

# Building an Accelerated Infrastructure for STAC-A2 with Red Hat OpenShift and NVIDIA DGX A100

Sebastian Jug  
Senior Performance Engineer

Performance and Scalability Team

# Red Hat OpenShift Container Platform

**Advanced Cluster Management**

## Multi-cluster Management

Discovery : Policy : Compliance : Configuration : Workloads

**OpenShift Container Platform**

Manage workloads

Build cloud-native apps

Data driven insights

Developer productivity

### Platform services

Service Mesh  
Serverless  
Builds : CI/CD Pipelines  
Log Management  
Cost Management

### Application services

Languages & Runtimes  
API Management  
Integration & Messaging  
Process Automation

### Data services

Databases : Cache  
Data Ingestion & Preparation  
Data Analytics : AI/ML  
Data Mgmt & Resilience

### Developer services

Developer CLI : IDE Plugins & Extensions :  
Cloud-native IDE :  
Local developer sandbox

**OpenShift Kubernetes Engine**

## Cluster Services

Automated Ops : Over-The-Air Updates : Monitoring : Registry : Networking : Router : KubeVirt : OLM : Helm

## Kubernetes

## Red Hat Enterprise Linux & RHEL CoreOS



Edge



Physical



Virtual



Private cloud



Multi-Arch



Public cloud



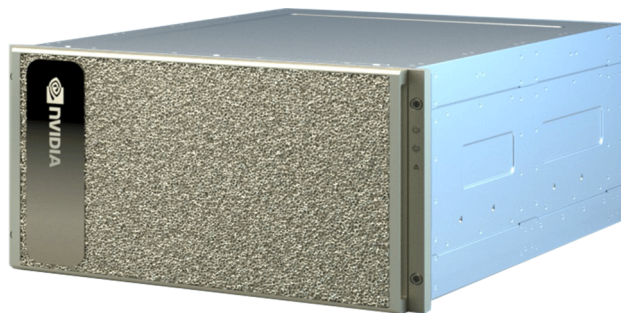
Managed cloud  
(Azure, AWS, IBM, Google)



# Red Hat OpenShift & STAC

- Tried & Tested Performance
  - OpenShift running N1, M3, A2
  - Variety of hardware configurations
- Enabled
  - Development of the NVIDIA GPU Operator, NFD Operator + others
  - Shared artifacts
- Supported
  - Multi-vendor collaboration / joint support

# STAC-A2 DGX A100 SUT



## *STAC-A2 Runtime*

NVIDIA-developed  
STAC-A2 Pack for CUDA  
(Rev G, r56)

CUDA 11.2

## NVIDIA DGX A100

### STAC-A2 Runtime

Red Hat OpenShift 4.8.3 (Red Hat CoreOS 48.84)

2x AMD EPYC 7742 64 core processors @ 2.25 GHz

8x NVIDIA A100 SXM4 80GiB GPUs

32x 64GiB Dual Rank ECC DDR4 DIMMs @ 3200 MT/s



**Red Hat**  
OpenShift

# Additional Resources

- **Blogs**

- [Red Hat, Partners Achieve Record-Setting STAC Benchmark on Financial Risk Analytics](#)
- [Red Hat and NVIDIA: Positioning Red Hat Enterprise Linux and OpenShift as Primary Platforms for Artificial Intelligence and other GPU-Accelerated Workloads](#)
- [Paving the way for Intelligent and Performance-Sensitive Applications on Kubernetes with Red Hat OpenShift Container Platform](#)

- **Prior OpenShift STAC Results**

- [STAC-A2 Pack for Intel® Parallel Studio XE \(Rev M\) with 2 x Intel® Xeon® Platinum 9242 \(Cascade Lake AP\) processors on Red Hat OpenShift 4.3 with RHEL 7.8 in an Intel Server System S9200WK SDP](#)
- [kdb+ v4.0 with Red Hat OpenShift 3.11 on a cluster of 10 x Supermicro SYS-6019U-TN4RT servers each with 2 x Intel® Xeon® Platinum 8280L \(Cascade Lake\) @ 2.70GHz and 6TB Intel® Optane DC Persistent Memory Modules](#)
- [UDP over 10GbE using Solarflare OpenOnload on Red Hat OpenShift 3.10 \(pre-release\) with RHEL 7.5 and Solarflare XtremeScale X2522 Adapters on Supermicro SYS-1029UX-LL1-S16 Servers](#)

- **Presentations and other results with OpenShift**

- [The Path to Cloud-Native Trading Platforms](#) - Jeremy Eder, Senior Principal Software Engineer, Red Hat Performance
- [OpenShift & STAC-A2 \(derivatives risk\) on Cascade Lake AP](#)
- [OpenShift vs bare metal on a modern low-latency stack](#)