MaxCompute使用手册

目录

[1. MaxCompute概述 2](#_Toc498936382)

[2. LogStore投递日志到MaxCompute 2](#_Toc498936383)

[2.1. 开启投递 2](#_Toc498936384)

[2.2. 配置投递规则 3](#_Toc498936385)

[3. MaxCompute使用 5](#_Toc498936386)

[3.1. MaxCompute网页使用说明 5](#_Toc498936387)

[3.1.1. DataWorks管理控制台 5](#_Toc498936388)

[3.1.2. 脚本开发 6](#_Toc498936389)

[3.1.3. 表查询 7](#_Toc498936390)

[3.2. MaxCompute插件使用说明 7](#_Toc498936391)

[3.2.1. MaxCompute Studio 7](#_Toc498936392)

[3.2.1.1. MaxCompute Studio介绍 7](#_Toc498936393)

[3.2.1.2. 安装步骤 7](#_Toc498936394)

[3.2.1.2.1. 环境要求 7](#_Toc498936395)

[3.2.1.2.2. 安装方式 8](#_Toc498936396)

[3.2.2. Eclipse插件 15](#_Toc498936397)

[3.3. SDK 15](#_Toc498936398)

[3.3.1. Java SDK 15](#_Toc498936399)

[3.3.2. Python SDK 15](#_Toc498936400)

[3.4. 相关下载 15](#_Toc498936401)

# MaxCompute概述

大数据计算服务(MaxCompute，原名ODPS)是一种快速、完全托管的TB/PB级数据仓库解决方案。MaxCompute向用户提供了完善的数据导入方案以及多种经典的分布式计算模型，能够更快速的解决用户海量数据计算问题，有效降低企业成本，并保障数据安全。

产品优势

*大规模计算存储*

MaxCompute 适用于 100GB 以上规模的存储及计算需求，最大可达 EB 级别。

*多种计算模型*

MaxCompute 支持 SQL、MapReduce、Graph 等计算类型及 MPI 迭代类算法。

*强数据安全*

MaxCompute 已稳定支撑阿里全部离线分析业务7年以上，提供多层沙箱防护及监控。

*低成本*

与企业自建私有云相比，MaxCompute 的计算存储更高效，可以降低 20%-30% 的采购成本。

详细帮助文档请查看：

<https://help.aliyun.com/document_detail/27800.html?spm=5176.doc50892.6.539.znN2hD>

# LogStore投递日志到MaxCompute

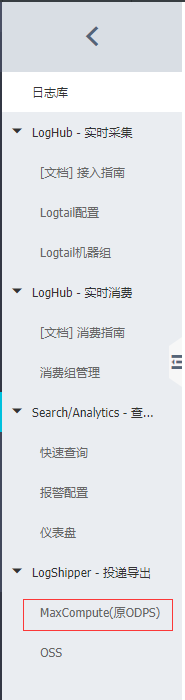
## 开启投递

通过二种方式可以进入开启日志投递界面：

1、在logstore界面点击需要开启投递的logstore的MaxCompute进去开启投递界面。



2、在logstore主界面的左侧点击[LogShipper-投递导出选]项下的[MaxCompute(原ODPS)]进入开启投递界面。



进入开启投递界面，点击[开启投递]。



点击[开启投递]之后进入投递配置界面。

## 配置投递规则

在**LogHub —— 数据投递**页面配置**字段关联**等相关内容。



选项含义：

| 参数 | 语义 |
| --- | --- |
| 投递名称 | 自定义一个投递的名称，方便后续管理。 |
| 项目名 | MaxCompute项目名称，该项默认为新创建的Project，如果已经是MaxCompute老客户，可以下拉选择已创建其他Project。 |
| 日志表名 | MaxCompute表名称，请输入自定义的新建的MaxCompute表名称或者选择已有的MaxCompute表。 |
| 字段关联 | 按序，左边填写与MaxCompute表数据列相映射的日志服务字段名称，右边填写或选择MaxCompute表的普通字段名称及字段类型。 |
| 分区字段 | 按序，左边填写与MaxCompute表分区列相映射的日志服务字段名称，右边填写或选择MaxCompute表的普通字段名称及字段类型。 |
| 时间分区格式 | \_\_partition\_time\_\_输出的日期格式，参考[Java SimpleDateFormat](https://docs.oracle.com/javase/6/docs/api/java/text/SimpleDateFormat.html)。 |
| 导入时间间隔 | MaxCompute数据投递间隔，有二种选择1800和3600，默认1800，单位：秒。 |

详细帮助文档请查看：

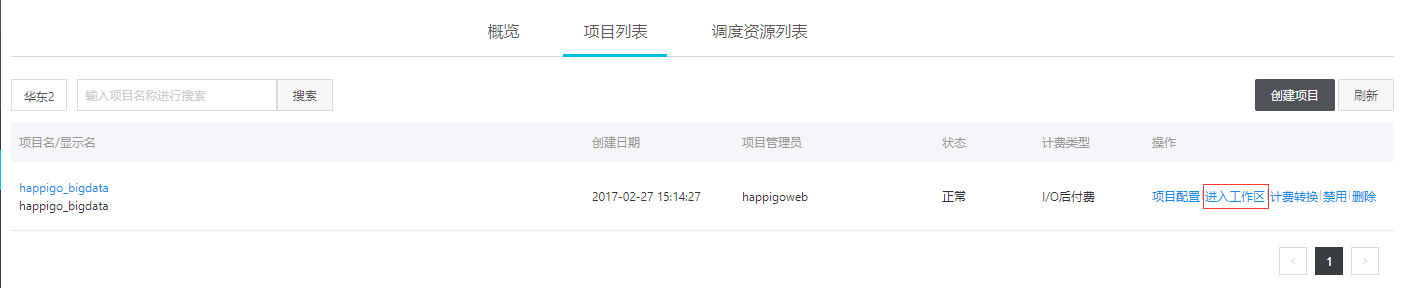
<https://help.aliyun.com/document_detail/29001.html?spm=5176.product28958.6.640.CTGZwu>

# MaxCompute使用

## MaxCompute网页使用说明

### DataWorks管理控制台

MaxCompute是在[DataWorks管理控制台](https://workbench.data.aliyun.com/console)中操作，项目名称为happigo\_bigdata,点击[进入工作区]。



进入DataWorks管理界面分为：任务开发、脚本开发、资源开发、函数开发、表查询。

### 脚本开发

在脚本开发界面可以使用SQL语句查询。

详细帮助文档请查看：

<https://help.aliyun.com/document_detail/27860.html?spm=5176.doc27810.6.611.IDTC0V>

MaxCompute SQL 中的所有保留字：

1. % & && ( ) \* +
2. - . / ; < <= <>
3. = > >= ? ADD AFTER ALL
4. ALTER ANALYZE AND ARCHIVE ARRAY AS ASC
5. BEFORE BETWEEN BIGINT BINARY BLOB BOOLEAN BOTH DECIMAL
6. BUCKET BUCKETS BY CASCADE CASE CAST CFILE
7. CHANGE CLUSTER CLUSTERED CLUSTERSTATUS COLLECTION COLUMN COLUMNS
8. COMMENT COMPUTE CONCATENATE CONTINUE CREATE CROSS CURRENT
9. CURSOR DATA DATABASE DATABASES DATE DATETIME DBPROPERTIES
10. DEFERRED DELETE DELIMITED DESC DESCRIBE DIRECTORY DISABLE
11. DISTINCT DISTRIBUTE DOUBLE DROP ELSE ENABLE END
12. ESCAPED EXCLUSIVE EXISTS EXPLAIN EXPORT EXTENDED EXTERNAL
13. FALSE FETCH FIELDS FILEFORMAT FIRST FLOAT FOLLOWING
14. FORMAT FORMATTED FROM FULL FUNCTION FUNCTIONS GRANT
15. GROUP HAVING HOLD\_DDLTIME IDXPROPERTIES IF IMPORT IN
16. INDEX INDEXES INPATH INPUTDRIVER INPUTFORMAT INSERT INT
17. INTERSECT INTO IS ITEMS JOIN KEYS LATERAL
18. LEFT LIFECYCLE LIKE LIMIT LINES LOAD LOCAL
19. LOCATION LOCK LOCKS LONG MAP MAPJOIN MATERIALIZED
20. MINUS MSCK NOT NO\_DROP NULL OF OFFLINE
21. ON OPTION OR ORDER OUT OUTER OUTPUTDRIVER
22. OUTPUTFORMAT OVER OVERWRITE PARTITION PARTITIONED PARTITIONPROPERTIES PARTITIONS
23. PERCENT PLUS PRECEDING PRESERVE PROCEDURE PURGE RANGE
24. RCFILE READ READONLY READS REBUILD RECORDREADER RECORDWRITER
25. REDUCE REGEXP RENAME REPAIR REPLACE RESTRICT REVOKE
26. RIGHT RLIKE ROW ROWS SCHEMA SCHEMAS SELECT
27. SEMI SEQUENCEFILE SERDE SERDEPROPERTIES SET SHARED SHOW
28. SHOW\_DATABASE SMALLINT SORT SORTED SSL STATISTICS STORED
29. STREAMTABLE STRING STRUCT TABLE TABLES TABLESAMPLE TBLPROPERTIES
30. TEMPORARY TERMINATED TEXTFILE THEN TIMESTAMP TINYINT TO
31. TOUCH TRANSFORM TRIGGER TRUE UNARCHIVE UNBOUNDED UNDO
32. UNION UNIONTYPE UNIQUEJOIN UNLOCK UNSIGNED UPDATE USE
33. USING UTC UTC\_TMESTAMP VIEW WHEN WHERE WHILE DIV

注意：

在对表、列或分区进行命名时，请不要使用保留字，否则会报错。

保留字不区分大小写。

### 表查询

在表查询界面，左侧显示该项目下已经存在的表的列表。选择某一张表之后，下方会出现列信息、分区信息、数据预览三个选项卡。

详细帮助文档请查看：

<https://help.aliyun.com/document_detail/56974.html?spm=5176.doc30284.6.629.8ZrAts>

## MaxCompute插件使用说明

### MaxCompute Studio

#### MaxCompute Studio介绍

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](https://www.jetbrains.com/idea/) 的开发插件，可以帮助您方便地进行数据开发。

#### 安装步骤

##### 环境要求

IntelliJ IDEA 支持在 Windows， Mac， Linux 操作系统上安装，硬件及系统环境要求请参见 <https://www.jetbrains.com/help/idea/2016.3/requirements-for-intellij-idea.html> 。基于 IntelliJ IDEA 平台的 MaxCompute Studio 也可以安装在这些操作系统的客户端上。

MaxCompute Studio 对用户环境有以下要求：

* Windows，Mac OS，或者 Linux 系统客户端。
* 安装 IntelliJ IDEA 14.1.4 以上版本 （支持 Ultimate 版本、PyCharm 版本和免费的 [Community 版本](https://www.jetbrains.com/idea/download/)）。
* 已安装 JRE 1.8 （最新的 IntelliJ IDEA 版本捆绑了 JRE 1.8）。
* 已安装 JDK 1.8 （可选：如果需要开发和调试 Java UDF，则需安装 JDK）。

##### 安装方式

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。

详细帮助文档请查看：

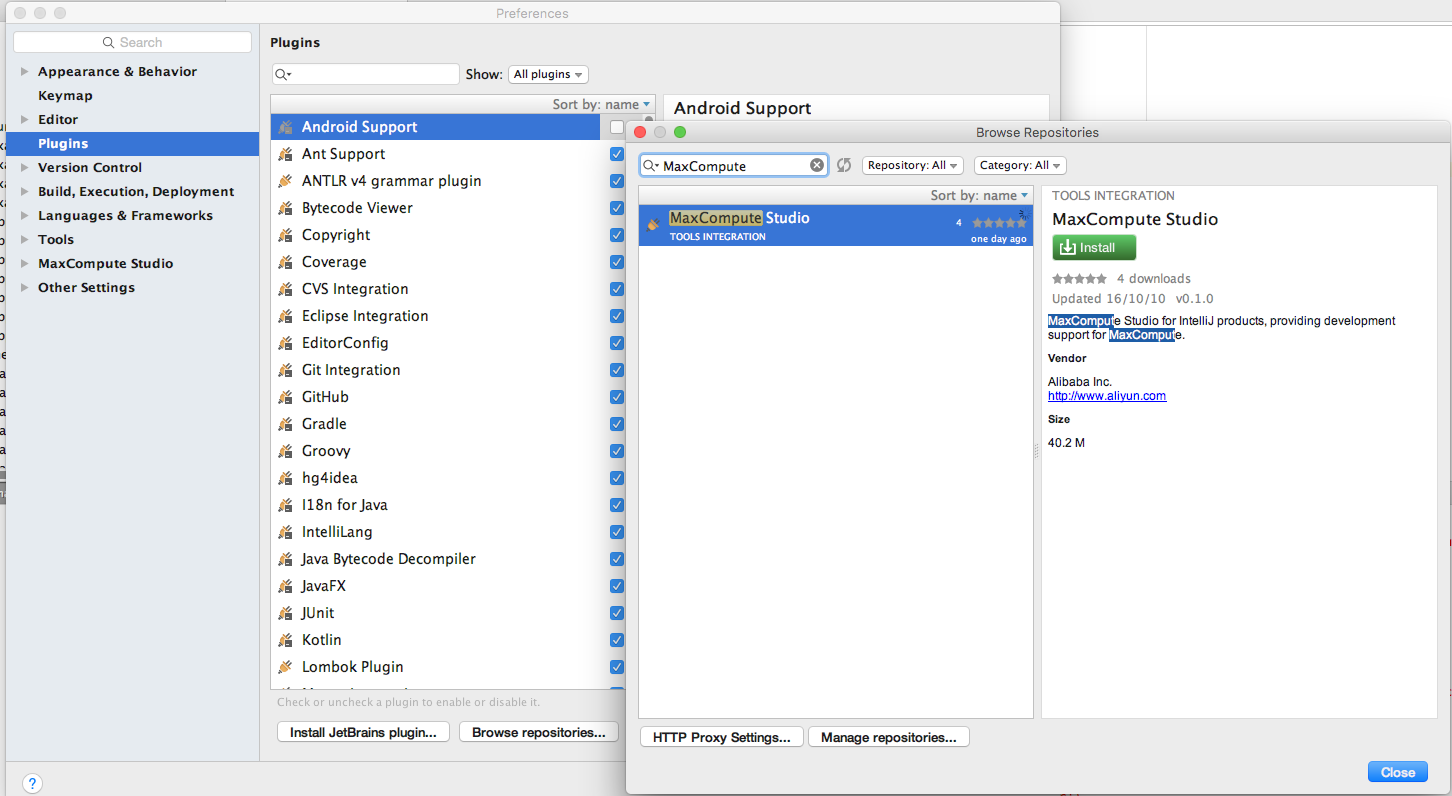
<https://help.aliyun.com/document_detail/50892.html?spm=5176.doc50891.6.712.1ZpZbi>

###### 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 官方插件库安装。

操作步骤

1. 在 IntelliJ IDEA 中打开插件配置页面 （Windows/Linux 用户导航至 **File > Settings > Plugins**， Mac 用户导航至 **IntelliJ IDEA > Preferences > Plugins** ）。
2. 单击 **Browse repositories…** 按钮，然后搜索 MaxCompute Studio。
3. 找到 MaxCompute Studio 插件页面，单击绿色 **Install** 按钮进行安装。
4. 确认安装后，重新启动 IntelliJ IDEA，完成安装。



###### 本地安装

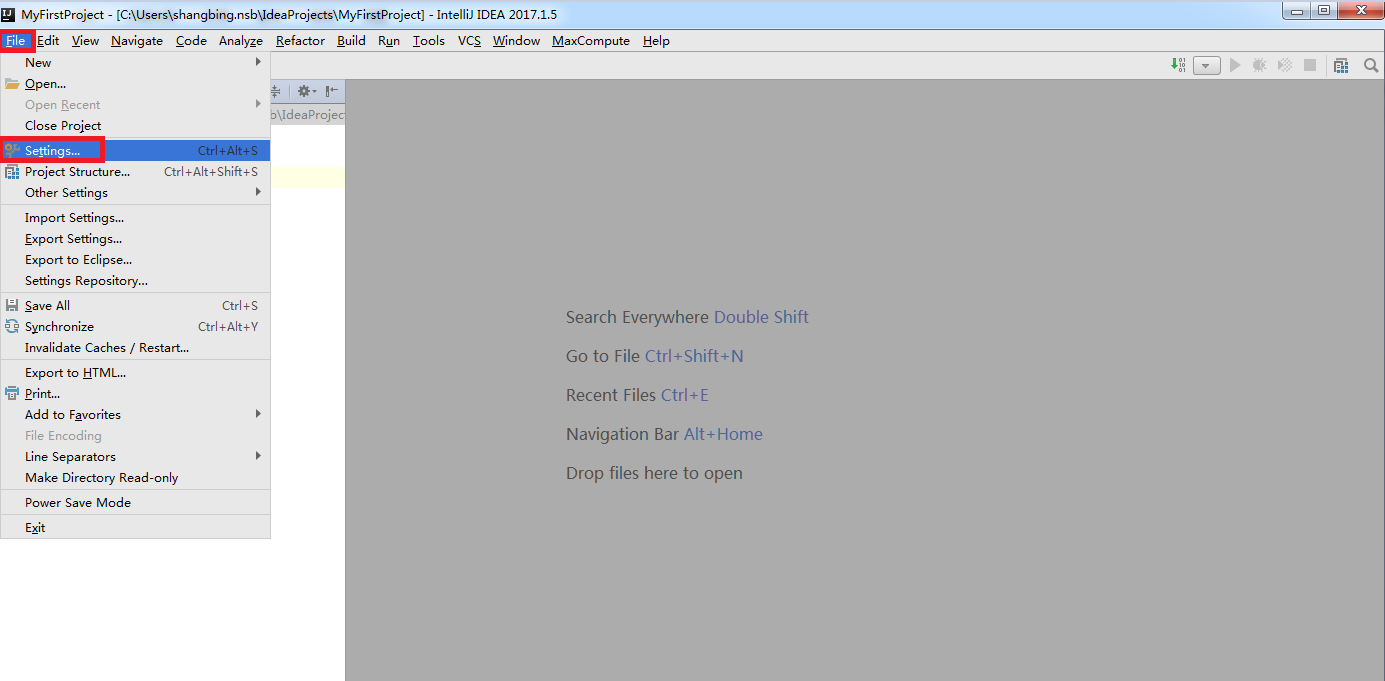
MaxCompute Studio 也可以在本地环境中进行安装。

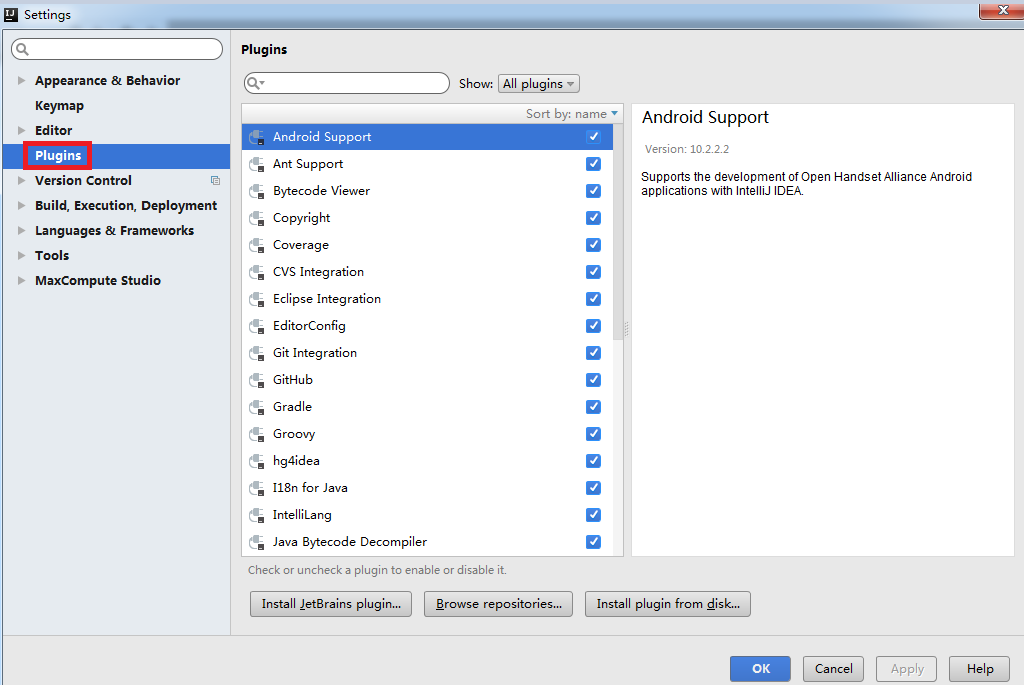
###### 操作步骤

1. 进入 [MaxCompute Studio 插件页面](https://plugins.jetbrains.com/plugin/9193?spm=5176.doc44555.2.1.4hXBG1) 下载插件包。
2. 运行 IntelliJ IDEA。
   * 如果是第一次，会出现欢迎界面，单击欢迎界面中的 **configure**（配置），选中弹出菜单中的 **Plugins**（插件），如下图所示：

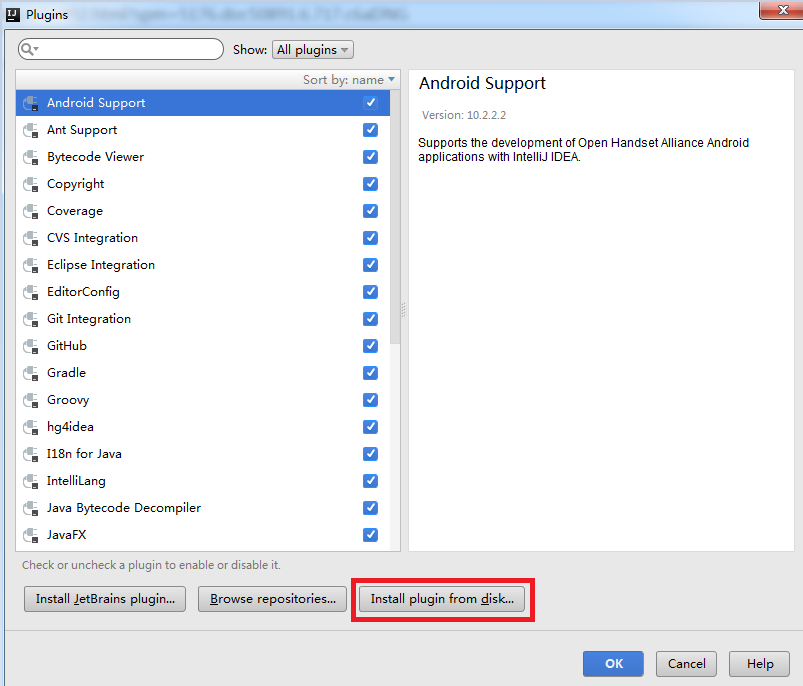


* + 如果不是第一次运行，可以依次单击菜单 **File > Settings > Plugins** 进入相同的界面，如下图所示：

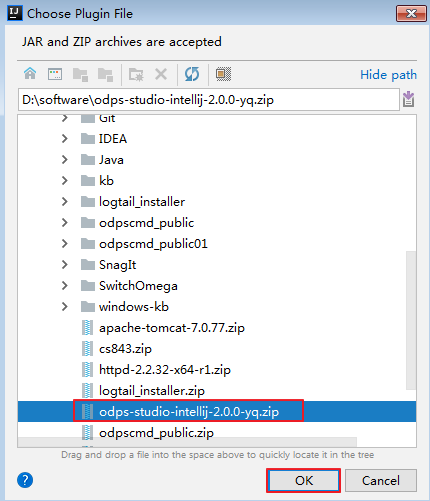




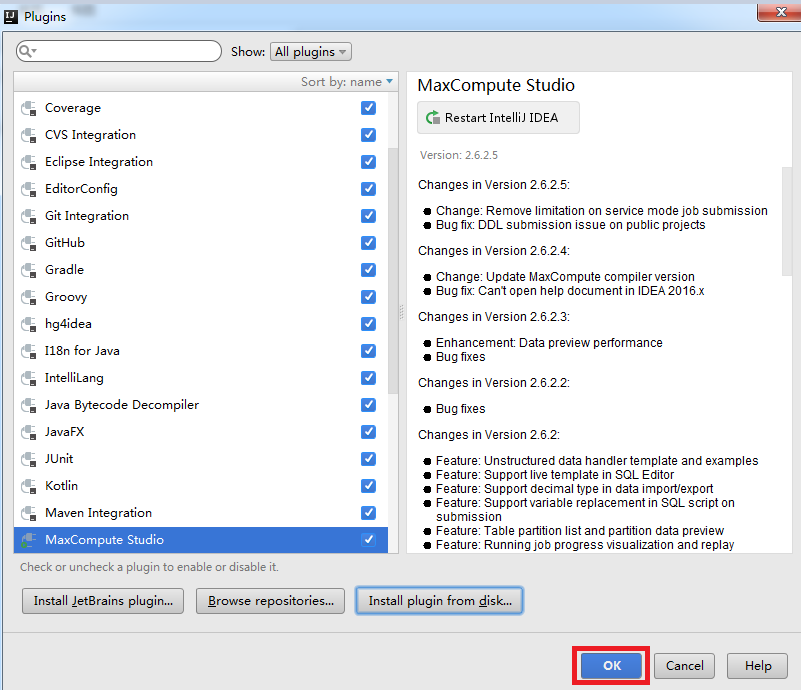
1. 在插件页面，单击 **Install plugin from disk…**（从本地磁盘安装插件），如下图所示：



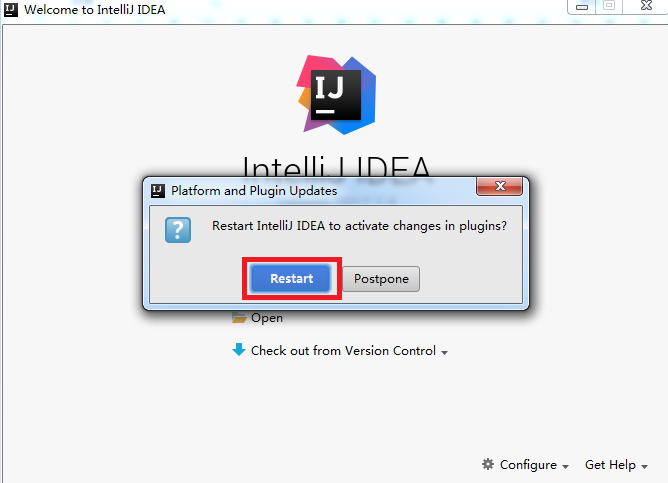
1. 在弹出窗口中，通过单击目录名称前的灰色图标进行导航，找到插件文件并选中，单击 **ok**。



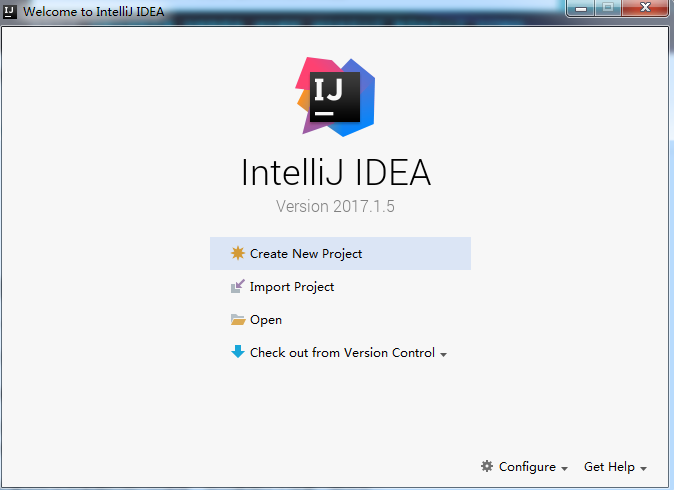
1. 回到插件首页后，单击 **OK**，开始安装本地插件。



1. 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。



1. 重新启动后，界面如下所示：



### Eclipse插件

详细帮助文档请查看：

<https://help.aliyun.com/document_detail/27981.html?spm=5176.doc27981.6.739.ErMMGG>

## SDK

### Java SDK

详细帮助文档请查看：

<https://help.aliyun.com/document_detail/34614.html?spm=5176.doc27861.6.686.OKh036>

### Python SDK

详细帮助文档请查看：

<https://help.aliyun.com/document_detail/34615.html?spm=5176.doc34614.6.687.iE4mnP>

## 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](http://search.maven.org/?spm=5176.doc27991.2.4.Cj7xxk) 中搜索 odps-sdk，获取不同版本的 Java SDK。
* 新版客户端：[点击此处](http://repo.aliyun.com/download/odpscmd/latest/odpscmd_public.zip?spm=5176.doc27991.2.5.Cj7xxk&file=odpscmd_public.zip) 即可下载新版客户端。
* Eclipse 开发插件：[点击此处](https://docs-aliyun.cn-hangzhou.oss.aliyun-inc.com/cn/odps/0.0.90/assets/download/odps-eclipse-plugin-bundle-0.16.0.zip?spm=5176.doc27991.2.6.Cj7xxk&file=odps-eclipse-plugin-bundle-0.16.0.zip) 即可下载 Eclipse 开发插件。
* InteliJ 开发插件：[点击此处](http://docs-aliyun.cn-hangzhou.oss.aliyun-inc.com/assets/attach/27991/cn_zh/1476766995338/odps-studio-intellij-2.0.0-yq.zip?spm=5176.doc27991.2.7.Cj7xxk&file=odps-studio-intellij-2.0.0-yq.zip) 即可下载 InteliJ 开发插件。