Built in Canada. Backed by Dell.

# From AI ambition to trusted action

The Dell AI Factory helps enterprises move from AI pilots to secure, scalable production.

**Nikki Matarazzo**

Artificial Intelligence GTM & Sales Leader

DELL Technologies



# The Canadian Customer Landscape

Three customer types across Canada — each with a different path to scaling AI.

## Neoclouds

AI infrastructure and GPU cloud providers

**FOCUS**

Scale capacity, efficiency, and differentiated services for AI Natives & Enterprises

## Enterprises

Established organizations operationalizing AI across the business

**FOCUS**

Move from pilots to secure, production-grade outcomes

## AI Natives

AI-first companies building products and businesses around data and models

**FOCUS**

Accelerate growth, performance, and time to market

Different customer models. **One Dell goal:** help each customer scale AI with the right infrastructure foundation.

Internal Use - Confidential

Copyright © Dell Inc. All Rights Reserved.

# Let’s make this a workshop.
# Think about your own organization. What is it doing well? What requires improvement? Could AI be the answer to augment, improve or fix it?


Copyright © Dell Inc. All Rights Reserved.

# Why organizations fail to deliver measurable impact with AI



Pilot projects stalling without clear goals



Workload, Data, and Security challenges



Existing infrastructure hindering innovation



Lack of AI tools slowing deployment



Skills gaps preventing AI adoption & success



Escalating AI cost and resource consumption at scale

No matter how big or small the disruption, together, **we can stay ahead of what's next.**

# Three Key Pillars for Successful Enterprise AI

Central AI Team

Governance & Metrics

Know Your Users



Copyright © Dell Inc. All Rights Reserved.

# Two very different AI workloads



## Reasoning-era

Single-turn. Human in the loop.

- User asks → model answers
- One prompt, one completion
- Latency budget: seconds
- Limited tools, memory, plan

~1–2K tokens per interaction



## Agentic workload

Multi-step. Model in the loop.

- Plan → act → observe → repeat
- Calls tools, reads files, runs code
- Latency budget: minutes
- Long context, retries, sub-agents

~10K–500K+ tokens per workflow

# More reasoning. More tokens. More cost.

As agentic AI scales, token costs surge.

## Bigger bills, even with cheaper tokens

While language model inference **costs are projected to fall more than 90%** by 2030[1], total token **consumption is projected to grow 24x**[2]

## Multiplying tokens per outcome

Today's agents plan, reason, leverage tools, and collaborate across multiple steps, which can cause them to **consume up to 1,000x tokens**[3]

## Shifting metrics

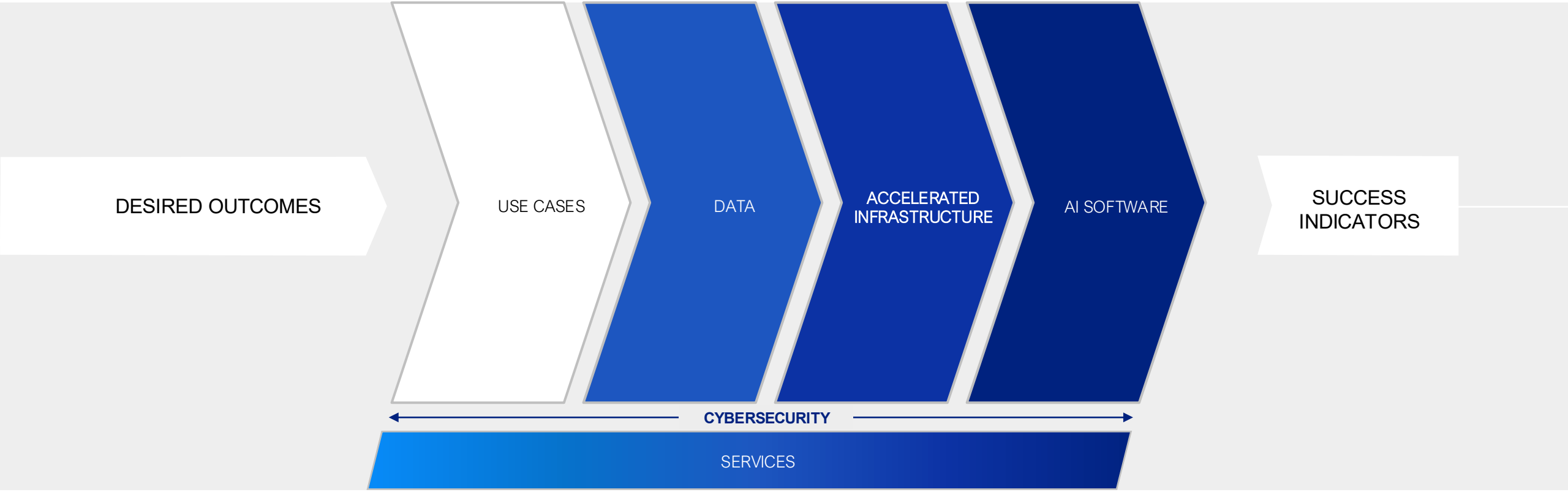
Measurement is moving from **cost per token** to **cost per outcome.** Impact is evaluated in terms of time, accuracy, and business value



[1] *"Navigating the Commoditization Trap as Token Costs Fall by Over 90% Through 2030," Gartner, March 2026*
[2] *"AI Agents Forecast to Boost Tech Cash Flow as Usage Soars," Goldman Sachs, May 2026*
[3] *"The Economics of Agentic AI: On-premises Deployments with Dell AI Factory vs. Cloud," Signal65, May 2026*


Copyright © Dell Inc.

# The Dell AI Factory – a proven track record of success




Real-world experience from over 6,500 implementations[1]

Flexible payment options to invest in AI on your terms

Expert AI guidance every step of the way

World's broadest AI portfolio for all your needs[2]

If there were three things in your organization that you could fix or improve today, what would they be?



Copyright © Dell Inc. All Rights Reserved.



# Desired Outcomes & Use Cases

Dell transitions your AI efforts from experiments to a structured, outcomes-driven program by aligning IT and business, defining use cases, creating a governed roadmap, and guiding secure, scalable deployments

## Establish a shared vision

Align business and IT on a few high-impact AI use cases with clear, measurable outcomes

## Build a governed AI roadmap

Create a repeatable intake and governance model that matches AI projects to your data, architecture, and risk profile

## Achieve production with confidence

Deploy critical AI use cases on hardened, secure architectures with operational runbooks



# Use Cases in Action

AI assistants and agents turn your data into real outcomes: natural-language answers from siloed content, faster and safer software delivery, real-time insights from image and video streams, and autonomous workflows that orchestrate complex processes on a secure, on-premises foundation.

**Knowledge Assistant**

**Code Assistant**

**Computer Vision**

**Agentic AI Platform**

# Use Case Prioritization - Standard Evaluation Criteria

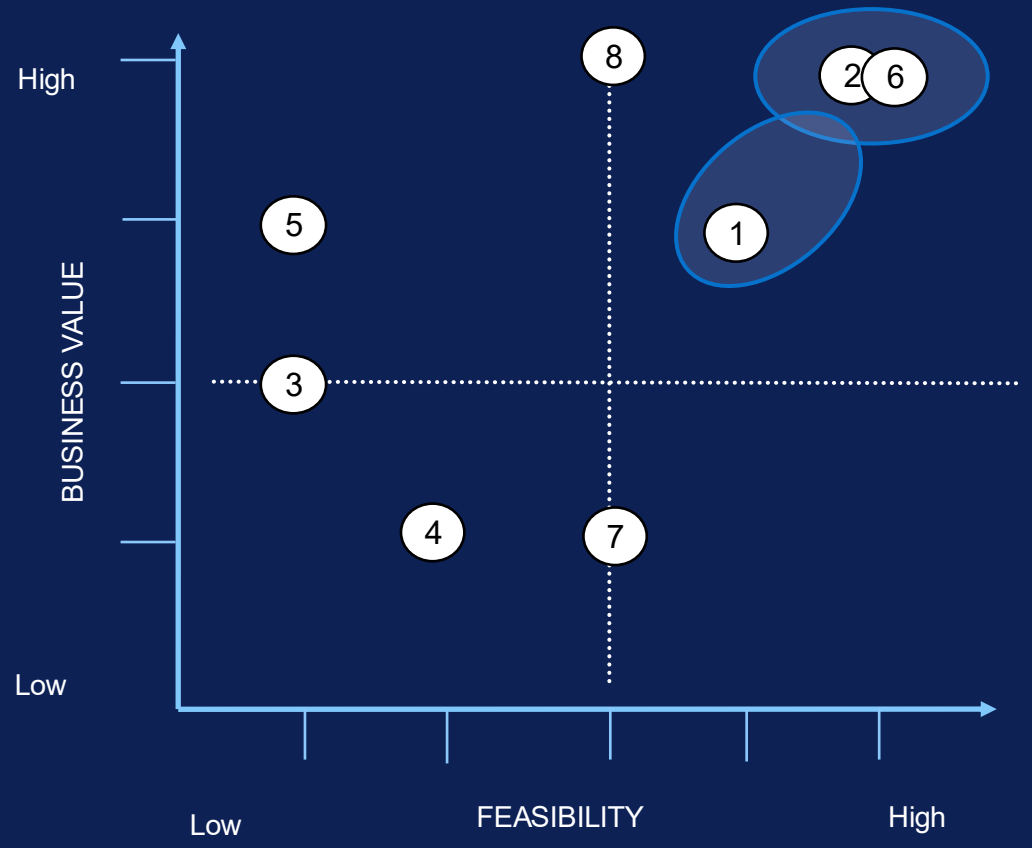
Criteria for consistently evaluating use cases

## Common Use Cases

1. Software Development
2. Customer Service Support
3. Enterprise Knowledge Access
4. Content Creation
5. IT Service Management
6. Finance Operations
7. Supply Chain Operations
8. Administrative Workflow Optimization

**BV**
MORE I BETTER
FASTER I CHEAPER

**F**
DATA I MODEL I PROCESS
PEOPLE I PLATFORM





Copyright © Dell Inc. All Rights Reserved.

# AI use cases inside Dell

Dell's priority use cases are aligned with our areas of differentiation, our business strategy, and our rigorous prioritization process.



Go to Market



Services



Product Development



Supply Chain


Copyright © Dell Inc. All Rights Reserved.

# Do you know your data? Does your data know you?



Copyright © Dell Inc. All Rights Reserved.

# Data

Dell enables you to run AI where it's most effective on a secure, AI-ready data foundation. With the Dell AI Data Platform, you can experiment quickly, refine successful approaches, and transition AI use cases into production faster with minimal integration effort



### Right place, Right time

Run each AI workload where it achieves the best cost, performance, and compliance

### AI-ready foundation

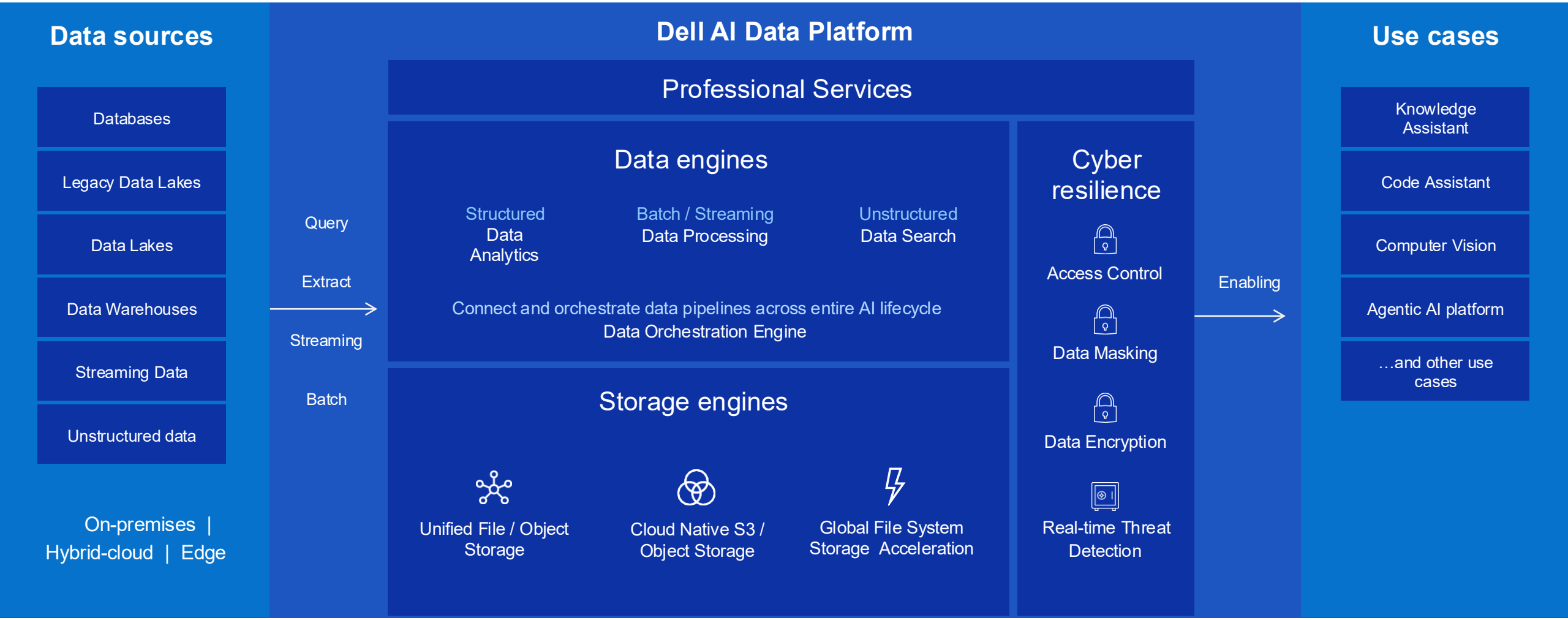
Unify, govern, and protect data end-to-end so models always access high-quality, secure datasets

### Faster path to value

Leverage expert services to experiment quickly and scale AI from pilot to production

# Dell AI Data Platform

Unify your data ecosystem to bring AI to your data

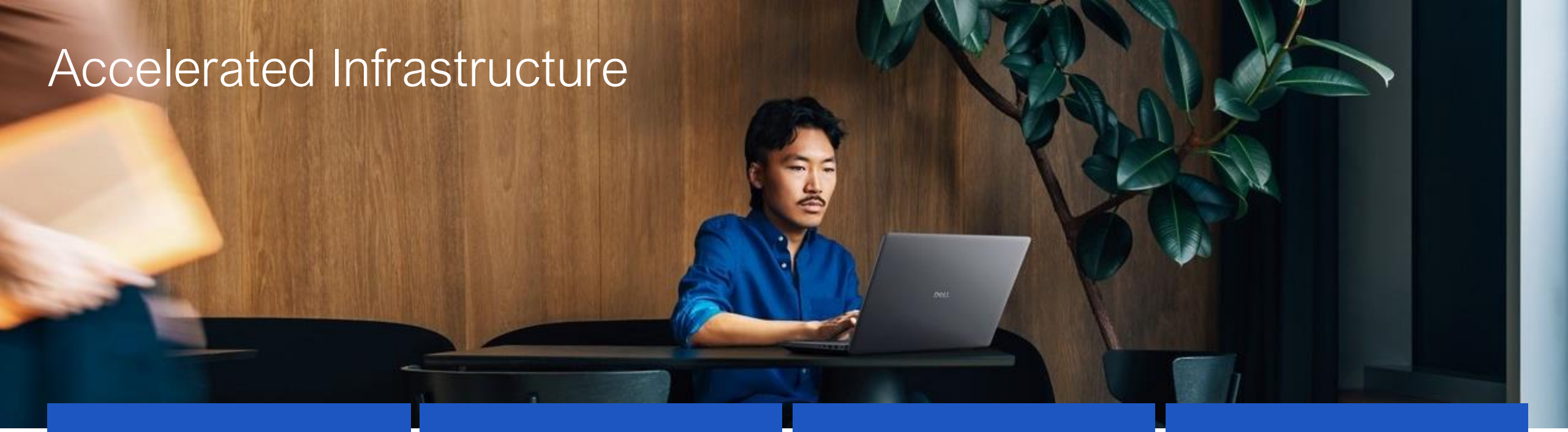


# What would become possible if compute were no longer the constraint?



Copyright © Dell Inc. All Rights Reserved.

# Accelerated Infrastructure



### Secure hybrid AI foundation

Keep sensitive data and IP in-house while running private and hybrid AI on your terms

### Scale with confidence

Grow on AI-enabled PCs and workstations and high-density server clusters using validated platforms and security controls

### From AI strategy to resilient platforms

Services that turn plans into secure, production-ready AI clusters with a clear path to scale

### Designed for what's next

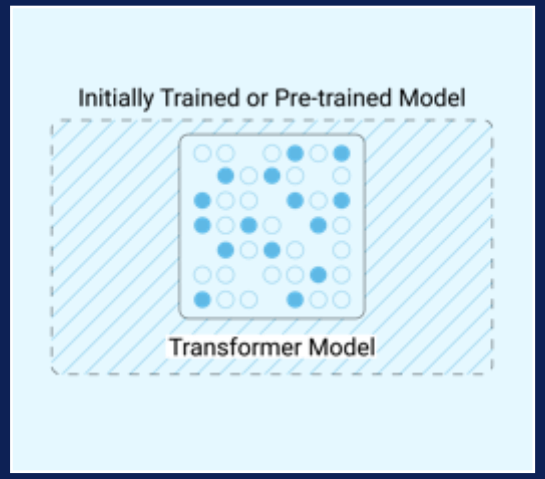
Efficient infrastructure built to run purpose-driven AI at enterprise scale

# AI deployment patterns and their implications
# Inferencing vs Training

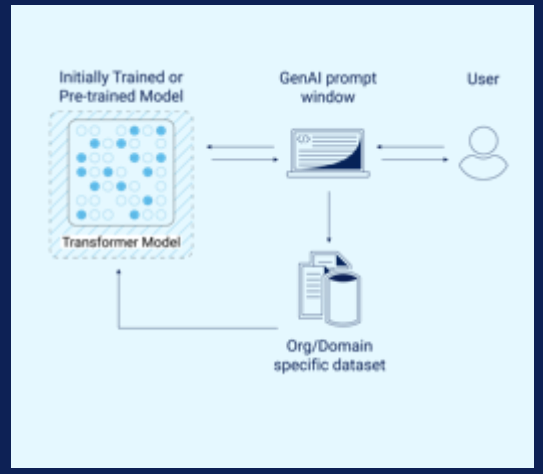
Value

Effort and cost

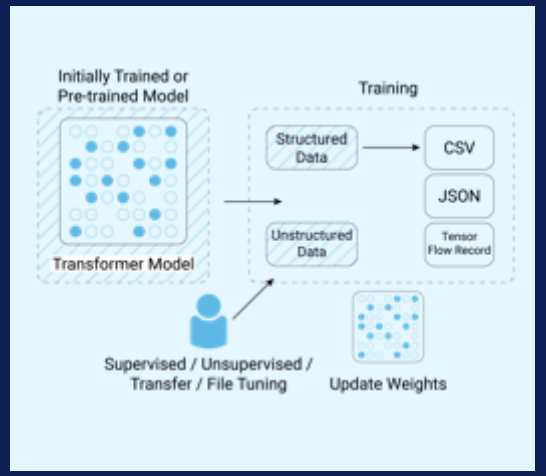
## Pre-trained model / Inferencing




## Model augmentation

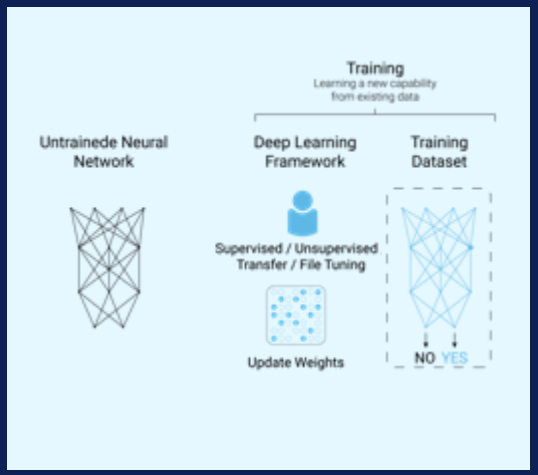



## Fine-tuning




*We think this is the sweet spot for most organizations*

## Model training



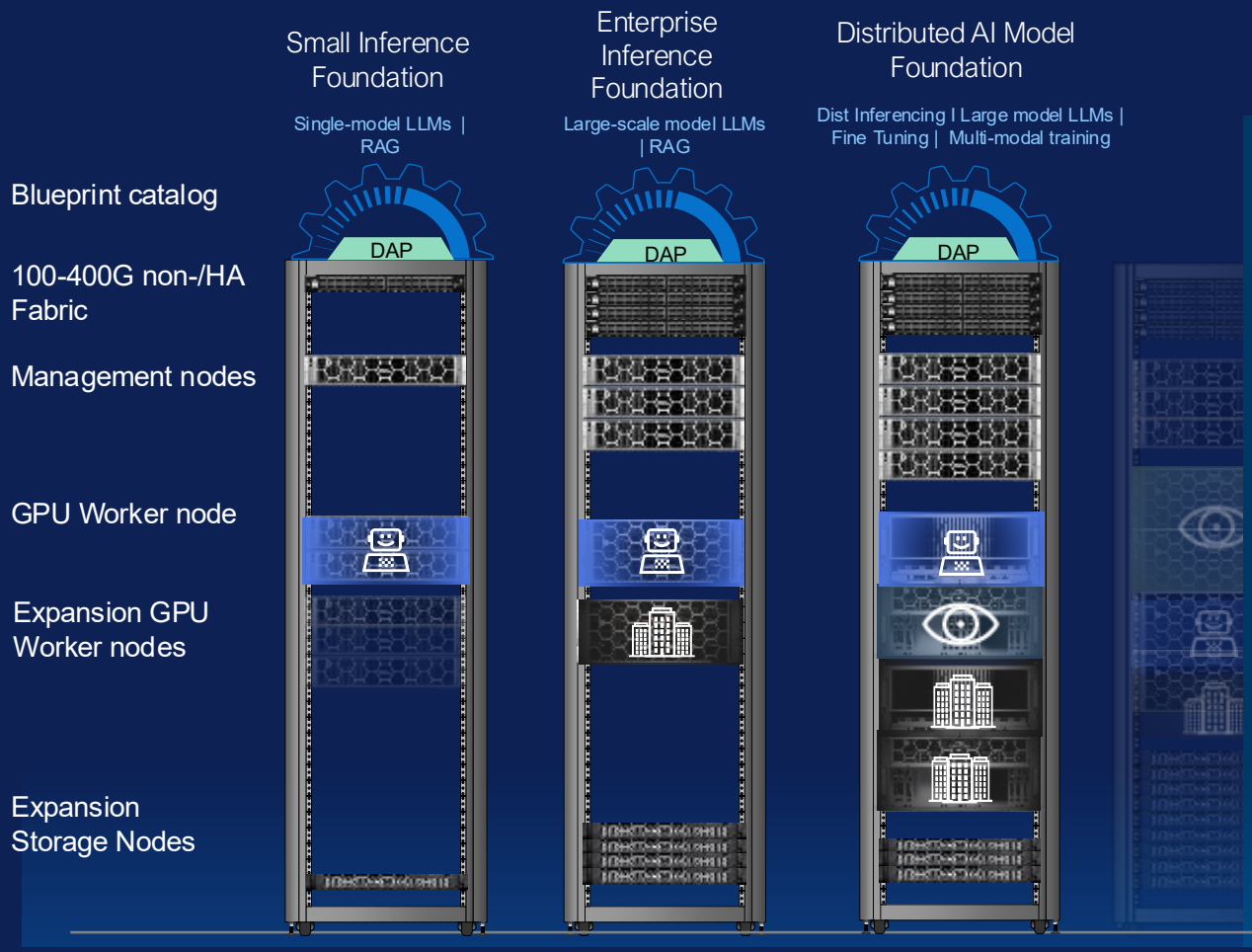

Developer(s) required

Data scientist(s) required

DELL Technologies


Copyright © Dell Inc. All Rights Reserved.

# One architecture. Multiple AI outcomes.




## The possibilities are endless…




Begin with one outcome • Build on a modular foundation • Scale out seamlessly


Copyright © Dell Inc. All Rights Reserved.

# The token economics of workload placement

| | 70B (Baseline Intelligence) | 120B–200B (Enterprise Default) | 300B+ (Frontier / Deep Reasoning) | Best For |
|---|---|---|---|---|
| **Cloud** | **$ - $$**<br>Fast to deploy<br>Margin heavy, latency considerations | **$$ - $$$**<br>• Primary cloud spend zone<br>• Costs escalate with context | **$$$$ - $$$$$$**<br>• Non-linear cost risk<br>• Use should have exec governance | **Bursty or specialized capacity workloads** |
| **Neocloud**<br>**Datacenter** | **$**<br>• Cheapest steady-state option<br>• Useful for concurrency | **$ - $$**<br>• Break-even sweet spot<br>• Capital-efficient at scale | **$ - $$$**<br>• Viable for continuous workloads<br>• CapEx intensive | **Most steady-state workloads** |
| **Deskside / Edge** | **$**<br>• Amortized hardware + power<br>• Limited concurrency | **$ - $$**<br>• Viable with quantization<br>• Throughput constrained | **$ - $$**<br>• Possible but may be slow<br>• Useful for sovereignty, not scale | **Pilots and localized workloads** |

Pricing approximation based on pricing calculated by iToolVerse "AI Pricing Table - Compare OpenAI, Claude, Gemini", updated Mar 2026.

# Do you need a Ferrari to drive to the grocery store?


Copyright © Dell Inc. All Rights Reserved.

# AI Software

**Software and services**

NVIDIA. run:ai

Create custom AI agents and boost accuracy with reasoning models

**Dell Managed Services**
Simplify AI lifecycle management

RED HAT OPENSHIFT
Extended partner platforms

**Ref designs**
Agentic AI

**Dell automation platform**

Catalog

AI Workload blueprints

cohere

Knowledge assistant

Code Assistant

Agentic AI platform

**Compute**
**Storage**
**Networking**



Dell Pro Max



PowerEdge XE Servers



PowerScale & Dell AI Data Platform



Dell PowerSwitch and NVIDIA Spectrum-X

# Major Hybrid Cloud Model Category (Examples)

Hosted on Dell AI Infrastructure

## Closed Frontier Models
(Public Cloud Only)

GPT-5

Claude

## On-Prem Frontier Models
(Private/Hybrid Cloud)

Gemini

Grok

OpenAI Codex

## Open Weight Models
(Public or Private Cloud)

KIMI

deepseek

Mistral

Reflection

Meta

Hugging Face

DELL Technologies


Copyright © Dell Inc. All Rights Reserved.

# Build Partners Aligned to Common Use Cases



Knowledge assistant



Coding assistant



Computer vision



Agentic AI platform



Data for AI


Copyright © Dell Inc. All Rights Reserved.

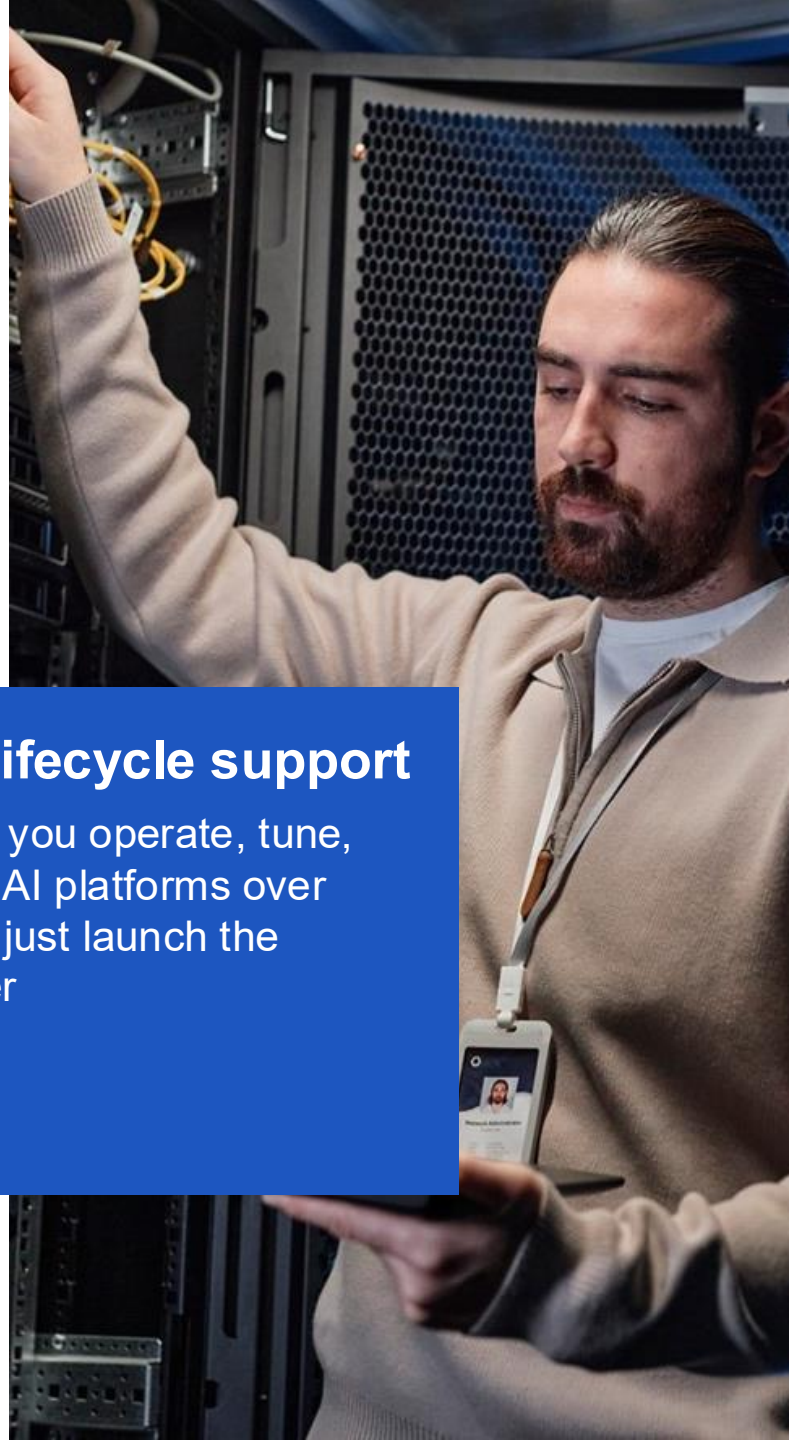
If your AI journey were a relay race, which legs could your team run confidently—and where would you need to hand the baton to a different kind of expertise?



Copyright © Dell Inc. All Rights Reserved.

# Deploy & Integrate AI solutions

Align the right services so AI use cases move from pilot to production faster—with less integration complexity and more predictable performance



## Maximize hardware investment

A cohesive model that aligns accelerators, networking, and software so training is faster, inference is higher-throughput, and GPUs stay busy—not idle

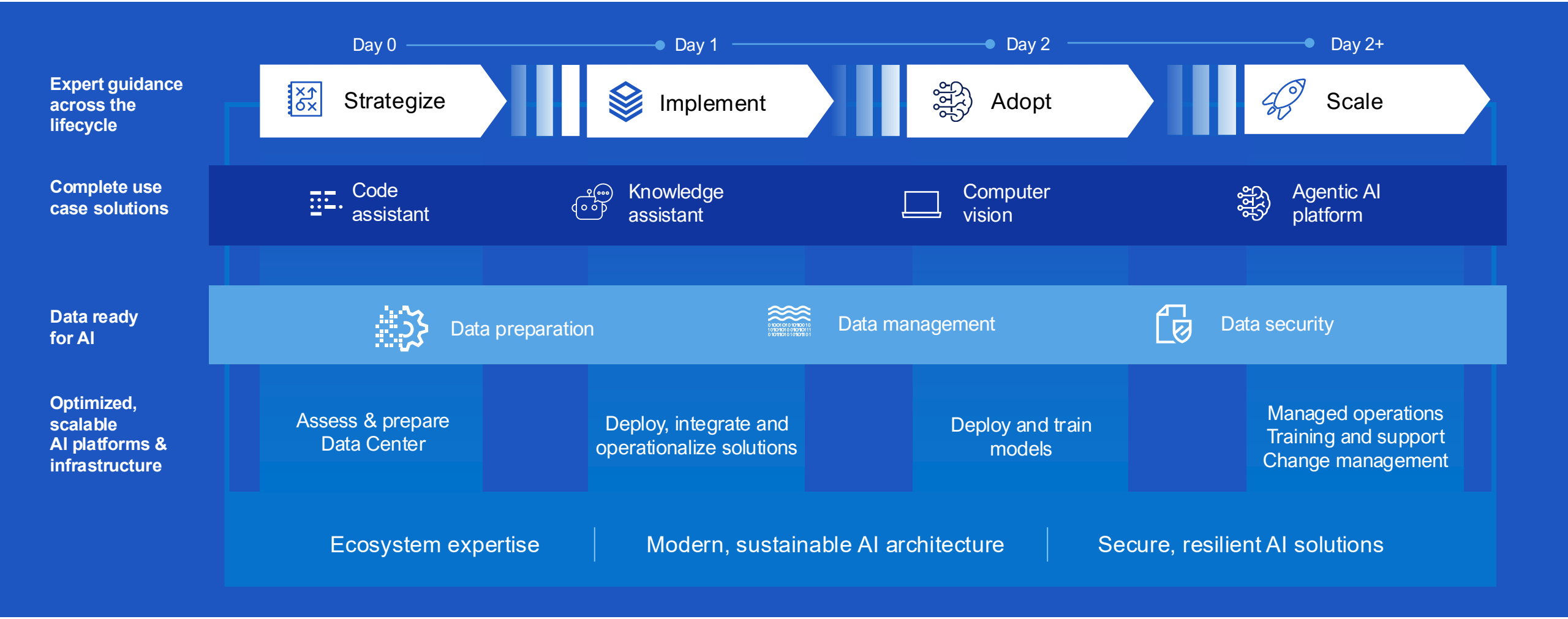
## Faster path to production

Pre-validated architectures and turnkey GenAI accelerators remove integration friction so real use cases go live quickly

## Full AI lifecycle support

Dell helps you operate, tune, and scale AI platforms over time—not just launch the first cluster

# Professional Services accelerate AI adoption from day zero and beyond





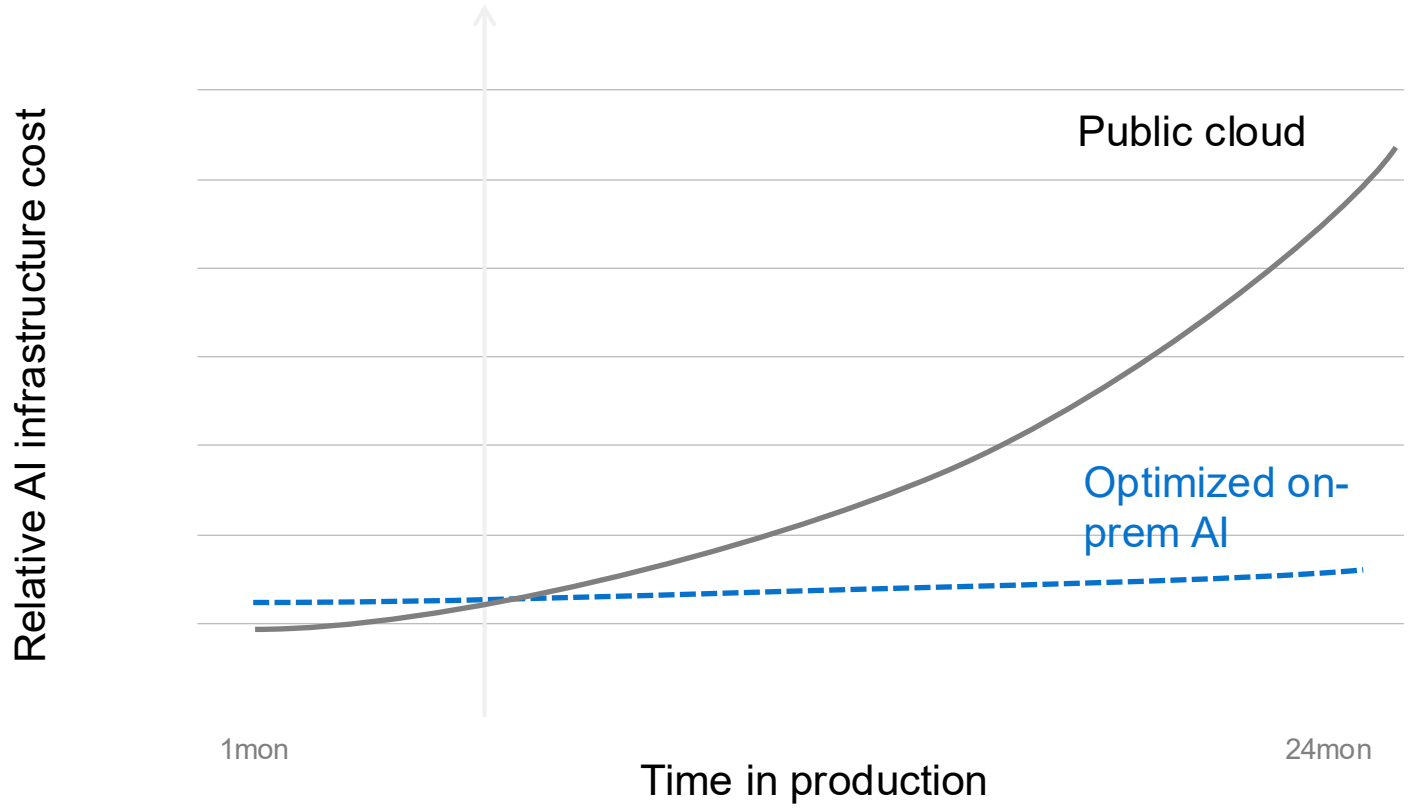

&




Copyright © Dell Inc. All Rights Reserved.

# The Dell AI Factory

## Delivering the broadest AI portfolio[1], with over 6,500 deployments




**8-90%**
lower token spend[2]

Breakeven[2] in
**2-22 months**

1 Prowess Consulting, "Which Vendor Offers the Broadest AI Portfolio for Infrastructure and Services?", October 2025. Prowess Consulting research shows Dell Technologies provides the world's broadest end-to-end AI portfolio from a single vendor.

2 Based on validated analysis by Signal65 and Futurum Group, "The Economics of Agentic AI: Cost Advantages of Dell AI Factory with NVIDIA vs Public Cloud from Pilot to Scale," September 2026. Based on publicly available API pricing and Dell solution pricing and performance data provided by Dell. Savings reflect a 2-year deployment across a range of Dell Pro Max and Pro Precision workstations and PowerEdge servers running knowledge, sales, and software development agentic workloads (knowledge and sales 260 days/yr; software development 350 days/yr). Ranges shown are vs. AWS Bedrock; savings reach up to 98% vs. leading frontier cloud APIs. Analysis is inclusive of estimated cloud discounts and overhead, plus infrastructure hosting, energy, infrastructure management, and Dell support services costs. Individual results may vary.



Copyright © Dell Inc. All Rights Reserved.



# Next steps

**While you're here**

Explore Dell's booth | Connect with us!

**When you get home**

Review use cases, solution stacks, and customer stories at Dell.com/AI

**When you're ready**

Schedule an AI strategy session with your Dell Business Advisor & MOBIA

Built in Canada. Backed by Dell.

# THANK YOU

**Nikki Matarazzo**
Artificial Intelligence GTM & Sales Leader

DELL Technologies

