

ALIBABA CLOUD

# 阿里云

## E-MapReduce

### 快速入门

文档版本：20210226

 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。
<b>粗体</b>	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.入门概述	05
2.准备工作	06
3.创建集群	07
4.创建并运行作业	09

# 1.入门概述

本教程指引您通过阿里云账号登录E-MapReduce控制台，使用一键购买，快速创建一个Hadoop集群并执行作业。

本教程包含以下操作：

- [准备工作](#)
- [创建集群](#)
- [创建并运行作业](#)

常用文件的安装路径，请参见[常用文件路径](#)。

## 2.准备工作

本文介绍创建E-MapReduce集群前所需的准备工作。

### 注册阿里云账号

如果您还没有阿里云的云账号，请注册阿里云账号，并完成实名认证。

### E-MapReduce服务账号授权

开通E-MapReduce服务后，您需要给E-MapReduce的服务账号授予名称为AliyunEMRDefaultRole和AliyunEmrEcsDefaultRole的系统默认角色，具体请参见[角色授权](#)。

### RAM用户授权

如果您需要使用RAM用户登录E-MapReduce控制台，并使用控制台上的功能，则需要您使用主账号登录访问控制RAM控制台，授予RAM用户相应的权限，具体请参见[RAM用户授权](#)。


### （可选）开通阿里云OSS服务

建议开通阿里云OSS服务，您可以将E-MapReduce作业日志和运行日志等数据保存在OSS存储空间中，方便使用并降低存储成本。开通阿里云OSS服务请参见[开通OSS服务](#)。在OSS上创建的存储空间需要与E-MapReduce集群地域相同，创建存储空间详情请参见[创建存储空间](#)。

### （可选）创建AccessKey

由于E-MapReduce调用访问的需要，您至少需要创建一个AccessKey，创建步骤如下：

1. 以主账号登录[阿里云管理控制台](#)。
2. 将鼠标置于页面右上方的账号图标，然后选择[AccessKey管理](#)。

 **注意** 如果出现如下提示框，请单击继续使用AccessKey。



3. 在安全信息管理页面，单击创建AccessKey。
4. 输入短信校验码，单击确定。

AccessKey创建成功。

### （可选）开通高配机型

如果您需要在按量的集群中使用8核及8核以上的机型时，则[提交工单](#)处理。

## 3. 创建集群

本文介绍如何在E-MapReduce控制台，通过一键购买方式，快速创建一个按量付费的Hadoop集群。

### 前提条件

注册阿里云账号，并完成实名认证。具体操作请参见[阿里云账号注册流程](#)。

#### 说明

根据阿里云ECS的规则，您在购买按量付费实例时，需要保证阿里云账户中可用余额（含现金、代金券、优惠券等）不得少于100元人民币。充值方式请参见[前往充值](#)。

### 操作步骤

1. 登录[阿里云E-MapReduce控制台](#)。
2. 在顶部菜单栏处，根据实际情况选择地域（Region）和资源组。
3. 单击上方的[集群管理](#)页签。
4. 在[集群管理](#)页面，单击右上角的[一键购买](#)。
5. 在创建集群页面，完成如下配置。本教程使用的配置如下，未提及的配置保持默认，更多配置信息详情请参见[创建集群](#)。

配置页面	配置项	示例	说明
基础信息	集群类型	Hadoop	默认集群类型。
	产品版本	EMR-3.30.0	默认最新的软件版本。
	集群名称	test	集群的名字，长度限制为1~64个字符，仅可使用中文、字母、数字、中划线（-）和下划线（_）。
	密码	自定义密码	请记录该配置，连接集群时您需要输入该密码。
	可用区	华东1（杭州）可用区I	集群创建后，无法直接更改地域和可用区，请谨慎选择。
	网络类型	专有网络	默认为专有网络，不可更改。
	VPC	VPC_Hangzhou(192.168.xx.xx/16)(ID: vpc-bp1f4epmkvncimpgs**)	选择对应区域下的VPC。如果没有，单击 <a href="#">创建VPC / 子网（交换机）</a> 前往新建。

网络配置配置页面	配置项	示例	说明
	交换机	vsw_i(192.168.xx.xx/24)(ID: vsw-bp1e2f5fhapl0g6p****)	选择在对应用VPC下可用区的交换机，如果在这个可用区没有可用的交换机，则需要新建一个。
	安全组名称	sg-bp1ddw7sm2risw****(ID:sg-bp1ddw7sm2riswb1****)	<div style="border: 1px solid #ccc; padding: 5px; background-color: #e6f2ff;"><p> <b>注意</b> 禁止使用ECS上创建的企业安全组。</p></div> 您可直接输入安全组名称来新建一个安全组。如果已有在使用的安全组，则可以直接选择使用。

6. 选中E-MapReduce服务条款复选框。

7. 单击创建。创建集群后可以通过刷新页面来查看进度，当集群状态显示为空闲时，集群创建成功。

### 问题反馈

如果您在使用阿里云E-MapReduce过程中有任何疑问，欢迎您扫描下面的二维码加入钉钉群进行反馈。





## 4. 创建并运行作业

本文为您介绍如何创建并运行Spark作业。

### 创建项目

1. 登录 [阿里云E-MapReduce控制台](#)。
2. 单击上方的数据开发页签。
3. 在数据开发页面，单击新建项目。
4. 在新建项目对话框中，输入项目名称为test和项目描述为test。

新建项目

[注意] 请使用主账号创建项目

\* 项目名称: test

\* 项目描述: test

\* 资源组选择: 默认资源组 刷新

如果创建新的资源组，您可以点击 [去创建](#)

创建 取消

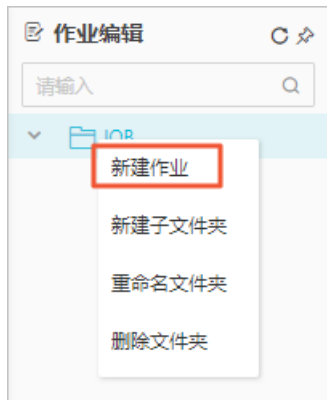
5. 单击创建。

### 新建作业

1. 在项目列表页面，单击新建项目所在行的作业编辑。

项目名称	描述	创建时间	管理员	操作
test	test	2020年7月29日 15:11:48	107	<a href="#">作业编辑</a> <a href="#">工作流设计</a> <a href="#">数据工作台</a> <a href="#">运行记录</a> <a href="#">删除</a>

2. 在作业编辑区域，在待操作的文件夹上单击右键，选择新建作业。



**说明** 您还可以通过在文件夹上单击右键，进行创建子文件夹、重命名文件夹和删除文件夹操作。

3. 在**新建作业**对话框中，配置各项参数。
  - i. 设置作业名称为Spark\_test。
  - ii. 设置作业描述为test。
  - iii. 从作业类型列表中，选择**Spark**。
  - iv. 单击**确定**。

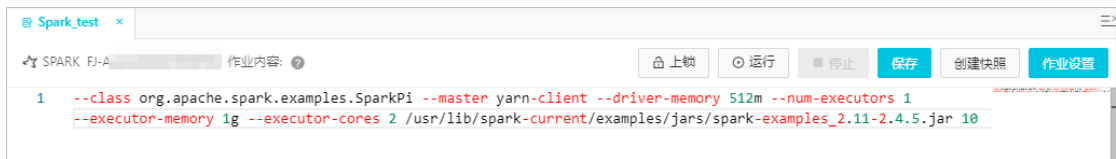
### 配置并运行作业

1. 配置作业内容。
  - i. 在**集群基础信息**页面，查看Spark的版本。



- ii. 输入如下作业内容。

```
--class org.apache.spark.examples.SparkPi --master yarn-client --driver-memory 512m --num-executors 1 --executor-memory 1g --executor-cores 2 /usr/lib/spark-current/examples/jars/spark-examples_2.11-2.4.5.jar 10
```



/usr/lib/spark-current/examples/jars/spark-examples\_2.11-2.4.5.jar 中的 2.4.5 是您集群中的Spark版本。

2. 单击右上角的**保存**。
3. 单击右上角的**运行**。
4. 在**运行作业**对话框中，单击**确定**。

### 查看日志

您可以单击作业下方的运行记录页签，查看作业的运行情况。

运行实例ID	开始时间	结束时间	状态	操作
FJI-9[...]	2020年7月29日 15:57:54	-	○ RUNNING	<a href="#">详情</a>   <a href="#">停止作业实例</a>
FJI-8[...]	2020年7月29日 15:45:12	2020年7月29日 15:45:50	● OK	<a href="#">详情</a>   <a href="#">停止作业实例</a>

单击详情，您可以查看作业的实例信息、提交日志和YARN容器日志。

## 问题反馈

如果您在使用阿里云E-MapReduce过程中有任何疑问，欢迎您扫描下面的二维码加入钉钉群进行反馈。

