



Lorem ipsum dolor sit amet, consectetur  
adipiscing elit, sed do eiusmod tempor  
incididunt ut labore et dolore magna  
aliqua. Exceperunt sint occaecat

# 小企业 AI 机器学习 / 深度学习解决方案

in voluptatem.

如果您是一家这样的小企业：

- 正在从事并提供图像识别、人脸识别、声音识别等 AI 技术研发和技术交付，或者正在以 AI 技术为核心进一步开展 AI 应用及解决方案的开发；
- 预算有限，没有专门的 IT 人员，或没有独立机房，但需要高性能的计算力和超大存储空间，以满足机器学习和深度学习持续稳定工作。



Gartner 的数据显示，未来几年人工智能的发展将进一步加速，到 2020 年人工智能和机器学习将成为超过 30% 的 CIO 的五大投资重点之一。各行各业都在通过 AI 来更好地了解客户并开发更好的产品。例如：交通运输业的自动驾驶、制造业的设备预防性维护、零售业的产品个性化推介，以及医疗业的病理人工智能。

基于机器学习和深度学习的 AI 技术开发将会是未来 5-10 年的重要发展领域。对于这些 AI 技术开发 / 研发型的初创小企业来说，在创业初期就需要部署稳定可靠且高算力的产品来支持研发，从而保障产品技术的顺利交付。由于机器学习和深度学习需要源源不断的计算性能和算力的提升，同时需要最大限度降低成本和数据迁移以提高生产效率，这些都是公有云平台无法满足的。如果企业正处于创业初期或高速发展时期，便更加需要从第一台服务器开始就满足伴随整个生命周期的强算力、高通量存储、成本优势，且 7x24 专业原厂 IT 服务。

# 机器学习 / 深度学习的 业务痛点与 IT 需求

## ① 海量数据，对生命周期内的存储容量及峰值算力需求高

机器学习 / 深度学习得以实现的必要条件之一就是：海量的数据，数据的数量和质量与学习结果的精准度成正比。面对大量来自边缘端、云端及核心端的数据进行机器学习，需要强大的高性能计算能力、高带宽和高并发能力的支撑。而深度学习往往需要对结构化和非结构化数据进行快速分析，实时或近实时交付结果，这需要基础架构具备高效的任务并行处理能力和快速处理复杂工作负载的能力（如图像识别、面部识别及自然语言处理）。对于小企业来说，面对高速增长的数据量，强大的高性能计算能力和大规模数据的存储能力尤为重要。

## ② AI 训练开支大，对架构弹性扩展力的要求高

释放数据潜能并转换为业务洞察力提升，推动了 AI 大规模部署和对实时 / 近实时分析与洞见的需求。但随着业务扩张用于 AI 训练的数据量也在呈几何倍数攀升，随之而来的基础架构的计算力和性能都面临着瓶颈，新增采购成本和大规模数据存储成本成为小企业的最大负担。

与此同时，机器学习的训练 (training) 和推理 (inference) 两个环节，通常需要数周甚至数月时间，训练起来既耗时又耗资，对于小企业资金和空间有限的现状而言，需要部署高存储量、高 I/O 性能、高响应速度、高可扩展且低运维成本的智能化基础架构，来满足不断增长的计算密集型应用程序工作负载对大规模训练、实时决策，以及弹性高可扩展的需求。

## ③ 对随叫随到的 1V1 原厂服务要求高

小企业在进行机器学习 / 深度学习的建模、训练等行为时，IT 系统的连续、稳定运营至关重要。然而小企业人力资源构成方面，以专业技术开发人员为主，专业的 IT 运维人员比较缺乏。一旦 IT 系统出现问题，无法及时修复补救，不仅影响企业基于 AI 的内容开发，还将给企业业务带来严重影响和经济损失。因此，随叫随到的专业且完善的 IT 原厂服务支持，对小企业未来业务的可持续发展至关重要。



# 小企业机器学习 / 深度学习解决方案



## 强算力、大存储伴随整个开发生命周期

为了满足人工智能、机器学习 / 深度学习、大数据分析工作负载对高性能计算力的需求，通过部署针对 GPU 和 FPGA 计算进行了优化的戴尔 PowerEdge AI 服务器，通过配置 NVMe 驱动器、多 CPU/GPU/FPGA 和大存储容量，提升数据并行或者模型并行的训练速度，为其提供强大的计算力支撑，满足 AI 建模、仿真和预测分析等大规模计算工作负载的需求，能够在眨眼间完成数百万个任务，显著加快机器学习 / 深度学习训练效率和计算力。

此外，戴尔 AI 基础架构平台 DSS8440 支持 10 块本地硬盘（NVMe SSD 和 SAS/SATA SSD），满足本地训练数据集存储容量要求和 KB 级别小文件（图像、语音）IO 性能要求的同时，提供高性能、低延迟的计算能力。

## 低成本弹性扩展，让机器学习和推理游刃有余

深度学习的训练和推理，意味着分析的数据集规模在不断增加，为了满足小企业对 IT 架构低成本高可扩展能力的要求。戴尔 PowerEdge 服务器采用了英特尔® 至强® 可扩展处理器和内存，不仅能够轻松扩展性能和容量，实现系统的更快部署，同时满足深度学习算法在处理资源数据库不断增长的数据量对计算负载、存储容量等方面的需求。

此外，戴尔 AI 基础架构平台 DSS8440 具有强大的 I/O 扩展能力，随着计算量攀升，通过灵活扩展可最大程度降低 GPU 服务器之间的通信延迟，提高机器学习 / 深度学习的实时决策力。

## 专业原厂服务，远程现场全覆盖

小企业在发展过程中需要全方位 IT 支持，以随时应对未知挑战。戴尔 OpenManage 能够监控和管理 IT 基础架构，借助无代理的 iDRAC（集成的戴尔远程访问控制器）可实现自动、高效、智能的管理，从而显著减少企业对专业 IT 运维管理人员的需求。同时，戴尔为成长型企业提供 7x24 端到端专业 IT 服务和专家指导，为小企业的业务拓展提供全方位保障。

此外，戴尔还提供 ProSupport Plus 白金级别原厂服务，用户可以享受到专业的售后客户服务经理以及精英工程师的服务，并通过戴尔 SupportAssist 的预测型监控、通知、自动化案例创建和前瞻性响应，加速故障的排除，对潜在问题做到防患于未然。

# 戴尔产品推荐

DSS8440/R740/T640/R940XA/R840/+V100/T4GPU 卡

推荐配置：6240\*2/16G\*16/H730P/1.2T/1.8T\*8/V10/T4

## Dell PowerEdge DSS8440

采用 4U2 路服务器设计，旨在为机器学习应用提供极高的性能。DSS8440 基于行业标准 PCIe 的开放式架构允许自定义内部组件，如加速器，存储选项和网卡，可配备 4 个、8 个或 10 个 Nvidia Tesla V100GPU 在满足企业用户灵活需求的同时，提供了非常优势的计算能力。其非常适用于机器学习、训练应用程序以及其他计算密集型工作负载。



## Dell PowerEdge R740

一款针对工作负载加速进行了优化的 2U 双路通用服务器，提供 GPU 卡、存储和计算能力的卓越组合。单台服务器最高支持 3 块双宽 GPU 卡（Nvidia Tesla），可最大限度地提高应用程序性能。可利用英特尔至强可扩展处理器扩展计算资源。非常适用于 AI 中小规模 Training 与 Inference 推理场景。



### Dell PowerEdge R940xa

一款功能强大的四路服务器，将 4 个 CPU 和 4 个 GPU 以 1:1 的大比率相结合，显著提高了应用性能，并且通过直连的 NVMe 驱动器减少数据延迟，从而实现工作负载加速。可配多达 6TB 内存，提供一致且快速的响应速度。适合计算密集型应用程序、机器学习和人工智能、GPU 数据库加速等应用。



### Dell PowerEdge T640

一款集强大性能和大规模内部存储容量于一体的双路服务器，提供 GPU 加速器，可扩展业务架构，并支持混合搭配 NVMe 和 SSD 驱动器，具有出色的计算性能和巨大的内部存储容量，可随工作负载变化而扩展，可利用灵活的存储、性能和 I/O 选项处理一系列工作负载。



### Dell PowerEdge R840

专为数据库内分析而设计，支持四个至强可扩展 CPU 和高达 6TB 的 DDR4 内存，满足大型数据集的性能和容量需求。此外，它还支持多达 12 个高性能 NVDIMM 模块，具有 6 个 PCI-Express (PCIe) 扩展插槽，可处理 2 个双宽度 GPU 或 2 个全高 FPGA，从而加快应用运行速度。

