ORACLE
for Research

TECH TALK HOUSEKEEPING

- Today's webinar is being recorded. We will share the link to the recording with you via email after the event. The recording will also be made available to the Oracle for Research community.

- We invite your comments and questions, both about the tech topic being discussed and about the series more generally. Questions may be submitted using the Q&A box on your screen or you may ask questions directly using your microphone. When not asking a question, please mute your microphone.

- Questions may be asked during the presentation and we will also have a Q & A time at the end of the presentation when you can ask questions directly and engage in discussion.

- At Oracle for Research, we believe that research and innovation happen best when a diverse and thoughtful community is free to engage in respectful, compassionate, and open dialog. To that end, when asking a question or providing feedback, we ask that all participants be respectful, collaborative, and constructive.

Agenda

|	
Recap and asks	1. Shape selection , storage and performance testing 2. Architectures and automation 3. Performance benchmarks & Data science (platform vs Image usage)
Architectures for Researchers	1. Standard researcher architecture (Demo) 2. Cloud bursting architectures
OCI Automation	1. Instance scaling (demo) 2. Resource stacks and Terraform (demo) 3. Programmatic automation with OCI CLI and API
Github and links	1. Technical How-Tos 2. Repositories and Images 3. Researcher collaboration
Q & A	Q & A What works best? Any researcher wish list?

A GENERIC WORKFLOW TO DECIDE OPTIMAL SHAPES
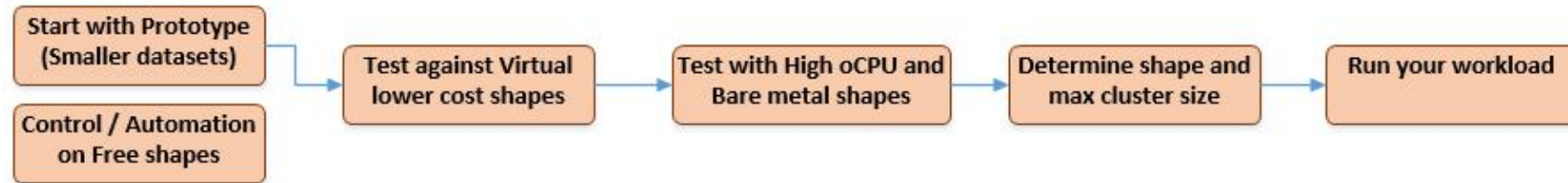
SHAPE TESTING HIERARCHY FOR NON-GPU USAGE USE-CASE
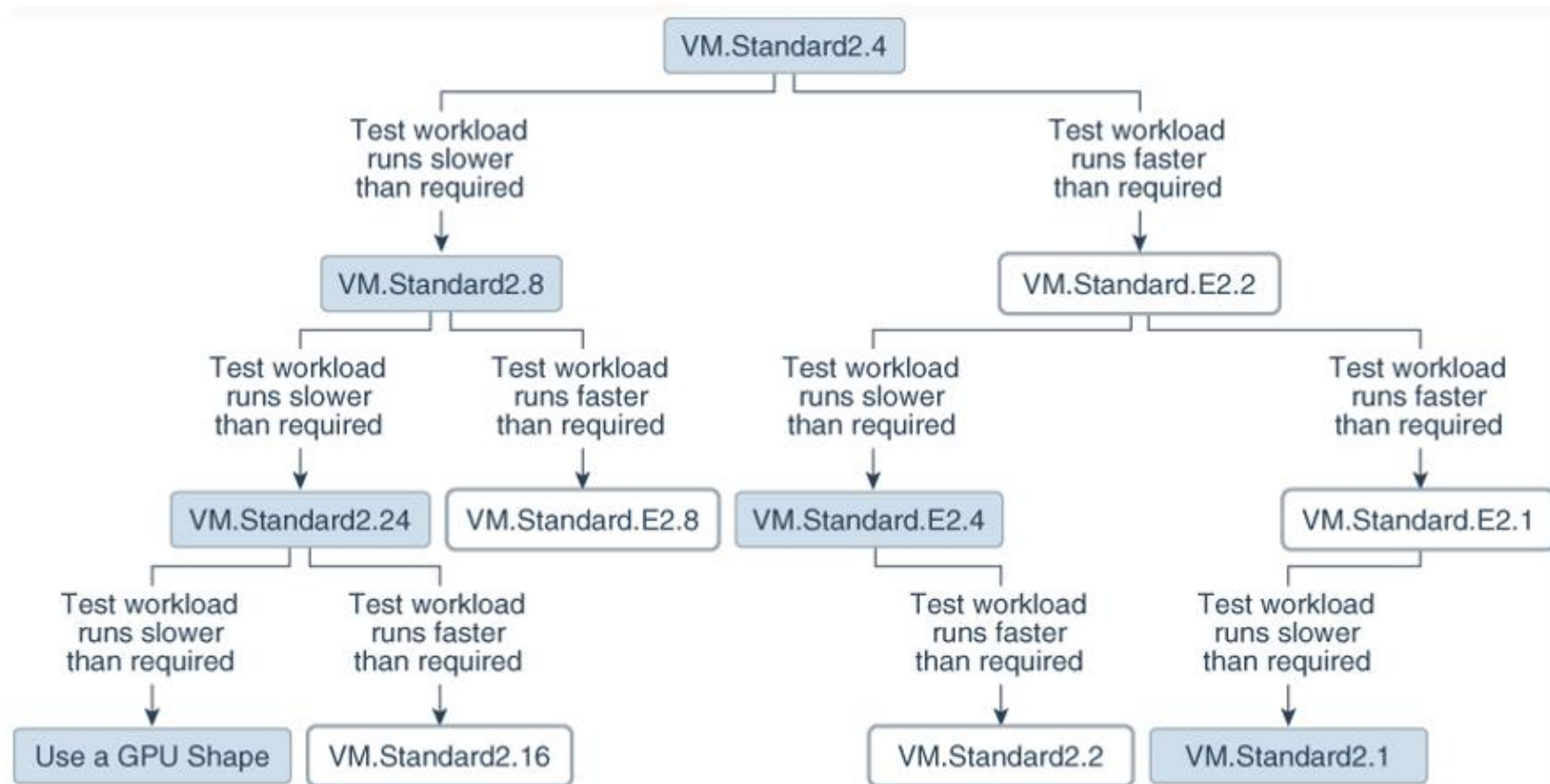
NOTE : This should be used as a guideline and may differ based on researcher data and computation scenario
Courtesy : Oracle architecture center
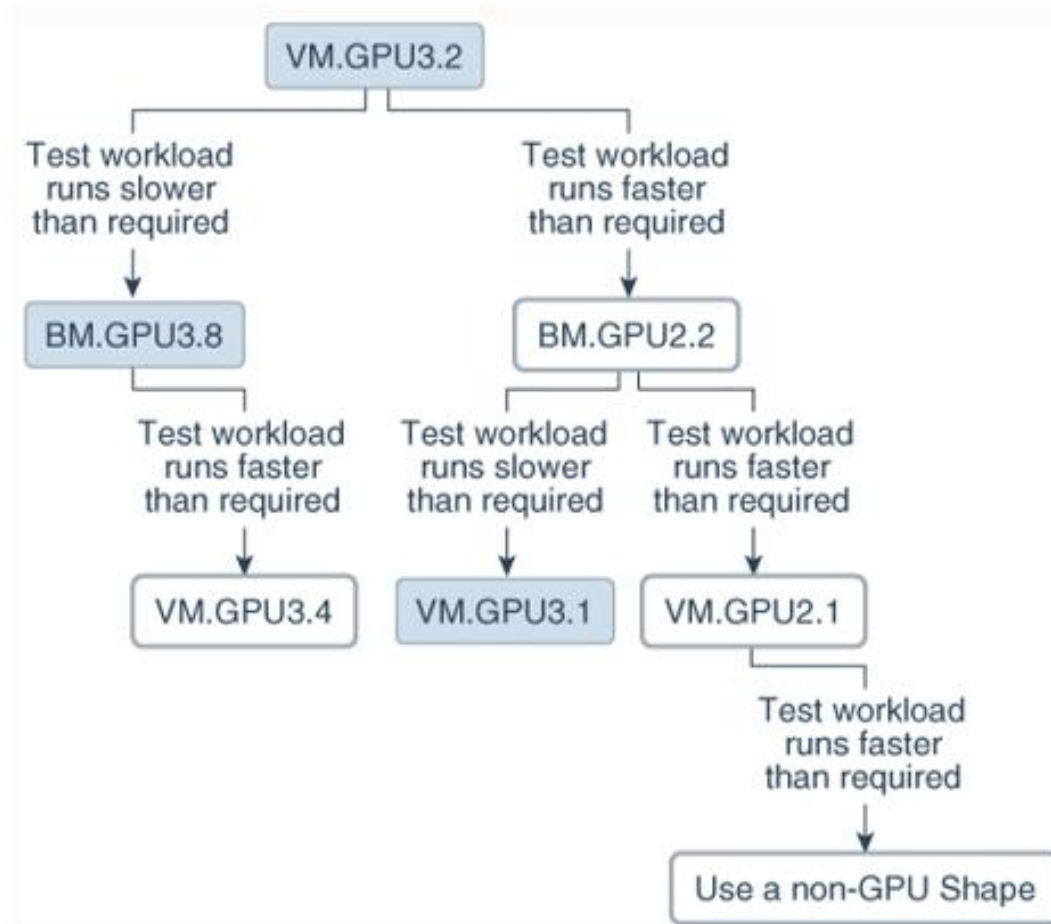
SHAPE TESTING HIERARCHY FOR GPU USAGE USE CASE

NOTE : This should be used as a guideline and may differ based on researcher data and computation scenario
Courtesy : Oracle architecture center

OCI Standard cluster architecture for Researchers
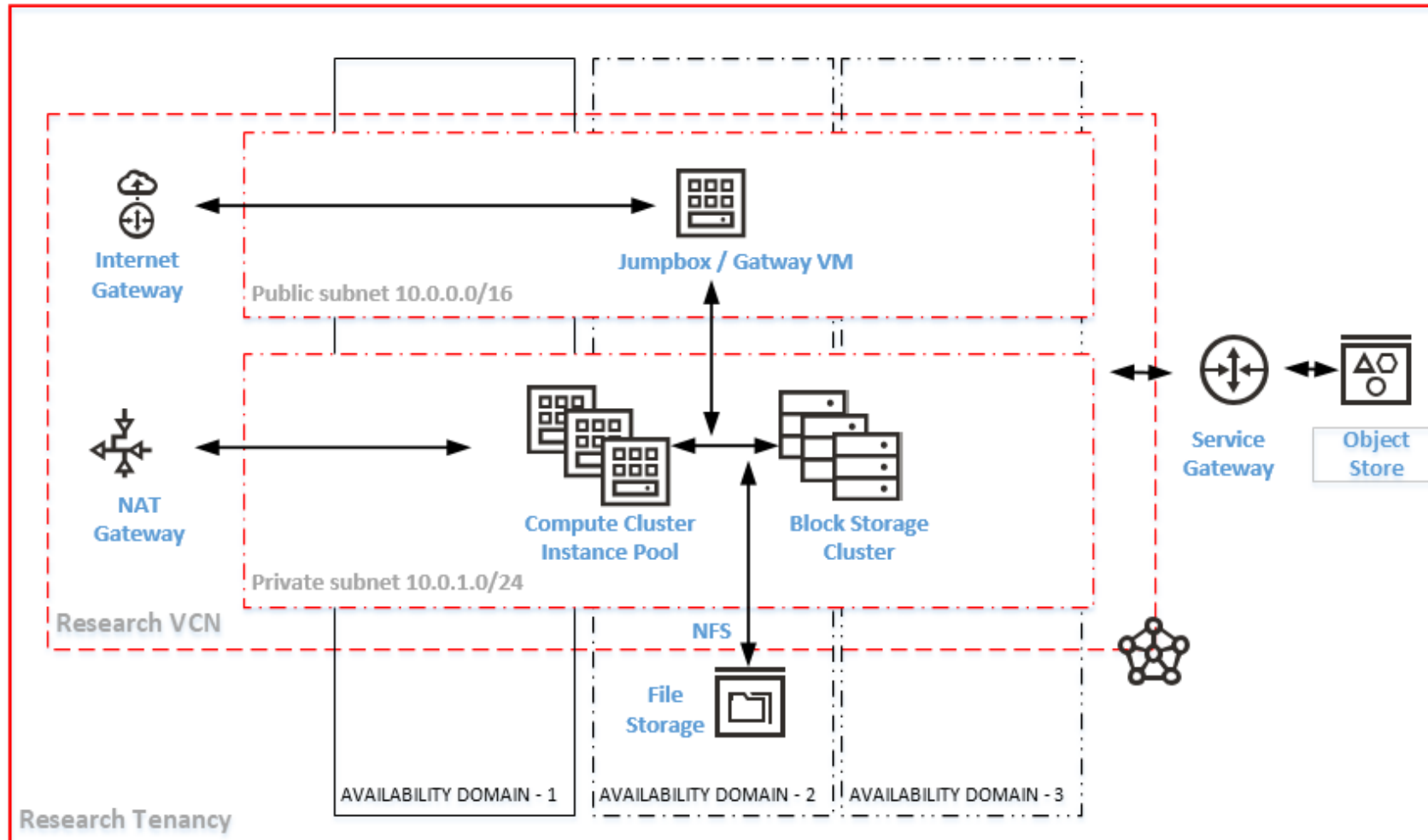
Component	Recommendation
Jumpbox VM	Use Free-Tier VM
CPU/GPU Cluster	Manual Build + Block storage
CPU/GPU Cluster	Automated Build + Block storage
HPC Cluster	Instance Pool + RDMA
NAT Gateway	Download software to pvt subnet
File storage	For cross-AD exports

Cloud bursting architecture for Researchers
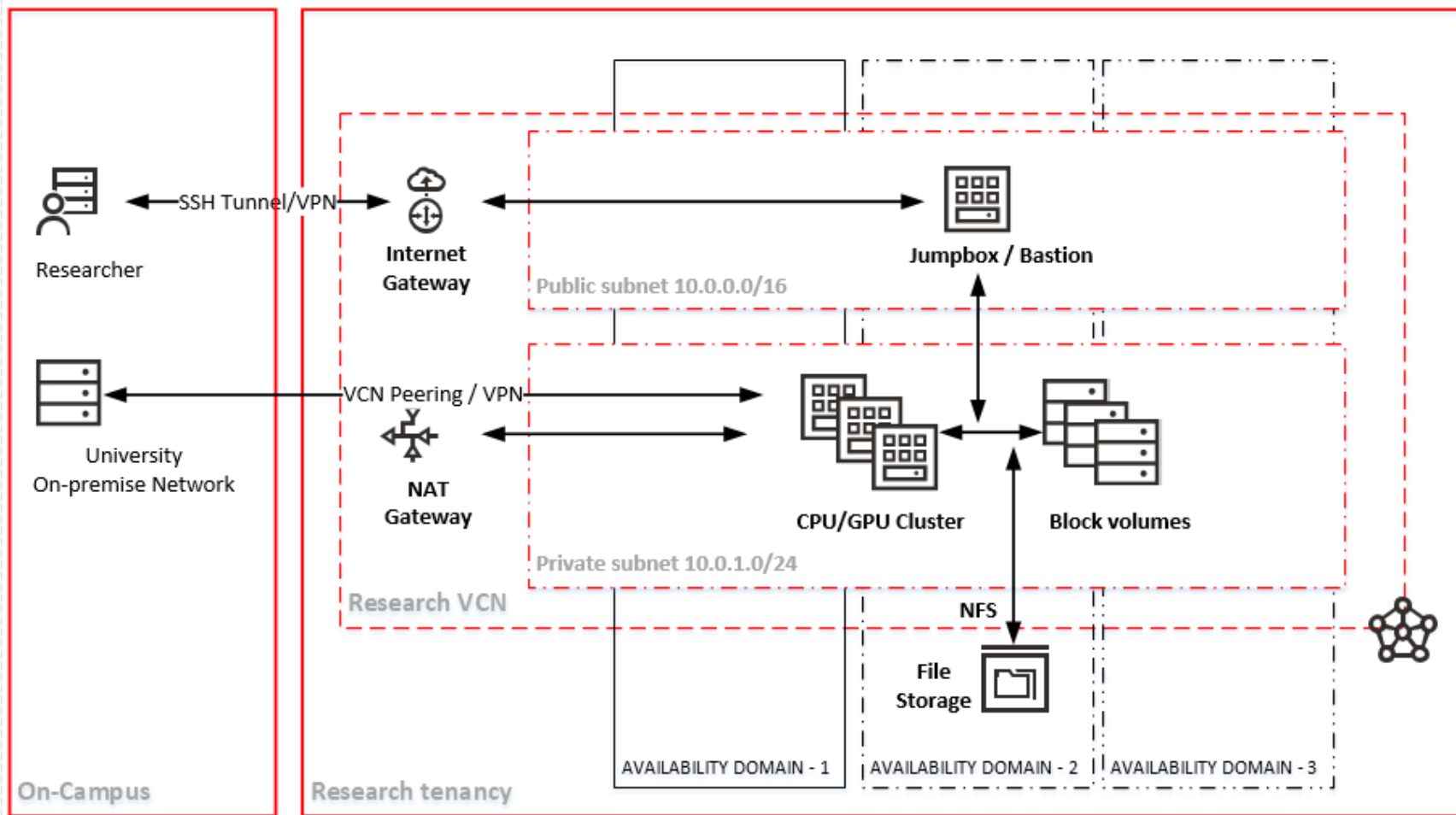
Key features
1. OCI access – (SSH Tunnel to Bastion to GPU or VPN)
2. CPU/GPU bursting (OCI Scheduling) - OCI Instance pooling + auto-scaling
3. CPU/GPU bursting (Campus Scheduler) - OCI CLI + Stacks or OCI API
4. NAT Gateway - Download software from public repositories
5. File storage - Cost effective data sharing within or across AD

Architecture use-cases

Standard Architecture	Cloud bursting Architecture

Standard Architecture

Recommended when
1. Quickly standup the application / cluster in Oracle cloud
2. Entire application would run in cloud
3. Database centric ML applications (Autonomous DB / Roracle)
4. Dynamic scaling of cluster nodes
5. Hybrid cluster with multiple shapes
6. Different workload types (Emb parallel / Tightly-coupled)
7. Simple SSH but secure access to computational VMs
8. All data contained in the cloud
9. Benchmarking against on-campus / clouds
10. Using Oracle cloud automation framework & scripts

Not recommended when
1. Computation in cloud but data is resident outside Oracle cloud
2. On-campus resources are sufficient for workloads
3. Non-research operational projects
4. Frequent transfer of code and data in/out from Oracle cloud

Cloud bursting Architecture

Recommended when
1. A percentage of the workload can be run on Oracle cloud
2. Infrequent cloud usage
3. On-premise workload scheduling
4. On-premise cluster is overloaded
5. Reduce workload by moving an application/data to OCI
6. Smaller computational data sets
7. Performance benchmarking against on-campus / other clouds
8. Scheduling automation is tested and working
9. Containerized applications & canary testing
10. Benchmarking against OCI shapes
11. Embarrassingly parallel HPC jobs

Not recommended when
1. Large data migration during bursting
2. Tightly-coupled apps (Cross node communication)

Why standard architecture is important?

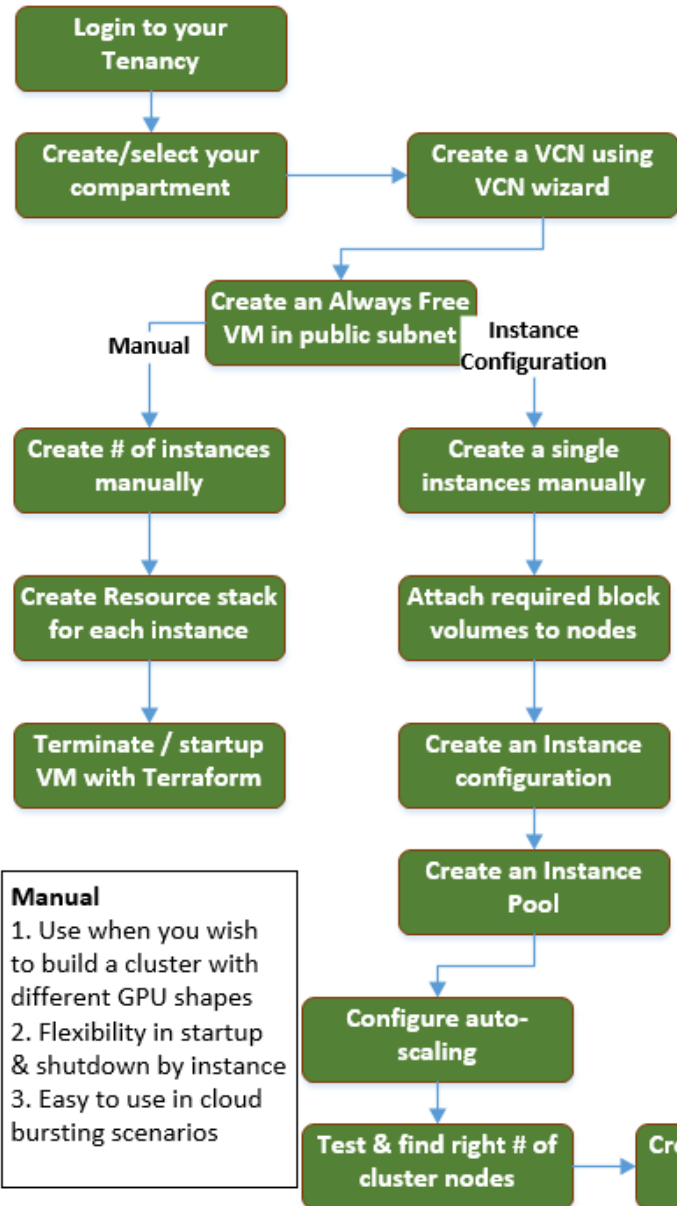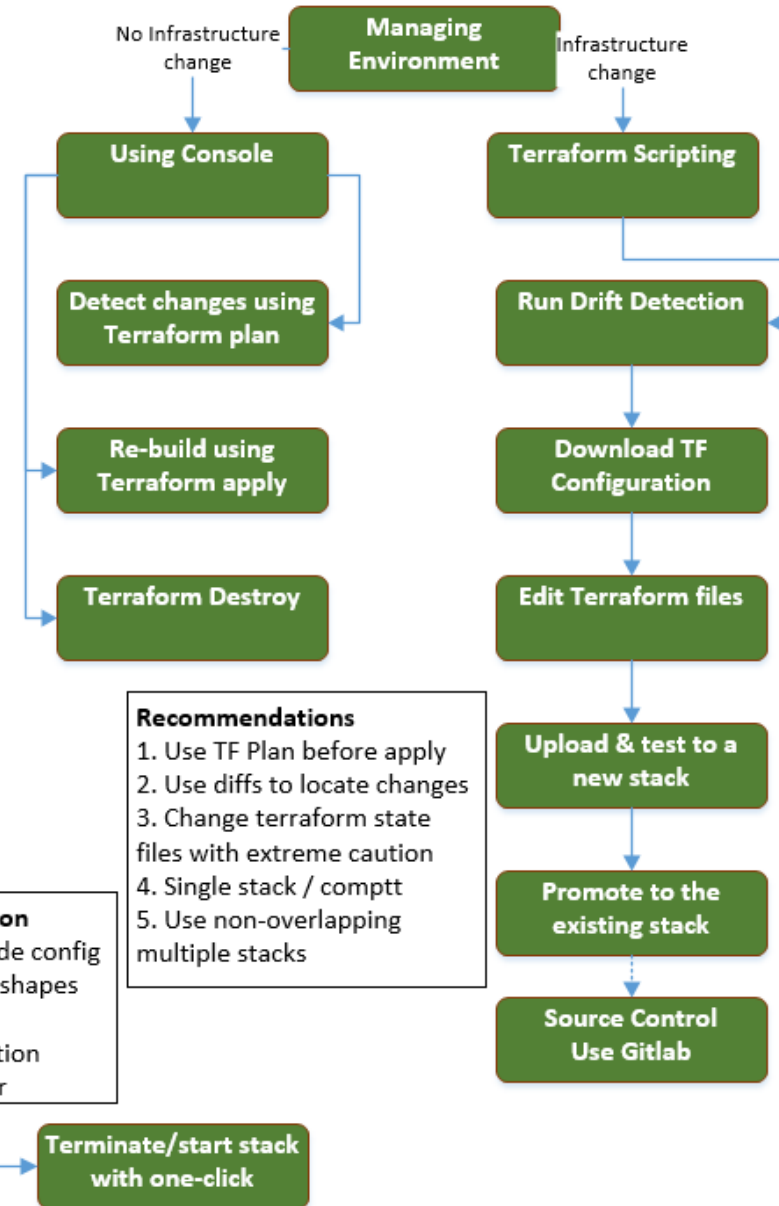|	
Being Standard	1. Quick and easy to implement 2. Easier resource management 3. Single & common reference for all researchers
Automation	1. One-click standup and shutdown of instance through console (Stacks and Terraform) 2. Quick and easy configuration management 3. Programmatic automation (CLI / API) through a Free-tier VM
Better resource utilization	1. Planned utilization of CPU/GPU/Storage resources 2. Effective cluster and computational planning 3. Better credit usage and service limit allocation
Better benchmarking	1. Benchmarking against similar hardware & computational specifications 2. Leverage performance tips from Oracle 3. Testing same workload against higher end shapes
Quicker support	1. Issues can ne resolved faster 2. Issues can be replicated/tested without access to researcher tenancies 3. Comparative benchmarks from the community can help tune your implementation
Security	1. Private subnet insulates your computation VM and data 2. Access only through a jump box / secure gateway VM 3. NAT access provides direct internet access for code / patch downloads

Automation Setup & management workflow

Initial Creation

Login to your Tenancy

Create/select your compartment → Create a VCN using VCN wizard

Create an Always Free VM in public subnet

Manual | Instance Configuration

Create # of instances manually

Create Resource stack for each instance

Terminate / startup VM with Terraform

Create a single instances manually

Attach required block volumes to nodes

Create an Instance configuration

Create an Instance Pool

Configure auto-scaling

Test & find right # of cluster nodes → Create compartment level stack → Terminate/start stack with one-click

Manual
1. Use when you wish to build a cluster with different GPU shapes
2. Flexibility in startup & shutdown by instance
3. Easy to use in cloud bursting scenarios

Instance configuration
1. Need identical node config
2. Config on various shapes
3. OCI Auto-scaling
4. One click termination
5. Right sizing cluster

Managing Changes

No Infrastructure change | Managing Environment | Infrastructure change

Using Console

Detect changes using Terraform plan

Re-build using Terraform apply

Terraform Destroy

Terraform Scripting

Run Drift Detection

Download TF Configuration

Edit Terraform files

Upload & test to a new stack

Promote to the existing stack

Source Control Use Gitlab

Recommendations
1. Use TF Plan before apply
2. Use diffs to locate changes
3. Change terraform state files with extreme caution
4. Single stack / comptt
5. Use non-overlapping multiple stacks

Oracle for Research github – https://github.com/OracleForResearch

Technology How-Tos	1. Short 2-page How-Tos (Categorized) 2. Important OCI links (Supplemental pages) 3. Aligned and dedicated to researcher needs
Image Sandbox	1. AIML Sandbox & AFNI (Base reference images with CPU/GPU versions on OL7 and Ubuntu) 2. Actively worked on by Oracle for Research 3. OFR will develop more based on active researchers participation and feedback
Researcher Collaboration	1. Contributions (benchmarks, test results, data) 2. OFR and OCI reviews 3. Researcher publications repository
Images & Applications	1. One stop place for Oracle provided images 1. Molecular dynamics 2. Oracle cloud Images for Research only 2. Open source and proprietary images 3. Researcher image contribution 3. Agro & farm data images
Technology Talks	1. Oracle for Research presentations 2. Relevant Tech presentations for researchers 3. Product updates
Coming up..	Benchmarking and guidelines Choosing the right data science platform. Researcher images for sharing and publication What we need? – Your feedback / what makes sense?

ORACLE
for Research

TECH TALK:

Architecture and Automation for Researchers

Questions? Comments? Feedback?

Contact us!

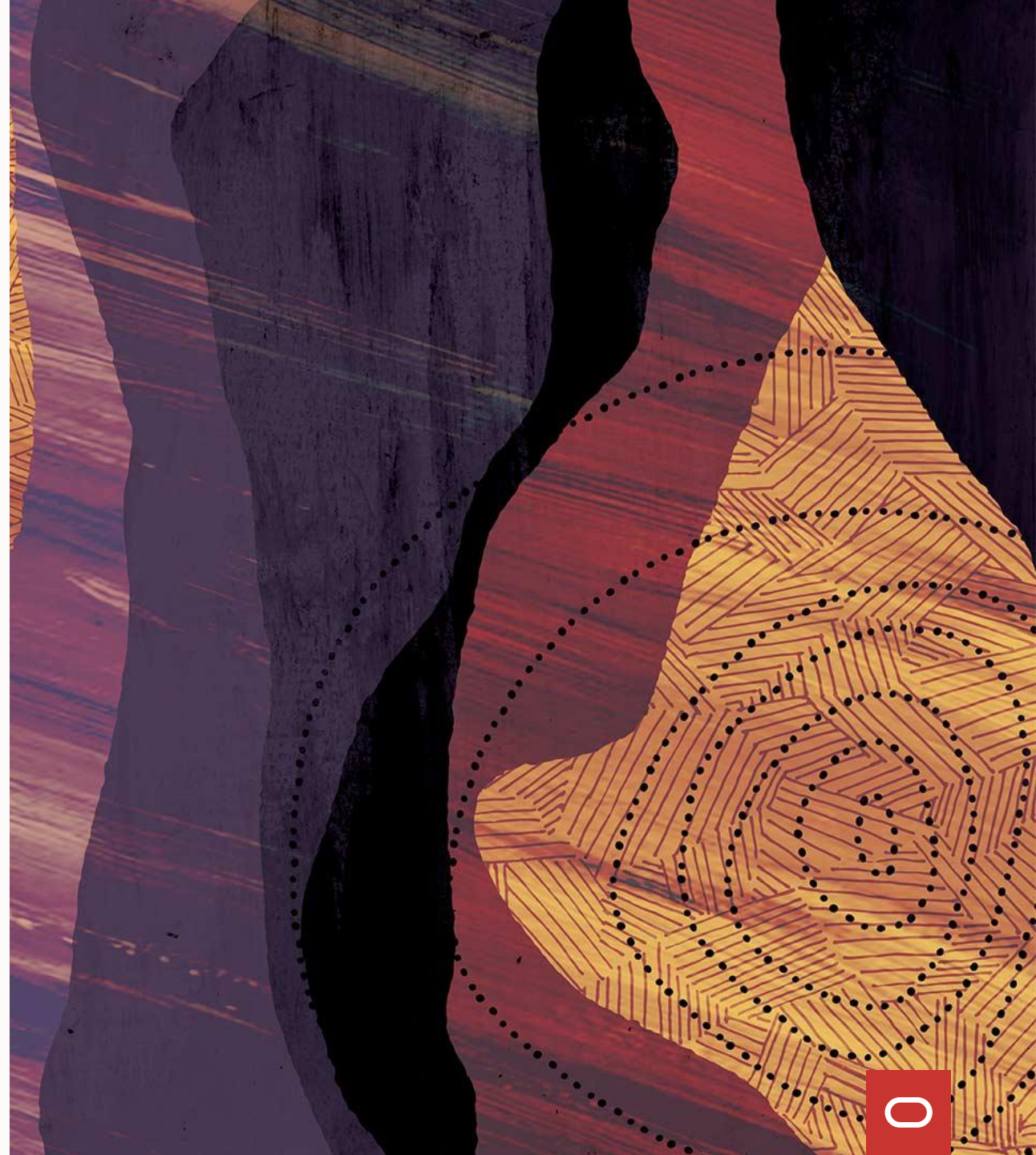**Website:** oracle.com/oracle-for-research/
Twitter: @OracleResearch
Email: OracleForResearchTech_ww@oracle.com

Next Tech Talk: November 13, 2020, 10:30AM EDT

Parking lot

For Reference

Oracle for Research Tech Talk Oracle Cloud Topics

Foundational	Oracle cloud – Getting you started and running Cloud instances and cloud storage options Migrating data and running computations
Architecture	Reference architecture patterns for researchers New features updates and recommended practices Performance benchmarks and data
Tools and automation	Tool selection, version and guidance Image repositories, Terraform and interfaces
HPC and cluster	High performance computing, workload classification, parallelization Cluster setup, utilization and monitoring
Machine and Deep Learning	Model selection guidance
Researcher guidance	Functional and data guidance and curation support Industry models with research computing

Instance Type	Shape series	Shape	Purpose
Virtual	Always Free	VMStandardE2.1Micro	Automation control, gateway,configurations
	Standard	VMStandard1.1~1.16	Low workload testing / Image builds / installs
	AMD (Gen 2)	VMStandardE2.1~2.8	Prototype workload testing
	DenseIO	VMDenseIO2.x (NVMe)	Heavy IO workload testing
	GPU (P100)	VM.GPU2.1	AI / ML or other GPU prototype testing
	GPU (V100)	VM.GPU3.1~3.4	Tensor core AI / DL workloads
	Intel Skylake (Fixed)	VM.Standard2.1~2.24	Workloads to save on credits
	AMD Rome (Flex)	VM.StandardE3.Flex	Benchmarking / price-performance
Bare metal	HPC	BM.HPC2.36 (NVMe)	CPU+high throughput for HPC workloads
	AMD (Gen 3)	BM.StandardE3.128	High CPU/throughput workloads
	Standard	BM.Standard1.36/B1.44	Low CPU/RAM utilization at lowest BM cost
	AMD (Gen 2)	BM.StandardE2.52	Best price-performance for BM workloads
	AMD (Gen 3)	BM.StandardE2.64	Best Gen3 price-performance for BM workloads
	DenseIO	BM.DenseIO2.52 (NVMe)	Best price performance for IO intensive workloads
	GPU (P100)	BM.GPU2.2	Benchmarking pascal based GPU workloads
	GPU (V100)	BM.GPU3.8	Best price performant for large GPU workloads
	GPU (A100)	BM.GPU4.8	Fastest GPU – large DL applications (pre-GA)

Oracle cloud storages

Storage type	Features	Recommended usage
On Campus storage	Good for on-campus data processing requirements Data could be distributed (laptops) or centralized Quick data retrieval for on-campus computations	Store data for on-campus computational purpose Store data if storage is available and it is already paid for
Object storage archive	Unlimited data storage in Oracle cloud Low cost ($0.0026/GB/Month) by consumption Better if cost/GB is lower than on-campus storage	Not so good for frequent large data retrievals Not readily available for computational purpose
Object storage standard	Unlimited data storage with faster access than OS archive Costs (0.0255/GB/Month) by consumption Good for frequently accessed data across cloud tenancies Secured and encrypted data at rest and in transit (https)	Store data backups for quick downloads in cloud / campus Store large data volumes at relatively lower cost Store data that is infrequently processed
Block volume	Most common storage for compute/databases Mountable across multiple instances within an AD Cost (0.0255/GB/month) – based on total volume in GB Supports parallel filesystem & best price performance	Leverage to store computational data for most loads Extend storage / instance as needed Fill up allocated capacity to save on costs
File system storage	NFSv3 unlimited file system storage mountable across AD Higher cost (0.3/GB/month) on consumption Good for file sharing across tenancies and OS Performs linearly or better with higher data set size	Not good for cross AD data transfer in computational cycle Good for moving large volumes of data quickly across AD Use sparingly
Local NVMe	Highest IOPS and throughput & good for IO intensive loads Higher cost (built into compute) Non-persistent data Part of DenseIO and HPC shapes	Use for IO intensive parallel CPU workloads

Useful links

Getting Started
Key concepts and terminology
Signing in to console, Sign-in options and changing your password
Setting up your tenancy
Tutorial – Launching your first Linux instance
Tutorial – Launching your first windows instance
Object storage and Pre-authenticated access
Image import and export
File storage system concepts
OCI Hands on labs
New features and navigation updates
Oracle cloud Free tier and FAQ
Custom key generation with puttygen or ssh-keygen
Frequently asked Questions
Getting help and contacting support

Identity federation
Federated identity for single sign-on
Migrating from on-premise to Oracle identity service

Databases and moving data
Oracle Autonomous databases and Tools
MySQL and NoSQL Services
Key concepts and terminology
Migrating databases to cloud
Loading data to autonomous with OCI Functions
Single-click move to autonomous

Data science and AI/ML
Oracle Data science platform and Tutorials
Genome analysis toolkit
Julia AI/HPC GPU Image
NVIDIA images and NVIDIA GPU image
Building a ML sandbox on Oracle cloud
Setting up an open-source ML and AI Environment
Machine learning autonomously

High performance computing (HPC)
Oracle HPC Cluster and Oracle HPC File system
NVIDIA GPU Cloud machine image
Oracle Linux 7 Cluster Networking Image
Oracle marketplace slurm image (HPC + Slurm combo)
Oracle cloud slurm image
Github OCI-HPC
Enabling HPC Cluster networking
Deploy High performance computing on Oracle cloud Infrastructure
Deploy scalable and distributed file system using Lustre
Deploy BEEGFS parallel file system
UoB Cluster in the cloud
Cluster in the cloud - github
Molecular dynamics NAMD runbook and GROMACS runbooks

Usage, billing and credit control
Oracle cloud storage costs
Resource billing for stopped instances
Oracle cloud universal credit PaaS and IaaS service descriptions

Oracle for Research Tech Talk Series Will Cover:

Technology training

Reference architectures
Best practices
Tools and automation
Cost control

Collaboration

Q & A Live discussions
Tips and tricks library
Community forum
Meet Oracle experts

Product announcements

OFR technology updates
OCI product updates
Images and containers
Public data

Researcher for researchers

Technology Innovations
Researcher publications
Benchmarks
Lessons learned

OCI Automation for Researchers

Common Researcher Issues	1. Where to start? 2. What shapes, images and storages do I need? 3. How to manage my credit allocation effectively?	1. Any automation to terminate idle instances? 2. How do I scale or burst my workload? 3. Can I use my on-campus identity to login?
OCI Architectures for Researchers	1. Standard researcher architecture – Bastion Free Tier VM + Private subnet for compute 2. Shape and storage selection guidelines 3. Scaling and cloud bursting architectures	
Automation and usage control	Instance scaling (demo) Resource stacks and Terraform (demo) Tooling with OCI CLI	
Credit control mechanisms	Cost analysis and cost reports Setting budgets and alerts	
Github and links	Overview of the repositories Collaboration with Oracle and Researchers	
Q & A	Q & A What works best and researcher wish list?	