Why Federal Agencies are Increasingly Deploying AI on Workstations

AI has become a critical differentiator across industries, driving demand for hardware capable of supporting AI workloads. While much of the conversation in tech revolves around the exponential growth of AI models—spanning billions of parameters, high-memory requirements, and HPC-class infrastructure—this level of computing remains an outlier, especially in enterprise settings.

Many federal agencies today are actively engaged in AI initiatives, including generative AI, without requiring supercomputers. In fact, a significant portion of AI development—and increasingly, AI deployment—takes place on powerful workstations. Workstations offer several advantages: they provide immediate access to GPU acceleration, eliminate the need to compete for server time, and are cost-effective compared to cloud or server-based alternatives. They also enable secure on-premises data storage, reducing concerns over escalating cloud costs during AI experimentation.

Additionally, the edge is emerging as a rapidly growing AI deployment environment, surpassing both on-premises and cloud-based implementations. Workstations play a crucial role in AI inferencing at the edge, often relying on software-optimized CPUs rather than GPUs. The range of edge AI applications is expanding rapidly, spanning AIOps, disaster response, radiology, oil and gas exploration, land management, telehealth, traffic management, and industrial monitoring.

This white paper explores the increasing role of workstations in AI development and deployment on workstations designed for AI applications within the context of the Federal Government.

The AI Boom and Infrastructure Evolution

AI adoption is accelerating across industries, transforming government operations through software driven by AI models. One key metric illustrating this growth is projected spending on AI-related server infrastructure: by 2026, businesses and cloud service providers will allocate $34.6 billion, nearly 22% of the global server market.

However, AI computing is not solely dependent on servers. A substantial portion of AI development—including data preparation, prototyping, and even deployment—is shifting to workstations. Federal Agencies are rapidly integrating AI into their applications, and workstations offer an accessible, high-performance platform for experimentation and model refinement.

Neural networks, a particularly successful AI approach, have gained traction due to two major factors:

1. The widespread availability of large, diverse, and cost-effective data sources, including unstructured and semi-structured data.

2. The emergence of massively parallel computing, allowing neural networks to process data efficiently within acceptable timeframes.

Traditional machine learning models often run effectively on workstation CPUs, which feature dozens of cores. However, deep learning (DL) models require specialized coprocessors—most commonly GPUs—to parallelize computations across thousands of cores. These accelerators have reshaped both server and workstation markets, enabling AI professionals to develop sophisticated models locally before scaling to enterprise infrastructure.

By 2026, the accelerated server market is expected to grow from $21.8 billion to $43.4 billion, with 57% dedicated to AI workloads. Simultaneously, sales of discrete GPUs for workstations reached 6.4 million units in 2022, and the market for AI-driven workstations is projected to approach $2 billion by 2026.

The AI Development Lifecycle

Developing AI models requires a structured approach, with six key stages:

1. **Data Management** – Identifying, ingesting, and managing relevant data across datacenters, the edge, and the cloud.
2. **Data Preparation** – Storing, cleaning, and transforming data into usable formats using tools like Spark or Pandas.
3. **Model Selection** – Choosing the optimal AI model based on error rate and performance metrics.
4. **Model Development** – Designing AI models using frameworks such as TensorFlow, PyTorch, and scikit-learn.
5. **Model Training** – Running models on compute infrastructure with sufficient processing power, often incorporating explainability and validation mechanisms.
6. **Model Deployment & Monitoring** – Deploying models in production environments for inferencing and performance tracking.
   Workstations play a vital role in all six stages, complementing datacenter, cloud, and edge infrastructure.

Data management | Data preparation | Modeel selection | Model development | Model training | Model hosting and monitoring
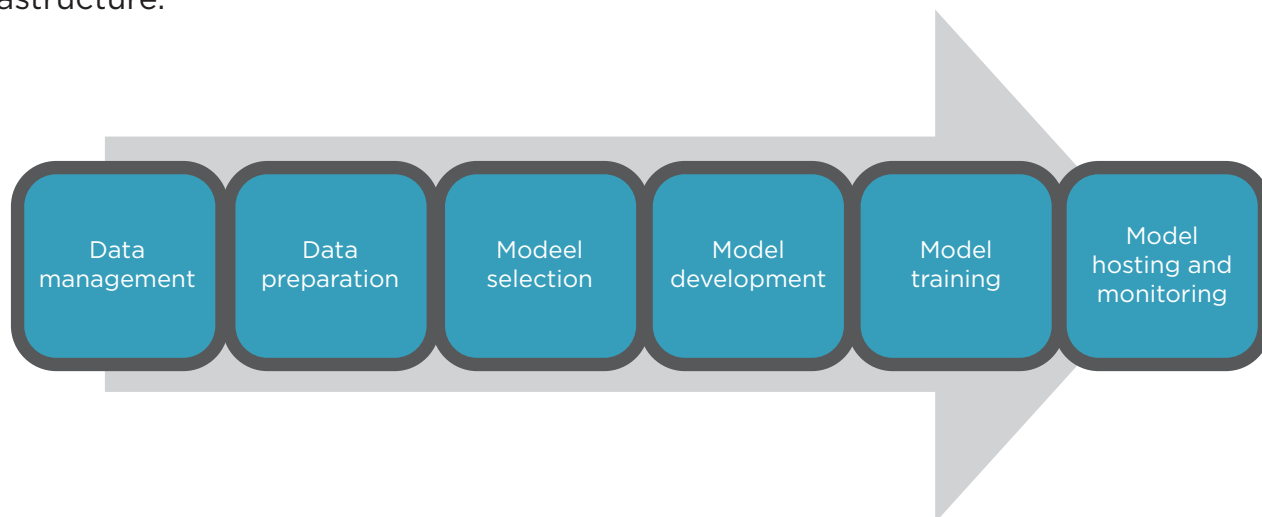
Workstations vs. Personal Computers

Unlike standard PCs, workstations are built for high-performance computing, making them ideal for AI development. Key features include:

- High-grade processors (e.g., Intel Xeon Scalable)
- Powerful GPUs (e.g., NVIDIA RTX 6000 Ada)
- Ample storage (up to 60TB) and high-speed I/O
- Large memory capacity (up to 6TB)
- Advanced cooling systems for sustained performance
- High-speed network interfaces for efficient data transfers
- Professional-grade displays for data visualization
- Error-correcting code (ECC) memory for data integrity during AI training
- Specialized silicon (e.g., Intel Movidius VPUs for vision processing)
- Optimization software (e.g., Intel OneAPI, NVIDIA CUDA)

Workstations vs. Servers: A Symbiotic Relationship

Organizations often balance their AI workloads between workstations, on-premises servers, and cloud instances, depending on the project stage.

Workstations offer a distinct advantage in AI development:

- **Portability and flexibility** – AI professionals can work from anywhere without relying on datacenter availability.
- **Unrestricted experimentation** – No need to request server access or deal with cloud billing concerns.
- **Optimized performance** – Workstations deliver interactive AI development with powerful GPUs, enabling faster iteration cycles.
- **Data locality** – Computing closer to the data reduces bandwidth costs and network congestion.

Additionally, while the market for accelerated servers is growing, only about 4% of enterprise datacenter servers currently feature AI accelerators. This scarcity makes workstations a valuable alternative for AI workloads.

Workstations vs. Cloud: Cost and Practicality

While the cloud offers scalable AI development, it comes with challenges:

- **Availability risks** – Cloud service outages can disrupt workflows.
- **Security & compliance** – Regulations like GDPR and industry policies may restrict data movement.
- **High costs** – Cloud compute fees can escalate rapidly, especially for AI workloads.
- **Psychological pressure** – AI teams may hesitate to experiment freely due to unpredictable cloud costs.

In contrast, workstations allow AI professionals to:

- **Work anywhere** - Ideal for secure, air-gapped environments.
- **Keep data local** - Avoid compliance risks by processing data on-site.
- **Experiment cost** - effectively – Avoid metered cloud expenses while iterating on AI models.

For reference, the cost of running a modest AI setup in the cloud (one NVIDIA T4 instance with 375GB SSD) for a standard workweek is approximately $140 per month. More intensive setups can exceed $2,700 per month, making cloud costs significantly higher than a high-end workstation's annual depreciation.

AI Prototyping and Deployment on Workstations

Beyond development, workstations are increasingly being used for AI inferencing at the edge. Edge AI deployments have grown significantly, with organizations expected to invest **$6.9 billion in AI compute at the edge by 2026.**

Workstations are well-suited for AI inference because:

- Many tasks do not require high-end GPUs; optimized CPUs can handle inference efficiently.
- Intel DL Boost enhances CPU-based inference performance, making it viable for edge deployments.
- Low-power AI processors, such as Intel Movidius Myriad, enable energy-efficient AI applications in remote locations.

As AI adoption accelerates, workstations are proving to be indispensable tools for both AI development and deployment, offering cost-effective, high-performance alternatives to servers and cloud computing.

Use Cases for AI on Workstations

Deploying AI models on workstations is particularly useful in scenarios that involve large volumes of machine-generated time-series data, video streams, or images. These deployments often require real-time analysis and human oversight. Key use cases include:

1. AIOps (AI for IT Operations)

AI helps IT teams transition from reactive troubleshooting to proactive monitoring. AI models can establish performance baselines, detect anomalies, and automate remediation, especially in edge environments where technical staff may be scarce.

2. Disaster Response

First responders rely on AI to assess emergency situations, track critical assets, and allocate resources efficiently. AI workstations process real-time data from drones, sensors, and communication channels, even in areas with limited network access.

3. Manufacturing Plant Monitoring

AI supports predictive maintenance, defect detection, and supply chain optimization in manufacturing plants. Workstations deployed on-site enhance operational efficiency by providing real-time insights.

4. Drones for Infrastructure Inspection

Drones capture images for monitoring power lines, pipelines, and construction sites. AI workstations process this data locally, enabling faster analysis and reducing reliance on remote servers.

5. Office Productivity Tools

AI-powered assistants like Microsoft Copilot improve workplace efficiency by automating tasks, summarizing documents, and enhancing collaboration.

6. Renewable Energy Management

AI models optimize energy production and maintenance at wind farms, solar power plants, and hydroelectric facilities. Local processing ensures uninterrupted monitoring and decision-making.

7. Scientists and Researchers

AI workstations accelerate complex simulations, data analysis, and model training for scientific research. Researchers can process vast datasets locally without waiting for cloud resources, which allows for faster experimentation, iterative testing, and real-time discovery; empowering scientists to push the boundaries of innovation.

Which Workstations for AI Deployment

Dell has partnered with NVIDIA to offer a range of AI-ready workstations under its Data Science Workstation (DSW) brand. These workstations are optimized for AI workloads, ensuring efficient model development and deployment. Key features include:

- **Powerful hardware configurations:** Dell Precision workstations offer multi-core processors, high RAM capacity, and multiple GPU options.
- **Scalability and customizability:** Users can tailor hardware configurations to specific AI workload requirements.
- **Certification and optimization:** Dell collaborates with NVIDIA to certify its workstations for AI tasks, ensuring seamless compatibility with NVIDIA RTX GPUs.
- **Preloaded AI software stacks:** Dell workstations come with optimized AI frameworks and libraries for faster deployment.

Unique Dell Technologies for AI Workstations

Dell Technologies and NVIDIA: Powering AI Workstations

Optimized Performance with Dell Workstations and NVIDIA GPUs
Dell's AI workstations are designed for peak performance, leveraging the power of **NVIDIA GPUs** to accelerate deep learning, machine learning, and data science workloads. With **NVIDIA RTX™ and data center-class GPUs**, Dell workstations provide the computational muscle required for AI model training, inference, and visualization—all within a desktop form factor.

Enhanced Reliability with RMT Pro
Dell's **Reliable Memory Technology (RMT) Pro**, working alongside **ECC memory**, detects and corrects errors in real time, minimizing crashes and maximizing uptime. This ensures AI workloads run smoothly without unexpected disruptions, crucial for long training cycles and high-performance computing tasks.

Precision Performance with Dell Optimizer

Dell Optimizer for Precision further enhances AI efficiency by **dynamically fine-tuning system performance**. This intelligent tool monitors and optimizes **GPU, CPU, memory, and storage utilization**, ensuring AI applications run at peak efficiency with minimal manual intervention.

Bridging the AI Adoption Gap

Many agencies struggle with the misconception that AI requires expensive, high-performance servers. In reality, Dell's AI workstations, powered by NVIDIA GPUs, provide a cost-effective solution for training and deploying AI models. These workstations serve as an ideal middle ground between cloud computing and large-scale AI infrastructure, making AI more accessible for agencies of all sizes.

For organizations scaling their AI initiatives, **Dell ensures a seamless transition from AI workstations to high-performance servers**, offering a complete ecosystem for AI development. By combining **Dell's workstation innovation with NVIDIA's cutting-edge GPUs**, agencies can confidently drive AI advancements without technological bottlenecks.

Conclusion

Workstations are an underappreciated yet powerful solution for AI development and deployment. They offer an affordable alternative to cloud-based AI infrastructure, providing agencies with lower capital expenditures, reduced operating costs, and greater flexibility. Federal Agencies developing AI models that do not require massive computational resources should consider workstations for both experimentation and edge deployment. By leveraging workstation AI, agencies can enhance productivity, reduce latency, and maintain greater control over their AI initiatives.

https://www.wildflowerintl.com/nvidia-ai-on-workstations