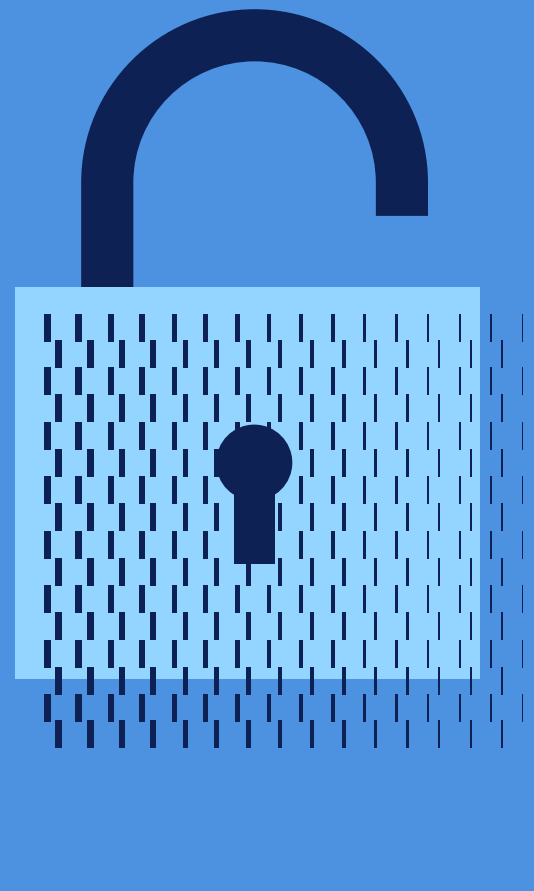


# Unlock better outcomes



## Dell Generative AI Solutions delivers everything you need for Large Language Model Inferencing

Generate higher quality, faster time-to-value predictions and outputs, while accelerating decision-making with a powerful full-stack GenAI solution from Dell Technologies and NVIDIA. This joint architecture delivers a modular, secure, and scalable design supporting a multitude of inferencing use cases and computational requirements.

Now you can simplify the adoption of GenAI across your organization and reduce the time it takes to get results with a proven solution.



Create new value with secure infrastructure for your business-critical operations



Improve IT value with strategic guidance



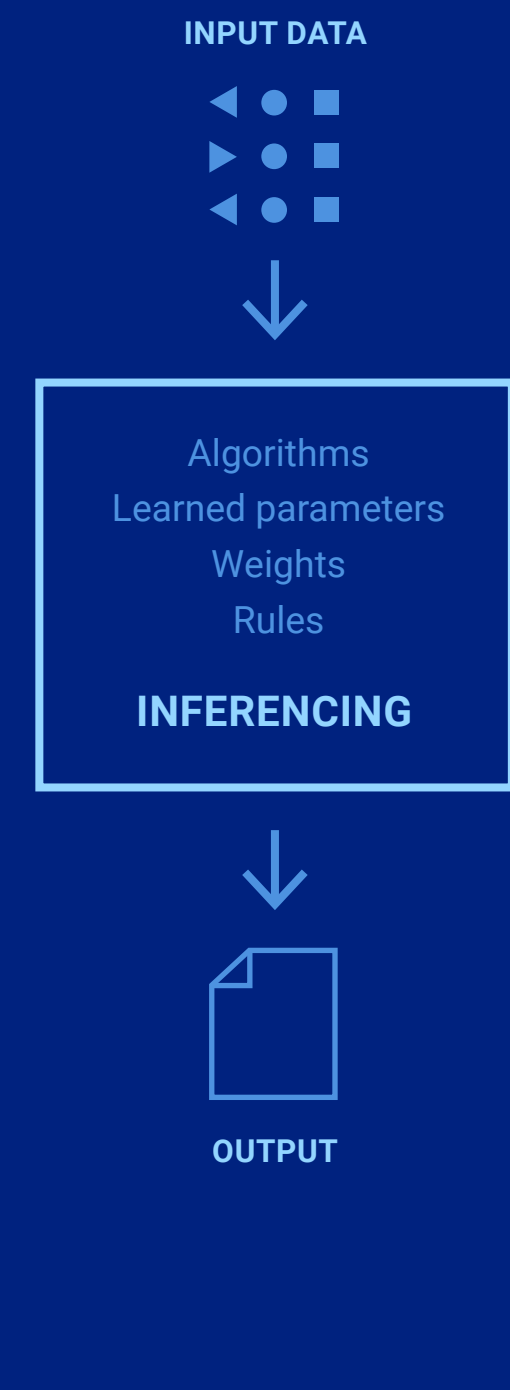
Right-size your infrastructure with flexible consumption models

## What is Inferencing?

Inferencing in AI refers to the process of using a trained model to generate predictions, make decisions, or produce outputs based on input data. It involves applying the learned knowledge and patterns acquired during the model's training phase to new, unseen data.

During inferencing, the trained model takes input data and processes it through its computational algorithms or neural network architecture to produce an output or prediction. The model applies its learned parameters, weights, or rules to transform the input data into meaningful information or actions.

Inferencing is a crucial stage in the life cycle of an AI system. After training a model on labeled or unlabeled data to learn patterns and correlations, inferencing allows the model to generalize its knowledge and make predictions or generate responses on real-world or unseen data.



## How are you using GenAI?

### Natural Language Generation:

Models can be used for text generation tasks such as document writing, dialogue generation, summarization, or content creation

### Chatbots and Virtual Assistants:

Power conversational agents, chatbots, and virtual assistants by generating natural language responses based on user queries or instructions

### Code Development:

Get assistance in software development with features like code completion, ability to generate unit tests, or a chat function for explain code

## Dell Validated Design for Generative AI with NVIDIA - Inferencing

Accelerate deployment and reduce risk with pre-tested and proven solutions designed to help you avoid design, planning, and adoption pitfalls. Components can be mixed and matched, and independently scaled depending on your application needs.



### Generative AI Framework

Conversational AI: NVIDIA Nemo

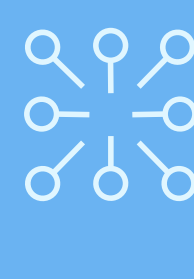
End-to-end enterprise framework for developers to build, customize, and deploy generative AI models with billions of parameters.



### AI Ops and ML Ops Platforms

NVIDIA AI Enterprise

Partner AI Ops software for a smooth end-user experience, including interactive notebooks, experiment management, pipelines, and more.



### Software Infrastructure

NVIDIA Base Command Manager Essentials

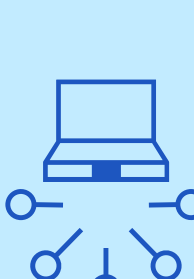
Orchestration & scheduling layer for running AI training, including multi-node jobs, and scaling inference.



### Infrastructure Management

OpenManage Enterprise OneFS | CloudIQ

Familiar Dell management tools delivering proactive monitoring and predictive analytics simplifying infrastructure operations.



### Hardware Infrastructure

Dell Servers, Storage, Networking | NVIDIA GPUs

### Compute with Accelerators

Dell PowerEdge R760xa servers with NVIDIA A100 or H100 GPUs

### Storage

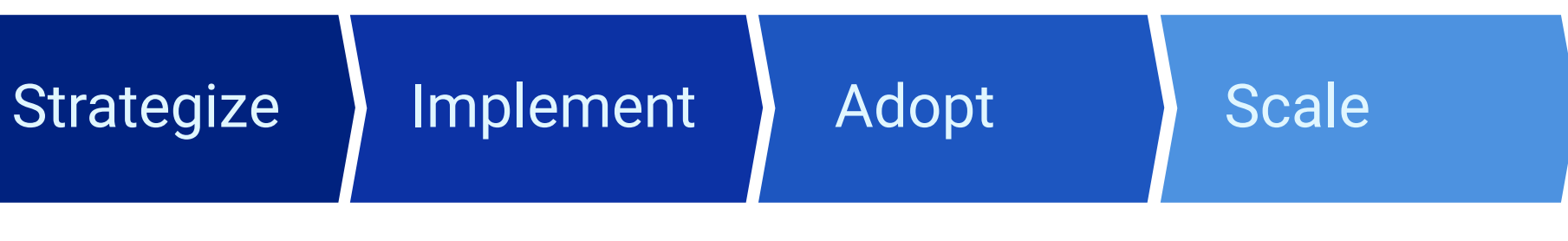
Support for Dell PowerScale, ECS, and ObjectScale

### Networking

Dell PowerSwitch S5232F-ON or S5248F-ON

## Deliver outcomes faster with our help

Dell Services experts can assist you at every stage of your GenAI journey:



### Strategize

Build your roadmap to achieve the innovation objectives of your IT and business stakeholders

### Implement

Establish your platform, leveraging Dell Validated Designs to implement GenAI inferencing hardware and software

### Adopt

Accelerate the value of your use cases by implementing a pre-trained inferencing model

### Scale

Manage your innovation portfolio with resident technical experts and training offers to develop the GenAI skills of your team

## Dell Technologies and NVIDIA

Dell Technologies and NVIDIA work together to enable and accelerate Generative AI workloads, deliver engineering-validated hardware and software to accelerate AI, ML and DL workloads to meet customer needs across all businesses and verticals. With Dell Technologies and NVIDIA, you can deploy AI solutions to accelerate your digital transformation through real-time data that improves key decision-making, with solutions optimized for fastest time to value from your AI initiatives.



[Learn more](#) about Dell solutions



[Contact](#) a Dell Technologies Expert



[View more](#) resources



Join the conversation with [#PowerEdge @DellTech](#)