

supercomputing made super human™

Next Generation Infrastructure Software for Hybrid Cloud HPC

About Nimbix

- Nimbix is the pioneer of purpose-built cloud computing for machine learning, Al and HPC applications
- Powered by JARVICE[™], the Nimbix Cloud provides high-performance software as a service, dramatically speeding up data processing for Energy, Life Sciences, Manufacturing, Media and Oil & Gas software applications
- Its JARVICE[™] XE Enterprise HPC platform brings the power of JARVICE[™] and the Nimbix Cloud to any on-premises cluster or hybrid cloud environment
- 2019 HPCWire Readers' Choice Award winner for Best Use of High-Performance Data Analytics & Artificial Intelligence
- Recently named the Official Compute Partner of the US Sailing Team
- Trusted by Fortune 500 companies and world-class partners







The Growing Demands of Computation



Source: Hyperion

- HPC demand is accelerating
- From traditional supercomputers and clusters to • HPC in the Cloud
- Enterprises adding new workloads, including deep • learning/machine learning which demand HPC and accelerated infrastructure
- Increasing complexity to consume optimized • software when its needed



Traditional Approach of Running HPC Applications

- Traditional HPC Clusters
 - Attach to local high-performance storage
 - HPC workload managers/schedulers: Slurm, LSF, SGE, PBS...
- User scripts, jump nodes, in-house workflow automation, customized environments





New Challenges to Traditional HPC



- Scaling Beyond Single Cluster, Single Location
- Replicating applications and workflows beyond single cluster
- Performance
 management on different
 infrastructures
- Increased complexity of multiple hardware architectures, multiple software stacks
- Converging with Kubernetes for AI/Deep Learning or traditional IT?





5

Early years of Cloud Computing for HPC

🎁 AWS 🛩 Ser	vices	* 🕕 EC2	📢 S3 🢡	IAM Edit -	mike udin 🗸	N. Virginia 🗸	Suppo	irt
EC2 Dashboard	Lau	inch Instance	Connect Ac	tions ~		Ð	• •	3
Tags	Q Filter by tags and attributes or search by keyword					1 to 20 of 20	> >1	
Reports								
Limits	01 -	Instance ID -	Instance Type -	Availability Zone ~	Instance State *	Status Checks		1
INSTANCES Instances Spot Requests Reserved Instances Commands Dedicated Hosts		i-cf2c3b7f	g2.2xlarge	us-east-1b	running	2/2 checks	pas	1
		1-262e3996	g2.2xlarge	us-east-1b	🔵 running	2/2 checks	pas	1
		i-c72d3a77	g2.2xlarge	us-east-1b	running	2/2 checks	pas	1
		i-b82d3a08	g2.2xlarge	us-east-1b	🔵 running	2/2 checks	pas	1
		i-bc2d3a0c	g2.2xlarge	us-east-1b	👴 running	2/2 checks	pas	1
		i-212e3991	g2.2xlarge	us-east-1b	running	2/2 checks	pas	7
IMAGES AMIs Bundle Tasks		i-f82d3a48	g2.2xlarge	us-east-1b	running	2/2 checks	pas	1
		I-c62d3a76	g2.2xlarge	us-east-1b	running	2/2 checks	pas	ì
		i-c42c3b74	g2.2xlarge	us-east-1b	running	2/2 checks	pas	1
 ELASTIC BLOCK STORE Volumes Snapshots 		i-c12c3b71	g2 2xlarge	us-east-1b	running	2/2 checks	pas	1
		i-be2d3a0e	g2.2xlarge	us-east-1b	running	2/2 checks	pas	1
		i-cf2d3a7f	g2.2xlarge	us-east-1b	🔵 running	2/2 checks	pas	1
NETWORK & SECURITY Security Groups Elastic IPs Placement Groups Key Pairs Network Interfaces LOAD BALANCING Load Balancers		i-fc2d3a4c	g2.2xlarge	us-east-1b	running	2/2 checks	pas	1
		i-3f2e398f	g2.2xlarge	us-east-1b	running	2/2 checks	pas	ï
		i-e32d3a53	g2.2xlarge	us-east-1b	🔵 running	2/2 checks	pas	1
		i-cd2d3a7d	g2.2xlarge	us-east-1b	running	2/2 checks	pas	ï
		i-b52d3a05	g2 2xlarge	us-east-1b	running	2/2 checks	pas	1
		i-ce2c3b7e	g2.2xtarge	us-east-1b	running	2/2 checks	pas	1
		i-fa2d3a4a	g2.2xlarge	us-east-1b	running	2/2 checks	pas	1
		1-252e3995	g2 2xlarge	us-east-1b	running	2/2 checks	pas	Ì
 AUTO SCALING Launch Configurations 							all	

- For the first decade, cloud was completely impractical for HPC
 - Highly inadequate compute resources
 - Very complex to set up
 - Cost a fortune for attempting to run real HPC apps
 - Data security and data movement challenges created friction
- On Premises workflows & applications highly customized and not easily portable to a cloud environment





Hyperscale Clouds Today

- Growing heterogeneity in cloud resources
 - New CPUs: ARM, AMD, Power
 - Accelerators: GPUs, FPGAs, AI Chips
- Higher performance infrastructure and interconnects: InfiniBand, bare metal
- Still complex to consume and engineer in house workflows without being tethered to single provider among hyperscalers











Desired State: Hybrid On-Prem/Cloud for Enterprise HPC

Balance and Optimize economics of on prem clusters
 and cloud use

- Preserve existing applications and workflows
- Limit customizations for any given infrastructure or cloud platform make it adaptable
- Optimize for policy-driven computing
- Full enterprise reporting across all resources





Modern Application Deployment for Accelerated HW



9



Proliferation of Containers

- "Inter-modal" method for packaging, distributing, and deploying applications and dependencies onto arbitrary infrastructure
 - Consistency and integrity across platforms
 - Repeatable mechanisms
- Displacing traditional VMs for application image management
- Every major platform provider supports
 Docker containers
- Enables applications to take advantage of bare-metal infrastructure capabilities for high performance workloads







Containerized HPC Application Comparison



* Low-latency fabric (e.g. InfiniBand, RoCE), requires HPC-capable container platform, such as JARVICE



Enterprise HPC Platform JARVICE™ XE



- Brings industry leading HPC container technology to your datacenter, your cloud or any cloud, anywhere in the world
- Delivers and synchronizes entire HPC/ML application ecosystems to your clusters via HyperHub[™]
- Rapidly deploys and synchronizes your applications and workflows to your HPC environments around the world
- Enables rapid adoption of new technologies and new infrastructure even before hyperscale clouds make them available





HPC on Kubernetes with JARVICE XE





13

JARVICE XE Features

- Unified Kubernetes infrastructure
- HyperHub application ecosystem syncs commercial, custom or in-house applications and workflows globally
- Enterprise security integrates with Active Directory and SAML for seamless, secure authentication and authorization
- Comprehensive enterprise analytics and reporting across all HPC workflows and jobs across all global clusters
- Seamless global updates of both platform and applications
- Secure containerized architecture
- Compatible with every major cloud provider and any HPC compute infrastructure



JARVICE XE Benefits

Simplified User Experience

• Simple point-click-run workflows on any infrastructure

Simplified Administration for HPC Admins and Enterprise IT

- Unified SaaS for HPC and Deep Learning
- Unified platform for multi-cloud, multi-datacenter deployments

Streamlined Application Distribution and Deployment

- Platform-as-a-Service (PaaS) continuous integration and deployment for compute intensive workflows using your in-house algorithms, applications, or customization of commercial applications
- Automatic synchronization with HyperHub™

Reduced Infrastructure Complexity

• Unified infrastructure layer with Kubernetes





Next Gen Platform Software Optimizing HPC Costs

- Replace alternative common solution stacks that combine commercial schedulers
- Deploy on open source Kubernetes software that works with existing Linux infrastructure, eliminating the need for expensive HPC-specific cluster management or infrastructure management solutions
- Reduce/Eliminate the need for dedicated HPC application engineering resources, since application management is selfservice and does not require "installing" anything on servers, everything is containerized
- Allow IT/HPC admins to focus resources on more general projects rather than having to specialize on HPC management
- Improve user experience and access from anywhere means fewer helpdesk calls and less management of end user clients, a web browser is all that is needed
- Seamlessly burst to other infrastructure when needed, avoiding the need to procure more HPC hardware to add temporary, project-based, or seasonal capacity to core clusters



Key Attribute Considerations for Next Gen HPC Software Platforms





HyperHub[™] Application Marketplace

- Distribute, deploy, and run applications and workflows globally on any cloud or on premises datacenter
- Built-in support for autoscaling apps for high performance clusters
- Native accelerator support for both GPU and FPGA infrastructure
- HPC and supercomputing applications as point-and-click workflows
- Automatic synchronization of marketplace to any JARVICE XE deployment









HyperHub[™] Ecosystem

- Offers a curated catalog of Simulation, AI/ML/DL software optimized for HPC
- Combined with JARVICE XE enables users to access the entire catalog of applications with virtually unlimited supercomputing power on any infrastructure
- Customers can customize the compute catalog with their own workflows and applications



Recommended Reading

Download the White Paper: HPC on Kubernetes

https://www.nimbix.net/case-studieswhite-papers NIMBIX

20

WHITE PAPER

HPC on Kubernetes

A practical and comprehensive approach

Leo Reiter CTO Nimbix, Inc.



Contact Us

THANK YOU!

- Experience the benefits of JARVICE XE today with a complementary trial
 - Reach out to your Nimbix or Partner representative for information on Nimbix Cloud and JARVICE XE availability: <u>sales@nimbix.net</u>
 - Visit us at <u>www.nimbix.net</u>



linkedin.com/company/nimbix



facebook.com/nimbix





