

SMART CLOUD

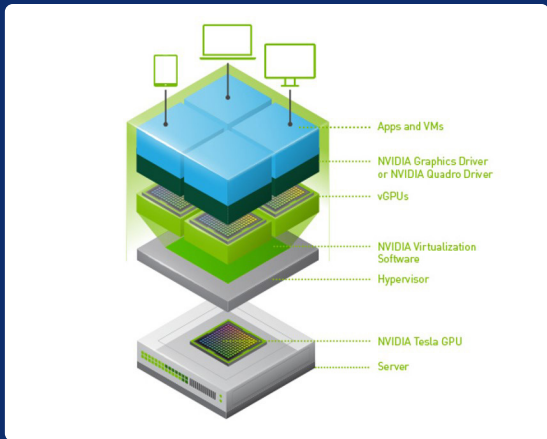


“Full Stack
DevOps Platform”

DevStack that is technology business company
at specialized OpenStack is
optimized in Edge and Fog computing and
effective deep learning processing using GPU.

DOS-LF™

OpenStack Powered Platform

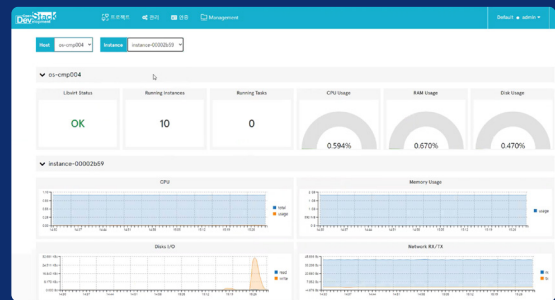
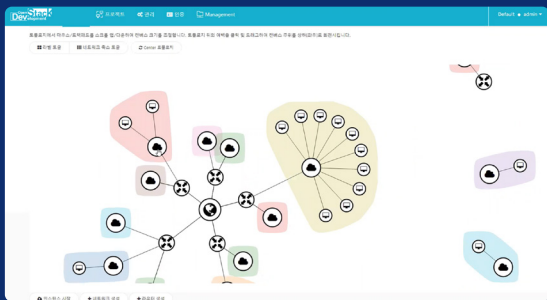


First in Korea to provide GPU virtualization-based public cloud system in an OpenStack environment



Providing hardware accelerators(GPU, FPGA, RDMA) in Container environments using a K8s Device Plugin Framework

Providing Web-based UI/UX



Integrating Helm Chart and Docker Image to deploy Big Data frameworks and various AI provided by **NGC(NVIDIA GPU CLOUD)***

Project / AI Service / AI Framework

AI Framework

TensorFlow
TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible...

Launch

PyTorch
PyTorch is a deep learning framework that puts Python first. It provides Tensors and Dynamic neural networks...

Launch

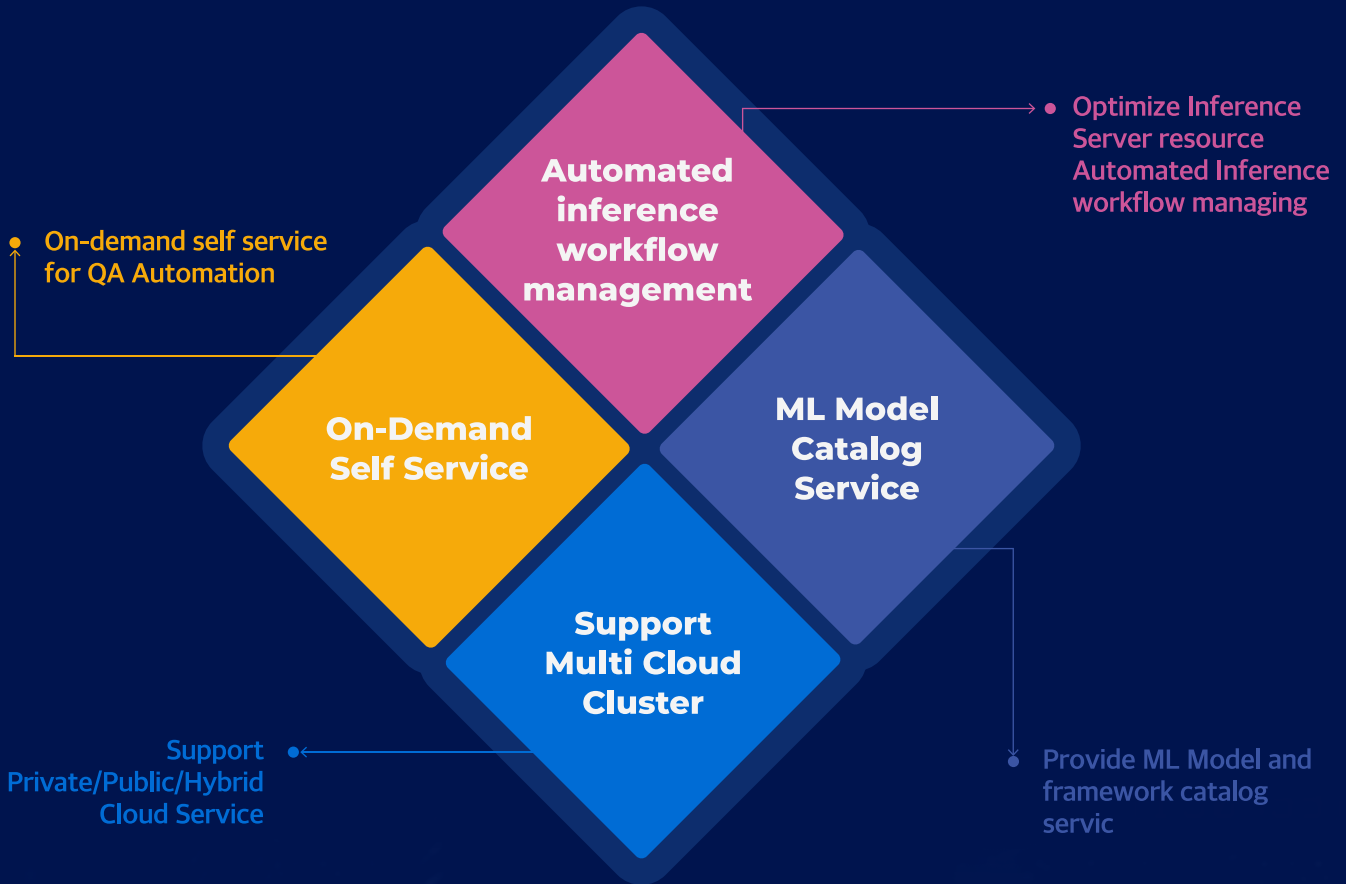
MXNet
Apache MXNet (incubating) is a deep learning framework designed for both efficiency and flexibility.

Launch

Instance Information

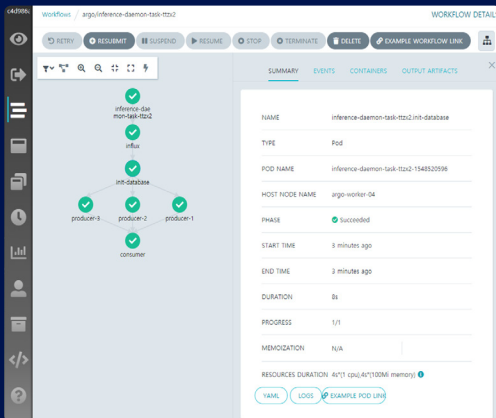
Name	Console	IP Address	Key Pair	Status	Power State
No items to display					

KEY FEATURES



DOS-DP™

Deep Learning Workflow Engine Platform



Suitable for job arrangement of the process and management Of dependent jobs

Optimization applicable to models developed with various frameworks

Reduce operating cost and increase efficiency of process management platform

Applied to cloud services in various environments

Automated inference workflow management



- Improve machine learning inference speed through batch processing optimization
- Efficient use of inference work resources through auto-scaling
- Inference task workflow creation and message-based workflow execution automation

ML Model Catalog Service

NAME	NAMESPACE	STARTED	FINISHED	DURATION	PROCESS	DETAILS
workflow-dept0	argo	a few seconds a...	a few seconds a...	0s	1/1	SHOW
preprocess	argo	4 minutes ago	4 minutes ago	0s	1/1	SHOW
inference	argo	4 minutes ago	4 minutes ago	0s	1/1	SHOW
preprocess	argo	2 minutes ago	2 minutes ago	7s	1/1	SHOW
pre-inf-workflow0	argo	2 minutes ago	2 minutes ago	0s	1/1	SHOW
pre-inf-post-workflow	argo	3 minutes ago	2 minutes ago	10s	1/1	SHOW
inference-workflow0	argo	3 minutes ago	3 minutes ago	0s	1/1	SHOW
inference-workflow-01	argo	3 minutes ago	3 minutes ago	0s	1/1	SHOW
workflow-01	argo	4 minutes ago	4 minutes ago	0s	1/1	SHOW

- Multiple Framework base Deep learning model inference processing -TensorFlow, PyTorch, onnx
- Inference model optimization using TensorRT
- Catalog service for configuring workflow -Custom container image for Workflow -NGC catalog
- Provides GUI for intuitive workflow configuration

Support Multi Cloud Cluster

Private Cloud Service (On-premise): Available in integrated inference server form
 Public Cloud Service: Register AWS/Azure Market Place (scheduled for 2020.12)
 Multi-Cluster services across multiple clouds

On-Demand Self Service

Support Multi tenancy, Multi User
 SaaS Service Portal for Automated QA
 Workflow Managing and Resource Metering/Monitoring