Document Version: 1.0
September 20, 2025
Jim Venuto

Phase 0 PoC Workload Identification

Core Requirement

Primary Need: Additional GPU bandwidth to address computational bottlenecks on LRI's on-premises HPC cluster
Phase 0 Scope: 5 researchers testing with **synthetic data only**
Purpose: Validate cloud GPU capability and justify investment in automated Phases 1-3

Identified Workloads for Phase 0 PoC (Synthetic Data)

1. Large Language Models (LLMs) & Text Analytics

- **Test Workload**: LLM zero-shot predictions using synthetic text data
- **Synthetic Data**: Generated text blurbs (non-PHI)
- **Validation Goal**: Confirm GPU performance for LLM inference
- **Manual Process**: Researcher provisions GPU VSI, transfers synthetic data, runs custom software

2. Single-Cell Omics Analysis

- **Test Workload**: Cell Ranger pipeline with synthetic genomic datasets
- **Synthetic Data**: Simulated single-cell genomic data
- **Validation Goal**: Test GPU acceleration for genomics workflows
- **Manual Process**: Manual data staging to Object Storage, GPU provisioning, pipeline execution

3. Machine Learning Frameworks

- **Test Workload**: TensorFlow/PyTorch model training on synthetic datasets
- **Synthetic Data**: Generated training datasets (images, numerical data)
- **Validation Goal**: Validate Python-based GPU ML capabilities
- **Manual Process**: Manual environment setup, data transfer via SFTP/Aspera, training execution

4. Large-Scale Data Analysis

- **Test Workload**: Batch processing of synthetic large datasets
- **Synthetic Data**: Generated computational datasets
- **Validation Goal**: Demonstrate cloud bursting potential
- **Manual Process**: Manual resource allocation, job submission, result retrieval

Phase 0 PoC Constraints

Synthetic Data Requirements

- **No PHI/Real Patient Data**: Eliminates compliance barriers
- **Representative Workloads**: Synthetic data mimics real computational patterns
- **Performance Validation**: Same GPU requirements as production workloads
- **Security Simplified**: Reduced security controls for PoC phase

Manual Process Testing

- Researchers manually provision GPU-enabled VSIs
- Manual data transfer using SFTP/Aspera
- Self-managed resource cleanup
- Manual cost tracking and monitoring

PoC Validation Objectives

Technical Validation

- GPU resources successfully provisioned and accessed
- Synthetic workloads complete successfully
- Performance benchmarks established
- Data transfer mechanisms functional

Process Documentation

- Time required for each manual step
- Complexity of manual provisioning
- Pain points requiring automation

- Support burden on researchers

Success Metrics for Phase 0 Workloads

Functional Success

- ✓ Each workload type runs successfully with synthetic data
- ✓ GPU acceleration confirmed for all test cases
- ✓ Data transfer methods validated (even if slow)
- ✓ Resource provisioning process documented

Business Case Development

- Quantified manual effort required per workload
- Identified automation opportunities for Phase 1
- Demonstrated need for integrated HPC solution
- Clear path from synthetic to production data requirements

Key Outcome

Phase 0 PoC with synthetic data demonstrates technical feasibility while exposing manual process inefficiencies, creating compelling justification for automated Phases 1-3 investment to support real research data and 240+ labs.