Building a National AI & Big Data Platform on SUSE Software Defined Infrastructure

SUSECON 2019 [CAS1366]

Justin Kao
Chief Architect
inwinSTACK

Derek So
Sr. Cloud Technologist, APJ
SUSE
Agenda

- About NARL
- The AI/Big Data Cloud Project
- The AI Cloud in Action
- Implementation Journey & Partnership with SUSE
- The Future
Who are we?

- **Justin Kao**
  - Chief Architect, inwinSTACK
  - 10+ years of experience in Cloud Solutions

- **Derek So**
  - Senior Cloud Technologist – APJ, SUSE
  - Help customers success through enabling cloud technologies
The AI Cloud Project
National Applied Research Lab (NARL)

Established in 1991 in Taiwan
Research Unit of National Center for High-performance Computing (NCHC)
Defining the future of Taiwan

- AI and IoT becomes more common in industrial and daily use.
- The government approved funding NCHC to build an AI platform with Open Infrastructure principles.
- The goals are
  - to accelerate AI technology development and,
  - to cultivate tech talent in Taiwan.

Ministry of Science and Technology
AI Cloud Project Overview

Big Data
250 Cloud Nodes
Taiwan’s largest Big Data Analytics Cloud

AI Cloud
2000 GPU
Total 9 PFLOPS
AI Frameworks for Training & Development

Data Storage
50PB
5x NCHC’s current storage capacity.
AI Cloud Project – Investment

2017 - 2020

US$ 1.6b

A-Team
Over 80+ Experts
AI Project Core Team

- **Data Center**
  - Taiwan Mobile

- **Hardware**
  - QCT (AI machine + HD osd)
  - ASUS (Compute node)
  - Mellanox, IBM, nvidia

- **Software**
  - ASUS (Prtoal)
  - inwinSTACK (ststem integration)
  - SUSE soc ses
**Hardware - whole system**
- 252 nodes / 9072 CPU cores / 2016 GPUs
- 193.5 TB memory
- 10 PB storage
- EDR InfiniBand 100 Gbps
- 1.2 PUE (Warm Water Cooling)

**Software Environment**
- Slurm / Kubernetes
- Nvidia NGC Docker
- Ceph
- Spectrum Scale (GPFS)
- CentOS

**Hardware - single node**
- Intel Xeon Gold CPU x 2
- Nvidia Tesla V100 w/32GB x 8
- 768 GB memory
- 240 GB SSD + 4TB NVMe

**AI Framework**
- Tensorflow
- Caffe / Caffe 2
- PyTorch / Torch
- ......and more
How does the AI Cloud work?
雲端物件儲存系統
Object Storage System
Use Case

Use over 1500 video cameras to predict traffic flow on Freeways.
AI pipeline steps

- AI key factors: Data, Algorithms, Computing
- Goal: To provide mega computing power for AI pipeline steps (Step: 2~5)
Data Processing Architecture

OpenStack

K8s / HPC
Roadmap of Vision Service

- Vision App. Serving
  - Model Zoo
    - Image classification
    - Image segmentation
    - Object classification
    - Object detection
    - Faces recognition and specific model
  - Data Pool
    - Open Images
    - COCO, ILSVRC
    - PASCAL, CIFAR
    - MNIST, KITTI
    - VGG, SVHN and specific set
- Deep Learning Frameworks
  - TensorFlow / Caffe2

- Docker
- Web serving
- API Serving
- Train Visualization
- Data Visualization
- NN Trainer
Implementation Journey & SUSE Partnership
SOC / SES

- **SES**
  - Total: 80 OSD nodes + 15 mon node + admin node
  - Raw Capacity: 30PB (object) 3PB (block)
  - Network IB (25G)

- **SOC**
  - Total: 250+ compute nodes + 10 controller nodes + admin node
  - GPU Passthrough
  - FWaaS, LBaaS, workflow, heat, run k8s
  - Network IB (25G), uplink (100G)
Implementation Challenges & Vendor Selection

● Q1 – Building AI Cloud is a new domain for everyone. Requirement is unclear.
● A1 – Only open architecture can adapt this kind of requirement – to adapt to changes. Proprietary approach to meet their requirement is not an option.
● Q2 - Performance
● Q3 – Scalability of SOC. Cost consideration. No vendor lock-in.
Why partnering with SUSE?

- Proven track records at NCHC
  - Big Data analytics cloud platform
- Excellent Partnership with local prestigious service partner.
- More focus on openstack than our competitors
- Proximity to local technical support in Taiwan
- Fast response to critical issues resolution
Why SUSE OpenStack Cloud & SUSE Enterprise Storage?

● SUSE Openstack Cloud 8
  – GitOps-based Cloud Lifecycle Manager (CLM) shortens deployment time and effort.
  – Proven configuration for large scale deployment
  – Complete documentation for operational needs

● SUSE Enterprise Storage
  – Proven reliability and well configured to integrate with SOC8
  – Stability, Reliability
  – Easy to manage and scale
Partnering with SUSE

- SUSE has been working as one of the key software partners in the A-Team to help customers building this AI Cloud Platform.
- Executive Support & Commitment
Got Questions?
THANK YOU