

计算加速套件 TACO Kit

产品简介



腾讯云

【 版权声明 】

©2013–2024 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。

文档目录

产品简介

产品概述

产品优势

应用场景

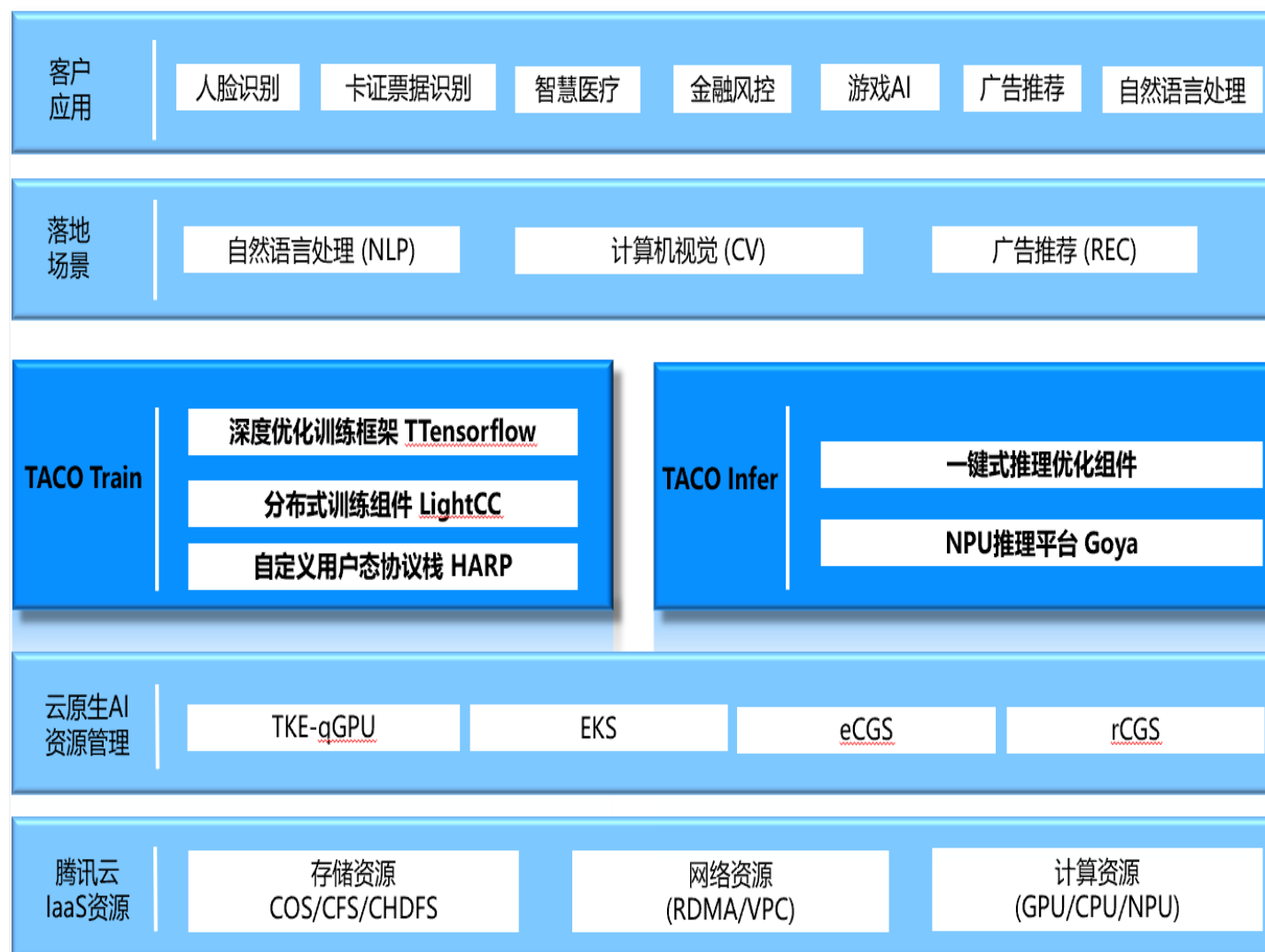
产品简介

产品概述

最近更新时间：2023-04-21 09:35:42

计算加速套件 TACO Kit (TencentCloud Accelerated Computing Optimization Kit) 是一种异构计算加速软件服务，具备领先的 GPU 共享技术和业界唯一的 GPU 在离线混部能力，搭配腾讯自研的软硬件协同优化组件和硬件厂商特有优化方案，支持物理机、云服务器、容器等产品的计算加速、图形渲染、视频转码各个应用场景，帮助用户实现全方位全场景的降本增效。

其中，AI 加速引擎 TACO Train 和 TACO Infer 是腾讯云虚拟化团队依托云帆团队，立足于腾讯内部丰富的 AI 业务场景，深耕训练框架优化、分布式训练优化、网络通信优化、推理性能优化等关键技术，携手打造的一整套 AI 加速方案。为了更好的服务用户，腾讯云决定将内部深度优化的加速方案提供给公有云用户，助力广大用户提高 AI 产品迭代效率。其架构示意图如下所示：



您可以通过以下文档，了解更多关于 TACO Train 及 TACO Infer 信息：

- [TACO Train 概述](#)
- [TACO Infer 概述](#)

产品优势

最近更新时间：2023-01-04 10:14:47

TACO Train AI 训练加速引擎和 TACO Infer AI 推理加速引擎通过软硬件协同优化，屏蔽底层硬件差异，适配 CPU、GPU、NPU 等不同加速硬件，降低用户使用计算资源的学习成本的同时提高计算效能。

TACO Train AI 训练加速引擎优势

多层次深度优化加速

提供从自底向上的网络通信、分布式策略及训练框架等多层级的优化加速组件，用户可以根据需要选择适配。

支持无侵入式业务迁移

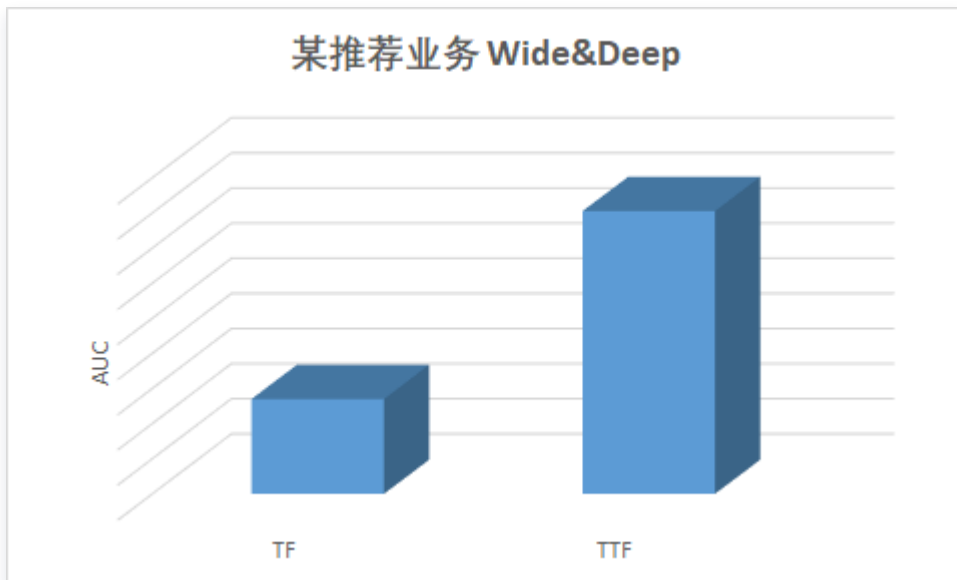
HARP、LightCC 等优化技术支持插件式集成，无需业务代码改动，即可加速分布式训练业务。

灵活扩展分布式训练场景

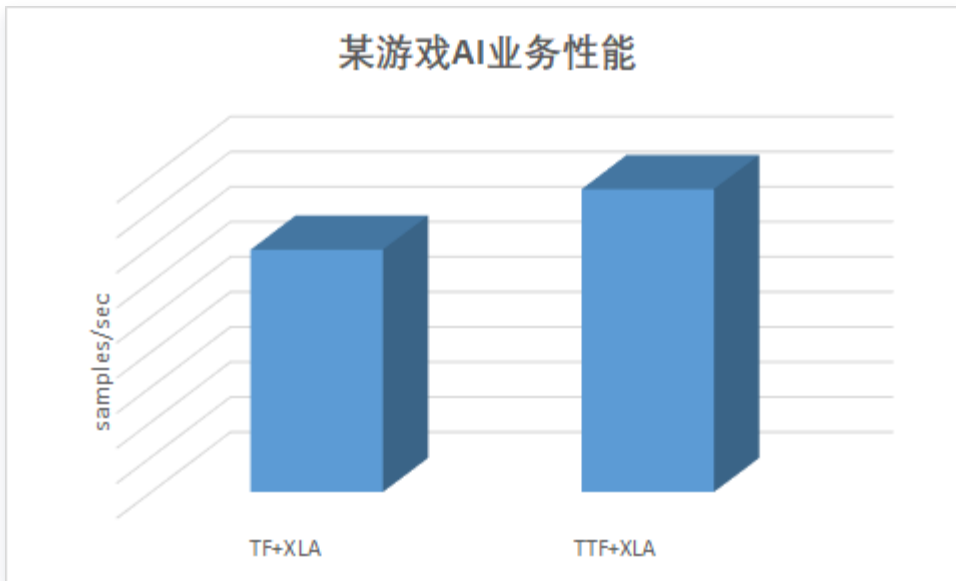
支持大规模多机多卡分布式训练场景，提高加速比和模型迭代效率。

训练性能提升数据

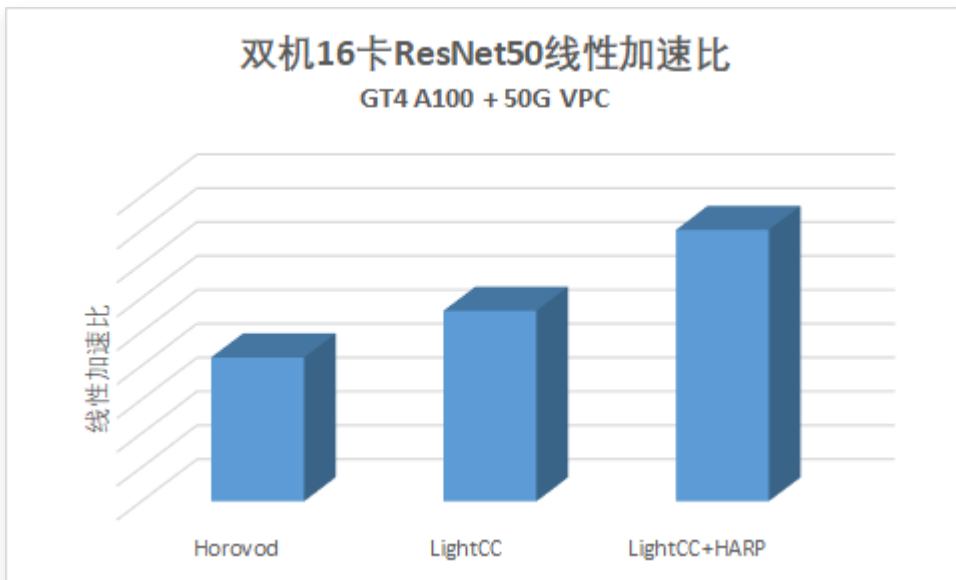
- TencentTensorflow（以下简称 TTF）动态 Embedding 在某推荐业务上对 AUC 的提升效果：



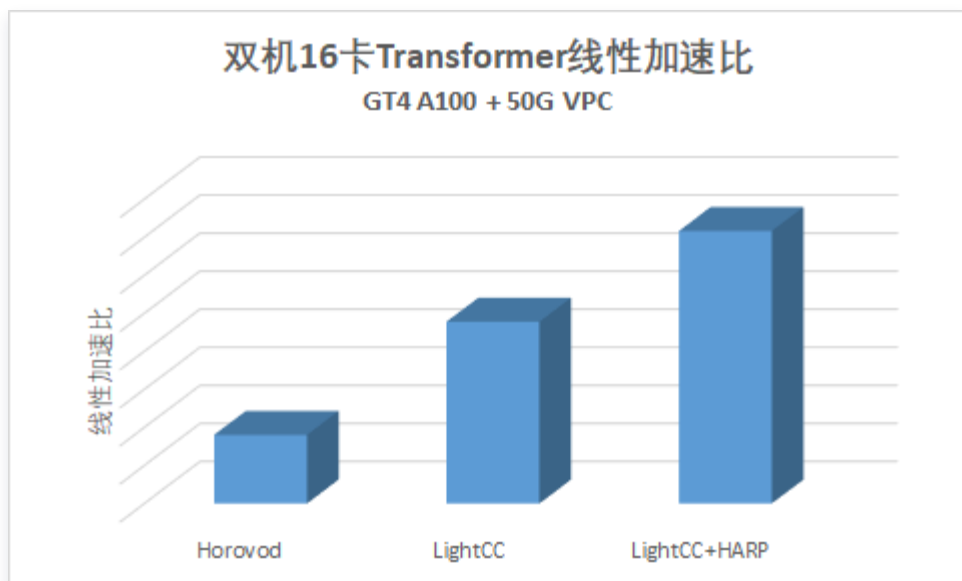
- TTF XLA 在某游戏业务上的性能加速效果：



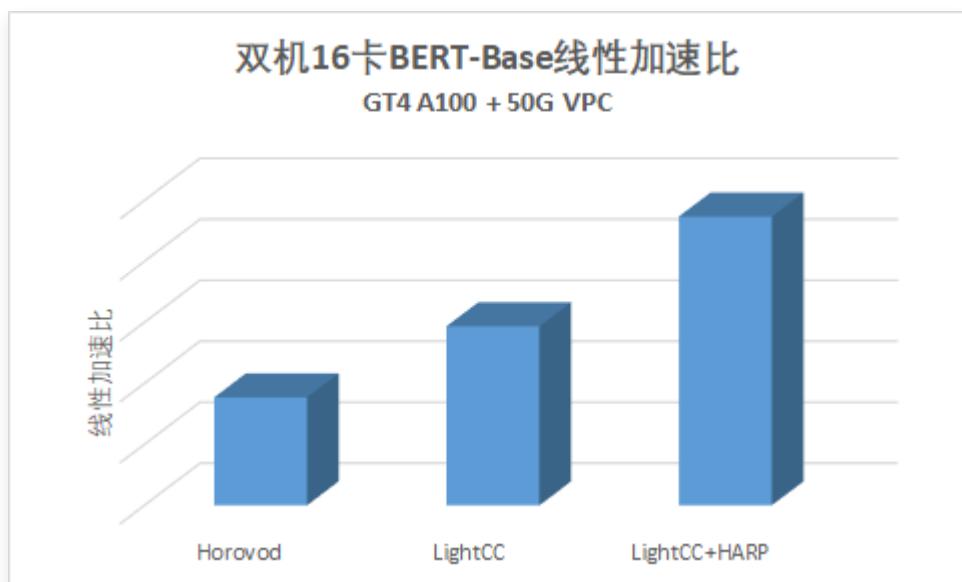
- 在腾讯云50G VPC 环境下，ResNet50的多机训练加速效果：



- 在腾讯云50G VPC 环境下，Transformer 的多机训练加速效果：



- 在腾讯云50G VPC 环境下，BERT-Base 的多机训练加速效果：



TACO Infer AI 推理加速引擎优势

部署简洁

TACO Infer 仅有一行简洁的优化接口，不会改变用户的模型格式。用户可以保持其一贯的使用和部署习惯，并提供插件式的第三方开发接口，支持适配不同业务场景。

软硬件兼容

支持多种框架模型和多种加速硬件，可运行在虚拟机、物理机、容器等各种环境。

一站式解决推理部署相关优化依赖

集成硬件厂商的定向开源的加速方案，整合先进的编译优化、图优化和算子优化技术。

应用场景

最近更新时间：2023-07-13 10:02:11

TACO Train AI 训练加速引擎和 TACO Infer AI 推理加速引擎目前支持但不限于以下场景：

- 计算机视觉
例如，ResNet、MobileNet、Inception、ViT 等。
- 推荐系统
例如，Wide&Deep、DeepFM 等。
- NLP 服务
例如，BERT、Transformer、ASR 等。