Document Version: 1.0
September 20, 2025
Jim Venuto

Scope guardrails (context, not BOM)

- **Manual, self-managed GPU burst** on IBM Cloud.
- **SFTP only** for data movement (SSH on GPU/utility VSI).
- **No** centralized scheduler, VPN, Aspera, PHI.

Summary counts (4 labs + Dev/MVP)

Item	Shared	Per-Lab	Total (4 labs)	Dev/MVP
Resource Group	1	–	1	–
Access Groups	2	–	2	–
API Keys	2–4	–	2–4	1–2
VPC	1	–	1	–
Subnets (AZs)	2	–	2	–
Public Gateways	2	–	2	–
Security Groups (baseline)	2	–	2	–
Security Group (lab)	–	1	4	1
SSH Key	1	–	1	–
COS Instance	1	–	1	–
COS Buckets	0 or 1	1	4	1 (or prefix)
GPU VSI	–	1	4	1
Boot Volumes	–	1	4	1

Item	Shared	Per-Lab	Total (4 labs)	Dev/MVP
Data Volumes	–	1	4	1
Public IPs	–	1	4	1
Schematics Workspaces	1	1	5	1 (dev apply)
Budget Alert Policies	1	–	1	–

If using **one shared COS bucket** with per-lab prefixes, set **Shared COS buckets = 1**, **Per-Lab COS buckets = 0**.

Detailed BOM

1) Identity, Access, Governance

- **Resource Group:** LRI-Phase0
- **Access Groups:** LRI-Phase0-Admins, LRI-Phase0-Researchers
- **IAM Roles:** Admin (Jim + designated CC/IBM), Editor/Viewer (researchers as needed)
- **API Keys:** Per-persona (admin, automation)
- **Tags (required):** org=LRI, phase=0, lab=<name>, owner=<PI>, env=lab, data=synthetic, cost-center=<id>
- **Budget Alerts:** 1 policy, tag-filtered per lab

2) Network (VPC)

- **VPC:** lri-phase0-vpc
- **Subnets:** 2 AZs (e.g., 10.10.1.0/24, 10.10.2.0/24)
- **Public Gateways:** 2 (one per subnet)
- **Security Groups (baseline):**
  - sg-default-outbound (egress allow)
  - sg-ssh-inbound (TCP/22 from approved CIDRs)
- **SSH Keys:** lri-phase0-key (uploaded once; reused)

- **Optional: Bastion host (CPU VSI)** for private-only pattern (1 lab demo)

3) Compute (per lab + dev)

- **GPU Virtual Server (VPC):** 1 per lab; 1 for Dev/MVP
    - **GPU profile:** to be chosen by capacity (e.g., A100/L40s/Gaudi 3 family)
    - **OS Image:** Ubuntu LTS or RHEL
    - **Public IP:** 1 (Phase 0 minimal path)
    - **SFTP:** OpenSSH on instance (SSH/22)

4) Block Storage (VPC)

- **Boot Volume:** 1 per VSI (auto-created)
- **Data Volume (per lab):** 1
    - **T-shirt size defaults:** S=256 GB / M=512 GB / L=1024 GB

5) Object Storage (COS)

- **COS Instance:** 1 (regional)
- **Buckets:**
    - **Option A:** 1 bucket per lab (×4) + 1 dev bucket
    - **Option B:** 1 shared bucket + per-lab prefixes + dev prefix
- **Lifecycle (optional):** Example rule (expire objects after 30 days) for non-prod

6) Data Movement (Phase 0 only)

- **Protocol: SFTP (SSH/22)** to GPU/utility VSI
- **Checksum:** sha256sum/openssl dgst for validation
- **COS CLI:** Upload/download sample commands in runbook

7) Tooling & Automation

- **CLI Workstation (local):** ibmcloud CLI + VPC/COS plugins
- **IBM Cloud Schematics:**
    - **Workspace: foundation** (VPC, subnets, SGs, key, COS)
    - **Workspace: lab** (GPU VSI, volumes, SG rules; per-lab variables)

- **Variables (lab workspace):** region, lab, gpu_profile, image, data_volume_gb, ssh_key_name, allowed_cidrs, cos_bucket/prefix, tags

8) Observability & Evidence

- **Billing & Usage:** Budget alerts (email); tag-filtered usage views
- **Evidence capture:** Screenshots/logs for provisioning, nvidia-smi/accelerator check, SFTP checksum, COS list/get, budget alert, cleanup inventory

Per-Lab stack (contents)

- 1× **GPU VSI**
- 1× **Public IP** (or bastion path)
- 1× **Security Group** (lab-scoped)
- 1× **Boot volume** (auto)
- 1× **Data volume** (size per t-shirt)
- 1× **COS bucket** *(or shared bucket prefix)*
- **Tags** (full set)
- **Runbook** (UI/CLI/Schematics + start/stop/terminate + SFTP + COS sync + cleanup)

Dev/MVP sandbox (contents)

- 1× **GPU VSI** (same profile family)
- 1× **Public IP**
- 1× **Boot volume** (auto)
- 1× **Data volume** (smaller t-shirt ok)
- **COS**: dev bucket or prefix
- **Schematics**: foundation apply once, lab apply with lab=dev
- **Runbook** authoring and dry-runs

IBM Cloud catalog & console paths (quick map)

Category	Service / Object	Console Path (short)
Compute	Virtual Server for VPC (GPU)	VPC Infrastructure → Instances
Network	VPC, Subnets, Public Gateways	VPC Infrastructure → VPC/Subnets/Gateways
Security	Security Groups, Rules	VPC Infrastructure → Security groups
Keys	SSH Keys	VPC Infrastructure → SSH keys
Storage	Block Storage (VPC volumes)	VPC Infrastructure → Volumes
Object Store	Cloud Object Storage	Storage → Object Storage
IAM	Users/Groups/Policies	Manage → Access (IAM)
Governance	Tags, Budgets/Usage	Manage → Account / Billing & usage
IaC	Schematics	Automation → Schematics

Naming (examples)

- **VPC:** lri-phase0-vpc
- **Subnets:** lri-phase0-subnet-a, lri-phase0-subnet-b
- **SG (baseline):** sg-ssh-inbound, sg-default-outbound
- **SG (lab):** sg-lab-<name>
- **GPU VSI:** lab-<name>-gpu-01
- **Volume:** lab-<name>-data-01
- **COS:** lri-phase0-<lab> or lri-phase0-shared + lab/<name>/