

洞察人工智能部署引擎

执行摘要

是什么。 人工智能部署引擎 (ade) 是insight的专有自动化框架，可在数据中心和云中加速和标准化人工智能平台的交付。通过打包经过验证的架构并自动化安装、配置和基线测试，ade帮助组织在几周内而非数月内从战略转向安全、高性能的人工智能平台，同时提高运营可靠性和治理。

它为什么重要。 当团队面临基础设施复杂性、工具链碎片化以及技能差距时，人工智能的成果会停滞不前。ADE通过打包成熟的参考架构、自动化配置以及集成MLOps/AIOps控制，消除摩擦——同时缩短价值实现时间，提高可靠性和治理能力。

你得到的。 一个符合经验证设计 (戴尔+NVIDIA) 的工作人工智能平台，具备基准安全、可观察性和性能护栏——通过Insight的PLAN-BUILD-MANAGE方法交付 (RADIUS™实施) © DEVSHOP)

→

→

1.

问题：速度、风险和规模

企业AI程序通常在从幻灯片到系统的过渡过程中变慢。计算、存储、网络、Kubernetes、安全、MLOps和AIOps之间的集成并非易事；不一致的模式和手动步骤会增加成本、延长时间表并创造运营风险。ADE解决了这些差距 **通过将最佳实践进行编码以及自动化平台的搭建**，所以稀缺的专家将时间用于用例和管理——而不是管道。

从商业角度来看，关键杠杆是 **投资回报周期，可预测性，和 合规性** 通过执行一个有偏见的参考架构和兼容性矩阵，ADE减少了部署中的差异，并支持跨站点和业务单元的可重复结果。

关键优势



通过自动化和推荐默认设置实现更快的价值实现



通过内置安全和可观察性实现运行信心



可扩展架构，与经过验证的设计保持一致 (戴尔科技 + 英伟达)



洞察人工智能部署引擎

2.

什么是ADE (商业术语)

ADE是Insight的 **软件加速器/部署引擎** 那个安装、配置并验证AI平台的。在交付过程中，ADE引入了部署工件（工具、配置、文件、设置），这些工件用于在部署后保持完整性——这是我们的知识产权，它确保了环境的持久性和可支持性。这在我们的服务文档中明确被称为 **洞察平台部署引擎**。

高管可以将ADE视为 **平台建设标准运营模式** 一个可重复、自动化的路径，在提高质量的同时减少经过时间和风险。它嵌入在我们 **人工智能基础设施与人工智能平台** 提供产品，为客户提供从准备到生产的简化端到端路径。

3.

如何 ADE 工作 (不使用行话)

ADE推动分阶段推广，按顺序进行以降低风险和确保每一步都有可衡量的进展：

1. **身体准备** (机架/堆叠，固件，织物检查)
2. **核心kubernetes** (安全集群基础)
3. **开发与部署工具** (CI/CD, 注册中心, 自动化)
4. **数据科学工具** (笔记本，精选框架)
5. **安全性与连接性** (身份，策略，混合链路)
6. **性能与可观察性** (SLOs, GPU/IO调优)

This “**战术平台—部署引擎**” - “这种方法是我们的内部手册和高层汇报的支柱。它确保快速进展，同时保留运营团队信任的一系列活动控制、测试和文档记录。”

4.

车底下的东西 (精选的，非定制的)

与其发明定制堆栈，ADE **与验证设计一致** 所以您能受益于最新的性能和操作指南：

- **生成式AI训练与平台模式** (Dell/NVIDIA): 训练和生产架构的全栈考虑。
- **大规模模型定制**：可扩展的、模块化的模式，用于监督微调和领域自适应。

对于高性能使用场景，ADE验证就绪状态 **AI织物网络** (e.g., RDMA/InfiniBand, Spectrum-X) 支持非阻塞数据路径——这对于多 GPU 训练和低延迟推理至关重要。

5.

安全、治理和可观察性 (内置)

安全并非在最后才添加；ADE引入 **策略基准、基于角色的访问控制、网络隔离和密钥管理** 作为标准引导的一部分。可观察性默认包含，带有遥测、日志和指标管道启用 **基于SLO的管理** 和 **AIOps驱动** 异常检测。这为扩展创造了更安全的路径，并为受监管环境提供了更简单的可审计性。

6.

交付模式 & 时间表

ADE嵌入在我们的PLAN-BUILD-MANAGE交付动作中：

- **计划 (RADIUS™)** — 战略、可行性以及与业务成果对齐的优先级待办事项列表。
- **构建 (实施)** — ade驱动平台安装、配置和针对slo的验证。
- **管理 (DEVSHOP®)** — 迭代增强、模型生命周期操作和变更管理。

7.

结果与证据

- **更快实现价值：** 标准化构建和自动化减少了人工工作量，并缩短了基础设施、平台和安全工作流中的关键路径依赖。
- **运行信心：** 经过验证的配置和内置的 observability 使得操作可预测，并且能够更顺畅地将工作移交给运行团队。
- **可扩展架构：** 与OEM承诺和合作伙伴堆栈 (NVIDIA) 对齐。® 在本地，Azure® 为未来保障和可移植性提供 AI 服务)

8.

常见问题解答

ADE如何契合我们的云战略？

ADE支持混合云和多云模式；它标准化了控制平面和流水线，以便工作负载可以根据合规性和经济性要求部署。

如果我们没有内部 MLOps/AIOps 成熟度会怎样？

ADE建立基线工具链和运行手册；我们的DEVSHOP模式填补了空白，同时让你的团队提升技能。

如何管理训练中的网络瓶颈？

我们在早期阶段验证面料就绪状态 (RDMA/InfiniBand , Spectrum-X)，以确保可预测的性能。

参考文献 (内部)

- [人工智能 2025 人工智能 洞察力量景观v1](#) 和 [洞察图谱 2025](#) — Ade定位，功能，提供。 [1][2]
- [人工智能平台V1高管阅读材料](#) 和 [人工智能基础设施解决方案 V1 高管阅读材料](#) — 阶段模型和合作伙伴对齐。 [10][4]
- [服务概览-人工智能赋能产品与服务 - 人工智能平台v0-1](#) — ade-ip/授权语言。 [9]
- [为生成式AI客户设计的解决方案演示文稿](#) — MLOps/AIOps 集成活动。 [3]
- [Nvidia AI顾问20250806](#) — 人工智能织物网络指南。 [12]
- [H20003-gen-ai-model-training-wp](#) 和 [gen-ai模型定制化设计指南 1](#) — 已验证设计。 [5][11]
- [最终 \[0625\] 甲板行走](#) 客户故事幻灯片汇编和 [RADIUS-AI 01 JCM](#) — 成果和客户证据。 [7][8]