

华为技术有限公司  
深圳龙岗区坂田华为基地  
电话: +86 755 28780808  
邮编: 518129  
www.huawei.com



加速AI行业化落地

# DCS AI 解决方案



#### 商标声明

是华为技术有限公司商标或者注册商标。  
在本手册中以及本手册描述的产品中, 出现的其他商标, 产品名称, 服务名称以及公司名称, 由其各自的所有人拥有。

#### 免责声明

本文档可能含有预测信息, 包括但不限于有关未来的财务、运营、产品系列、新技术等信息。由于实践中存在很多不确定因素, 可能导致实际结果与预测信息有很大的差别。因此, 本文档信息仅供参考, 不构成任何要约或承诺, 华为不对您在本文档基础上做出任何行为承担责任。华为可能不经过通知修改上述信息, 恕不另行通知。

版权所有©华为技术有限公司2025。保留一切权利。

非经华为技术有限公司书面同意, 任何单位和个人不得擅自摘抄、复制本手册内容的部分或全部, 并不得以任何形式传播。

# 01 / DCS AI解决方案架构概述

DCS AI解决方案通过整合AI软硬件产品并进行系统级优化与易用性提升，以提供领先的AI全栈方案。主要包含ModelEngine AI全流程工具链、eContainer AI容器平台、DME智能资源管理平台等软件，组合算存网等AI硬件，为行业客户提供AI开发全流程所需的数据工程、模型工程、应用编排能力，加速AI行业化落地。



# 02 / DCS AI解决方案组件介绍



## 软件组件

### 数据工程

数据处理组件，提供50+数据处理算子，实现多模态数据的归集、清洗、标注、向量化等，缩短大模型数据准备周期80%（开源组件）

### 模型工程

模型管理与训推组件，提供了模型的管理、开发、训练、推理等模型全生命周期管理能力

### eContainer容器

AI容器平台，实现xPU算力资源切分、跨节点聚合与训推任务调度，提升xPU资源可用度30%

### 应用编排

AI应用编排组件，可视化开发、0代码编排，降低RAG应用、AI智能体的开发门槛，缩短应用上线周期80%（开源组件）

### FusionCompute虚拟化

虚拟化平台，将服务器、存储和网络资源虚拟化成资源池，以承载AI方案管理面与更多应用

### DME智能管理

AI软硬件全栈管理平台，南向统一接入、北向API开放，端到端可视化管理，智能运维



## 硬件组件

- ✓ 存储：全闪高性能AI存储，KV Cache推理加速、NPU直通
- ✓ 服务器：高密AI算力、极致能效比、高速网络带宽
- ✓ 交换机：高性能无损RoCE，高密度接入

## 03 / DCS AI解决方案技术优势



### 轻量灵活 快速上线

- 算力8卡起配、管理3节点轻量起配，同等算力规模配置更轻量
- 全方位集成，100+相关组件集成交付，AI平台快速上线、交钥匙工程
- AI应用可视化编排、0门槛开发，应用上线时间缩减80%
- 多租户、数据管理、安全等能力按需积木式叠加部署

### 高效存储 快捷供数

- AI存储推理加速，KV Cache多级缓存，长序列推理并发提升2倍，首Token时延降低90%，吞吐提升1倍，NPU直通存储缩短模型训练时长
- AI数据湖，全局可视可管，数据免拷贝共享互通，千亿级文件秒级检索
- 大规模专家并行集群推理（大EP），单卡推理性能3倍提升
- 数据工程支持多种格式数据源，模型训练准备周期缩短80%

### 生态开放 极简易用

- 兼容业内主流大模型，多模态服务满足用户模型多样化需求
- 模型API网关实现模型服务开放，助力应用快速对接
- 一站式全流程AI工具链，非AI工程师无门槛上手
- 软硬全栈统一管理，可视化运维，问题定界定位时间缩短至分钟级

## 04 / DCS AI解决方案参数规格

分类	规格描述	规格数量
整体规格	AI集群规模	2048卡
	AI租户个数	128
	智算节点规模	256
	租户资源池规模	128节点
	资源池数量	128
服务规格	网络规格	样本面支持25G/100G/200G RoCE无损网络
	任务调度策略	1. 装箱(Binpack)和分散(Spread)调度策略 2. 分时调度策略，支持定时任务 3. 优先级调度与抢占 4. 推理实例弹性伸缩
	模型兼容范围	DeepSeek R1 671B、Qwen3 235B、Kimi K2、GLM 4.5、Qwen2.5、Qwen3、GLM4、Llama3、Llama3.1等
	训练框架	mindspeed-LLM、mindspeed-MM、llama Factory、mindspeed-RL
	推理框架	MindIE, vLLM-ascend Nvidia: vLLM, Sglang