KubeCon | CloudNativeCon

OPEN SOURCE SUMMIT

China 2019
HDFS CSI: Speed Up Kubernetes in On-Premises Big Data Cluster

Yi Chen & Junping Du
About us

• Yi Chen – Software engineer from Tencent, Apache Hadoop Committer & PMC, working on Apache Hadoop HDFS/Ozone.

• Junping Du – Chair of Tencent Open Source Committee, Apache ASF member, Apache Hadoop Committer & PMC.
Outline

• Big Data Introduction
• Challenges
• CSI
• Solution Options
Big Data Stack

Applications
- Batch
- Streaming
- Interactive
- Graph Analytics
- Machine Learning

Data Processing & Analysis
- DataFrame
- SQL
- SparkR
- Streaming
- Mlib
- GraphX

Resource Management & Co-ordination
- YARN
- ZooKeeper

Storage
- HDFS
- Ceph
- S3
- Parquet
- ORC
- Iceberg
Big Data Trend

- Move to cloud
- Containerization
- Compute and storage disaggregation
- Hybrid deployment
Benefit of Containerization

- Improve hardware utilization
- Strong resource isolation without much virtualization overhead
- Better application packaging
- Deployment agility and infrastructure independence
Kubernetes Architecture

Master
- Etcd
- API Server
- Scheduler and Controller Manager

Node
- Kubelet
- Proxy

Pod
- Container
- Container
- Container

Pod
- Container
- Container
- Container

Pod
- Container
- Container
- Container

Pod
- Container
- Container
- Container

Pod
- Container
- Container
- Container
Challenges

Kubernetes
• Optimized for containerized microservices
• Running Big Data workloads along with microservices together
• Better tuned for cloud use cases

Big Data
• Improving capabilities for persistent volumes
• Containerize, microservices
CSI (Container Storage Interface)

- Specification
- Interoperable
- Vendor neutral
- Control plane only
- Idempotency
• Identity Service
  • Provide CSI plugin version and capability information

• Controller Service
  • Create/Delete/Manager volumes

• Node Service
  • Mount/Unmount volumes on node
Kubernetes CSI Support

- Introduce CSI since v1.9
- GA v1.13
- 45+ CSI drivers supported
HDFS Architecture

• Strong consistency distributed file system
• 3 Replica, data high safety
• 10+ production adoption
HDFS Cluster Scale

- Single cluster 4000+ nodes
- Federation cluster 20K+ nodes
HDFS - Ozone

- Object storage
- Scaling to trillions of objects
- S3 gateway
- Ozone File System
Solution Options

• HDFS NFS Gateway
• HDFS Fuse
• HDFS CSI based on NFS Gateway
• OZone CSI
• HDFS Quadra CSI
HDFS CSI

- Name Node
- Storage
- Node
  - CSI Driver Node
  - HDFS NFS Gateway
- Node
  - CSI Driver Node
  - HDFS NFS Gateway
- Node
  - CSI Driver Node
  - HDFS NFS Gateway

Datanode
Resources

• CSI Spec
  • https://github.com/container-storage-interface/spec

• Kubernetes CSI Support
  • https://kubernetes-csi.github.io/docs/

• HDFS Ozone
  • https://hadoop.apache.org/ozone/