高性能应用服务

产品简介

腾讯云

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

产品简介

产品概述

最近更新时间：2023-11-10 16:17:51

快速了解高性能应用服务 HAI

高性能应用服务（Hyper Application Inventor，HAI）是一款面向 AI 和科学计算的 GPU/NPU 应用服务产品，提供即插即用的强大算力和常见环境。它可以帮助中小企业和开发者快速部署语言模型（LLM）、AI 绘图、数据科学等高性能应用，原生集成配套的开发工具和组件，大大提升应用层的开发生产效率。

与 GPU 云服务器的区别

若需了解高性能应用服务 HAI 与 GPU 云服务器的区别，详情可参见 产品对比 。

如何使用高性能应用服务

腾讯云为您提供如下方式进行高性能应用服务的配置和管理：

- **控制台**：高性能应用服务可使用 控制台 ，对算力资源进行管理。
- **API**：腾讯云也提供了 API 接口方便您管理高性能应用服务。关于 API 说明，请参见 API 概览 。

> **说明**
> 如果您未使用过高性能应用服务，可参见 通过高性能应用服务一键创建应用 开始使用。

计费概述

有关高性能应用服务的计费相关说明，详情请参见 计费概述 。

产品优势

最近更新时间：2023-11-10 14:45:42

简单易用

通过简化计算、网络和存储等基础设施的配置流程，大幅降低了云服务操作和管理的复杂度。

应用环境快速部署

支持多种 AI 环境快速部署，如 ChatGLM-6B、StableDiffusion 等，使用户可专注业务及应用场景创新。

高灵活性

支持用户登录实例，对 AI 模型及实例环境进行灵活配置。可进行内部开发、业务测试，或对外提供业务服务。

多种登录方式

除传统连接方式外，支持通过 jupyterlab、WebUI 等方式一键启动，提供更贴合使用场景的登录方式。

算力种类丰富

提供多种算力套餐选择，未来还将加入更多种类供用户选择。

产品对比

最近更新时间：2023-11-16 09:25:51

高性能应用服务 HAI 相比传统 GPU 云服务器的主要区别和优势请参考下表：

功能类别	GPU 云服务器	高性能应用服务 HAI
交付形态	基础的虚拟机	即插即用的应用
机型选择	需要了解 GPU 型号，自行选择合适机型，有不匹配风险	基于 AI 应用，自动匹配合适套餐
环境部署	需要自行部署驱动、CUDA、Python、Notebook 等环境依赖	分钟级快速启动，直接交付可用应用环境
资源配置	需要额外购买合适的云硬盘、带宽或流量	打包 GPU、云硬盘、带宽及网络，一键启动
产品入口	需要具备一定运维知识，登录命令行界面进行操作	提供 WebUI 等可视化连接方式，一键进入服务，可视化配置
模型甄别	各类模型版本繁多，难以自行挑选	预置最新版本的主流模型，适配套餐机型
资源下载	部分访问可能遇到网络拥塞问题	跨境线路自动择优，支持学术资源平台访问、下载加速

应用场景

最近更新时间：2023-11-10 14:45:42

AI 作画/设计

设计师和开发者可以使用高性能应用服务快速地部署和优化 AI 绘画模型。高性能应用服务预置 Stable Diffusion 等主流 AI 作画模型及常用插件，提供 GUI 图形化界面即开即用，大幅降低上手门槛。

AI 对话/写作

研究者和开发者可以使用高性能应用服务快速部署和运行大型语言模型，如 LLAMA2、ChatGLM 等，进行自然语言处理任务，如文本生成、情感分析、文本分类等。高性能应用服务提供的算力支持和优化环境确保了语言模型可以在最短的时间内进行部署，同时还能保持高稳定性和可靠性。

AI 开发测试

高性能应用服务的预配置环境支持大多数流行的 AI 框架和工具，如 TensorFlow、PyTorch 等，使得开发者可以专注于算法设计和模型优化。AI 研究者可以在高性能应用服务上进行模型的开发、训练、测试和优化，无需担心硬件兼容性和软件配置问题。如新算法的原型开发、模型微调与迁移学习、深度学习框架的交叉测试等。

数据科学

数据科学家们可使用高性能应用服务，快速进行数据分析和图标处理。高性能应用服务预置了 Notebook、Python 环境，以及主流分析软件。