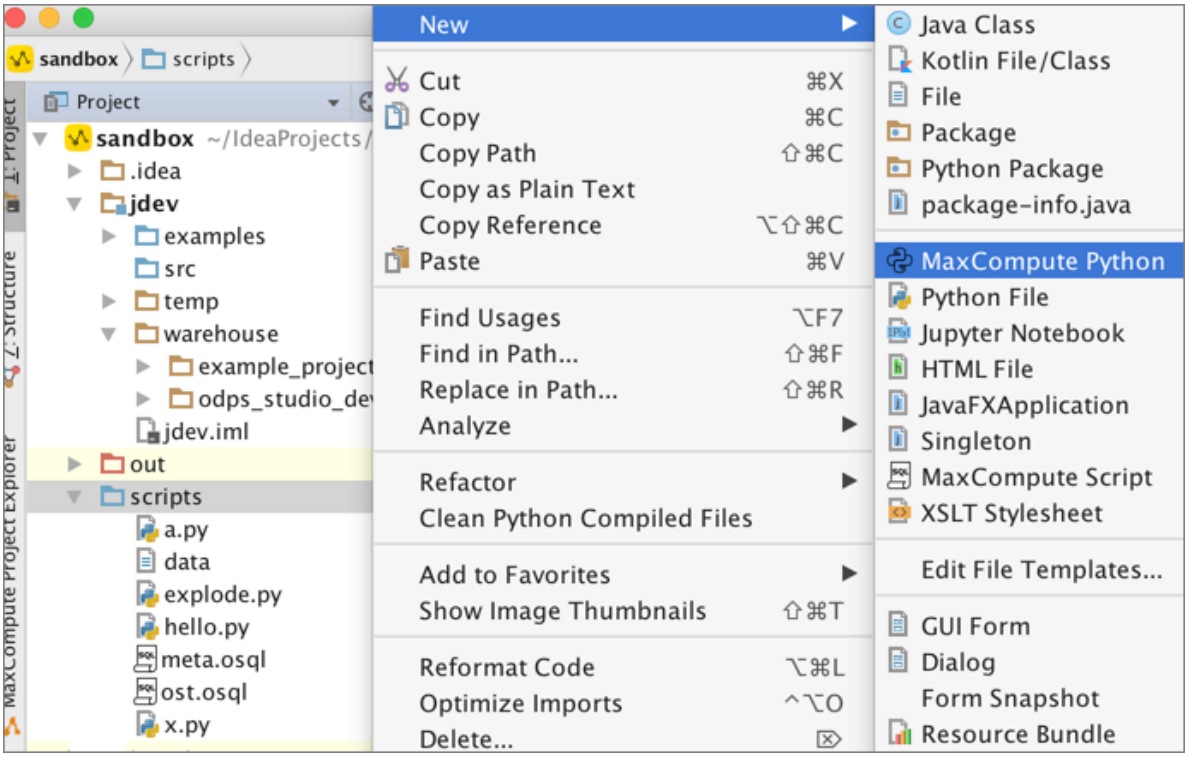


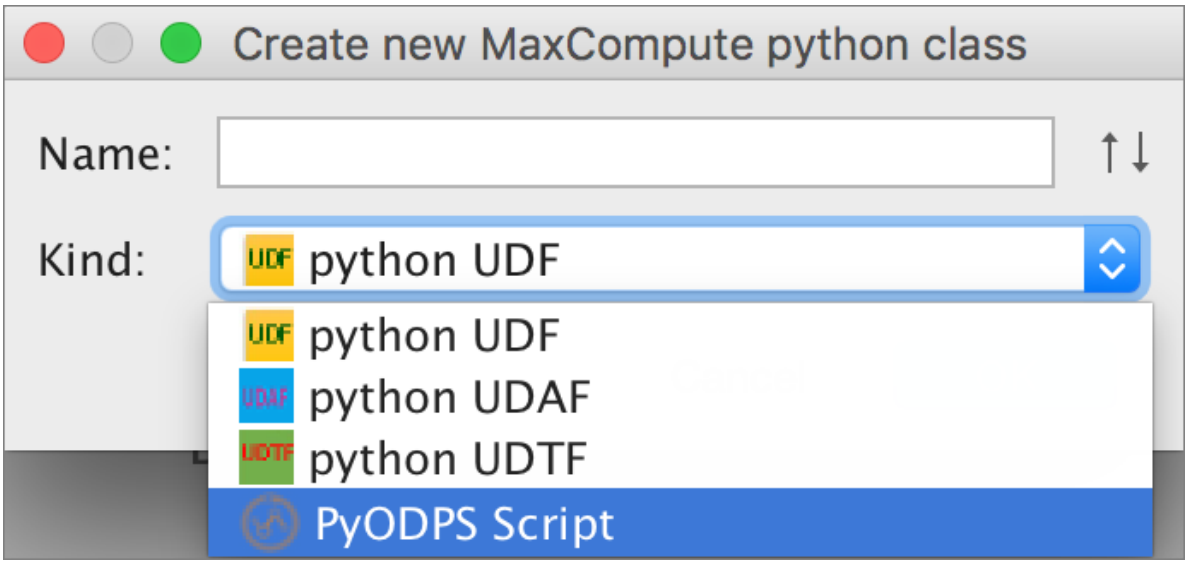
2.3.3 开发PyODPS脚本

开发PyODPS脚本

PyODPS是MaxCompute Python版本的SDK，它提供了对MaxCompute对象的基本操作，并提供了DataFrame框架，您可以在MaxCompute上进行数据分析。

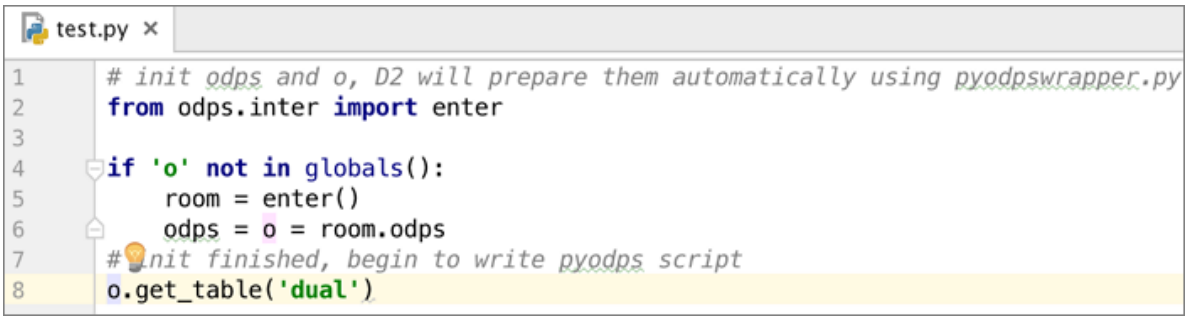
1. 使用PyODPS开发脚本，首先需要安装*PyODPS*及*Python*插件，安装完成后即可直接新建一个MaxCompute PyODPS脚本。





2. 新建成功后，模板会通过PyODPS *room对象*，自动初始化odps和o这两个对象。

> 说明：
> 在公共云Dataworks上运行时会自动在后台创建，所以为了IDEA编译通过，需要显式创建。



2.4 Studio视频介绍

- *Studio 安装介绍*
- 通过*Studio*管理数据
- 通过*Studio*编辑*SQL*
- *Studio SQL Scripting*
- 通过*Studio*开发*UDF*
- 通过*Studio remote debug*
- 通过*Studio*查看所有*job*

3 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。
- 新版客户端：点击此处 即可下载新版客户端。
- Eclipse 开发插件：点击此处 即可下载 Eclipse 开发插件。
- InteliJ 开发插件：idea工具点击此处 ，Studio插件点击此处 即可下载 InteliJ 开发插件。
- JDBC：MaxCompute提供开源JDBC，您可以在GitHub下载*JDBC*，也可以在云栖社区查看发布信息或提问。
- PHP SDK：您可以在GitHub下载。