



NVIDIA DGX BasePOD for the Telecommunications Industry

Streamline AI development and deployment.



Reducing Costs, Mitigating Risk, and Enhancing Customer Experiences

With the constant need to provide efficient connectivity, telecommunications companies invest billions of dollars each year in building and operating their networks, fostering technological advancements, developing new products and services, modernizing infrastructure, and enhancing operational efficiency. In this pursuit, AI has become an essential component of their investment roadmap. Telcos recognize the transformative potential of AI, as they strive to revolutionize their cost structure and introduce groundbreaking applications that will drive significant changes. In a [recent survey](#) conducted by NVIDIA with 400+ telecom industry professionals, 65 percent of respondents agreed that AI is important to their company's success.

Many telco use cases require on-premises AI and [large language model](#) (LLM) training to safeguard sensitive customer data, ensure regulatory compliance, reduce latency for real-time applications, lower cloud costs, maintain independence from providers, and enforce data governance, all while retaining control over data and infrastructure.

NVIDIA has made it easier, faster, and more cost-effective for telecommunications companies to train LLM models and deploy mission-critical AI use cases. By combining the proven performance, scale, and manageability of the [NVIDIA DGX BasePOD™](#) architecture with industry-tailored software and tools from the NVIDIA AI Enterprise software suite, enterprises have a trusted, full-stack platform for building and deploying their AI applications. This solution includes proven, open-source containers and frameworks that have been certified to run securely, both on premises and in the cloud, on the most demanding telecommunications workloads, such as for large language models (LLMs), conversational AI, routing optimization, and video analytics.

Benefits

- > Ability to train LLM models on premises
- > Eliminates design complexity
- > Accelerates deployment
- > Delivers predictable performance at scale
- > Includes the NVIDIA AI software stack
- > Comes with full-stack expertise from NVIDIA Enterprise Support.

Accelerate Telco Workloads With NVIDIA DGX BasePOD and NVIDIA AI Enterprise

- > NVIDIA NeMo™ for large language models
- > NVIDIA® Riva for multi-language agent assists
- > NVIDIA Merlin™ for recommenders
- > cuOpt™ for dynamic dispatching*
- > NVIDIA Triton™ Inference Server for AI inferencing
- > NVIDIA Spark RAPIDS for accelerated data science

*See the [NVIDIA AI Enterprise packaging, pricing, and licensing guide for details](#).

The Value of NVIDIA AI Enterprise

DGX BasePOD includes the **NVIDIA AI Enterprise** software suite*, which contains the key building blocks required to develop and deploy domain-specific, end-to-end AI workflows—from data prep and training to inference and deployment. Users can develop and train their models on DGX BasePOD with the flexibility to deploy anywhere. The suite’s proven containers, applications, and frameworks include **NVIDIA TAO Toolkit** for document automation and **NVIDIA Triton Inference Server** for streamlining and standardizing AI inference, enabling teams to deploy, run, and scale AI models from any framework on DGX BasePOD. The combination of DGX BasePOD with NVIDIA AI Enterprise software provides access to a fully integrated solution of AI-accelerated software and hardware that lets developers quickly deploy, streamline, and accelerate their AI workloads and achieve faster results. And because enterprise-class support is included, organizations get the transparency of open-source backed by the assurance that the global NVIDIA Enterprise Support team will help AI projects stay on track.

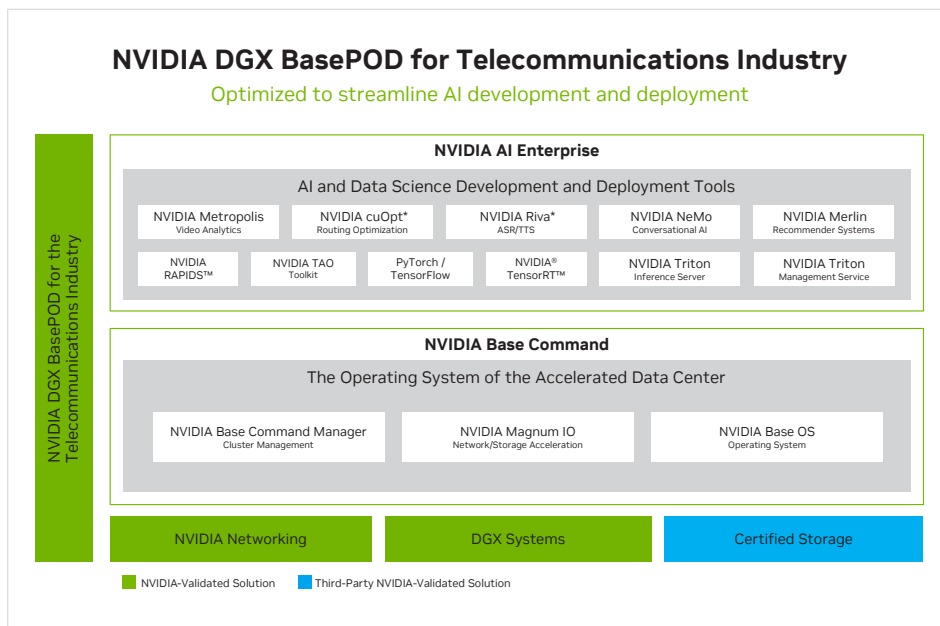


Figure 1. DGX BasePOD for the telecommunications industry

Powered by NVIDIA Base Command

Included with DGX BasePOD is **NVIDIA Base Command™**, the software engine of the DGX platform, which includes enterprise-grade orchestration and cluster management, libraries that accelerate compute, storage and network infrastructure, and an operating system optimized for AI workloads. This fully integrated solution delivers the highest performance and utilization in the industry. By providing the AI software, compute power, tools, and support needed, it gives organizations of any size access to enterprise-class, accelerated infrastructure, so they can focus on creating business value from AI.

*See the [NVIDIA AI Enterprise packaging, pricing, and licensing guide for details](#).

A Strong Ecosystem of Proven Partners

NVIDIA DGX BasePOD for Telco solutions are certified by NVIDIA and include a qualified and proven ecosystem of storage partners. They use the **NVIDIA Magnum IO™** portfolio for intelligent data center input/output (IO) and include technologies like NVIDIA GPUDirect® Storage, which provides the highest-performance IO directly to the GPUs powering the AI infrastructure, to accelerate jobs like image processing. DGX BasePOD for Telcos, fully integrated and tested with NVIDIA's partner ecosystem, also simplifies the deployment of on-prem accelerated AI infrastructure for enterprise IT organizations.

Supported by NVIDIA

With NVIDIA DGX BasePOD, both AI practitioners and IT administrative teams have access to **NVIDIA experts** globally. This provides coordinated support across the full solution, including partner products, control over upgrade and maintenance schedules with long-term support options, and access to instructor-led customer training and knowledge base resources.

Ready to Get Started?

To learn more about DGX BasePOD, visit:

nvidia.com/dgx-basepod

To learn more about NVIDIA AI Enterprise, visit:

nvidia.com/ai-enterprise-suite

