

计算加速套件 TACO Kit

常见问题



腾讯云

【版权声明】

©2013–2025 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【商标声明】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【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。

常见问题

最近更新时间：2025-08-20 16:04:52

计算加速套件 TACO Kit 包含哪些产品？

包含 TACO Train AI 训练加速引擎、TACO LLM 推理加速引擎、TACO DiT 推理加速引擎。

计算加速套件 TACO Kit 适用于哪些场景？

适用于训练加速和推理加速，目前支持以下场景及模型：

- 大语言模型。例如 Qwen2.5、DeepSeek R1、DeepSeek V3、Llama3 等。
- 生图模型。例如 Flux.1-dev、Wan2.1 等。
- 多模态模型。例如 QwenVL 等。

TACO Train AI 训练加速引擎包含哪些组件？

TACO Train 目前提供了三个训练加速组件：

- Tencent TensorFlow 1.15：基于 TensorFlow 1.15 深度优化的训练框架（简称 TTF）。
- LightCC：基于 Horovod 深度优化的分布式训练框架。
- HARP：自研用户态网络协议栈。

TACO DiT 支持哪些特性？

支持多卡并行加速，通过聚合多张卡的算力，降低单次推理耗时。同时支持单卡加速，通过算子优化、图编译优化等方式降低推理耗时。

TACO LLM 支持哪些场景？

TACO LLM 适用于大语言模型的推理加速业务，可满足多种业务场景下推理提效的需求。典型业务场景包括客户服务、内容创作与编辑、机器翻译等。

使用 Keras + Horovod 训练过程中，发现内存使用量随着 epoch 增加而增加，最终导致内存溢出应该怎么处理？

建议使用 [TCMalloc](#) 进行内存优化，可以使内存在多个 epoch 之间基本保持不变。

哪些 GPU 机型支持 HARP 网络协议？

所有的 GPU 实例类型均支持 HARP 网络协议。

如何在公有云上申请使用 TACO Kit 的推理加速引擎？

如您在大模型场景有推理加速需求，可联系 [技术支持](#)，提供 TACO 定制化加速服务。