

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



轻量灵活 快速上线 高效存储 快捷供数 生态开放 极简易用

华为AI训推超融合（DeepSeek一体机）

内置**DCS AI**全流程软件栈



需要更多信息？

邮箱：enquiry@asl.com.hk

电话：+852 2608 6399

商标声明

 HUAWEI, HUAWEI  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 AI训推超融合（DeepSeek一体机）架构概述

华为DCS AI解决方案针对DeepSeek本地部署推出FusionCube A3000训/推超融合一体机，深度适配DeepSeek V3&R1及蒸馏模型，有效解决大模型行业化落地中的关键难题。方案通过数据工程工具化(ModelEngine)，模型与应用对接标准化，显著提升行业专家模型的训练、推理及精调过程的效率，同时降低成本。训/推超融合一体机支持本地私有化部署，一站方案轻、专业能力强、交付实践专和应用上线快，实现大模型的行业专业化，使能大模型在千行百业快速落地。

智能办公 内容创作 智能客服 医疗健康 教育科研 经营问数



DeepSeek V3/R1 | DeepSeek R1-Distill-Qwen | DeepSeek R1-Distill-Llama

华为AI训推超融合（DeepSeek一体机）

软件

ModelEngine
AI使能平台（数据使能、模型使能、应用使能）

DCS eContainer
容器软件

DCS FusionCompute
虚拟化软件

DME
统一运维

资源
发放

资源
调度

全栈
监控

智能
运维

硬件



FusionCube A3000

智算节点 | 昇腾

管理节点 | 超融合

存储节点 | OceanStor

网络节点 | CloudEngine

02 AI训推超融合（DeepSeek一体机）组件介绍



软件组件

数据使能

AI数据使能工具，提供数据一站式集成、开发、分析/透视、共享开放

应用使能

大模型应用一站式开发与部署工具，低代码编排，降低AI应用开发难度，加速应用落地

模型使能

AI模型使能工具，面向AI开发提供了模型的开发、训练、微调、推理部署及管理全流程服务能力

DCS FusionCompute

虚拟化软件，实现对服务器、存储和网络的虚拟化，形成VM弹性资源池，实现资源自动调度和管理

DCS eContainer

容器软件，提供裸机容器、虚拟机容器、XPU池化、任务调度、容器集群生命周期、容器应用生命周期管理、容器镜像仓、容器运维监控等

DME

数据中心全栈管理软件，面向IT管理员提供全景统一监控、故障自愈、运维等；以及面向租户提供统一的运营管理



硬件组成

- ✓ 存储：全闪+混闪，兼顾性能&成本，高密NVMe全闪
- ✓ 服务器：Atlas集群，高密算力，大模型算力唯二选择
- ✓ 交换机：高性能无损RoCE，8K组网0丢包

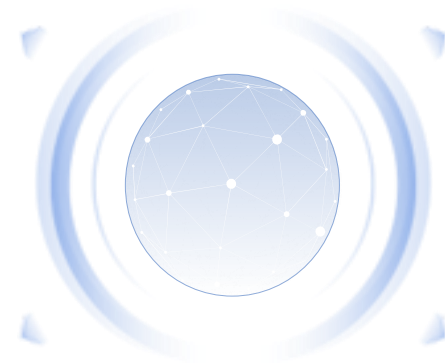
03 AI训推超融合（DeepSeek一体机）技术优势



一站方案 轻

轻量灵活：单节点起步，按需平滑扩展；

全栈集成：智算、通算、存储和网络全集成，一站满足AI应用需求。



专业能力 强

数据工程工具化：内置全流程工具链，知识库构建效率提升10倍；

模型对接标准化：CoT QA数据自动生成，快速蒸馏垂域模型。

应用上线 快

模型训练高效：通用模型到行业专家模型，训练和精调过程高效，低成本；

应用对接标准化：可视化应用编排，开放生态，应用小时级发布。

交付实践 专

一键部署：预适配DeepSeek大模型，推理服务一键配置；

专业保障：华为专业团队深耕行业，具备海量AI工程交付经验，平滑交付。

04 AI训推超融合（DeepSeek一体机）典型配置

产品型号		FusionCube A3000 Ultra(满血版)	FusionCube A3000 Pro(蒸馏版)	FusionCube A3000 Lite(蒸馏轻量版)
模型		DeepSeek-R1 (671B) DeepSeek-V3 (671B)	DeepSeek-R1-Distill-Qwen-32B DeepSeek-R1-Distill-Llama-70B	DeepSeek-R1-Distill-Qwen-1.5B/7B/14B DeepSeek-R1-Distill-Llama-8B
支持模型能力		推理（INT8）	微调/推理	推理
应用场景		科研 高端企业服务 (复杂逻辑推理)	企业问答问数 内容创作 (成本与性能平衡)	个性推荐 智能办公 (低成本和低时延)
吞吐率 token/s		671B: 1911	32B: 4940 70B: 3300	14B: 2920 7/8B: 3824
硬件	智算节点	2*Atlas800I A2 1*Atlas800 3000	1*Atlas800I A2	1*Atlas800 3000 (4*Atlas 300I Duo)
	存储节点	OceanStor Dorado 5500	OceanStor Dorado 2100	-
	管理节点	超融合节点	超融合节点	-
	网络节点	1*XH9210 1*CE6885 1*CE5855	1*CE6885 1*CE5855	1*CE6885 1*CE5855
软件	AI软件栈	DCS ModelEngine、DCS eContainer DCS FusionCompute、DCS DME		
服务	实施服务	硬件设计规划与实施、软件部署服务（含DeepSeek模型部署）等		
	AI使能服务	ModelEngine使能支持		