QingCloud
Technologies

www.qingcloud.com

青云
金融行业
智算解决方案

懂客户
更懂云上金融创新

CONTENTS

目录

01

金融行业
数智化发展趋势

02

如何选择
AI 数字基础设施

03

金融行业
智算解决方案

04

案例实践

QingCloud Technologies

01

金融行业
数智化发展趋势

FINANCIAL INDUSTRY
DEVELOPMENT TREND OF DIGITALIZATION AND INTELLIGENCE

QINGCLOUD
FINANCIAL INDUSTRY
COMPUTING SOLUTIONS

金融行业正在经历一场深刻的变革，从业务数字化逐渐步入深度智能化阶段。随着互联网技术的不断进化和普及，金融行业也在不断地探索和应用新技术，以实现业务的数字化与智能化。在这一进程中，云计算、大数据、人工智能等技术正在逐步改变着金融行业的运作方式和服务模式。特别是 AI 大模型的出现，为金融行业带来了无限可能，推动着金融行业向智能化、自动化与个性化的方向发展。

金融行业数智化转型历程

01

内部数字化

金融行业在内部推行业务的计算机处理、金融数据的入网与联网，建立起数字化处理业务的基础设施，统一和规范了内部业务处理流程，提高了业务处理效率。

02

外部经营数字化

随着移动互联网与移动智能设备的普及，金融行业开始通过线上响应，实现客户随时随地办理金融业务的需求。这一转变进一步扩大了金融行业的经营群体和业务受体，提高了经营效率。

03

深度、全面的智能化

在人工智能、区块链、云计算、大数据等金融科技技术的推动下，金融行业开始进入深度、全面的智能化阶段。智能风控、信贷审批模型、信用分析系统、智能机器人客服等应用越来越广泛，成为金融行业数智化转型的重要方向。

大模型在金融行业的应用与发展

大模型功能强大

大模型以其强大的内容创作和人机交互能力，支持文字、图片、视频等多模态交互。同时，大模型还具备强大的上下文学习能力（In-Context Learning，简称 ICL），能够根据上下文信息理解用户的意图和需求。此外，大模型还具备推理能力，能够通过一系列中间推导过程的思维链（Chain-of-Thought，简称 CoT）实现数学推理、常识推理等复杂任务。零样本学习（Zero-Shot Learning）和生成式 AI 等技术的运用，使得大模型能够处理未见过的数据和任务，展现出强大的适应性和可扩展性。

数智化场景应用广泛

大模型在金融行业的应用场景广泛。大模型可以自动化和优化程序，承担商业银行中的行政管理、日常辅助决策等工作。大模型还可以在信贷审批、风险评估、投资组合管理、市场分析等关键业务活动中发挥重要作用，提高决策的准确性和效率。此外，在客户服务、个性化产品推荐和员工招聘等活动中，大模型也能够发挥更加具有创造性和吸引力的作用。

金融行业大模型应用趋势

金融行业大模型的应用正呈现出以下几个趋势：

先内后外、从易到难、场景迁移

金融机构首先会在内部场景中应用大模型，逐步扩展到外部场景，实现金融业务的全面智能化。

大模型与小模型协同进化

随着技术的不断发展，大模型与小模型将实现协同进化，共同推动金融行业的智能化升级。

多模态金融大模型的发展与应用

多模态金融大模型将能够更好地处理金融领域的复杂问题，为金融机构提供更加全面、高效的服务。

AI Agent * 成为金融业信息基础设施

AI Agent 作为一种能够感知环境、进行决策和执行动作的智能实体，将成为金融业信息基础设施的重要组成部分，推动金融行业实现智能化转型。

*AI Agent（人工智能体）一种计算实体或程序，具备感知、决策、执行以及自适应等能力，能够自主完成特定任务。不同于传统的人工智能，AI Agent 具备通过独立思考、调用工具去逐步完成给定目标的能力。

金融行业智算挑战与机遇

金融行业在智能化转型中面临着诸多挑战，如技术更新迭代快、数据安全和隐私保护、监管政策变化等。然而，随着大模型技术的不断发展和应用，金融行业也在不断地探索和创新，以实现更加智能化、自动化和个性化的服务。

面临的挑战

技术更新迭代快

金融行业需要不断跟进新技术的发展，以适应快速变化的市场需求。

数据安全和隐私保护

在数智化转型过程中，金融机构需要确保客户数据的安全和隐私，防止数据泄露和滥用。

监管政策变化

金融行业的监管政策不断变化，金融机构需要时刻关注政策动向，确保业务合规。

智能化转型带来的机遇

提升业务效率

智能化转型能够大幅提高金融业务的处理效率和准确性，降低运营成本。例如，智能风控系统能够实时识别潜在风险，提高信贷审批效率；智能客服系统能够 24 小时不间断地为客户提供服务，提高客户满意度。

优化客户体验

智能化转型能够为客户提供更加便捷、个性化的服务体验。例如，通过智能投顾系统，客户可以获得全天候、个性化、低门槛的投资建议；通过智能支付系统，客户可以享受更加安全、便捷的支付服务。

创新服务模式

智能化转型为金融行业带来了新的服务模式和创新机会。例如，利用大模型技术，金融机构可以开发新的金融产品和服务，满足客户的多样化需求；通过区块链技术，可以实现去中心化的金融交易和清算，提高金融系统的效率和安全性。

QingCloud Technologies

02

如何选择 AI 数字基础设施

HOW TO CHOOSE
AI DIGITAL INFRASTRUCTURE

QINGCLOUD
FINANCIAL INDUSTRY
COMPUTING SOLUTIONS

选择合适的部署模式

	私有化部署	行业云部署	公有云部署
说明	金融机构将大模型部署于自有服务器，由金融机构负责维护和管理。	由行业内起主导作用的组织建立与维护，在确保数据安全的前提下，向行业内部或相关组织提供云服务。	通过标准接口调用部署在公有云上的资源与服务，由云服务提供商负责维护和管理，具备更低的成本、更高的灵活性和可扩展性。
适用对象	有一定技术实力的金融机构。	由行业节点类金融机构建设，中小金融机构作为使用者采用。	仅适合金融机构内部使用的非敏感类场景。
优缺点	**优点：**数据更安全，满足合规要求，便于拥有更灵活高效的资源调度。 **缺点：**成本相对高昂，需要响应的技术人才储备，维护成本较高。	**优点：**低门槛，起步成本低，同时满足数据安全和合规要求。 **缺点：**资源受平台可用规模所限。	**优点：**成本低，随需所用。 **缺点：**监管合规风险。
行业现状	国有银行、股份制等大中型金融机构，以及部署中等规模及以上的金融机构采用私有化部署。	中小型金融机构以较低成本，快速获得大模型能力的方案之一。	更适用于互联网类的金融企业或非敏感类业务场景。
选择青云	青云 AI 智算平台	青云 AI 智算平台	青云 AI 算力云

鉴于个人隐私保护和数据不出域等相关要求，私有化部署仍是金融机构部署大模型的主要选择方式。对于不涉及数据保密性的场景，比如证券公司基于公开数据生成投资策略及研报撰写，行业云或公有云部署具有一定优势。

选择青云的理由

北京青云科技股份有限公司（简称：青云科技，股票代码：688316），是一家技术领先的企业级云服务商与数字化方案提供商。自 2012 年创立以来，坚持核心代码自研，以顶尖的技术实力见长，构建起端到端的数字化解决方案，全面布局 AI 算力与服务生态，持续打造云原生最佳实践，以中国科技服务数字中国。青云科技自 2014 年开始布局混合云市场，无缝打通公有云和私有云，交付一致功能与体验的混合云，并于 2021 年 3 月登陆上交所科创板，被称为 "混合云第一股"。

青云科技坚持自主创新、中立可靠、灵活开放的理念，立足企业现实需求，围绕 "数字化、AI 算力、信创、云原生" 四大场景，打造核心业务线，帮助企业构筑坚实的数字基石，实现全场景自由计算。

当 AI 成为数字经济的创新驱动力、算力成为社会发展必不可少的关键生产力之时，青云科技抢先布局，全面开展 AI 算力云服务、AI 智算平台等业务，携手 AI 产业生态合作伙伴，让 AI 真正能释放出业务价值。

AI 智算平台

打造智算中心的建设与运营新模式，像管理本地资源一样管理 AI 基础设施。

对 AI 算力进行动态监控调整，以满足不同业务的需求，提高 AI 算力的整体使用效率和管理效率，已在国家超算济南中心等算力中心成功落地并投入使用。

AI 算力云

面向人工智能场景的资源与服务，实现云上开发与训练。

包括 AI 裸金属 GPU 主机、AI 训练集群、并行文件存储、镜像仓库在内的 AI 专用产品，实现租户隔离，满足安全、可靠的云上开发与训练需求。

在当今金融行业的数智化转型大潮中，选择一个可靠的合作伙伴至关重要。青云科技凭借其深厚的行业经验、创新的技术能力和丰富的产品矩阵，为金融机构提供了全面、可靠、灵活的解决方案，成为金融数智化转型的坚实基石。

青云助力 300+ 金融机构云上创新，包括 3 家大型国有商业银行、12 家股份制商业银行、百余城镇商业银行、TOP 5 保险机构、60+ 证券机构等。

银行	保险	证券 / 基金	泛金融
中国银行 BANK OF CHINA	泰康人寿 TAIKANG LIFE	CICC 中金公司	中银金科 BOC FINTECH
中国农业银行 AGRICULTURAL BANK OF CHINA	中国太平 CHINA TAIPING	中邮证券 CHINA POST SECURITIES	东风金融 DONGFENG FINANCE
招商银行 CHINA MERCHANTS BANK	PICC 中国人民保险	民生证券 MINSHENG SECURITIES	中移金融科技 CM Financial Technology
Bank 中国光大银行	大家保险	国融证券 GUORONG SECURITIES	倍特期货 BRILLIANT
华夏银行 HUAXIA BANK	阳光保险集团 Sunshine Insurance Group	易方达	建信融通 CCB SCF CO., LTD.

灵活多样

青云在 AI 算力领域拥有完整而领先的布局，产品架构一致，因此青云不仅提供传统的算力公有云服务及算力租赁服务，还具备构建专属云、私有云的能力。从 3 节点的小型私有云到上千节点的大型集群，青云都能轻松应对，确保金融机构能够根据自身的发展阶段和业务特点，选择最适合自己的 AI 算力服务模式。

中立开放

青云解决方案采用松耦合架构，提供开放的 API 接口和 SDK，不与硬件绑定，可 OEM。同时，支持公私混托、多种部署方式和多元异构，为金融机构提供了极大的灵活性和可扩展性。此外，青云还积极与开源社区和 AI 生态伙伴合作，共同推动金融行业的创新发展。

落地	AI 应用场景生态	金融 行业应用生态	交通 行业应用生态	制造业 行业应用生态	能源 行业应用生态	自然资源 行业应用生态
共筑	AI 算力服务生态	SENSORO	感易智能 Sensedeal.ai	商汤 sensetime	数字绿土 GreenValley	GEOVIS
MaaS	AI 模型生态	智谱清言	紫东太初2.0	商量 SenseChat	百川智能 BAICHUAN AI	
调度	AI 算力调度	QingCloud Technologies				
适配	AI 算力组件生态	intel.	NVIDIA	AMD	HUAWEI Ascend	中科曙光 Sugon
		Enflame 燧原科技	天数智芯 Iluvatar CoreX	海云科	摩尔线程 MOORE THREADS	

自主可控

青云科技核心代码 100% 自主研发，积极适配国产自主可控生态，包括 CPU、GPU、DPU、NPU、操作系统、国产应用等。青云拥有经过大规模金融行业实践检验的全栈云能力，与 100+ 国产合作伙伴完成产品兼容互认证。这使得金融机构在选择青云科技时能够轻松应对数智化转型及信创技术双重挑战，快速实现业务创新。

深入行业

拥有十余年金融行业云计算平台建设经验，全面支持金融数智化场景创新与信创技术实践。已部署的金融云项目，覆盖金融监管、银行、保险、证券行业的数百家金融机构，包括中国人民银行、中国银行、农业银行、招商银行、光大银行、泰康保险、中金公司、易方达等头部金融机构。

Gartner

全球服务器虚拟化代表厂商

IDC ANALYZE THE FUTURE

IDC 金融行业云原生实践典范案例

金融电子化

中国金融 IT 服务商优秀解决方案

甲子光年 JAZZYEAR

甲子光年 中国 AI 算力层创新企业

信通院 算力互联互通首批成员单位

首批入选《金融信创解决方案（第一批）》

证券基金行业信息技术应用创新联盟

《证券基金行业信息技术应用创新联盟》成员单位

CAICT 中国信通院 EDCC

政企信息技术应用创新促进中心 成员单位

03

金融行业智算解决方案

QINGCLOUD
FINANCIAL INDUSTRY
COMPUTING SOLUTIONS

金融 GPU 算力池化解决方案

概述

GPU 池化解决方案是针对金融行业高性能计算需求而设计的解决方案。该方案通过整合多台 GPU 服务器，构建一个高效共享的 GPU 资源池，通过资源管理和调度系统，实现 GPU 资源的统一管理和动态分配。这不仅降低了金融机构的运维成本和风险，还大幅提升了 GPU 资源的利用率，灵活支持金融行业中各种不同场景的 GPU 计算需求，轻松应对金融机构对于数据处理、RAG 优化和模型推理等 AI 应用的挑战。

痛点

资源利用率低

传统 GPU 部署方式下，GPU 直接绑定到特定的服务器或应用上，导致 GPU 资源利用率低下，特别在业务多样性和波动性强的金融行业中，GPU 资源往往存在闲置和浪费。

管理复杂

金融行业 CPU 和 GPU 资源众多，传统的管理方式需要单独配置和维护每一台服务器，导致管理效率低下。

成本高昂

大量购买、部署和维护 GPU 资源需要高昂的成本，对于金融机构来说是一笔不小的负担。

解决方案介绍

GPU 算力池化解决方案通过集中管理多台同构或异构 GPU 服务器，形成 GPU 资源池。该资源池通过资源管理和调度系统，实现 GPU 资源的统一管理和动态分配。方案主要提供以下几个能力：

异构 GPU 支持

兼容国内外主流的 GPU 产品，如英伟达、海光、昇腾、寒武纪、海飞科、天数智芯等，满足多样化需求。

GPU 资源池化

将多台 GPU 服务器集中部署，形成 GPU 资源池。通过虚拟化技术将物理 GPU 转化成多个 vGPU，实现资源的灵活分配。可分配给多个应用任务使用，支持构建专属资源池和共享资源池。

资源管理调度

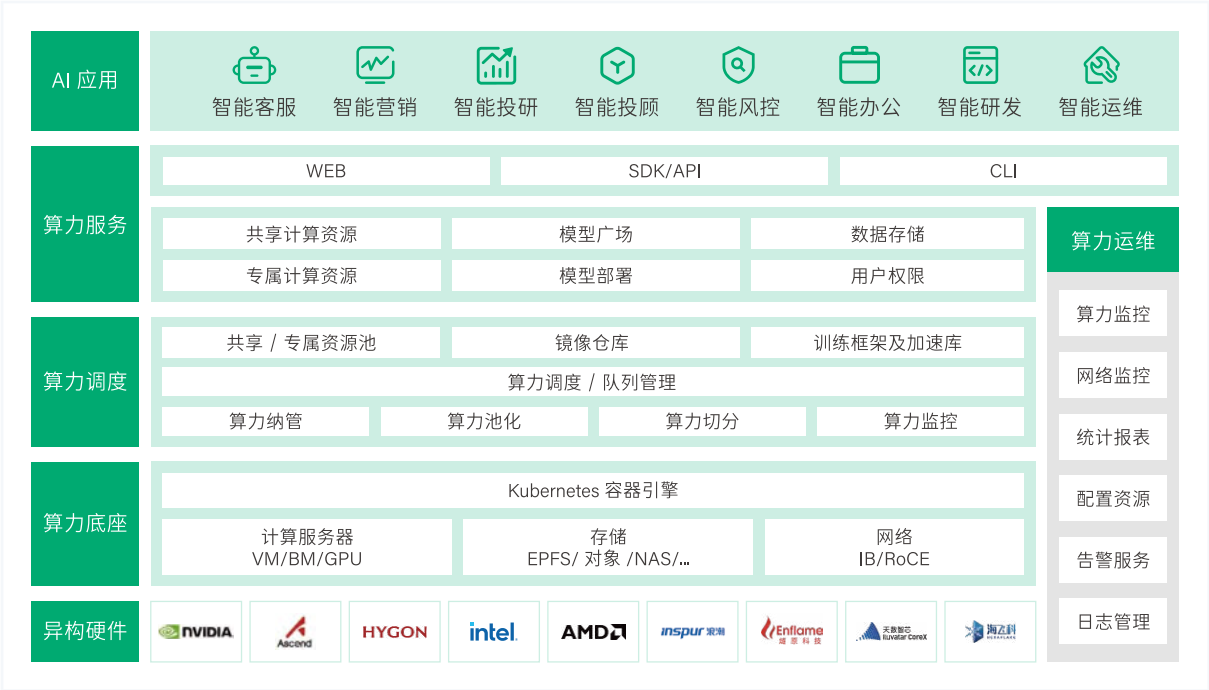
通过先进的资源管理和调度系统，实现 GPU 资源的统一管理和动态分配，包括算力纳管、算力池化、算力切分、算力监控等功能，降低管理复杂性，提高资源利用率。

镜像仓库

提供深度学习常用镜像，在平台进行代码开发、模型训练时可以通过镜像快速构建运行环境。同时还提供自定义镜像管理功能，支持用户根据基础镜像或 Dockerfile 自主开发镜像、管理镜像版本、在线构建。

模型服务

支持用户在模型广场选择模型进行快速部署，提供模型推理服务。模型广场提供常用的开源模型，如 LLama 系列、ChatGLM、Baichuan 等。支持用户快速部署采购自第三方的或者自建的模型镜像，并且对外提供在线推理服务。



适用场景

主要针对 GPU 服务器节点数量较少，以推理服务需求为主的各类 AI 应用场景，如风险评估、交易分析、客户画像等。

解决方案优势

高效资源利用

通过 GPU 资源池化和智能管理调度，实现资源的充分利用，避免资源浪费。

降低成本

支持异构 GPU，优化资源配置和管理维护，降低管理复杂度和总体成本。

提高业务效率

借助镜像仓库和模型广场，快速交付模型服务，提升业务处理效率。

灵活扩展

根据业务需求灵活调整 GPU 资源池规模，满足不断变化的 AI 应用需求。

金融 AI 算力调度解决方案

概述

随着人工智能技术的不断发展，尤其大语言模型等 AI 技术的快速普及，金融行业对于 AI 算力的需求日益增长。为了满足金融行业在营销、客户服务、投研投顾、风控、交易分析等领域对智能应用的高算力需求，青云金融 AI 算力调度解决方案凭借高效的算力调度、对丰富 AI 计算框架的支持，以及涵盖分布式模型训练、在线推理等一站式解决方案的能力，助力金融机构在激烈的市场竞争中脱颖而出，实现智能化转型的飞跃。

痛点

资源交付难

GPU 算力资源的快速交付成为模型预训练、微调和推理的瓶颈。

资源利用率低

如何高效利用 GPU 算力资源支持更多 AI 项目任务，成为挑战。

计算流程复杂

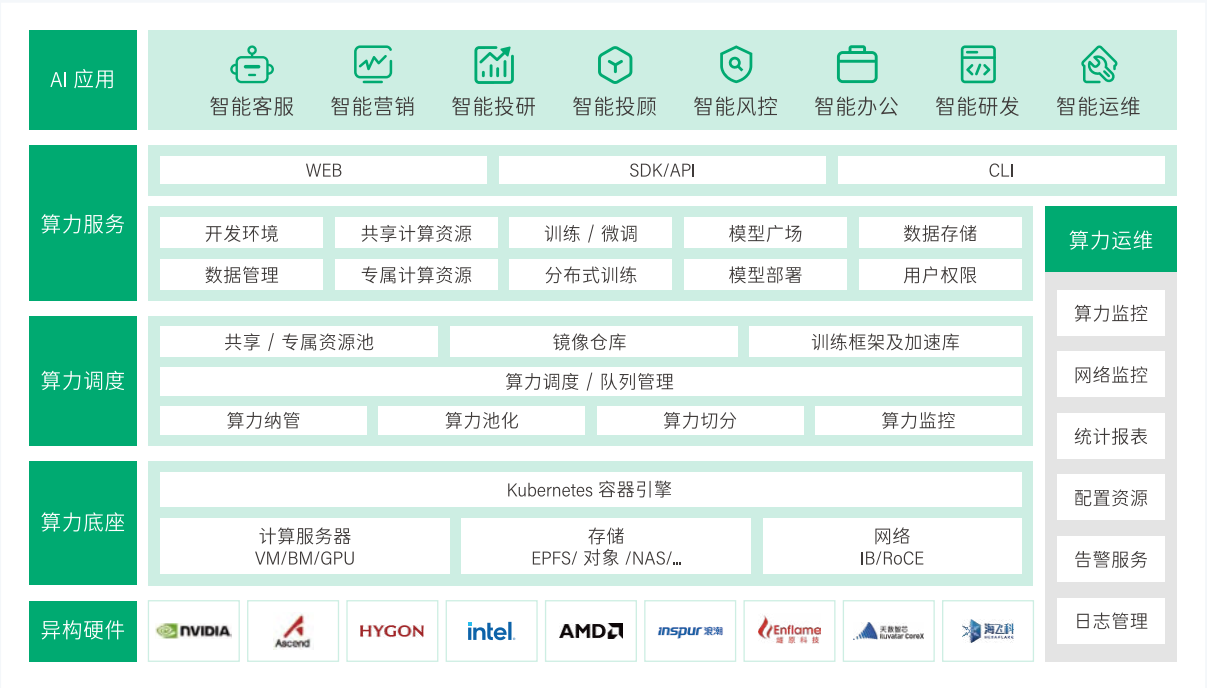
AI 计算流程涉及数据准备、大模型预训练、模型微调和推理，流程周期长，技术门槛高，故障处理难度大。

性能瓶颈

AI 算力中心对网络、存储、服务器乃至机柜布局均有严格要求，整体系统复杂，易产生性能瓶颈。

解决方案介绍

青云金融 AI 算力调度解决方案在 GPU 算力池的基础上，强化了数据管理能力与 AI 开发环境能力。方案不仅提供分布式训练能力，支持模型预训练和模型微调任务，还针对部分金融机构的 AI 算力平台建设经验不足，提供一站式 AI 算力平台规划与设计服务，通过精准计算资源规划，确保算力、网络、存储与机房条件达到最佳平衡，从而消除潜在的性能瓶颈。



适用场景

主要适用于企业级 AI 算力平台的建设，满足金融机构内部各类 AI 应用场景的实践需求，支持从模型预训练、模型微调到模型推理的全流程服务，为金融机构提供强大的 AI 算力支持。

解决方案优势

高效算力调度

通过服务化方式快速响应并高效调度算力资源，满足金融行业对高性能计算的需求。

全栈集成技术

采用先进的网络架构设计与全栈集成技术，确保系统高可靠性与高可用性，降低系统复杂性及维护成本。

灵活扩展

具有良好的可扩展性，可根据金融机构业务需求而快速扩展与升级，满足业务发展的动态需求。

一站式服务体验

提供从数据准备、模型训练、模型微调到模型推理及模型服务等 AI 全流程一站式服务，让金融机构轻松享受 AI 带来的便捷和高效。

金融智算行业云解决方案

概述

随着金融行业的数字化、智能化转型，AI 技术已成为推动其发展的关键力量。然而，AI 算力建设的高昂成本与技术门槛，让许多中小金融机构望而却步。青云金融智算行业云解决方案应运而生，旨在降低算力成本，打破技术壁垒，让中小金融机构也能轻松享受 AI 技术带来的红利。该方案通过共享 AI 算力资源，实现科技普惠，助力金融机构在数智化转型的道路上迈出坚实步伐。

痛点

成本效益考量

大型机构在为中小金融机构提供金融智算行业云服务时，如何在保证服务质量的同时，合理控制成本，提供性价比高的解决方案，是大型机构面临的重要挑战。

金融合规难题

随着监管趋严，大型机构需确保金融智算行业云服务既技术领先又完全符合国内外金融法规，以保障客户数据安全，规避合规风险。

运维运营能力弱

需要实现算力资源的云服务化管理，并提供多租户管理、计量计费管理等运营能力，否则难以正常开展智算行业云服务业务。

解决方案介绍

金融智算行业云解决方案在提供算力调度的同时，着重增强了算力运营能力，包括：卡时计费、实例计费、客户折扣、优惠券等多种计费方式，以及规模定义、消费统计、发票开具等运营能力，帮助运营方快速为其租户提供专业、便捷的 AI 算力云服务。

适用场景

主要面向大型金融机构或金融关键行业节点机构，打造符合金融合规要求的金融智算行业云，为中小金融机构提供功能全面、价格合理、交付便捷的智算行业云服务，助力其实现数智化转型。

解决方案优势

成熟的云服务运营体系

凭借青云多年公有云和行业云的丰富经验，提供功能完备的行业云运营体系能力，确保项目顺利实施并达成设计要求。

专业咨询服务

拥有大型公有云、行业云以及 AI 算力平台建设经验，能够为客户提供智算行业云规划设计及建设服务，确保项目高质量完成。

稳定可靠的运营支持

提供稳定可靠的运营平台与专业的运营团队，确保智算中心的长期稳定运行，并为客户提供持续的技术支持与服务保障。

金融 AI 应用解决方案

金融智能文档

金融智能文档运用 AI 技术，实现金融文档的自动化处理、风险智能控制、数据深度洞察和知识智能管理，助力金融机构提升业务效率、决策精度与客户满意度，同时确保数据安全。

痛点

效率低谷

人工处理海量合同、报告、交易记录等文档，耗时长、易出错，严重影响工作效率。

风控难题

金融机构准确性和合规性要求高，任何文档处理失误都可能引发法律与财务风险。

洞察盲区

海量文档中关键信息的提取与分析困难，影响精准决策。

知识孤岛

金融机构知识资源于信息分散，难以集中管理及有效共享，阻碍了工作效率与决策质量的提升。

能力

文档自动化	高效处理合同、报告等，减少人工干预。
风险智能控制	确保合规，降低法律风险。
信息提取分析	快速准确提取关键信息，支持决策。
知识智能管理	构建智能知识库，提升团队协作效率。
用户服务友好	提供便捷界面与高效服务，增强市场竞争力。
数据安全保护	确保数据安全与用户隐私安全。

优势

提升效率 降低成本	强化风控 确保合规	深度洞察 辅助决策	提升客户体验 优化知识管理

应用

合同智能审查与管理　信贷资料智能审核　投资研究报告生成

金融编码助手

金融编码助手以先进的 AI 技术为核心，针对金融行业软件开发中的效率、质量与检索痛点，提供一体化的解决方案。通过自动化代码补写、代码质量优化与提升检索效率等功能，助力企业开发者显著提升开发效能，确保软件项目的高质量交付。

痛点

风险管理与合规性

复杂的金融产品与严格的监管要求导致合规风险增加。

数据安全

敏感的客户数据和交易信息需要严格保护，避免泄露。

业务敏捷性

市场变化迅速，要求金融机构快速响应并推出新产品。

代码质量困境

在追求提效的过程中，测试、调试、注释等关键活动被忽视，导致代码质量下降。

能力

自动代码补写

毫秒级生成速度，无缝融入开发流程。

代码解释与翻译

快速解析历史代码逻辑，提升团队协作效率。

自动生成测试用例与注释

确保代码健壮，提升代码可读性。

自动纠错与调优

实时检测代码缺陷，提出改进建议。

RAG 代码仓库与私有函数

自动识别并调用私有接口与函数，降低错误风险，确保代码准确性。

统一交互界面

集成多样化功能，避免频繁切换工具。

优势

- 提高风险管理和合规性
- 增强数据安全与隐私保护
- 提升业务敏捷性与响应速度
- 提高代码质量并降低维护成本

企业知识库

企业知识库，以先进的向量数据库与大模型技术，构建一站式知识管理与决策支持平台。无缝整合数据资源，实现高效存储、精准检索与智能问答，消除信息孤岛，降本增效，助力企业在海量知识中洞察先机、决策制胜。

痛点

知识孤岛

企业内部知识分散存储，缺乏统一管理和有效整合，导致信息难以共享与协同利用。

检索低效

面对海量数据，传统的关键词搜索方法往往无法精准定位所需信息，导致知识查找耗时长、准确度低。

知识更新滞后

知识库的维护与更新不及时，无法实时反映最新业务动态与最佳实践，影响决策时效性。

人工处理负担重

知识整理、分类、解析等工作主要依赖人工，耗费大量时间和人力成本。

能力

向量化存储	海量数据高效压缩，一键构建向量索引，确保信息处理的速度和准确性，存储效率提高 90%。
多路召回	先进的匹配算法与召回框架，快速找到最精准的数据和信息，命中效率 95%+。
检索增强生成	深度分析洞察文本，精准回答各类问题并提供建议，沟通分析成本降低 60%。
自定义助手	自动拆解需求，快速生成符合业务的个人助手，0 代码拥有个性化业务助理。

优势

- 解锁知识价值
- 提升工作效率
- 优化知识生态
- 引领创新潮流

应用

规则制度查询 \ 员工助手 \ 知识库管理 \ AI 仿写

QingCloud Technologies

04 案例实践

CASE
STUDIES

QINGCLOUD
FINANCIAL INDUSTRY
COMPUTING SOLUTIONS

中国农业银行

背景与挑战

随着金融科技的不断发展，中国农业银行致力于构建智慧银行和数字生态银行，特别是在“三农”普惠领域，力求通过科技手段提升服务质量和效率。在智慧化转型过程中，中国农业银行面临以下挑战。

传统运营模式挑战

传统网点运营管理成本高，依赖“人盯人”方式，效率低下。

客户体验提升需求

客户粘度不足，潜客数据挖掘不足，需要提升客户体验。

大规模节点管理挑战

遍布全国的网点涉及大规模异构边缘节点，需要实现高效管理与监控。

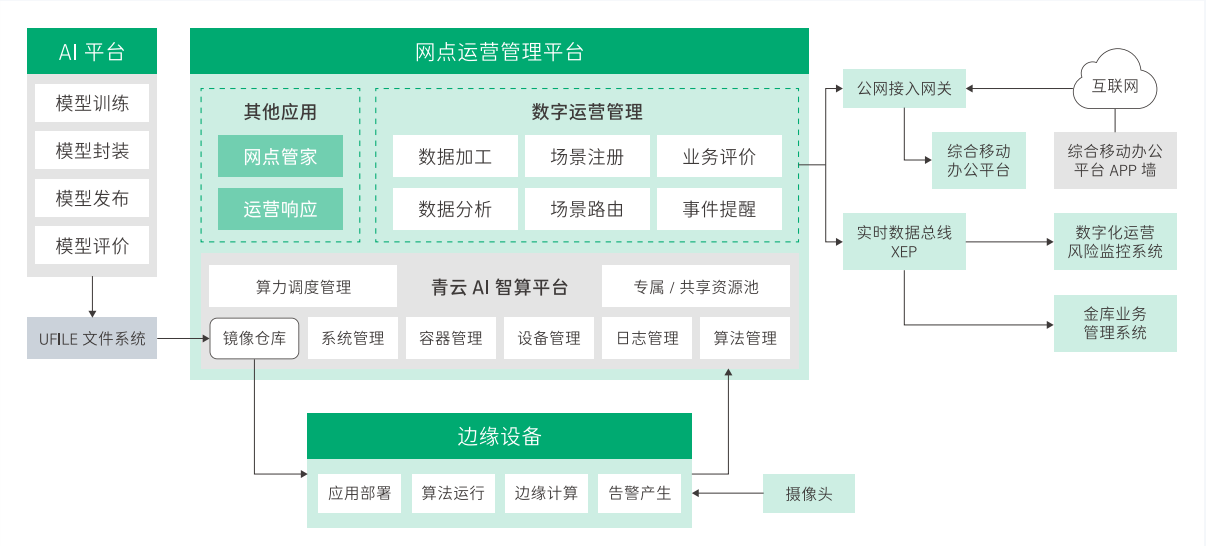
组织架构复杂性

金融行业组织架构复杂，需要在各分行 / 支行进行权限的精细化管理。

为应对这些挑战，中国农业银行决定引入青云作为技术合作伙伴，以提升客户体验，优化运营效率。

案例方案

中国农业银行采用青云 AI 智算平台，利用云原生调度架构结合边缘计算框架，构建总分一体化云边协同体系。该方案以分支机构为组织单元，以边缘节点为单位，实现统一管理、统一调度、统一监控，从而达成云边协同，并显著提升边缘计算效率。



稳定性架构构建

- 总行中心云上的边缘计算平台采用主备管理集群模式，确保灾备能力。
- 边缘计算平台核心服务分布式部署，支持横向扩容，提高服务承载能力。
- 边端与中心云对接采用负载均衡方式，避免单点故障和性能瓶颈。
- 分行边缘节点发生故障时，中心云调度策略可及时调整，确保业务连续性。

大规模异构边缘节点自动化纳管

- 实现异构边缘节点的自动化纳管，提高管理效率。
- 构建边缘节点镜像缓存能力，降低网络带宽占用。

分布式可观测性体系

- 采集分析各层面监控指标，包括分行边缘节点、边缘应用及云端平台数据。
- 实现可视化支持，设置阈值策略，提前发现故障点。

安全隔离能力

- 总行统一管控，各子机构资源逻辑隔离。
- 各子机构权限隔离，项目资源隔离，应用资源隔离。

客户收益

为全国 430 余家网点和 40+ 业务场景提供边缘算力资源，覆盖超柜代客操作、三方驻场监测、数字孪生、智慧畜牧等不同业务场景，为中国农业银行智慧化转型提供有力支撑。

提升客户体验

加快业务处理速度，提供个性化服务。

优化运营效率

降低运营成本，提高故障处理速度。

增强数据挖掘能力

精准挖掘客户需求，提升营销效率。

提升安全性

通过安全隔离和资源隔离，降低安全风险。

广发证券

背景与挑战

在数智化转型的浪潮中，广发证券积极响应，努力探索并应用新技术来提升业务处理效率和客户体验。尤其在人工智能领域，广发证券面临着算力资源不足、利用率低、管理复杂等挑战。为了解决这些问题，广发证券决定引入青云 GPU 资源池化解决方案，以实现高性能、高效率、高可用的 AI 算力支持。

案例方案

广发证券采用的青云 AI 智算平台实现 GPU 资源池化，具备一池多芯、异构算力资源池的特点，能够支持多种主流 AI 框架和硬件加速器。

一池多芯

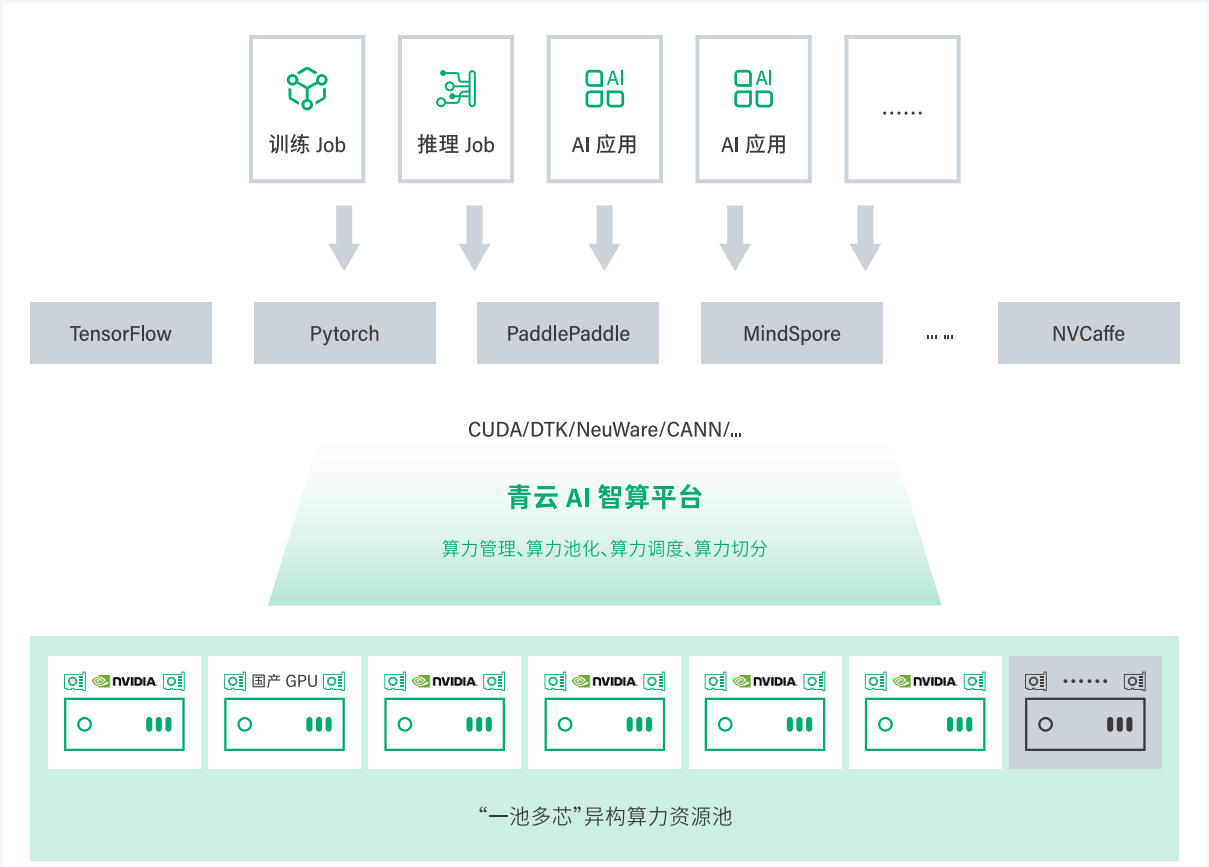
支持一池多芯的异构算力资源池管理，能够同时管理多种不同型号的 GPU 加速器。这种设计能够充分利用各种 GPU 加速器的性能优势，提高整体算力效率。同时，平台还支持国产算力与英伟达算力的异构池化管理，实现国产化的平稳替代。

算力管理

能够实现算力资源的动态分配和调度，还支持算力优化和切分功能，可以根据业务特点进行精细化管理与优化。

支持多种 AI 框架

如 TensorFlow、PyTorch、PaddlePaddle 和 MindSpore 等，满足不同业务场景的需求。



客户收益

广发证券取得了以下显著的成效，这为银行数智化转型提供了有力支撑，也为未来的业务发展奠定了坚实基础。未来，广发证券将继续深化对 AI 技术的应用和探索，不断提升业务处理效率与客户体验。

01

算力利用率大幅提升

平台通过算力调度和优化算法，实现了算力资源的精细化管理和优化，大幅提升了算力利用率，降低了成本，提高了业务处理效率。

02

业务响应速度加快

平台提供高性能、可扩展的 AI 算力资源，能够满足不同业务场景的需求，确保业务的高效运行。

03

安全性增强

强大的安全防护能力，能够保护业务数据和用户隐私。同时，平台还支持多租户隔离和权限管理功能，确保不同租户之间的数据安全和隐私保护。

广西北部湾银行

背景与挑战

在金融行业积极拥抱数智化转型中，广西北部湾银行作为地区领先的金融机构，积极拥抱人工智能（AI）技术，推动业务创新和服务升级。为了推进广西北部湾大模型项目的进一步研究与实际应用，银行在已部署多种计算资源的前提下，亟需建设一个能够对异构算力进行调度、模型训练开发，推理一体化平台。

单卡 GPU 算力利用率低

由于缺乏有效的资源调度策略，单张 GPU 的算力往往无法得到充分利用，导致资源浪费。

资源动态分配能力不足

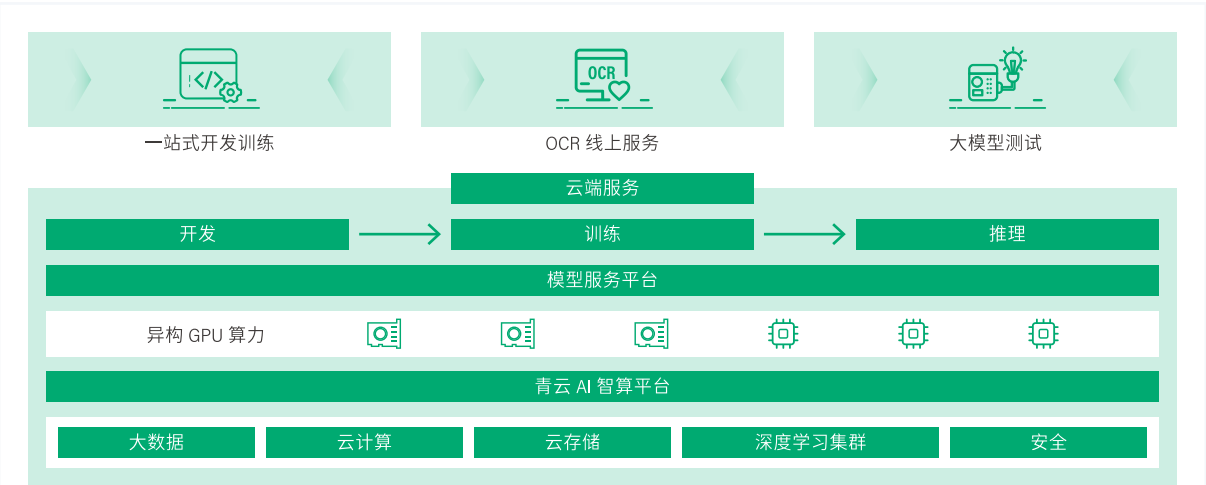
缺乏动态分配、多机聚合、动态挂载 / 释放等能力，使得资源无法根据业务需求进行灵活调整。

开源方案局限性

虽然开源方案在一定程度上能够实现部分功能，但其稳定性和隔离性欠佳，管理精细度不足，无法满足广西北部湾大模型项目的需求。

案例方案

携手青云科技，广西北部湾银行成功打造了一款为金融行业量身定制的 AI 模型训练平台。该平台将客户现有算力资源进行集成，通过构建统一调度的管理资源池化层，实现了 GPU 资源的统一调度、灵活分配、弹性伸缩等云化能力。这一创新解决方案不仅满足了不同时段、不同业务对资源的需求，更为上层全栈云平台提供了稳定、高效的 GPU 算力资源。



客户收益

01 GPU 共享能力提升

平台在云原生环境下实现了 GPU 的共享能力，使得多个应用能够同时访问和使用 GPU 资源，提高了资源的利用率。

02 资源使用灵活性增强

通过统一调度和灵活分配，银行能够根据不同业务需求，快速调整资源配比，实现了资源的最大化利用。

03 无缝对接现有应用

平台兼容当前 AI 应用代码和使用习惯，无需对现有应用进行大规模改造，即可享受平台带来的便利。

04 国产芯片支持

平台不仅支持虚拟 GPU 和物理 GPU 的调度与共享使用，还实现了对国产芯片及硬件的统一调度和管理，展示了强大的兼容性和前瞻性。

05 运维效率提升

平台通过自动化、智能化的管理手段降低了运维复杂度，提高了运维效率和工作质量，进一步释放了人力资源。

06 开源方案技术短板弥补

相比于开源方案，平台在池化能力、国产芯片支持、安全隔离性等方面具有显著优势，为广西北部湾银行的大模型项目提供了强有力的技术支撑。

QINGCLOUD

持续创新，开放共赢，与客户共筑数字世界新未来

QingCloud Technologies

Tel : 400-8576-886 E-mail : contactus@yunify.com

印刷时间：2024 年 06 月