

Minimum Azure Bill of Materials

Private Cloud — Lean Private System-to-System Setup

VM Operating System: Linux | Period: 24 Months | Region Assumption: Azure South Africa North

Date: 4 June 2026

1. Azure Bill of Materials

No.	Azure Service	Azure Product / SKU Code	Qty / Capacity	Purpose
1	Linux GPU Virtual Machine	Standard_NC4as_T4_v3	1 VM; Linux; 4 vCPU; 28 GiB RAM; 1 x NVIDIA T4 GPU; 16 GB GPU memory	Runs Ollama and private AI models
2	Linux Processing Virtual Machine	Standard_D8as_v5	1 VM; Linux; 8 vCPU; 32 GiB RAM	Runs Qdrant, Docling, OCR, n8n, API/orchestration and document processing
3	GPU VM OS Disk	Premium SSD LRS P15 or equivalent	256 GiB	Linux OS, Ollama runtime, model cache and container runtime
4	Processing VM OS Disk	Premium SSD LRS P10 or equivalent	128 GiB	Linux OS, Docker, processing services and system packages
5	Processing / Qdrant Data Disk	Premium SSD LRS P30 or equivalent	1 TiB	Qdrant vector data, embeddings, metadata and temporary processing workspace
6	Temporary Object Storage	StorageV2 Standard_LRS Hot	500 GiB	Temporary processing files, failed-job retention, exported logs and transient document artefacts
7	Azure Bastion	Azure Bastion Basic	1 Bastion instance	Secure admin access to Linux VMs without public SSH exposure
8	Virtual Network	Microsoft.Network/virtualNetworks	1 VNet	Private network for AI, processing, storage and management subnets
9	Subnets	Microsoft.Network/virtualNetworks/subnets	Minimum 4 subnets: AI, Processing, Management, Private Endpoints	Network segmentation
10	Network Security Groups	Microsoft.Network/networkSecurityGroups	3–4 NSGs	Restrict traffic between merSETA systems, AI server, processing server and admin subnet
11	Private Connectivity	VNet Peering or Private Endpoint / Private Link	Approved merSETA systems only	Private system-to-system access from approved subscriptions
12	Public IP for Bastion	Standard Public IP	1 public IP for Bastion only	Required for Azure Bastion management access
13	Key Vault	Azure Key Vault Standard	1 vault	Stores API keys, secrets, certificates and service credentials
14	Log Analytics Workspace	Log Analytics Workspace	50 GB/month ingestion; 90-day retention	VM logs, API logs, admin access logs, AI processing logs and security events
15	Azure Monitor	Azure Monitor Metrics and Alerts	Basic alert rules	VM health, disk, GPU/CPU/memory, failed jobs and queue monitoring
16	Recovery Services Vault	Recovery Services Vault	1 vault	Backup vault for VM and configuration recovery
17	Azure Backup	Azure Backup for Azure VM	2 VMs + approximately 1.5 TiB protected data	Daily backups of VM configuration, processing data and critical service state

No.	Azure Service	Azure Product / SKU Code	Qty / Capacity	Purpose
18	Container Runtime Allowance	Docker on Linux VMs	Included in VMs	Runs Ollama, Qdrant, Docling, n8n and supporting services
19	Bandwidth / Data Transfer Allowance	Azure bandwidth	Low-volume private system-to-system traffic	Internal traffic, API calls and limited outbound updates
20	Azure DNS / Private DNS Zones	Private DNS Zone	1–3 private DNS zones	Internal name resolution for AI API, Qdrant, storage and private endpoints

3. Sizing and SKU Basis Notes

Item	Note
VM operating system	Both virtual machines must be Linux-based.
GPU VM	Standard_NC4as_T4_v3 is included as the minimum GPU compute SKU for modest private AI inference, including smaller text models and selected vision workloads.
Processing VM	Standard_D8as_v5 is included as the minimum CPU/RAM SKU for Qdrant, Docling, OCR, n8n, API/orchestration and document processing.
Storage sizing	The AI engine is not the authoritative document repository. The BOM therefore includes 500 GiB temporary object storage and 1 TiB processing/Qdrant data disk capacity.
Qdrant storage	Qdrant is expected to store embeddings and approved metadata. Original documents remain in source merSETA systems unless explicitly agreed otherwise.
Bastion	Azure Bastion Basic is included to avoid public SSH exposure for Linux administration.
Connectivity	Private connectivity is included for controlled system-to-system access from approved merSETA systems or subscriptions. Public WAF/Application Gateway is excluded from this minimum private BOM.
Pricing	Costs are indicative planning estimates. Suppliers must provide current Azure calculator output and disclose reserved instance, savings plan or pay-as-you-go assumptions.