Cloud Native IBM Spectrum Scale

Ted Anderson, John Dunham, Sanjay Gandhi, Deepak Ghugh, Abhishek Jain, Monica Lemay, John Lewars, Muthu Muthiah, Aaron Palazzolo, Smita Raut, Wayne Sawdon, Dimitris Skourtis, **Vasily Tarasov**, Digvijay Ukirde, Yadavendra Yadav

Presentation at the 2019 HPCXXL Summer Meeting

2019 September 24

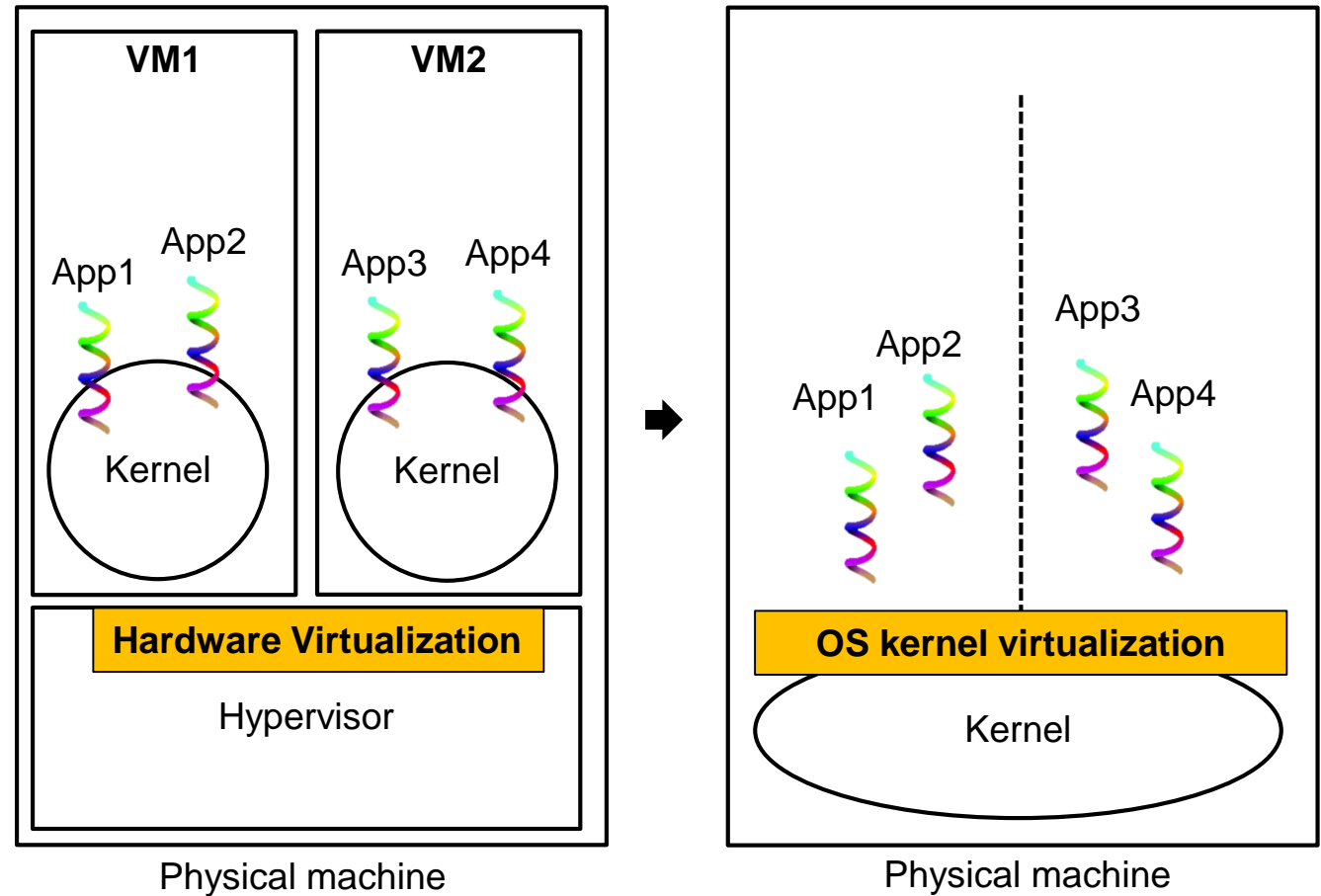**"Cloud native is a term used to describe container-based environments"**

Agenda

1. Background
   – Containers, Runtimes, Orchestrators

2. Using IBM Spectrum Scale with containerized applications
   – CSI Driver for IBM Spectrum Scale

3. Running IBM Spectrum Scale in containers
   – Containerization
   – Orchestration and management

Containers vs. Virtual Machines

- Hardware virtualization ➡ OS kernel virtualization

- Lower overhead to create virtual environments

- Enabler for microservices

- Linux provides primitives to create containers
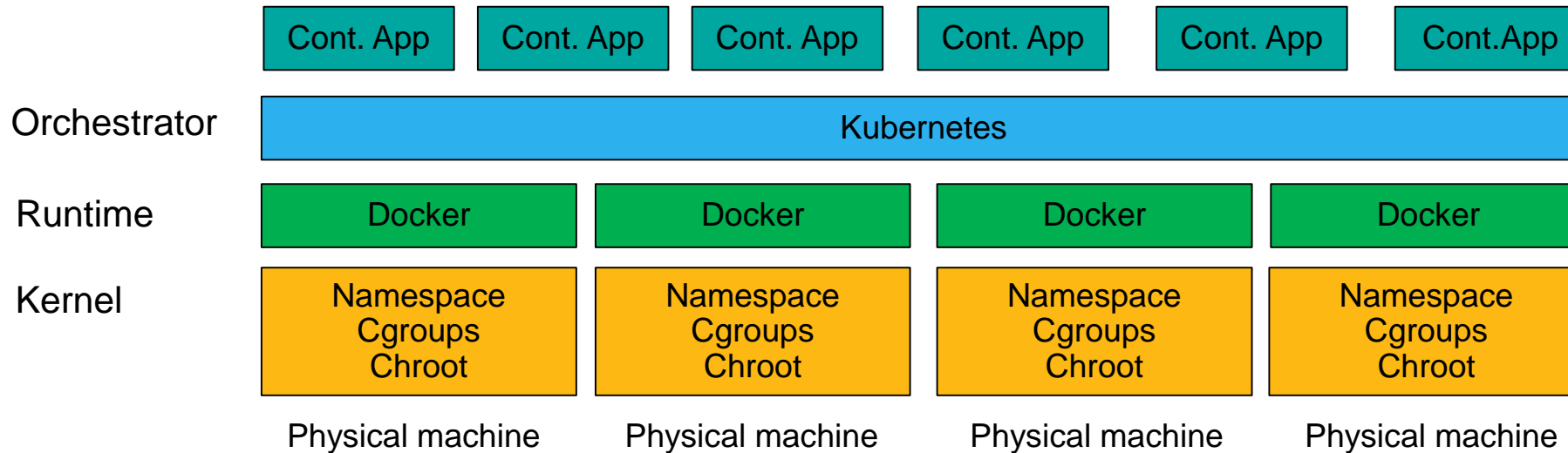 - namespaces
  - cgroups
  - chroot

VM1

VM2

App1 App2

App3 App4

Kernel

Kernel

Hardware Virtualization

Hypervisor

Physical machine

App1 App2 App3 App4

OS kernel virtualization

Kernel

Physical machine

Container runtimes

- Docker, CRI-O, containerd, rkt, …

- Use kernel primitives to create containers

- Container images
  - Start containers from images
  - Self-contained
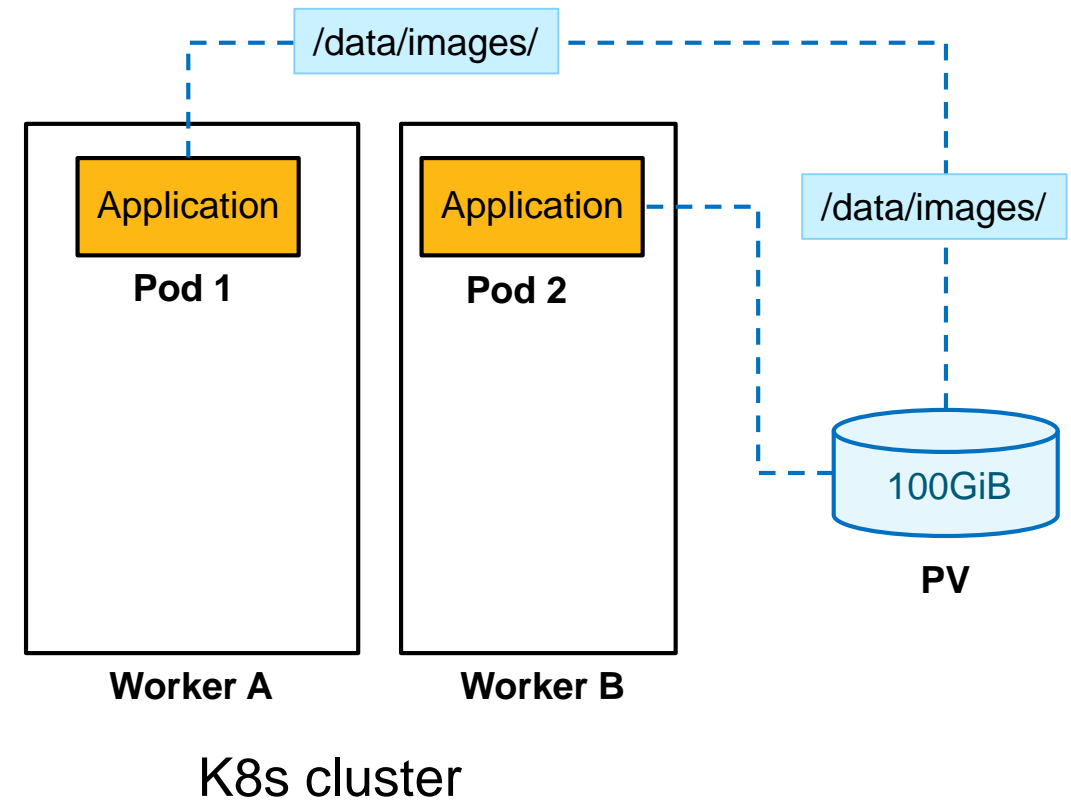 - Build, pull from, and push images to registries

pull/push

Image
Registry

Docker

Namespace
Cgroups
Chroot

Physical machine
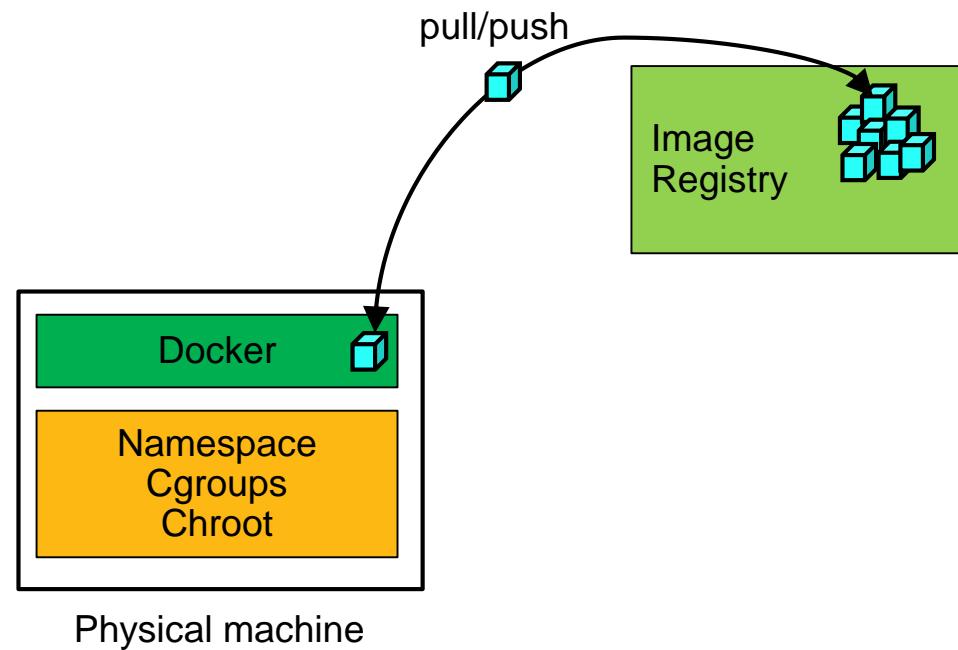
Container orchestrators

- **Kubernetes** (K8s), Docker Swarm, Mesos, ...
  – 50+ distributions of K8s, e.g., Red Hat OpenShift

- Manage containerized applications and compute resources in a cluster

- Additional features
  – Load balancing, health monitoring and healing, rolling updates, etc

| Cont. App | Cont. App | Cont. App | Cont. App | Cont. App | Cont.App |

Orchestrator	Kubernetes

Runtime	Docker	Docker	Docker	Docker

| Kernel | Namespace Cgroups Chroot | Namespace Cgroups Chroot | Namespace Cgroups Chroot | Namespace Cgroups Chroot |

Physical machine Physical machine Physical machine Physical machine

Persistent storage in K8s

- Originally K8s did not have a systematic approach to persistent storage

- Persistent Volume (PV) is a unit of storage that K8s manages
  - Provision, Deprovision, Attach, Detach

- PV provisioning can be static or dynamic

- Users typically interact with PVs through PV Claims (PVCs)
  - "I want 100GiB PV of type Spectrum Scale"

- Attached PVs are visible in the container as a directory tree
  - Specified when a container starts

/data/images/

Application

Pod 1

Application

Pod 2

/data/images/

100GiB

PV

Worker A

Worker B

K8s cluster

Step 0: CSI Driver for IBM Spectrum Scale

- CSI – Container Storage Interface

- CSI Drivers implement the specifics of a PV provisioning, deprovisioning, attachment, detachment
  – The standard is being adopted by K8s, Mesos, CF, Swarm
  – Run in containers across the nodes of the cluster; simple deployment

- CSI Driver for IBM Spectrum Scale to be GA in Dec 2019 (Open-source beta in Sep 2019)
  – Native spectrum Scale clients are pre-deployed on the hosts
  – CSI Plugin communicates to Scale through the REST API server
  – Dynamic and static provisioning

- PVs are be mapped to directories or filesets

https://github.com/IBM/ibm-spectrum-scale-csi-driver
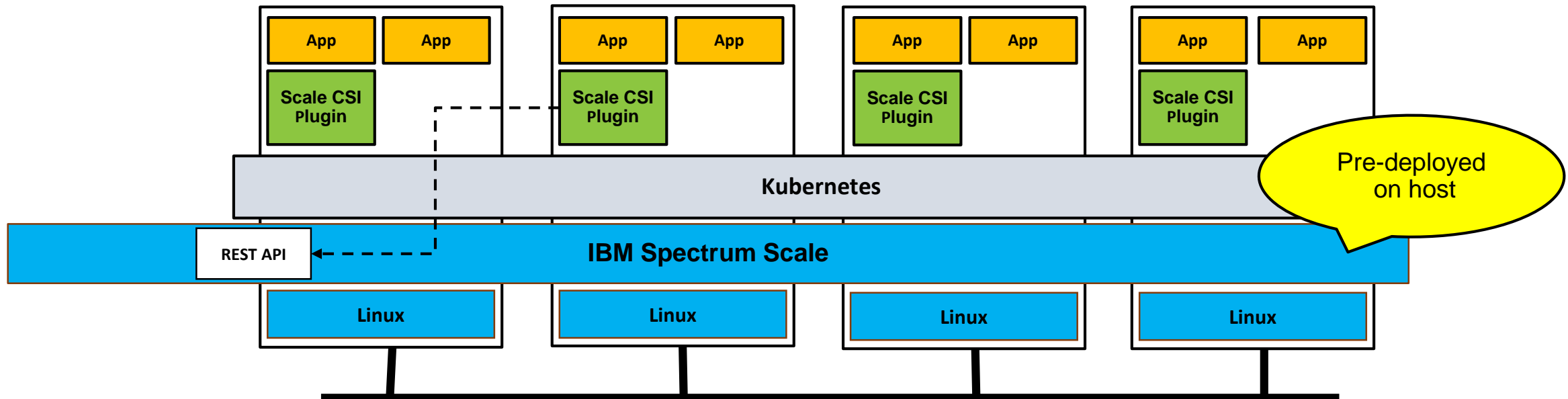
Step 1: Deploying and Running Spectrum Scale in K8s

- Hyperconverged setup with ECE

- OpenShift 4.*

- Technical Preview and later GA in 1H 2020
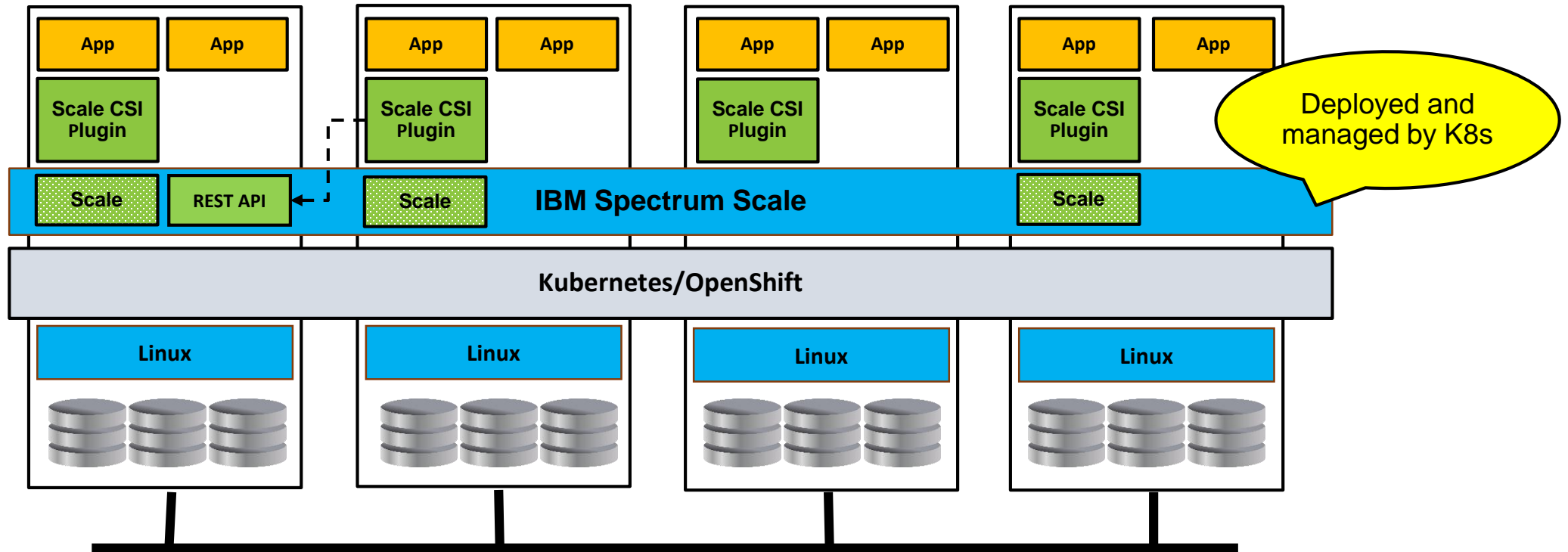
IBM Spectrum Scale Container

- The container image includes Scale's native client, server (ECE mode), and all dependencies

- Notable properties of the Scale container
    - Privileges to insert kernel extension and mount file system
    - Access to block devices
    - Uses host network (not overlay network)
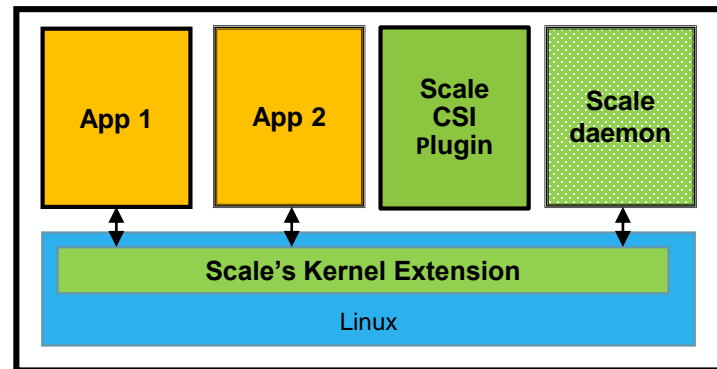
Deployment and management

- DaemonSet to start Scale containers on every node

- Initial deployment
  - Cluster creation
  - Block device discovery; NSD and ECE setup
  - File system creation

- Management
  - Dynamic cluster scale-up and scale-down
  - Reaction to node reboots and failures
  - Rolling upgrades

- Scale Operator to automate the initial deployment and management

- PV provisioning, deletion, attachment, detachment via CSI Plugin

Step 1+

- Support of other Scale features and deployment models
  - Protocol nodes
  - AFM
  - Dedicated storage servers

- Increase the number of supported filesets

- QoS in a multi-tenant environment

- Integration with cloud native stack
  - Prometheus, Fluentd, Jaeger

- Storage-aware container scheduling

- Splitting Scale in multiple containers

- Support of a no-kernel-extension mode

- Deployment in the public clouds with virtual block devices

- Hybrid Cloud model
  - Aspera

- Backups and snapshots

- Overlay network

Cloud Native IBM Spectrum Scale

Ted Anderson, John Dunham, Sanjay Gandhi, Deepak Ghugh, Abhishek Jain, Monica Lemay, John Lewars, Muthu Muthiah, Aaron Palazzolo, Smita Raut, Wayne Sawdon, Dimitris Skourtis, **Vasily Tarasov**, Digvijay Ukirde, Yadavendra Yadav

Presentation at the 2019 HPCXXL Summer Meeting

2019 September 24

Thank you. Q&A.

1) CSI Driver for IBM Spectrum Scale (Beta):
https://github.com/IBM/ibm-spectrum-scale-csi-driver

2) Describe your use case to us!