

如何快速部署大模型接口管理和分发系统：One-API

[什么是One API?](#)

[One API 的使用场景是什么?](#)

[如何使用One API?](#)

[如何在阿里云上快速部署One API](#)

简介： One API 是一个开源的接口管理与分发系统，支持多种大模型平台如 OpenAI、Google PaLM 2、百度文心一言等。通过统一接口访问不同大模型服务，简化工作流程并提高效率。适用于多模型集成项目、开发代理服务、教育研究及快速原型制作等多种场景。阿里云计算巢提供了快速部署方案，简化了部署过程。

目前业界内有很多的大模型服务，比如OpenAI’s ChatGPT、Google’s PaLM2/Gemini、百度的文心一言等多个平台的大模型，如果要接入大模型的话，需要针对每个特定的AI平台独立编写和维护代码。当需要集成多个大模型时，工作成本就高了，这个时候如果能通过统一的接口访问不同的大模型服务，可以极大地简化工作流程和提高效率，[One API](#) 就是这样一个工具。

什么是One API?

[One API](#) 是一个开源的OpenAI接口管理与分发系统，支持 Azure、Anthropic Claude、Google PaLM 2 & Gemini、智谱 ChatGLM、百度文心一言、讯飞星火认知、阿里通义千问、360 智脑以及腾讯混元，可用于二次分发管理 key，仅单可执行文件，已打包好 Docker 镜像，一键部署，开箱即用。

One API 的使用场景是什么?

- 多模型集成项目：** 当你的项目需要集成多个大模型，而这些模型分布在不同的AI平台时，One API 可以作为统一的接口，简化你的集成工作。
- 开发代理服务：** 如果你需要构建一个自己的AI服务，而后端需要调用不同的AI模型，One API 可以作为后端服务，你只需要与One API交互，而不需要处理与各个AI平台的具体交互细节。
- 教育和研究：** 在AI的学习和研究中，经常需要对比不同模型的表现。One API 可以简化这一流程，使研究者可以轻松切换不同的模型进行对比分析。

4. **小型企业或个人项目**：对于资源有限的小型企业或个人开发者，可能没有足够的时间和精力去适配和管理多个AI平台的API。One API提供了一个简单便捷的解决方案。
5. **在敏捷开发中快速原型制作**：当你需要快速验证一个涉及AI模型的想法时，One API可以帮助你省去了配置和接入多个AI服务的时间，让你更专注于原型的开发。

One API在处理需要访问多个AI模型的项目时尤为有用，它通过提供统一的接口来简化开发流程，节省时间和资源。不论是在开发复杂的综合AI服务、进行AI模型对比研究，还是简单地快速验证AI相关的项目想法，One API都能提供强大的支持。

如何使用One API?

One API 可以用来管理模型池，其可以兼容 OpenAI 、 Azure 、 国内主流模型和本地模型等。

1. 使用root 账号登录，默认密码：123456
2. 在渠道页面中添加新的渠道



3. 然后在令牌页面中新增访问令牌



4. 客户端使用令牌访问 One API。
5. 根据请求中的 model 参数，匹配对应的渠道（根据渠道里的模型进行匹配，必须完全一致）。如果

匹配到多个渠道，则随机选择一个（同优先级）

6. One API 向真正的地址发出请求，并将结果返回给客户端

如何在阿里云上快速部署One API

阿里云计算巢提供了快速部署的One API的解决方案，只需要简单填写配置就可以一键部署到自己的VPC 内，非常简单快捷，操作步骤如下：

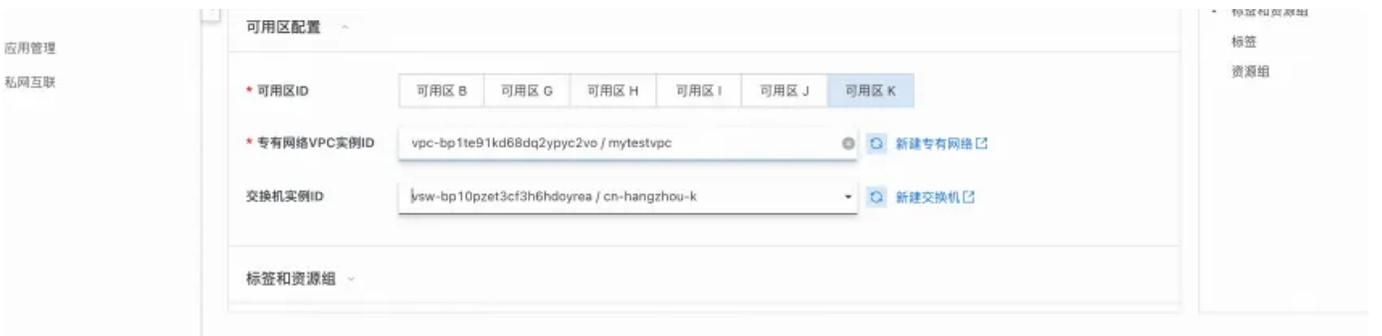
1. 单击**部署链接**，进入计算巢的服务实例部署界面。
2. 根据界面提示，填写参数完成部署。



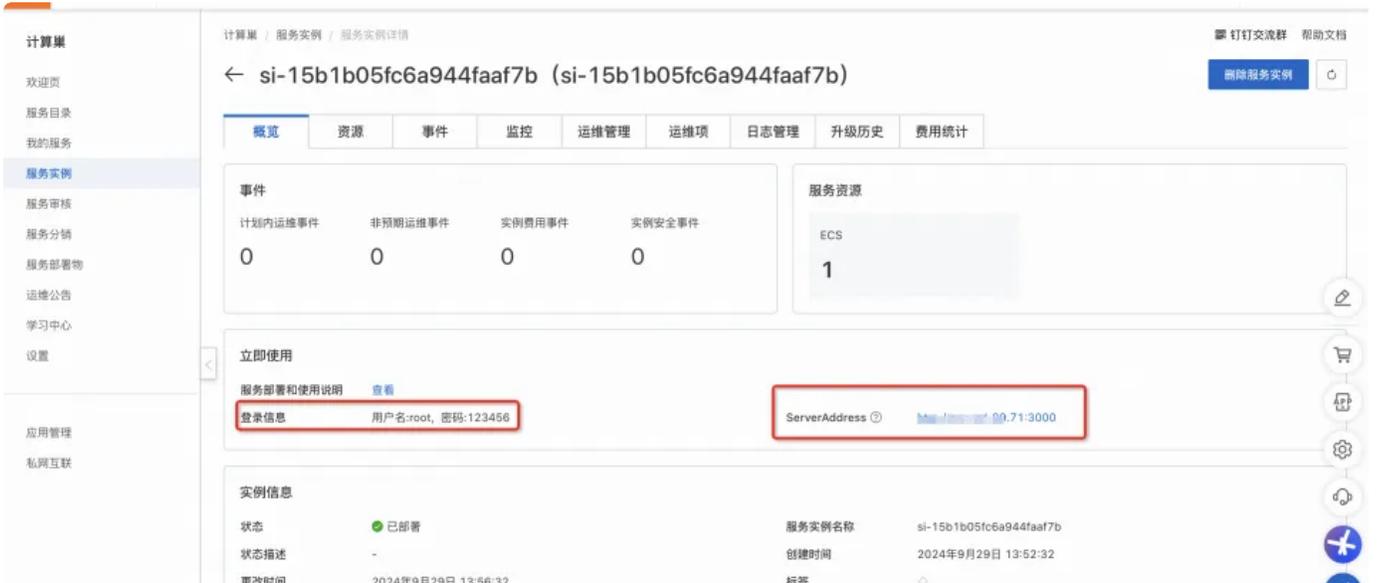
选择实例规格，并设置实例密码



设置可用区和VPC信息



3. 设置完参数后, 点击下一步确认订单, 点击立即创建, 等待服务实例创建完成。
4. 服务实例创建成功后, 进入服务实例详情页。在概览页可获取one-api登录信息。



使用初始账号登录，用户名为 `root`，密码为 `123456`。



The screenshot shows the One API dashboard interface. At the top, there is a navigation bar with the One API logo and several menu items: 首页 (Home), 令牌 (Tokens), 充值 (Recharge), 日志 (Logs), 设置 (Settings), and 关于 (About). A '登录' (Login) button is located in the top right corner. The main content area is titled '系统状况' (System Status) and is divided into two columns. The left column, '系统信息' (System Information), lists: 系统信息总览 (System Information Overview), 名称: One API (Name: One API), 版本: v0.6.9-alpha.2 (Version: v0.6.9-alpha.2), 源码: https://github.com/songquanpeng/one-api (Source: https://github.com/songquanpeng/one-api), and 启动时间: 2024-09-27 11:30:46 (Start Time: 2024-09-27 11:30:46). The right column, '系统配置' (System Configuration), lists: 系统配置总览 (System Configuration Overview), 邮箱验证: 未启用 (Email Verification: Not Enabled), GitHub 身份验证: 未启用 (GitHub Authentication: Not Enabled), 微信身份验证: 未启用 (WeChat Authentication: Not Enabled), and Turnstile 用户校验: 未启用 (Turnstile User Verification: Not Enabled). At the bottom of the dashboard, a footer states: 'One API 由 JustSong 构建，源代码遵循 MIT 协议' (One API built by JustSong, source code follows MIT license).

参考文档：

<https://docs.geekai.me/config/chat/one-api.html#geekai-%E9%85%8D%E7%BD%AE>

<https://aibard123.com/newtutorial/2024/02/one-api/>

查看更多计算巢服务 (<https://computenest.aliyun.com>)