



W H I T E P A P E R

Point and Click HPC: The XTREME-Stargate IaaS Platform

Point and Click HPC: The XTREME-Stargate IaaS Platform

Executive Summary

High Performance Computing (HPC), Data Analytics (DA), and Deep Learning (DL) applications offer leading edge capabilities to researchers and scientists, but these users don't always have the higher-level knowledge or the bandwidth required to configure a cluster to run them. Cluster administrators are often needed to deliver and maintain complex hardware and software environments in order to support HPC/DA/DL applications.

Cloud computing offers one point of entry for these applications as it eliminates the need to set up actual hardware, but cloud vendors typically only go so far in terms of resource capabilities (i.e., getting your applications to run is still your job). Users and administrators are still tasked with the problems of final configuration and handling support issues.

In light of this, XTREME-D developed XTREME-DNA to keep the user out of the cluster management business. Using a web-based interface, users can build HPC/DA/DL clusters in both private and public clouds (i.e., Azure and AWS) in as little as 10 minutes.

XTREME-D's next step in offering seamless HPC/DA/DL computing is the XTREME-Stargate gateway platform, or cluster portal. This on-prem device acts as a cluster "head node" and provides secure and fast access to baremetal clusters that are configured using enhanced XTREME-DNA technology. XTREME-Stargate offers a "super head node" for HPC cloud clusters, the ability to access shared or dedicated bare metal cloud resources, support for HPC accelerators and container technology, dedicated data centers, hybrid cloud capability, and enhanced security.

The following cost savings can be expected when deploying a local XTREME-Stargate appliance:

1. Rapid cluster spin-up/spin-down
2. A "pay-as-you-go" cloud model
3. Choice of price-to-performance by selecting between public cloud, hybrid cloud, and private baremetal hardware, including GPU, FPGA, and NVMe
4. Elimination of virtually all cluster administration costs
5. Dramatically increased user efficiency

High Performance, Not High Maintenance

When mentioning High Performance Computing, images of server racks, networking, system administration, package installation, and text terminals often come to mind. These “features” have long been the hallmark of high end parallel computing. The growth of scalable Data Analytics (i.e., Apache Spark) and Deep Learning (i.e., Tensorflow) have widened the playing field for typical uses of cluster computing.

One challenge that HPC, DA, and DL end users face is to keep focused on their science and engineering and not get bogged down with system administration and platform details when ensuring that they have the clusters they need for their work. It has often been said that if scalable cluster computing can become more turnkey and user-friendly (and less costly), then the market will expand to many new areas.

Cloud computing presents yet another option for accessing clustered computing, bringing with it its own challenges. The on-demand nature of cloud offers an attractive pay-as-you go option for some users (i.e., a “pay for only what you use” option). With the exception of quick instance spin-up, building clusters by hand in the cloud does not necessarily make cloud-based clusters any cheaper to use and maintain than on premises systems.

Keeping the end user closer to the application, with less wasted time, should be a priority in the HPC, DA, and DL spaces.

Point and Click HPC: XTREME-DNA

XTREME-D developed XTREME-DNA to keep the user out of the cluster management business. Using a web-based interface, users can build HPC/DA/DL clusters in both private and public clouds (i.e., Azure and AWS) in as little as 10 minutes. In addition to using a cloud pay-as-you-go cost model, XTREME-DNA eliminates the need for expensive administrators and HPC architects. This capability allows end users to manage the entire spin-up, configure, execute, and spin-down with just a few mouse clicks.

XTREME-DNA works by using HPC templates that describe how the cluster and application are to be configured. Plug-in templates are pre-configured and a wide selection of templates are available, including for applications such as Tensorflow, OpenFoam, and OSPRay. XTREME-D can also create custom templates for users.

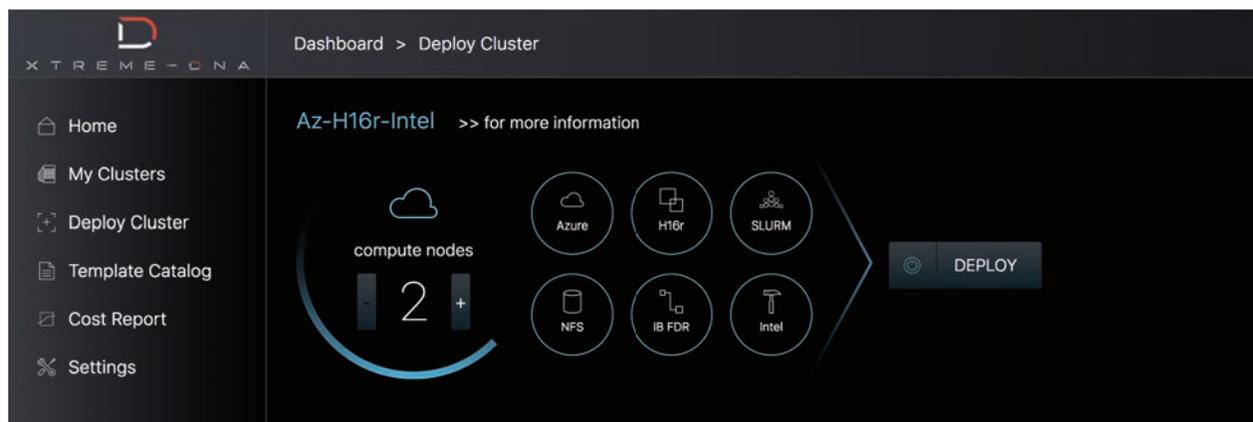


Figure 1: Xtreme-DNA web interface showing easily constructed cluster ready to deploy.

Users construct a fully functioning supercomputer by first selecting a template and a cloud provider. A well-designed web interface allows users to select the number of compute nodes, password, hostname, and shared disk size (See Figure 1). In addition, the cluster can easily be monitored for job status, budget usage, security state, job scheduler, etc. (See Figure 2) A virtual head node console is also provided as part of the cluster.

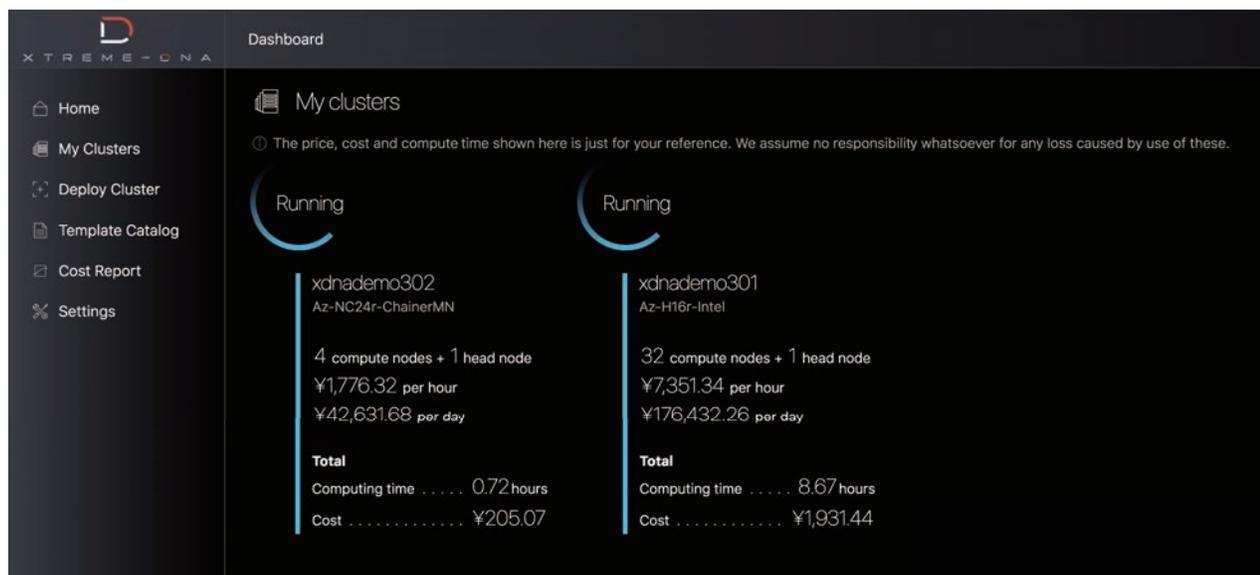


Figure 2: Example of cost monitoring for cluster as part of the XTREME-DNA interface.

Secure Baremetal Performance: XTREME-STARGATE

The next step in XTREME-DNA's evolution was to place the cluster "head node" in a local gateway appliance, providing better access and management options than are available when using XTREME-DNA over the Internet. Ideally, the Stargate appliance is located on your premises and provides secure access to offsite baremetal HPC cluster platform. Other HPC resources can be used as well, including local, private, and shared public clouds (i.e., Azure and AWS). XTREME-Stargate provides an instant HPC resource without integration headaches, allowing connections to almost any type of cloud.

The XTREME-Stargate gateway is a physically small set-top Linux appliance that provides a web portal to cluster resources. The appliance presents an even easier-to-use interface than the web GUI found in the XTREME-DNA cluster tool (See Figure 3). The XTREME-Stargate platform features:

- > A "super head node" for HPC cloud clusters
- > Ability to access shared or dedicated baremetal cloud resources
- > Support for HPC accelerators and container technology, GPU, FPGA, NVMe, and more.
- > Dedicated data centers in the US, Japan, and Singapore
- > Hybrid cloud capability
- > High-speed data transfer
- > Enhanced security including encrypted download and upload

XTREME-Stargate removes security concerns when using public or hybrid cloud by providing secure hardware-based data management and access to two types of dedicated HPC baremetal instances at the data center – shared and dedicated. In today’s climate security is paramount, so all XTREME-D clusters are deployed in data centers with a security level of Tier-3 or higher.

XTREME-Stargate provides Infrastructure as a Service (IaaS) and offers two options. The first is a shared baremetal cloud that utilizes public cloud services such as Azure and AWS. The second and more optimal solution is to use dedicated baremetal cloud run by XTREME-D. These clouds are optimized for HPC, DA, and DL and offer more flexibility and accelerator options, including GPU, FPGA, and NVMe hardware. In addition to an array of performance levels, these two cloud options can accommodate various user budgets.

Who Should Use XTREME-Stargate?

The XTREME-Stargate web portal was developed to bring more features, performance, and larger-scale functionality to HPC and enterprise customers. It is suitable for longer-term use, and allows for use of both private and hybrid cloud, whereas XTREME-DNA is a product designed for small- and mid-sized needs requiring public cloud clusters. *Table 1* compares the two products.

	 XTREME-DNA	 XTREME-Stargate	
Infrastructure	Public Cloud (Azure, GCP, AWS)	XTREME-Stargate Baremetal Cloud (Dedicated)	XTREME-Stargate Baremetal Cloud (Shared)
Hybrid Cloud Capability	Public Cloud * On-premise (Customer's Assets)	XTREME-Stargate Baremetal Cloud * Public Cloud	
Performance	Good	Excellent	
HPC Accelerators (NVMe, Xeon Phi, FPGA, InfiniBand, etc.)	-	✓	
Dedicated Datacenter	-	✓	
HPC Templates	✓	✓	
Container Technology	✓	✓	
Data Transfer	Standard Secure File Transfer Protocol	Managed Security with Hardware Key	

Table 1: The difference between XTREME-DNA and XTREME-Stargate.

What is Stargate?

XTREME-Stargate is an IaaS platform that connects users to HPC resources using an intuitive interface. Its main components are as follows:

- **An HPC Gateway** that manages the connection to HPC cloud clusters in the cloud.
- **A Secure Data I/O** component that uses a hardware key and high performance object storage for safe data movement.
- **A Cluster Dashboard** with an easy-to-use, template-based interface for easy application deployment. Includes real-time cluster cost reporting, with alarms for tight budget control.

A Walking Tour of XTREME-Stargate

When a user logs into XTREME-Stargate, the entire HPC pathway is handled through an easy-to-use, web-based graphical interface. The entire cycle, from spin-up and provisioning, to application compute and spin-down, is handled through the use of templates. Figures 3–6 provide a quick overview of the XTREME-Stargate interface.

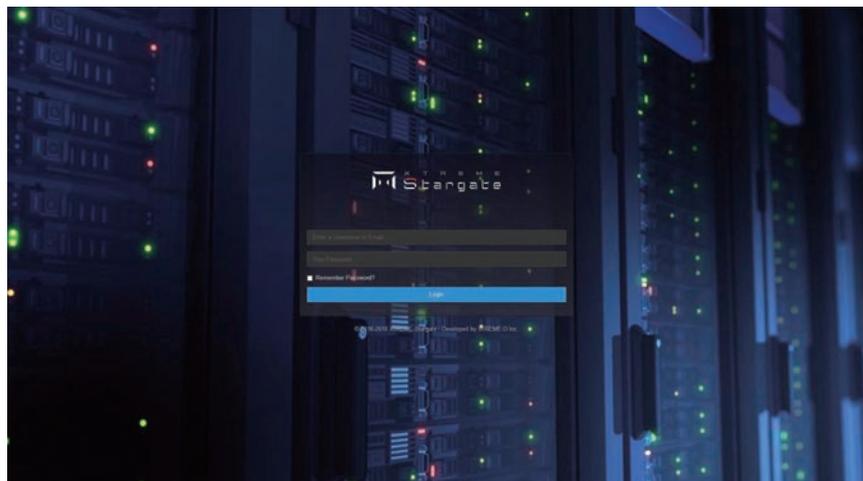


Figure 3: XTREME-Stargate login screen.



Figure 4: XTREME-Stargate home screen displaying Cost Report, Job Summary, Resource Summary, and Data Summary.

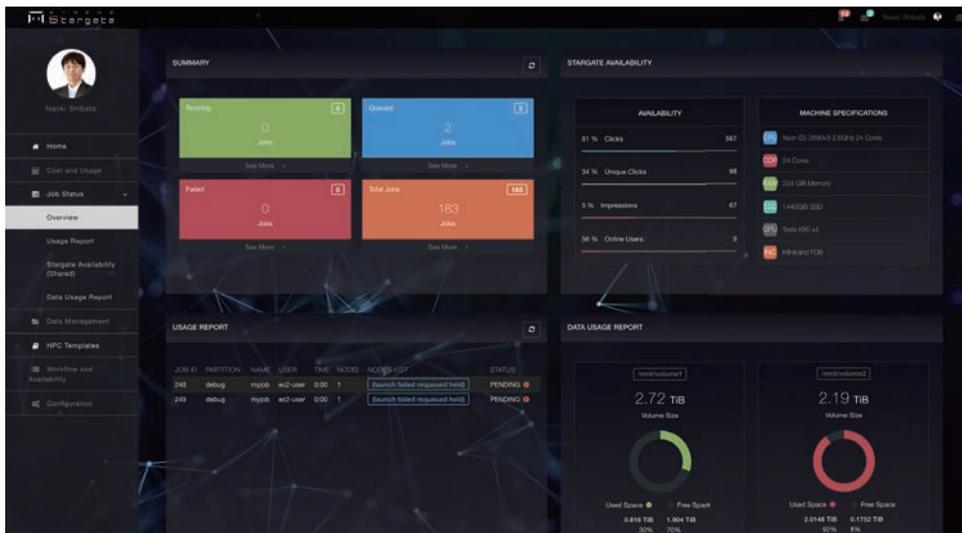


Figure 5: The Job Status overview window with Summary window, Usage Report, Availability, and Data Usage Report.

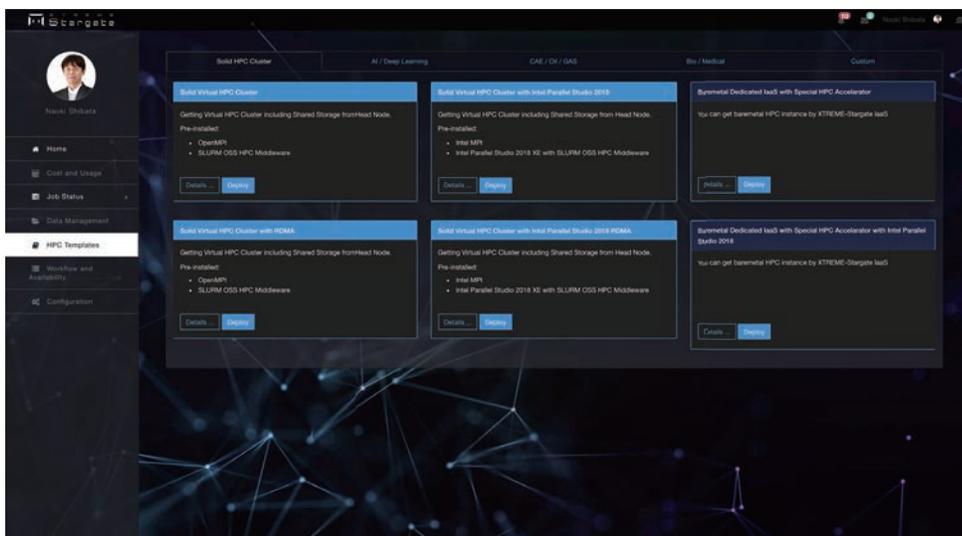


Figure 6: Sample HPC templates for constructing clusters.

Ease of Use Translates to ROI

Any new cluster solution should provide Return on Investment (ROI). Fancy GUIs only make sense if they help reduce costs. XTREME-D has elevated HPC, DA, and DL to a new level of simplicity, offering the following cost savings when deploying XTREME-Stargate:

- › A fast spin-up “pay-as-you-go” cloud model.
- › Choice of the best price-to-performance option by selecting the level of performance you need, from public cloud to private baremetal hardware, including GPU, FPGA, and NVMe.
- › Elimination of an entire level of administration costs, because a whole layer of tasks have been absorbed by cluster templates.
- › A dramatic increase in user efficiency (throughput) because there is no waiting for hardware procurement or configuring, building, and testing of application software

In essence, XTREME-Stargate puts end users as close as they have ever been to their applications. The platform also ensures that many time consuming “cluster environment issues” do not exist for XTREME-Stargate users.

Conclusion

Both the XTREME-DNA and XTREME-Stargate technologies offer a newly unprecedented level of HPC/DA/DL point-and-click capability. Users can actually design, spin up, compute, and spin down large leading-edge clusters – local, shared, cloud, or baremetal – with just a few mouse clicks.

The XTREME-D approach provides an ultra simple and intuitive template-based web GUI for cloud-based cluster computing that eliminates virtually all infrastructure/application design, configuration, and support issues. When the user is elevated above the minutiae of cluster computing, the cost/time to solution is reduced through a pay-as-you-go cloud model, price-to-performance budget options, reduced administration, and a dramatic increase in user efficiency.

The XTREME-D solutions provide a new portal into HPC, Data Analysis, and Deep Learning that is light years ahead of everyone else. Your next discovery can launch right from your desk.

<https://xtreme-d.net/>