

DeepSeek_Dify社区版工具使用指南

1 商品说明

DeepSeek-R1是一个高性能的AI推理模型，专注于数学、代码和自然语言推理任务，通过Ollama在云服务器中部署DeepSeek-R1蒸馏版轻量模型，快速打造您的私人AI助手。适用场景包括：自然语言处理（NLP）：能够理解和生成自然语言文本，适用于对话、翻译、摘要等任务；文本生成：能够生成连贯、逻辑清晰的文本，适用于内容创作、故事编写等；问答系统：能够回答用户提出的问题，适用于客服、知识库查询等场景；情感分析：能够分析文本中的情感倾向，适用于市场调研、舆情监控等；文本分类：能够对文本进行分类，适用于垃圾邮件过滤、新闻分类等；信息抽取：能够从文本中提取关键信息，适用于数据挖掘、知识图谱构建等。

Dify 是一款开源的大语言模型(LLM) 应用开发平台。它融合了后端即服务（Backend as Service）和 LLMOps 的理念，使开发者可以快速搭建生产级的生成式 AI 应用。即使你是非技术人员，也能参与到 AI 应用的定义和数据运营过程中。

由于 Dify 内置了构建 LLM 应用所需的关键技术栈，包括对数百个模型的支持、直观的 Prompt 编排界面、高质量的 RAG 引擎、稳健的 Agent 框架、灵活的流程编排，并同时提供了一套易用的界面和 API。这为开发者节省了许多重复造轮子的时间，使其可以专注在创新和业务需求上。

本商品在鲲鹏云的AArch64架构上Ubuntu24.04和HCE2.0系统中进行安装后以镜像提供给用户使用。

2 商品购买

您可以在云商店搜索“DeepSeek_Dify社区版”。

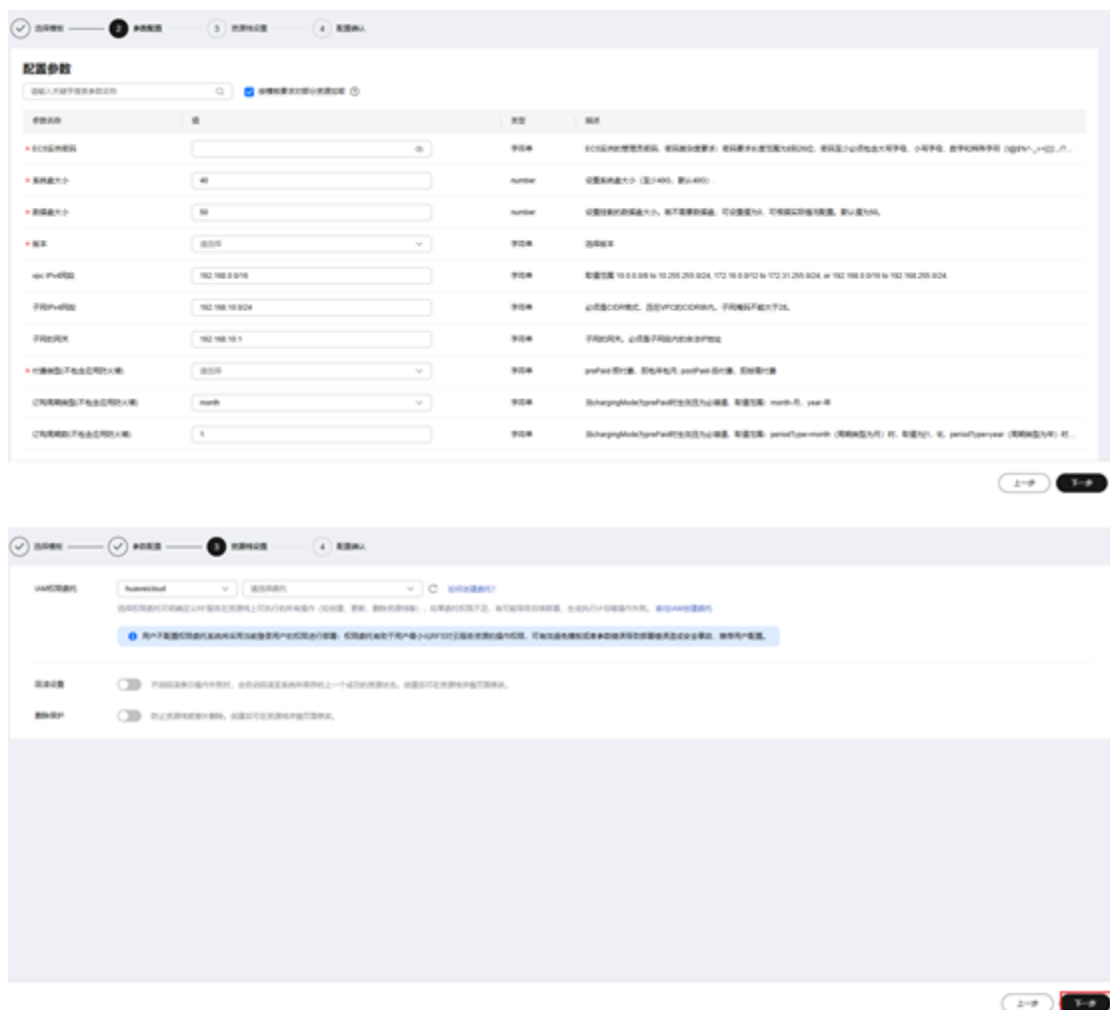
其中，地域、规格、按照推荐配置使用，购买方式根据您的需求选择按需/按月/按年，短期使用推荐按需，长期使用推荐按月/按年，确认配置后点击“立即购买”。

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



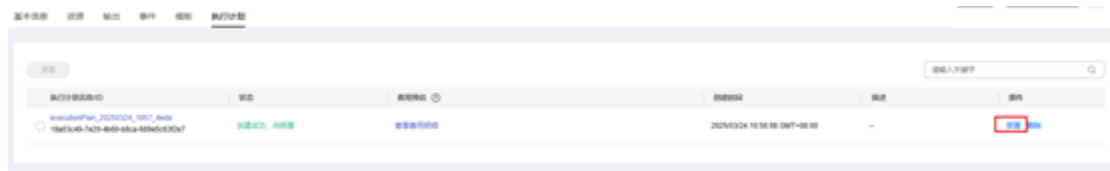
必填项填写后，点击 下一步



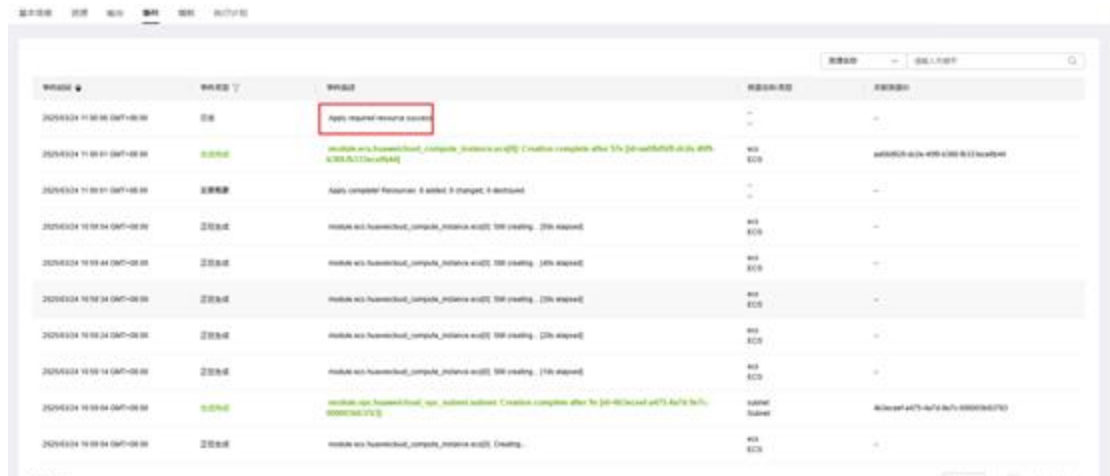
创建直接计划后，点击 确定



点击 部署



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



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80，**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：

基础配置

计费模式 

包年/包月  按需计费 竞价计费

按需计费实例不支持备案。[了解备案限制](#) 

区域 

📍 华北-北京四 👤 推荐区域 华北-北京四 华南-广州 华东-上海一  华北-乌兰察布一  西南-贵阳一

云服务器创建后无法更改区域；不同区域之间内网互不相通，请就近选择靠近您业务的区域，减少网络时延。[如何选择区域](#) 

可用区 

随机分配 可用区1 可用区2 可用区3 可用区7 ☐ 随机至多可用区

实例

规格类型选型

业务场景选型

CPU架构 

x86计算 鲲鹏计算

实例筛选 

--请选择vCPUs-- --请选择内存-- 请输入规格名称模糊搜索 

☒ 隐藏售罄的规格

鲲鹏通用计算增强型 鲲鹏内存优化型 鲲鹏超高I/O型

CSDN @p_xcn



操作系统

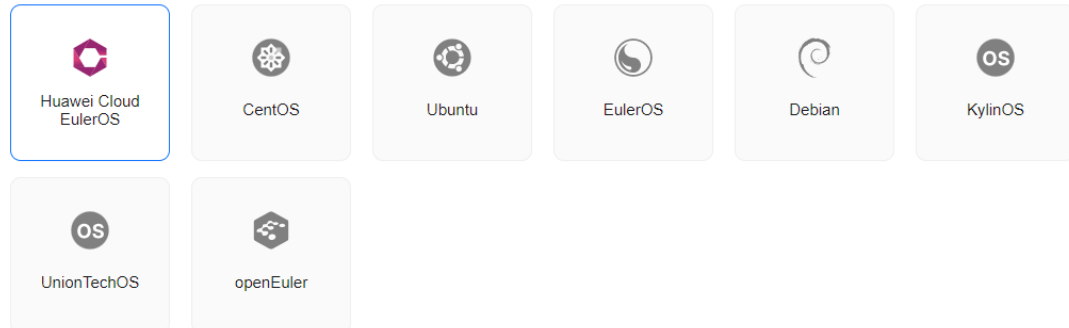
镜像 ?

公共镜像

私有镜像

共享镜像

市场镜像



Huawei Cloud EulerOS 2.0 64bit for kAi2p with HDK 23.0.1 and CANN ...

存储与备份

系统盘 ?

磁盘类型

系统盘大小(GiB)

通用型SSD

40

IOPS上限2,280, IOPS突发上限8,000 高级设置

⊕ 增加一块数据盘

您还可以挂载 23 块磁盘 (云硬盘)

☐ 开启备份

CSDN @p_xcn

云服务器的名称: ☐ 允许重名

描述:

登录凭证:

密钥对:

云备份:

云备份策略 (可选):

高级选项: ☒ 现在配置

实例自定义数据脚本:

脚本内容:

```
#!/bin/bash
echo 'root:xxx' | chpasswd
bash homeint.sh
```

购买: 台 配置费用 ¥0.3988/小时 + 弹性公网IP费用 ¥0.80/GB

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
- 高级配置需要在高级选项支持注入自定义数据, 所以登录凭证不能选择“密码”, 选

择创建后设置

- 其余默认或按规则填写即可。

4 商品使用

4.1 登录服务器查看 Dify 进程

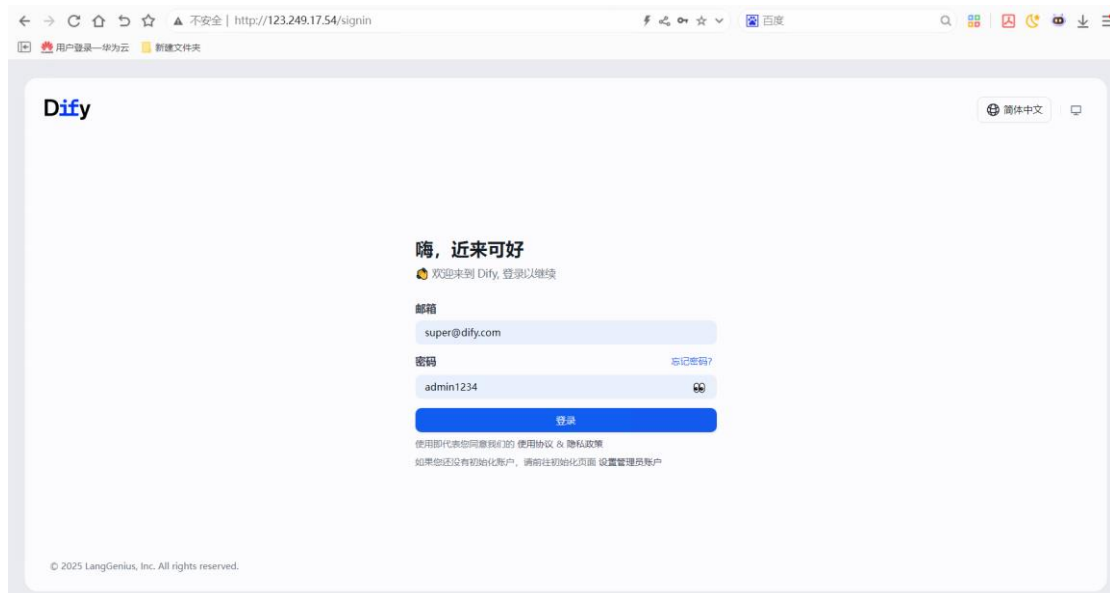
进入系统后通过命令 `docker ps` 查看 Dify 进程如下。

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS
0d085b08061c	nginx:latest	"sh -c 'cp /docker-e..."	52 minutes ago	Up 46 minutes	0.0.0.0:80->80/tcp
f2308e0a0e08	langgenius/dify-plugin-daemon:0.0.0-local	/bin/bash -c /app/e..."	52 minutes ago	Up 46 minutes	0.0.0.0:3003->3003/tcp
4919dc9c28cf	langgenius/dify-api:1.5.1	/bin/bash /entrypoi..."	52 minutes ago	Up 46 minutes	5001/tcp
c23894740b9	langgenius/dify-api:1.5.1	/bin/bash /entrypoi..."	52 minutes ago	Up 46 minutes	5001/tcp
054fcb272b04	redis:8-alpine	docker-worker-l..."	52 minutes ago	Up 46 minutes (healthy)	6379/tcp
74008f8d1f45	langgenius/dify-web:1.5.1	docker-redis-l..."	52 minutes ago	Up 46 minutes	3000/tcp
d427fe0ea9e9	semitechnologies/weaviate:1.19.0	/bin/sh /entrypoi..."	52 minutes ago	Up 46 minutes	
1b7dc1a517b	ubuntu/squid:latest	"sh -c 'cp /docker-e..."	52 minutes ago	Up 46 minutes	3128/tcp
754622682acf	postgres:15-alpine	docker-sacr_proxy-l..."	52 minutes ago	Up 46 minutes (healthy)	5432/tcp
a80951b17fb0	langgenius/dify-sandbox:0.2.11	docker-db-l..."	52 minutes ago	Up 46 minutes (healthy)	
		/main" docker-sandbox-l..."			

4.1.1 通过浏览器登录 Dify 平台

初次登录平台地址：<http://your ip/install>，已经默认注册用户了

默认邮箱地址：`super@dify.com` 密码：`admin1234`



4.1.2 配置 DeepSeek-R1 推理服务

- 通过命令 `systemctl status ollama` 查看服务状态，确保服务已经启动，模型名称 `deepseek-r1:7b-qwen-distill-q8_0`
- 已经默认配置DeepSeek-R1推理服务，预计开机后3到5分钟可以看到配置的服务信息

The image shows two screenshots. The top screenshot is a terminal window displaying the output of the command `systemctl status ollama`. It shows that the `ollama.service` is loaded, active (running), and has been running since Tuesday, 2025-06-03 at 15:39:51 CST. The main PID is 2283. Below this, there is a log showing several messages from the `ollama` service, including information about the LLM model being loaded and the GPU usage.

The bottom screenshot is a browser window showing the Dify interface. The browser address bar shows `http://123.249.17.54/plugins`. The Dify interface displays the 'Plugins' section, where the 'Ollama' plugin is listed. The plugin details show it is version 0.0.6 and is provided by 'langgenius / ollama'. The right sidebar shows the 'Ollama' plugin is installed and the model 'deepseek-r1:7b-qwen-distill-q8_0' is selected.

如果 5 分钟后还没有配置成功，可以登录 ECS 服务器查看脚本是否报错

【`systemctl status auto_config.service`】

The image shows a terminal window displaying the output of the command `systemctl status auto_config.service`. It shows that the `auto_config.service` is loaded, inactive (dead), and has been inactive since Wednesday, 2025-07-30 at 17:06:00 CST. The main PID is 21598. Below this, there is a log showing the execution of the `auto_config.sh` script, which successfully configured the LLM model and the GPU usage.

如果执行失败，可以手动执以下命令，进行配置

【sh /opt/auto_config.sh】

4.1.3 完整参考 Dify 手册

[Dify手册](#)