CHECKLIST

vCluster

Exemplar-Ready AI Cloud Checklist

A Self-Assessment for NVIDIA Cloud Partners

Use this checklist to evaluate whether your Kubernetes platform architecture supports Exemplar-level outcomes.

1 Benchmarking Readiness

□ We can create clean benchmark environments on demand

□ Benchmark environments are standardized and version-pinned

□ We can validate platform changes safely before broad rollout

□ Benchmark runs are isolated from production tenants

□ We can reproduce benchmark results across multiple runs

□ Benchmark environments can be deleted and recreated cleanly

Why it matters:

Exemplar validation requires repeatability.

2 Tenant Isolation

□ Tenants operate independently without requiring separate physical clusters

□ Each tenant has an isolated Kubernetes API

□ Tenants can manage RBAC and install CRDs independently

□ We can offer multiple isolation tiers (Shared, Dedicated, Private Nodes)

□ Dedicated workloads do not require dedicated clusters

Why it matters:
Isolation should not require multiplying clusters. Control plane isolation and flexible node allocation are foundational for scalable multi-tenancy.

3 GPU Efficiency

□ We can safely share GPU infrastructure where appropriate

□ We can provide Dedicated or Private Nodes for predictable performance

□ Dedicated node pools scale dynamically based on workload demand

□ Idle GPU capacity is minimized through automated node provisioning

Why it matters:
Exemplar-ready platforms balance standardization. Drift at the Kubernetes layer undermines benchmarking integrity.

4 Platform Standardization

□ Kubernetes versions are consistent across tenant environments

□ Upgrades are centralized and controlled

□ Policy enforcement is consistent across tenant environments

□ Tenant onboarding is automated and template-driven

Why it matters:
Exemplar readiness requires operational leverage. If cluster count grows with tenant count, scalability will eventually stall.