

A child wearing a brown aviator helmet and goggles sits on the shoulder of a large, white, humanoid robot. The robot's head is turned towards the child, and its right hand is raised, pointing towards the child's helmet. The background features a stylized world map with a grid pattern, set against a light blue sky with streaks of light. The overall scene conveys a sense of global connectivity and technological advancement.

DCS AI 解决方案

加速AI行业化落地，推动千行万业迈向智慧化



目录

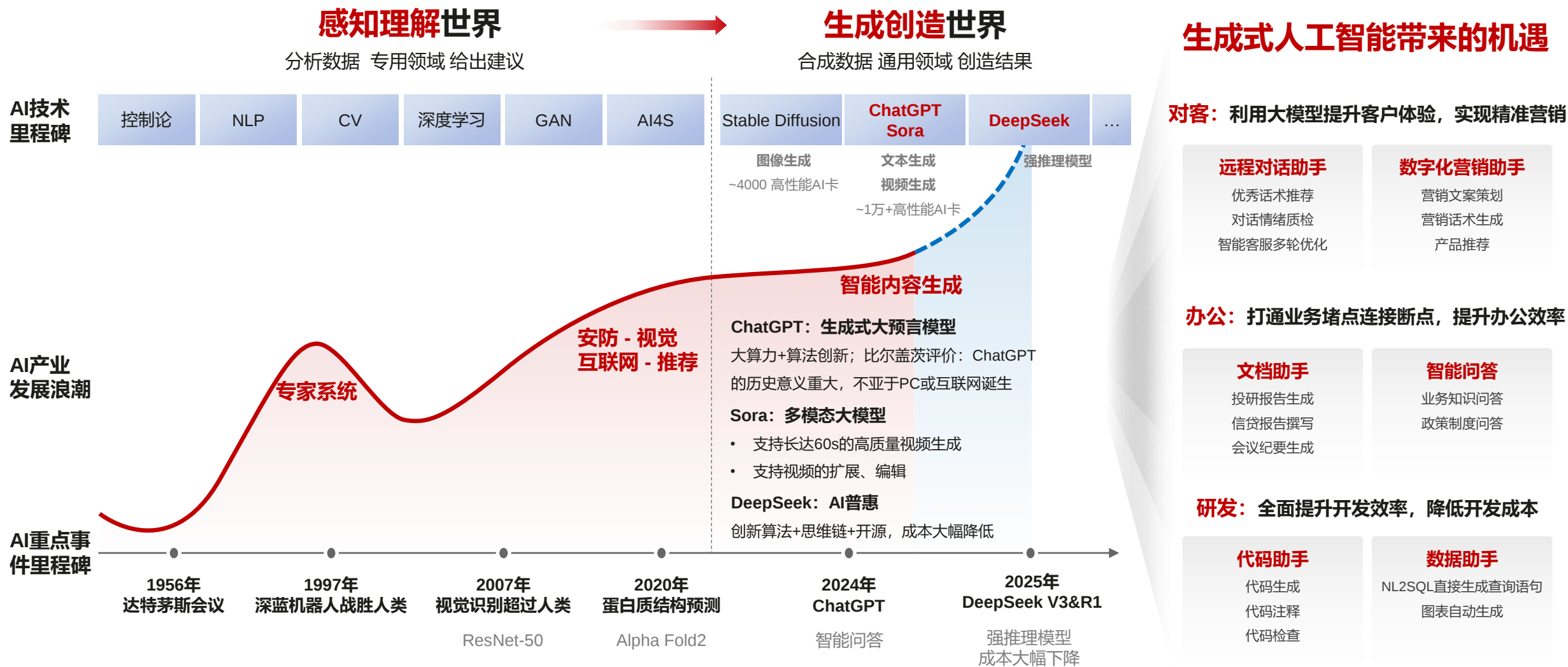


01 AI大模型发展趋势和挑战

02 DCS AI解决方案

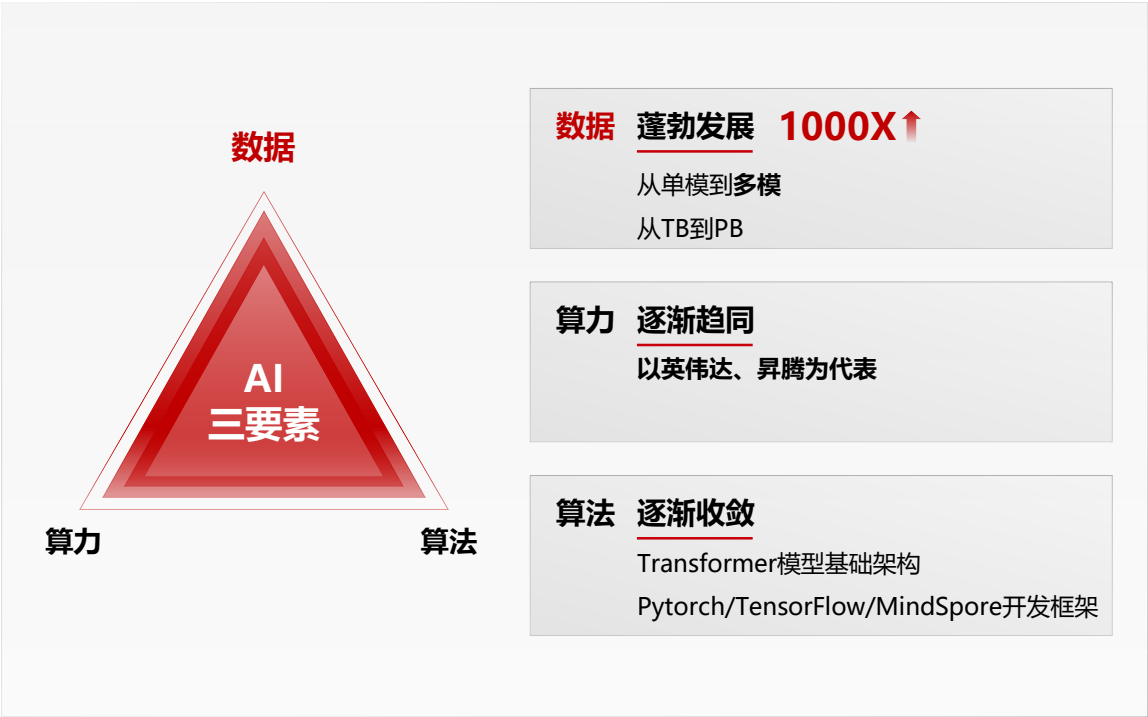
03 典型案例

AI趋势：从感知到内容生成，从简单预测到复杂逻辑推理，AI普惠化带来新的机遇

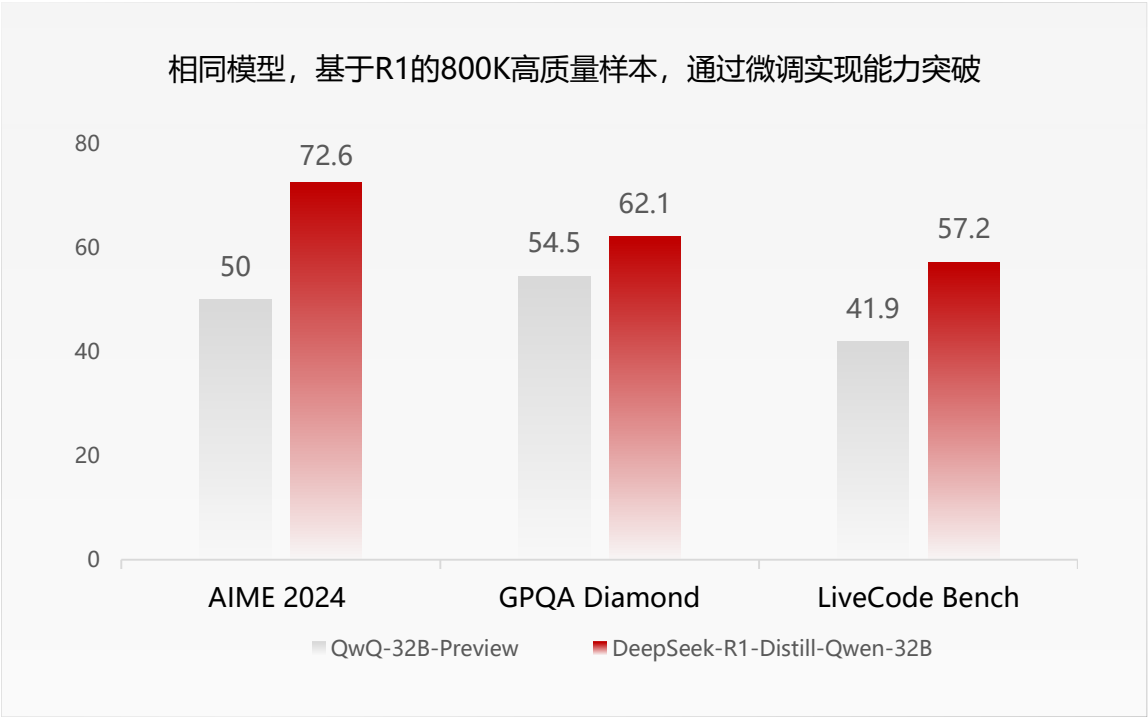


影响大模型效果的核心要素

数据是AI大模型的核心要素



AI高度取决于数据规模和质量



* <https://huggingface.co/Qwen/QwQ-32B-Preview>

训练&推理是大模型在行业落地必须经历的过程

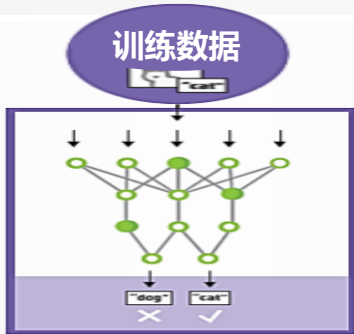
大模型训练及推理基本概念

模型训练 研发（开发态） 从现有的数据中学习新能力，形成算法（模型）

未训练的
神经网络模型



训练框架
(开发环境/工具)



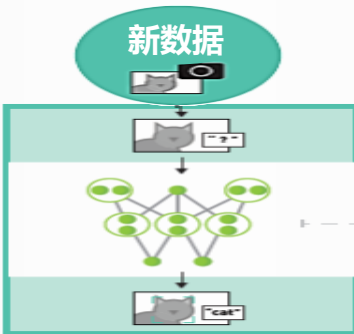
训练好的模型算法
(具有新的能力)



高算力
长时间
反复迭代

模型推理 生产（部署态） 把模型部署到应用或服务中，完成具体任务

训练好的模型算法
(具有新的能力)



训练好的模型算法

具备推理能力的
应用或服务

低算力
短时间
单次计算

模型训练



预训练
(义务教育)

微调/知识库
(专业课)

模型推理



推理/预测
(行业工作)

大模型在行业落地面临三大挑战

数据工程耗时长



数据工程占模型开发时长60%

数据孤岛、格式多样、资产繁多，
语料收集、清洗、标注等耗时长

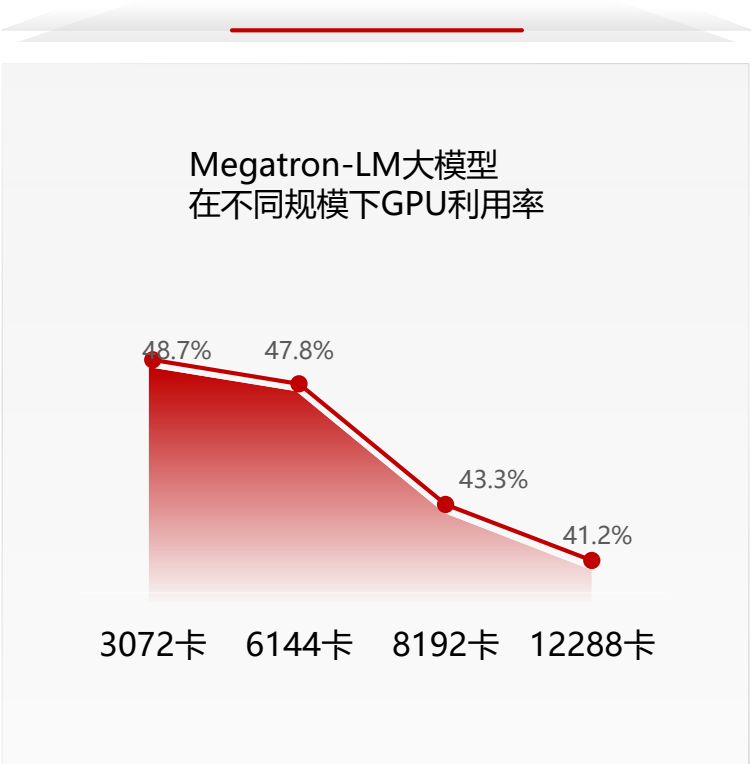
模型训练和应用落地难



开发周期不可控

项目开发难度大，人员技术要求高

集群可用度低



集群可用度不足50%

因算力等待、任务潮汐、资源碎片化等原因

目录



01 AI大模型发展趋势和挑战

02 DCS AI解决方案

03 典型案例

DCS AI提供训推集群、训推一体机解决方案，满足不同场景用户AI使用需求



面向企业训推场景（百卡以内）

DCS AI训推超融合一体机

单节点起步，快速上线

- 灵活轻量：单节点起步，一柜式全集成，提供计算存储网络
- 统一管理：全栈管理，硬件+软件平台 4小时开局
- 工具链：集成数据处理套件、模型微调套件、应用开发套件，加速大模型场景应用快速落地

面向智算中心场景（百卡以上）

DCS AI训推集群方案

一站式预集成、高性能训推集群

- 高效算力调度：统一纳管异构算力资源，提高算力利用率50%
- 超宽无损网络：200G RoCE网络，负载均衡有效带宽98%
- 高性能存储：百万IOPS，百G大带宽，智能分级存储
- 一站式AI开发工具链，高效AI训推开发、部署、算法迁移

企业训推场景：DCS AI训推超融合一体机助力大模型快速行业落地

客户挑战

设备部署运维难

硬件“七国八制”部署繁杂，设备运维、故障定界难

数据工程耗时长

数据格式复杂，清洗难度大；海量数据集标注，耗时耗人

模型训练和应用落地难

大量AI组件，手动部署适配繁杂；应用开发周期长

AI集群可用度低

资源碎片，造成资源浪费；集群业务存在明显的潮汐现象

目标架构

典型场景

科研、高端企业服务
复杂逻辑推理

问答问数场景
成本与性能平衡

智能办公场景
低成本和低时延



FusionCube A3000

模型层

DeepSeek | ChatGLM | LLaMA | Qwen | ...

AI使能平台

AI全流程工具链 | ModelEngine

数据使能



应用使能



模型使能



资源平台

裸金属容器平台 | eContainer

基础软硬件

AI开发框架

昇腾算力 OceanStor存储 CE交换机

运维管理 DME

资源监控

告警日志

性能监控

大屏报表

拓扑分析

健康检查

DCS AI训推超融合一体机

方案价值

部署运维易

- 一键式安装部署，业务上线从天->小时级
- 存算网全集成，5合1全融合管理

数据处理强

- 数据清洗效率提升60%
- 问答对自动生成，留用率可达60%

应用上线快

- 3步完成DeepSeek模型部署，快速拉起
- 应用上线从天到小时

系统效能高

- 系统任务并发提升30%
- 基于负载的资源弹性伸缩

部署运维易：一键式安装部署，5合1融合管理，降低部署和运维门槛

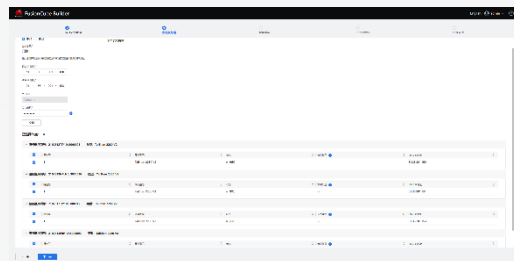
SmartKit 整柜开局

DME统一全栈管理

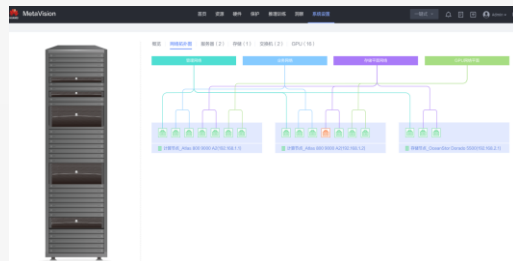
安装/部署/初始化



资源管理 + 设备管理 + 可视监控 + 系统管理 +



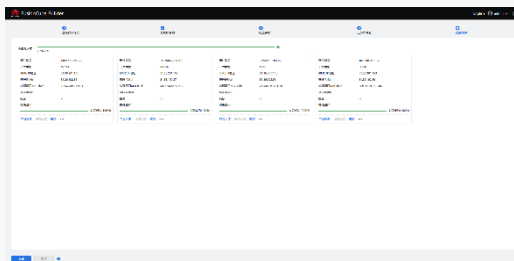
设备自动发现



硬件组网可视



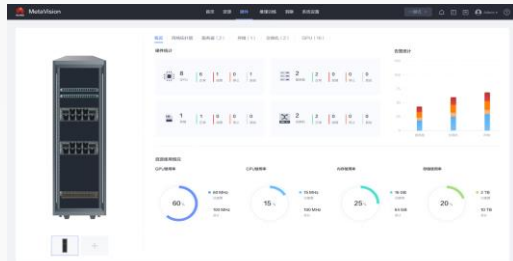
资源一眼可视



向导式安装



集群统一管理



设备可视监控&故障一眼定界

提供一站式部署工具

SmartKit

一键式安装部署，AI业务快速上线

2周

85%

2天

提供统一管理平台

DME

5合1全融合管理



数据处理强：数据清洗效率提升60%，问答对留存率达60%

ModelEngine 数据使能



统一数据归集

- 支持NFS v3、OBS协议和第三方存储
- Dorado与Pacific异构存储数据可归集



清洗效率提升60%

- 内置50+算子，多维数据处理
- 1000+小文件处理时长1小时



留存率可达60%

- 原始语料自动生成QA对，问答对留用率可达60%，领先业界50%



评估精度80%以上

- 多维自动化评估：完整性，一致性，有效性等

应用上线快：三步完成模型部署，应用上线天级->小时级



方案价值

三步完成模型部署 (以DeepSeek为例)

- Step1 DeepSeek社区原始权重下载
- Step2 上传权重至ModelEngine
- Step3 下发推理服务部署任务

应用上线天级->小时级



模型部署复杂

100+ AI组件, 部署流程复杂, 人工适配, 效率低

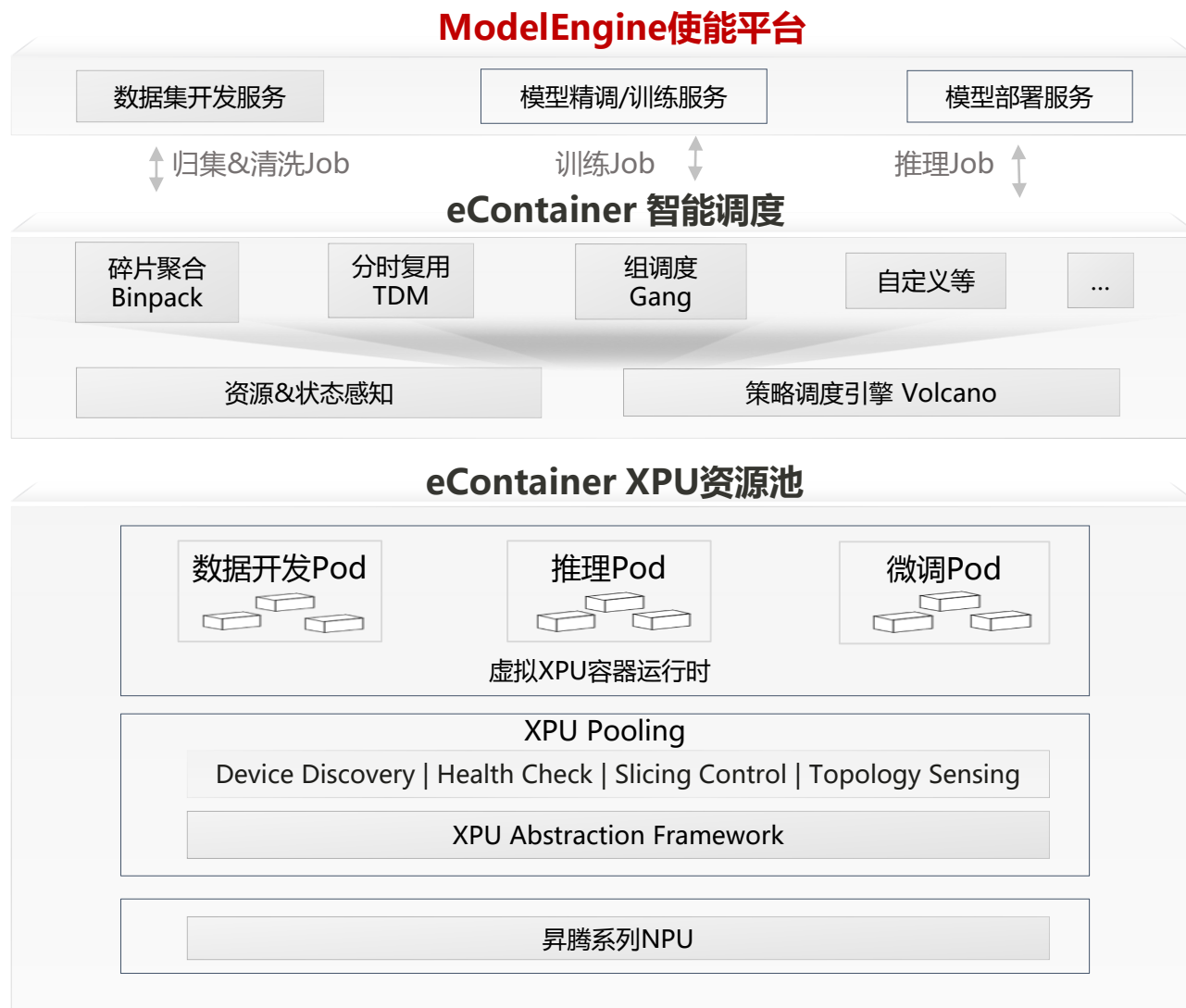
技术门槛高

需要跨领域的技能 (数据, 算法等), 技术门槛高

业务上线慢

需解决数据格式、通信协议等兼容性问题, 工作量大

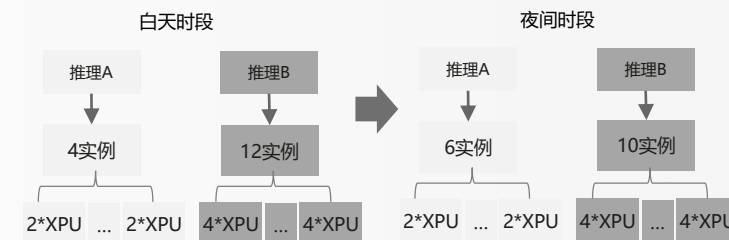
系统效能高：资源按需伸缩，系统任务并发提升30%



资源需求
随需伸缩

系统并发
提升30%

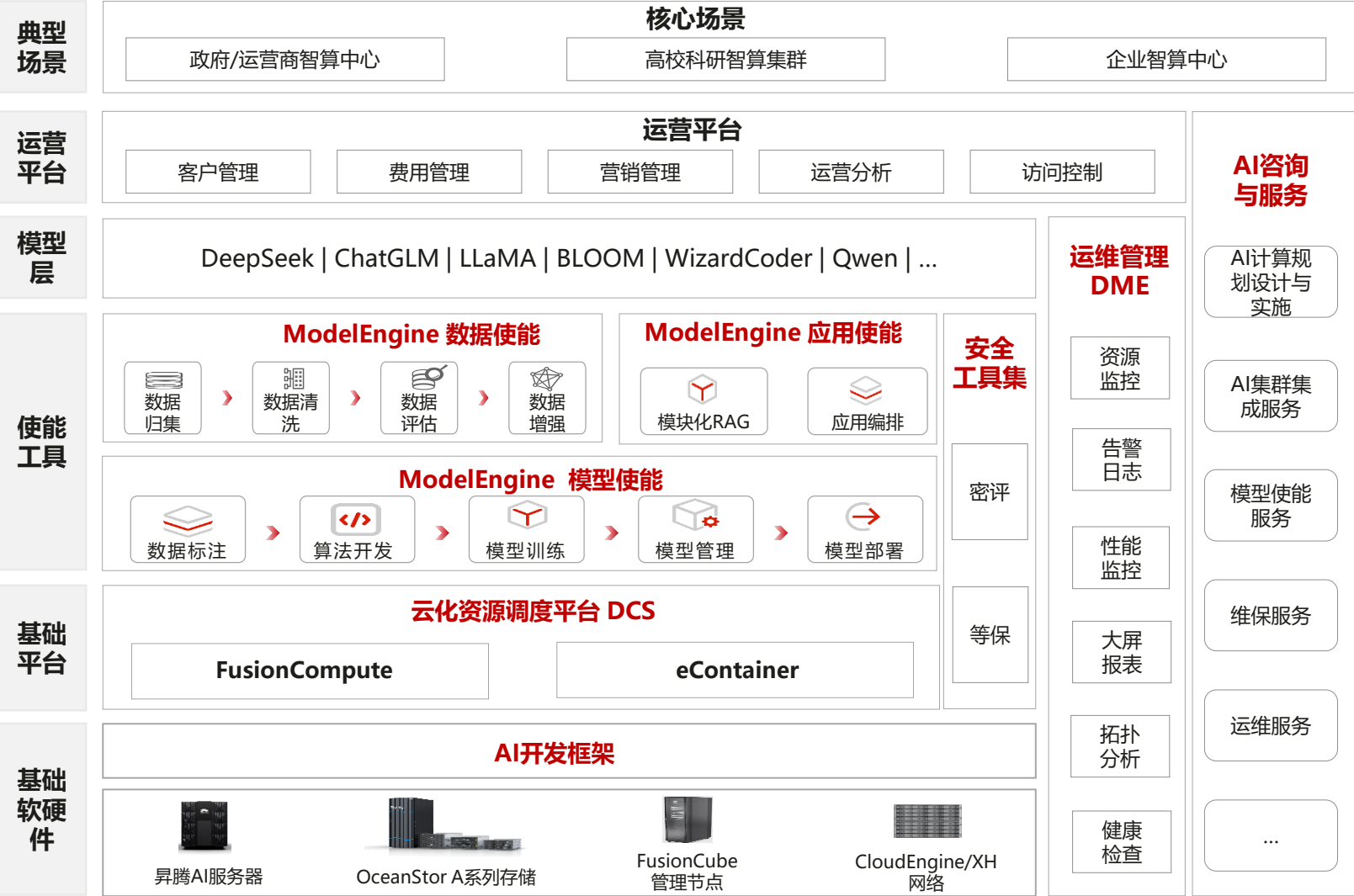
业务错峰错谷，基于任务负载的实例数弹性伸缩



多维智能调度，系统任务并发提升30%



智算中心场景：DCS AI一站式方案助力模型高效训推开发



轻量起步，平滑扩展

算力**8卡**起步、管理**4节点**起步，**乐高式服务**叠加扩展



极简易用，全栈使能

一站式全流程AI工具链，支持数据使能，模型使能，应用使能，**零代码**开发训练,应用落地时间缩短**3倍**



存算协同，极致性能

软硬协调优化，集成DataTurbo 文件加速技术，KVCache缓存推理加速，**性能提升30%**



智能调度，高利用率

XPU池化与高效调度，构筑业界最小**1:20**算力切分粒度

AI专业服务：从建设到运维，助力客户建好用好算力平台，释放算力价值



目录

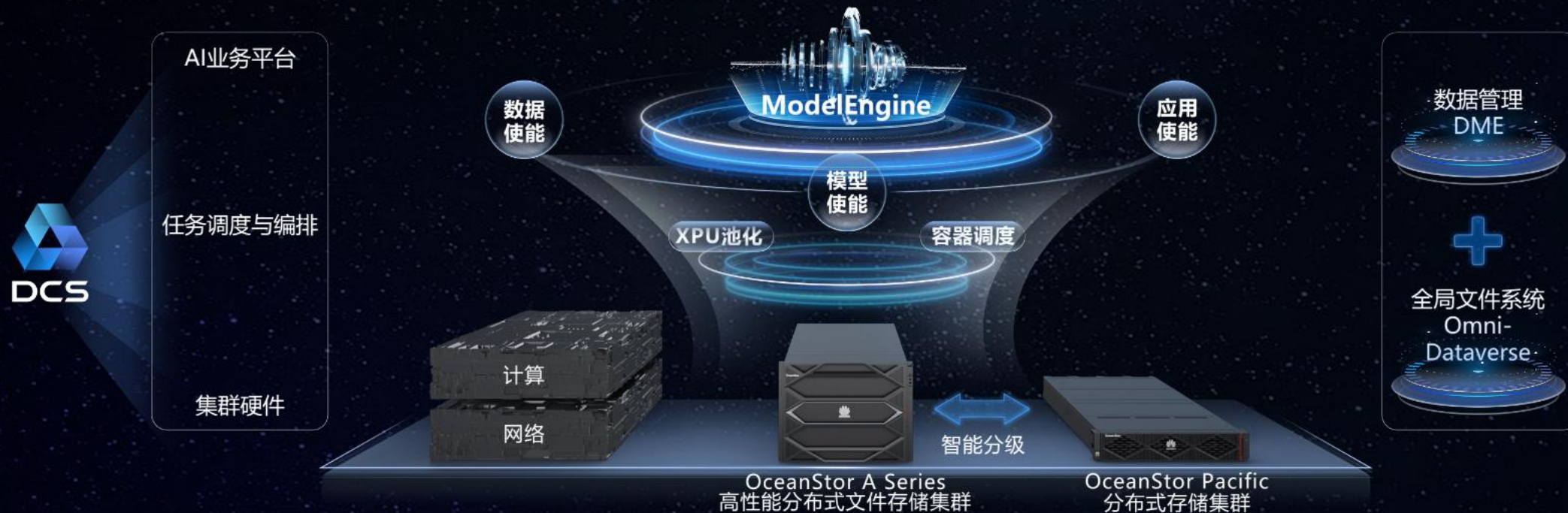


01 AI大模型发展趋势和挑战

02 DCS AI解决方案

03 典型案例

华为DCS AI解决方案



Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home and
organization for a fully connected,
intelligent world.

**Copyright©2018 Huawei Technologies Co., Ltd.
All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

