Nebius B.V.,
Gustav Mahlerlaan 308,
1082 ME Amsterdam

AI Solutions Architect Certification Exam Guide

Role description

A Nebius Certified AI Solutions Architect is able to design end-to-end AI infrastructure solutions on Nebius AI Cloud by bridging machine learning workload requirements with the underlying infrastructure.

They evaluate the workload, team, budget, and timeline, then produce the architecture and cost estimate. They also choose the orchestrator, GPU type, storage tier, and network design.

Solutions Architects, whether in an internal cloud customer team or an integrator/vendor team, are accountable for architectural decisions while serving as strategic advisors to stakeholders.

Target candidates

Infrastructure engineers, solutions architects, and other professionals with 3+ years of cloud experience. Regardless of job title, candidates may represent either the cloud customer side or the integrator/vendor side.

Domains & weighting

The exam includes the following content areas and their weightings. This guide doesn’t cover every topic, but it provides helpful context to prepare for the exam.

Domain	Exam weight
1. Design high-performing architectures	~25%
2. Design resilient architectures	~25%
3. Design efficient & cost-aware architectures	~25%
4. Design secure architectures	~25%

Exam format

The assessment is delivered remotely and monitored with AI-assisted proctoring. It is delivered via a partner exam platform.

All questions in the exam follow the same format:
Multiple choice – one correct answer and three incorrect options (distractors).

There are no strict prerequisites. No pre-tests or additional eligibility checks are required.

Duration is limited to 1 hour.

Content outline

1. Design high-performing architectures

Skills in:

- Selecting GPU architecture and platform-level building blocks for a workload
- Designing the storage architecture for AI workloads
- Designing the training-platform support stack and user environment

Expertise in:

- Defining workload outcomes and the right deployment shape
- Architecting orchestration and workload-framework overlays
- Recommending distributed-training architecture and its tradeoffs

2. Design resilient architectures

Skills in:

- Designing checkpoint and restart strategies
- Designing observability scope across logs, metrics, and traces
- Designing the validation, benchmark, and acceptance plan

Expertise in:

- Designing failure handling and the operational support model
- Designing the migration and handover path

3. Design efficient & cost-aware architectures

Skills in:

- Planning capacity, quotas, and onboarding cost
- Reducing operational effort through reuse and automation
- Applying FinOps practices: governance, cost monitoring, and decommissioning

Expertise in:

- Articulating tradeoffs and constraints, where cost can outweigh raw performance
- Applying cost dimensions to performance and right-sizing choices
- Designing multi-cloud connectivity and data-ingestion pipelines

4. Design secure architectures

Skills in:

- Designing the identity and access model
- Designing private connectivity for control-plane and operator access
- Designing the secrets, audit, and compliance posture

Expertise in:

- Applying the Shared Responsibility Model, tenant isolation, and least-privilege principles
- Designing data classification, encryption, and incident-response posture