Simplifying Cluster Operations

Running a cluster can be complex.

Having a deep understanding of your organization’s hardware architecture, software stack, and unique needs is critical to the task at hand.

To simplify cluster operations and maximize GPU/Network-enabled cluster utilization, NVIDIA recommends working with a Site Reliability Engineer (SRE).

An NVIDIA Site Reliability Engineer (SRE) is a DevOps engineer who helps customers manage and maintain their cluster remotely by assisting with day-to-day operational cluster management, user-facing, and vendor-facing tasks.

What to expect from your Nvidia SRE

- Provides guidance on cluster management, including how to manage scheduling software, such as Slurm, and monitoring the state of individual nodes.
- Works with the NVIDIA Technical Account Manager (TAM) to resolve cluster operation issues.
- Helps organizations face the complex challenges associated with technology adoption.
- Provides expert insight into cluster deployment.
- Enables you to build future-proofed infrastructure that meets today’s business objectives while adapting to growing workloads.

TERMS AND CONDITIONS

- Up to 10 weekly hours.
- An expert team, available in the customer’s time zone, provides coordinated service.
- Service will be provided during business hours (customer time zone).
- 6-month renewable service.
- The service is available only for clusters that have been deployed by Nvidia Professional Services.
- The Service start date is usually the system hand-over date (can’t be delayed).

OUT OF SCOPE

- Managing non-NVIDIA IT operations – storage / 3rd party network equipment.
- Code edits/changes in 3rd party cluster management tools or scripts.
- New tool development to help with day-to-day activities (paid service).
- The SRE does not provide post-sales on-call support, but will assist to obtain a real-time response from the post-sale support on-call team.
Scope of Responsibilities

- Cluster deployment practice advisor.
- Cluster administration and operation best practice.
  - Create/Remove user accounts.
  - Create or update automation [e.g., Ansible playbooks] for site-specific needs.
- Testing scenario advisor.
- Post-installation cluster management stack related questions/ issues.
  - Hosts “office hours” to field questions about cluster usage and best practices.
  - Opens and tracks trouble tickets with the vendor [NVIDIA] in coordination with the Parallel Tracking and Mapping System (PTAM).
- System/ Cluster monitoring advisor.
  - Monitors all aspects of the state of the cluster, including Compute, Networking, Storage, Power, and Software.
- New applications / software deployment.
  - Keeps installed software and firmware up to date in accordance with vendor best practices.
  - Trains customer administrators on best practices [lead by example]

Learn More

To learn more about NVIDIA Professional Services for the data center, contact: nbu-services-sales@nvidia.com