Accelerating Every Workload

Modern applications are transforming every business. From AI for better customer engagement, to data analytics for forecasting, to advanced visualization for product innovation, the need for accelerated computing is rapidly increasing. Because new compute demands are outstripping the capabilities of traditional CPU-only servers, enterprises need to optimize their data centers—making this acceleration a must-have. The NVIDIA data center platform is the world’s leading accelerated computing solution, deployed by the largest supercomputing centers and enterprises. It enables breakthrough performance with fewer, more powerful servers, driving faster time to insights, while saving money.

The platform accelerates a broad array of workloads, from AI training and inference to scientific computing and virtual desktop infrastructure (VDI) applications, with a diverse range of GPUs from the highest performance to entry-level, all powered by a single unified architecture. For optimal performance, it’s essential to identify the ideal GPU for a specific workload. Use this as a guide to those workloads and the corresponding NVIDIA GPUs that deliver the best results.
## Choose the Right Data Center GPU

<table>
<thead>
<tr>
<th>WORKLOAD</th>
<th>DESCRIPTION</th>
<th>NVIDIA A100 SXM</th>
<th>NVIDIA A30</th>
<th>NVIDIA A2</th>
<th>NVIDIA A40</th>
<th>NVIDIA A16</th>
<th>NVIDIA A100X</th>
<th>NVIDIA A30X</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Highest Perf Compute</td>
<td>Mainstream Compute</td>
<td>Entry-Level Compact AI</td>
<td>Highest Perf Graphics</td>
<td>Optimized for VDI</td>
<td>Highest Perf Converged Accelerator</td>
<td>Mainstream Converged Accelerator</td>
</tr>
</tbody>
</table>

### Recommended Number of GPUs or Converged Cards per Server

- **Deep Learning (DL) Training and Data Analytics**
  - For the absolute fastest model training and analytics
  - **SXM**
    - 4 or 8 GPUs (PCIe | SXM)
    - > 80GB: Bn+ parameter models (DLRM, GPT-3)
  - **PCIe**
    - 1-2 GPUs (PCIe | SXM) with multi-instance GPU (MIG)
    - > 80GB: large batch size constrained models (RNN-T)
- **DL Inference**
  - For batch and real-time inference
  - **SXM**
    - 2-4 GPUs with MIG
  - **PCIe**
    - 1-4 GPUs
- **High-Performance Computing (HPC) / AI**
  - For Higher Education Research and scientific computing centers
  - **SXM**
    - 2-4 GPUs (PCIe | SXM) with MIG
  - **PCIe**
    - 2-4 GPUs
- **Render Farms**
  - For batch and real-time rendering
  - **PCIe**
    - 1-4 GPUs for entry-level virtual workstations*
    - 2-4 GPUs for midrange to high-end virtual workstations*
    - 2-4 GPUs for highest virtual desktop and workstation user density**
- **Graphics**
  - For the best graphics performance on professional VDI
  - **PCIe**
    - 1-4 GPUs for mobile Android
    - 4-8 GPUs for (4K resolution)
- **Cloud Gaming**
  - For 4K resolution / Android
  - **PCIe**
    - 1-4 GPUs for inference and video workloads
- **Enterprise Acceleration**
  - For mixed workloads, including graphics, ML, DL, analytics, training, and inference
  - **PCIe**
    - 1-2 GPUs with MIG for compute workloads
    - 1-4 GPUs for balanced workloads*
    - 1-2 GPUs for graphics-intensive workloads*
- **Edge Acceleration**
  - For differing use cases and deployment locations
  - **PCIe**
    - 1-2 GPUs with MIG
    - 1-4 GPUs for graphics-intensive workloads & AR / VR*
- **5G vRAN**
  - For low-latency GPU-network communication
  - 1 card
- **AI-Based Security**
  - For GPU-powered network processing
  - 1 card

---

To learn more about Data Center GPUs, visit [www.nvidia.com/ampere](http://www.nvidia.com/ampere)

* NVIDIA RTX Virtual Workstation (vWS) software license required for virtual workstation workloads.
** NVIDIA Virtual PC (vPC) software license required for VDI workloads.

© 2021 NVIDIA Corporation. All rights reserved. NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All other trademarks and copyrights are the property of their respective owners. NOV21