

Spark分布式计算引擎使用指南

1 商品说明

Apache Spark 是开源的分布式计算引擎，核心优势在于内存计算（比 Hadoop 快数十倍）和统一处理框架（批处理、流计算、机器学习等）。它通过弹性数据集（RDD）实现高效容错，支持多语言 API，广泛应用于实时分析、大规模 ETL、AI 训练等场景，是当前大数据生态的核心组件之一。

本商品通过鲲鹏服务器+Euler2.0 部署。

2 商品购买

您可以在云商店搜索“Spark分布式计算引擎”。

其中，地域、规格、推荐配置使用默认，购买方式根据您的需求选择按需/按月/按年，短期使用推荐按需，长期使用推荐按月/按年，确认配置后点击“立即购买”。

2.1 商品支持自定义 ECS 购买，具体见章节 3.1.1

2.2 使用 RFS 模板直接部署



必填项填写后，点击 下一步

参数名称	值	类型	描述
ECS实例名称	<input type="text"/>	字符串	ECS实例的管理员名称。密码策略要求：密码要求长度范围为8到16位，密码至少必须包含大写字母、小写字母、数字和特殊字符 (@!#\$%^&*~+ ()_{}~)
系统盘大小	40	number	设置系统盘大小（至少40G，默认40G）。
数据盘大小	50	number	设置注册的数据盘大小，不需要数据盘，可设置值为0，可依据实际情况配置，默认为50。
版本	请选择	字符串	选择版本
vpc IPv4网段	192.168.0.0/16	字符串	取值范围：10.0.0.0/8 to 10.255.255.0/24, 172.16.0.0/12 to 172.31.255.0/24, or 192.168.0.0/16 to 192.168.255.0/24
子网 IPv4网段	192.168.10.0/24	字符串	必须是CIDR格式，且在VPC的CIDR块内，子网掩码不能大于24。
子网的网关	192.168.10.1	字符串	子网的网关，必须是子网段内的合法IP地址
计费类型(不包含应用防火墙)	请选择	字符串	prePaid-预付费，即包年包月；postPaid-后付费，即按量付费
订购周期类型(不包含应用防火墙)	month	字符串	当chargingMode为prePaid时生效且为必填项，取值范围：month、year、年
订购周期数(不包含应用防火墙)	1	字符串	当chargingMode为prePaid时生效且为必填项，取值范围：periodType=month（周期类型为月）时，取值为1、6；periodType=year（周期类型为年）时...

上一步 下一步

IAM角色选择: huaweicloud 请选择角色? 如何新建角色?

选择角色后，可明确定义角色在资源栈上可执行的所有操作（如创建、更新、删除资源栈等）。如果角色权限不足，有可能导致资源部署失败。前往IAM新建角色。

用户不配置IAM角色时，系统将采用当前登录用户的权限进行部署。权限委托有助于用户最小化对云服务的操作权限，可有效避免特权账号滥用导致数据丢失或安全事件，推荐用户配置。

回滚设置: 开启回滚表示在操作失败时，会自动回滚至系统所保存的上一个成功的资源栈状态。创建后可在资源栈详情页查看。

删除保护: 防止资源栈被意外删除。创建后可在资源栈详情页查看。

上一步 下一步

创建直接计划后，点击 确定

创建执行计划

通过执行计划，可以预定的资源变更信息。

执行计划名称: executorPlan_20250324_1057_4eda

描述: 请输入对执行计划的描述

确定 取消

点击 部署

基本信息 资源 输出 事件 模板 执行计划

执行计划名称ID	状态	操作策略	创建时间	描述	操作
executionPlan_20250324_1957_4e8a 18a83c49-7d29-4b69-b8ca-689e5c3f2e7	创建成功, 待部署	查看部署详情	2025/03/24 10:58:08 GMT+08:00	-	部署 删除

如下图“Apply required resource success.”即为资源创建完成

基本信息 资源 输出 事件 模板 执行计划

事件时间	事件类型	事件描述	资源名称/类型	关联资源ID
2025/03/24 11:00:06 GMT+08:00	日志	Apply required resource success.	-	-
2025/03/24 11:00:01 GMT+08:00	生成成功	module.ecs.huaweicloud_compute_instance.ecs[0]: Creation complete after 57s [id=aa08d928-dc2e-4095-b388-8b333ec8b844]	ecs ECS	aa08d928-dc2e-4095-b388-8b333ec8b844
2025/03/24 11:00:01 GMT+08:00	完成概要	Apply complete! Resources: 0 added, 0 changed, 0 destroyed.	-	-
2025/03/24 10:59:54 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [50s elapsed]	ecs ECS	-
2025/03/24 10:59:44 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [40s elapsed]	ecs ECS	-
2025/03/24 10:59:34 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [30s elapsed]	ecs ECS	-
2025/03/24 10:59:24 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [20s elapsed]	ecs ECS	-
2025/03/24 10:59:14 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Still creating... [10s elapsed]	ecs ECS	-
2025/03/24 10:59:04 GMT+08:00	生成成功	module.vpc.huaweicloud_vpc_subnet.subnet: Creation complete after 9s [id=fb3ec9ef-a475-4a7d-9e7c-00000316c3763]	subnet Subnet	4b3aceef-a475-4a7d-9e7c-00000316c3763
2025/03/24 10:59:04 GMT+08:00	正在生成	module.ecs.huaweicloud_compute_instance.ecs[0]: Creating...	ecs ECS	-

3 商品资源配置

商品支持ECS控制台配置，下面对资源配置的方式进行介绍。

3.1 ECS 控制台配置

3.1.1 准备工作

在使用ECS控制台配置前，需要您提前配置好安全组规则。

安全组规则的配置如下：

- 入方向规则放通端口8080，源地址内必须包含您的客户端ip，否则无法访问
- 入方向规则放通端口7077，源地址内必须包含您的客户端ip，否则无法访问
- 入方向规则放通CloudShell连接实例使用的端口22，以便在控制台登录调试。
- 出方向规则一键放通

3.1.2 创建 ECS

前提工作准备好后，选择ECS控制台配置跳转到购买ECS页面，ECS资源的配置如下图所示：

The screenshot displays the ECS console configuration interface. It includes sections for region selection, payment mode, availability zone, instance type selection (with a table of instance specifications), network configuration (VPC and Subnet), security group configuration, and public IP configuration. The total price is shown as 0.3988 CNY per hour.

实例类型	规格名称	vCPU核数	内存	CPU	基准/最大带宽	内网峰值包	规格等级
通用计算型s6	s6.medium.4	1vCPU	4GB	Intel Cascade Lake 2.5GHz	0.1 / 0.8 Gbit/s	10万PPS	40.34小时
通用计算型s6	s6.large.2	2vCPU	4GB	Intel Cascade Lake 2.5GHz	0.2 / 1.5 Gbit/s	15万PPS	40.36小时
通用计算型s6	s6.large.4	2vCPU	8GB	Intel Cascade Lake 2.5GHz	0.2 / 1.5 Gbit/s	15万PPS	40.68小时

值得注意的是：

- VPC您可以自行创建
- 安全组选择3.1.1章节中配置的安全组
- 弹性公网IP选择现在购买，推荐选择“按流量计费”，带宽大小可设置为5Mbit/s
- 其余默认或按规则填写即可。

4 商品使用

4.1 Spark 使用

IP+8080访问UI

运行示例程序：【spark-submit --class org.apache.spark.examples.SparkPi

`\${SPARK_HOME}/examples/jars/spark-examples_2.12-3.4.4.jar 10】

```
[root@spark ~]# spark-submit --class org.apache.spark.examples.SparkPi $SPARK_HOME/examples/jars/spark-examples_2.12-3.4.4.jar 10
25/04/27 15:28:02 WARN Utils: Your hostname, spark resolves to a loopback address: 127.0.0.1; using 192.168.0.35 instead (on interface eth0)
25/04/27 15:28:02 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
25/04/27 15:28:03 INFO SparkContext: Running Spark version 3.4.4
25/04/27 15:28:03 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
25/04/27 15:28:03 INFO ResourceUtils: =====
25/04/27 15:28:03 INFO ResourceUtils: No custom resources configured for spark.driver.
25/04/27 15:28:03 INFO ResourceUtils: =====
25/04/27 15:28:03 INFO SparkContext: Submitted application: Spark Pi

25/04/27 15:28:07 INFO TaskSchedulerImpl: Killing all running tasks in stage 0: Stage finished
25/04/27 15:28:07 INFO DAGScheduler: Job 0 finished: reduce at SparkPi.scala:38, took 1.305911 s
Pi is roughly 3.13980713980714
25/04/27 15:28:07 INFO SparkContext: SparkContext is stopping with exitCode 0.
25/04/27 15:28:07 INFO SparkUI: Stopped Spark web UI at http://192.168.0.35:4040
25/04/27 15:28:07 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
25/04/27 15:28:07 INFO MemoryStore: MemoryStore cleared
25/04/27 15:28:07 INFO BlockManager: BlockManager stopped
25/04/27 15:28:07 INFO BlockManagerMaster: BlockManagerMaster stopped
25/04/27 15:28:07 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
25/04/27 15:28:07 INFO SparkContext: Successfully stopped SparkContext
25/04/27 15:28:07 INFO ShutdownHookManager: Shutdown hook called
25/04/27 15:28:07 INFO ShutdownHookManager: Deleting directory /tmp/spark-f99a89b7-64c5-4b13-8e8f-6bfaa97bb509
25/04/27 15:28:07 INFO ShutdownHookManager: Deleting directory /tmp/spark-4885056b-94b1-45a1-b4c7-37ebe6252497
```

4.2 参考文档

- [spark 文档](#)