

IPS K1 Power 企业级容器云CSI方案

李相林

浪潮商用机器有限公司 技术支持部



目录

- 1 K8S CSI基本概念
- 2 K1 Power容器云CSI介绍
- 3 GPFS CSI vs NFS测试

容器是什么?



它取决于您询问的对象

INFRASTRUCTURE

APPLICATIONS

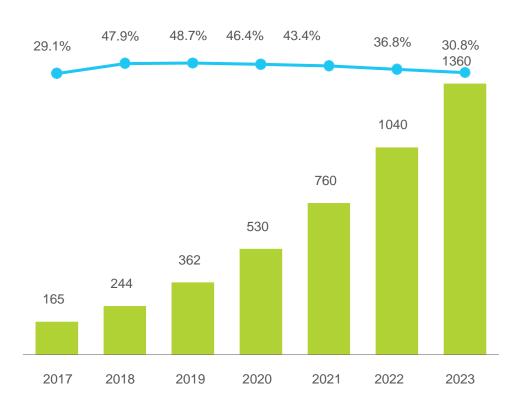
- Linux操作系统特性
- Linux包含在其中
- 比虚拟机更简单、更轻量化
- 高密度
- 需要编排和管理

- 独立包装
- 轻松添加至已构建的应用
- 轻松共享容器化组件
- 应用需运行的所有工作都在内部执行
- 开发人员具有更高控制权

容器云市场前景广阔

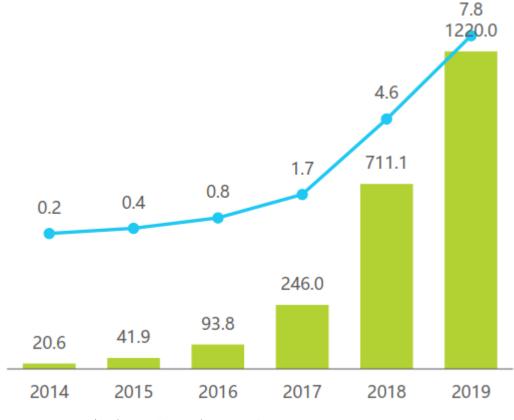
IPS

2017-2023年中国企业级容器云市场规模



- 一 中国企业级容器云市场规模(亿元)
- → 中国企业级容器云市场规模年增速(%)

2014-2019年中国移动互联网接入流量统计



- 移动互联网流量 (亿GB)
- ◆ 单户月均移动互联网流量(亿GB)

Kubernetes概述(K8s)



- ➤ Kubernetes 是容器的编排层,用于跨一组物理服务器或VM管理容器.
- > Kubernetes专门用于通过以下方式管理成千上万个短暂性的容器:
 - 部署容器化的应用程序
 - 按需求扩展容器的规模
 - 缩小或终止容器规模
 - 版本控制
 - 内部和外部容器通信
 - 存储创建和挂载
 - 监视-当主机服务器/ VM发生故障时进行故障切换



> Kubernetes在本地传统基础架构堆栈上的运行与在第三方服务提供商和公共云环境相似。



使用容器存储的场景及其特点



配置

无论集群配置信息还是应用配置信息,其特点是并发访问,也就是ROX/RWX,在不同集群或者不同节点,都能够访问同样的配置文件,分布式文件存储是最优选择。

日志

在容器场景中,日志是很重要的一部分内容,其特点是高吞吐,有可能会产生大量小文件。如果有日志分析场景,还会有大量并发读操作。分布式文件存储是最优选择。

• 应用(数据库/消息队列/大数据)

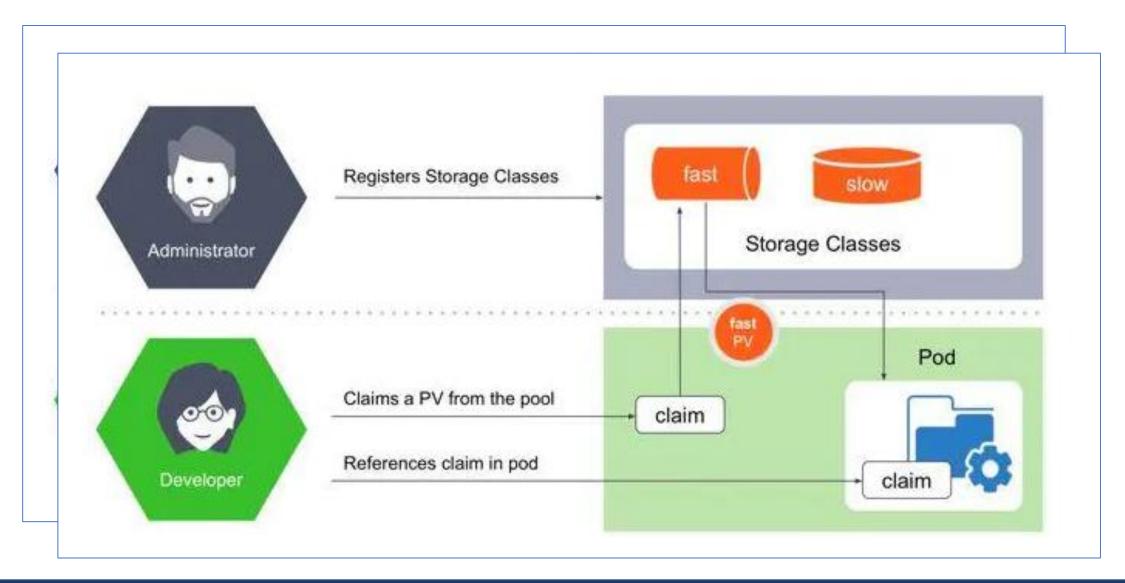
Kafka, MySQL, Cassandra, PostgreSQL, ElasticSearch, HDFS等应用,本身具备了存储数据的能力,对底层存储的要求就是高IOPS,低延迟。底层存储最好有数据冗余机制,上层应用就可以避免复杂的故障和恢复处理。以HDFS为例,当某个datanode节点掉线后,原有逻辑中,会选择启动新的datanode,触发恢复逻辑,完成数据副本补全,这段时间会比较长,而且对业务影响也比较大。如果底层存储有副本机制,HDFS集群就可以设置为单副本,datanode节点掉线后,启动新的datanode,挂载原有的pv,集群恢复正常,对业务的影响缩短为秒级。高性能分布式文件存储和高性能分布式块存储是最优选择。

备份

应用数据的备份或者数据库的备份,其特点是高吞吐,数据量大,低成本。文件存储和对象存储最优。

PV&PVC





Types of Persistent Volumes



- •非持久性存储
 - emptyDir
 - HostPath (只在单节点集群上用做测试目的)
- •网络连接性存储
 - SAN: iSCSI、ScaleIO Volumes、FC (Fibre C hannel)
 - NFS: nfs, cfs
- •分布式存储
 - Glusterfs
 - RBD (Ceph Block Device)
 - CephFS
 - Portworx Volumes
 - Quobyte Volumes

•云端存储

- GCEPersistentDisk
- AWSElasticBlockStore
- AzureFile
- AzureDisk
- Cinder (OpenStack block storage)
- VsphereVolume
- StorageOS
- •自定义存储
 - FlexVolume

PersistentVolume Plugin



Volume Plugin	ReadWriteOnce	ReadOnlyMany	ReadWriteMany
AWSElasticBlockStore	✓	-	-
AzureFile	✓	✓	✓
AzureDisk	✓	-	-
CephFS	✓	✓	✓
Cinder	✓	-	-
CSI	depends on the driver	depends on the driver	depends on the driver
=C	✓	✓	-
FlexVolume	✓	✓	depends on the driver
Flocker	✓	-	-
GCEPersistentDisk	✓	✓	-
Glusterfs	✓	✓	✓
HostPath	✓	-	-
SCSI	✓	✓	-
Quobyte	✓	✓	✓
NFS	✓	✓	✓
RBD	✓	✓	-
/sphereVolume	✓	-	- (works when Pods are collocated)
PortworxVolume	✓	-	✓
ScaleIO	✓	✓	-
StorageOS	✓	-	-

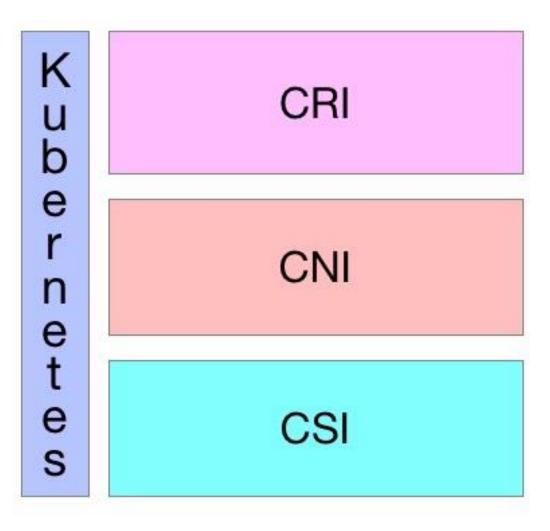
K8S 开放接口



Kubernetes作为云原生应用的基础调度平台,相当于云原生的操作系统,为了便于系统的扩展, Kubernetes中开放的以下接口,可以分别对接不同的 后端,来实现自己的业务逻辑:

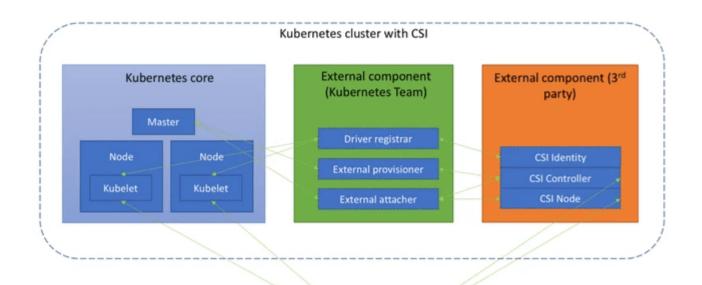
- •CRI (Container Runtime Interface) : 容器运行时接口,提供计算资源
- •CNI(Container Network Interface): 容器网络接口, 提供网络资源
- •CSI (Container Storage Interface) : 容器存储接口, 提供存储资源

以上三种资源相当于一个分布式操作系统的最基础的几种资源类型,而Kuberentes是将他们粘合在一起的纽带。



K8s容器存储接口 (CSI) 介绍





Container Storage Interface是由来自 Kubernetes、Mesos、Docker等社区 member联合制定的一个行业标准接口 规范,旨在将任意块和文件存储系统暴 露给诸如Kubernetes之类的容器编排系统(CO)上的容器化工作负载的标 准。

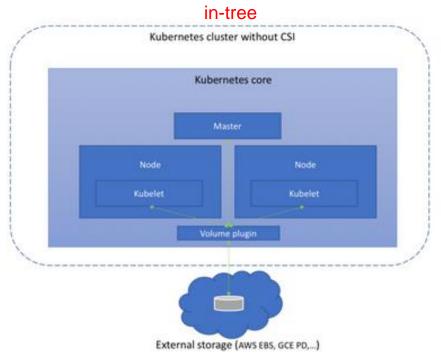
使用CSI的第三方存储提供商可以编写和部署在Kubernetes中公开新存储系统的插件,而无需接触核心的Kubernetes代码。



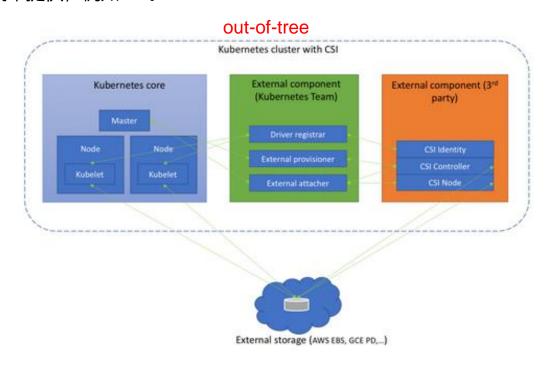
CSI机制-K8S与CSI



- · 之前,K8S里提供存储服务是通过一种称为 in-tree 的方式来提供,例如flexvolume。
- · 现在,K8S里提供存储服务是通过一种称为 out-of-tree 的方式来提供,例如CSI。



- 存储插件需要一同随K8S发布。
- 2. K8S社区需要对存储插件的测试、维护负责。
- 3. 存储插件的问题有可能会影响K8S部件正常运行。
- 4. 存储插件享有K8S部件同等的特权存在安全隐患。
- 5. 存储插件开发者必须遵循K8S社区的规则开发代码。



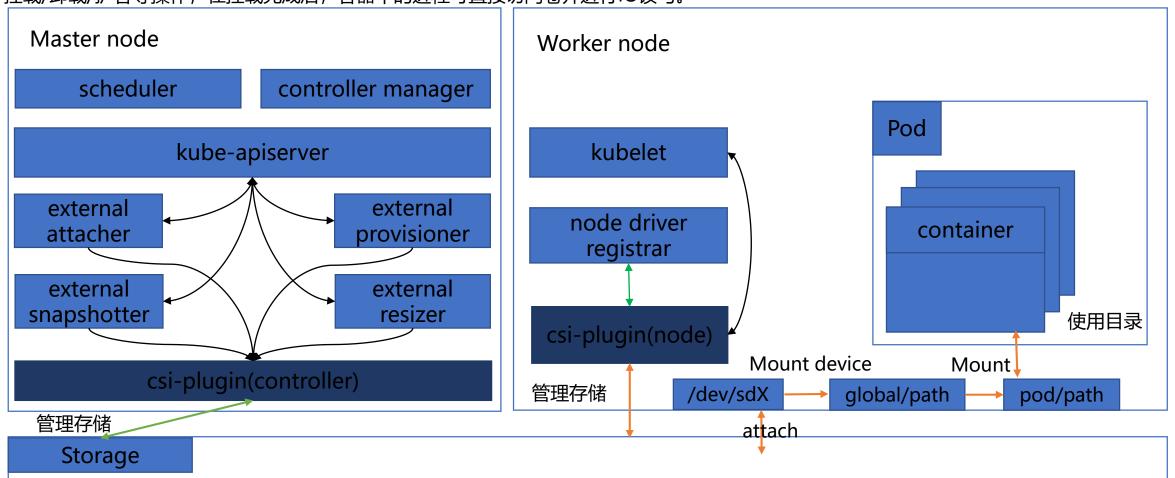
- External Provisioner: 监控Kubernetes系统里的PVC对象,调用对应CSI的Volume 创建、删除等接口。
- 2. External attacher:监控Kubernetes系统里的VolumeAttachment对象,调用对应 CSI的接口。
- 3. node-driver-registrar: 通过Kubelet设备插件机制,使用kubelet注册CSI驱动程



CSI存储插件在系统中的位置



CSI接口插件实现的是标准的CSI接口协议,在控制层面,Kubernetes在保持原有apiserver设计的条件下,通过引入了辅助服务来适配 CSI接口,并由辅助服务部分以CSI协议的方式与驱动交互;在节点业务层面,则是插件完成注册后,kubelet服务直接与插件交互,完成卷的 挂载/卸载/扩容等操作,在挂载完成后,容器中的进程可直接访问卷并进行IO读写。





目录

- 1 K8S CSI基本概念
- 2 K1 Power容器云CSI介绍
- 3 GPFS CSI vs NFS测试

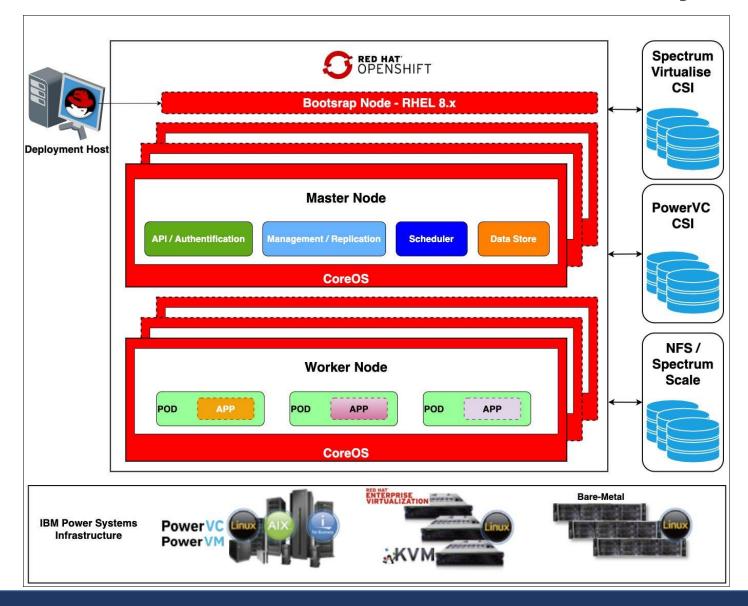
CSI 举例



- ✓ PowerVC CSI
- ✓ Spectrum Scale CSI
- √ Ceph CSI
- ✓块存储CSI: 华为fusionstorage、 浪潮G5存储等

Red Hat OpenShift Container Platform for IBM Power Systems





OpenShift 4.x Storage Options



Storage	Туре	ppc64le (Power Systems)	S390x (z/LinuxONE)	amd64 (Intel, AMD)
Raw Block	RWO	1Q 20	1Q 21	2Q19
NFS (Spectrum Scale CES, no dynamic provisioning)	RWX	2Q 20	1Q 20	2Q19
PowerVC CSI (container storage interface)	RWO	2Q 20	N/A	N/A
IBM Spectrum Virtualize (CSI) Block	RWO	2Q 20	1Q 20	4Q 19
IBM Spectrum Scale File	RWX	4Q 20*	4Q 20*	4Q 19
FC	RWO	1Q 21	1Q 21	1Q 20
iSCSI	RWO	4Q 20	4Q 20	1Q 20
HostPath	RWO	4Q 20	4Q 20	1Q 20
CephFS (OCS) File	RWX	4Q 20**	4Q 20**	1Q 20
CephRBD (OCS) Block	RWO	4Q 20**	4Q 20**	1Q 20
VMware vSphere	RWO	N/A	N/A	2Q 19

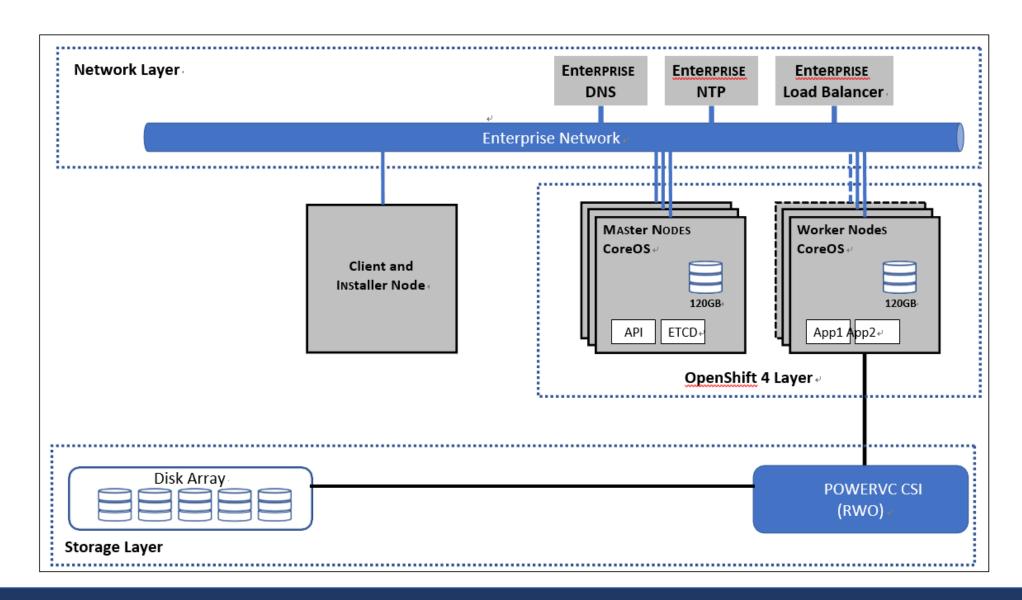
RWX = read-write many (shared) RWO = read-write only (dedicated)

*Dependent on containerized Spectrum Scale (GA in 4Q)

** Planned for OCP/OCS 4,6

PowerVC CSI

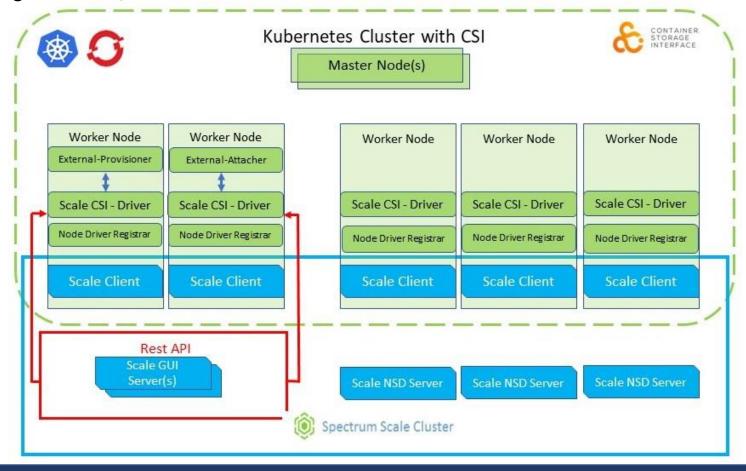




Spectrum Scale Container Storage Interface

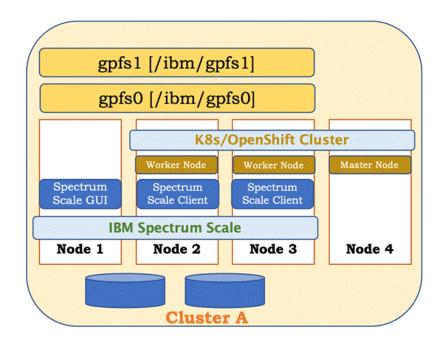


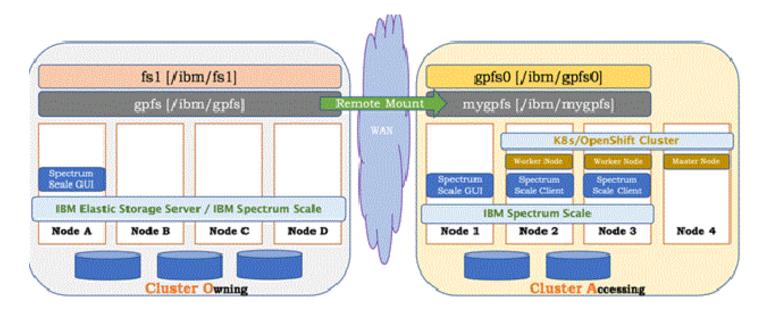
Spectrum Scale容器存储接口允许Spectrum Scale用作Kubernetes集群中运行的有状态应用程序的永久存储。通过Spectrum Scale容器存储接口驱动程序,可以从IBM Spectrum Scale设置 Kubernetes永久卷 (PV)。容器本质上可以与有状态的微服务一起使用,例如数据库应用程序 (MongoDB, PostgreSQL等)。



Spectrum Scale CSI





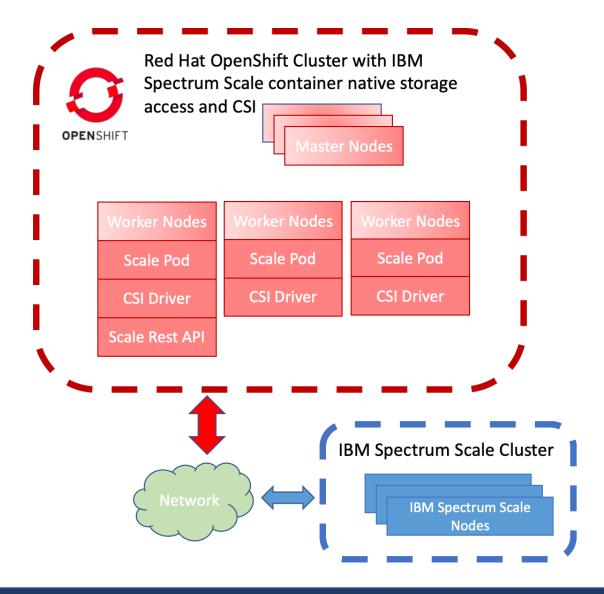


One Spectrum Scale clusters

Two IBM Spectrum Scale clusters with remote-mounted file systems

Spectrum Scale Container Native Storage Access

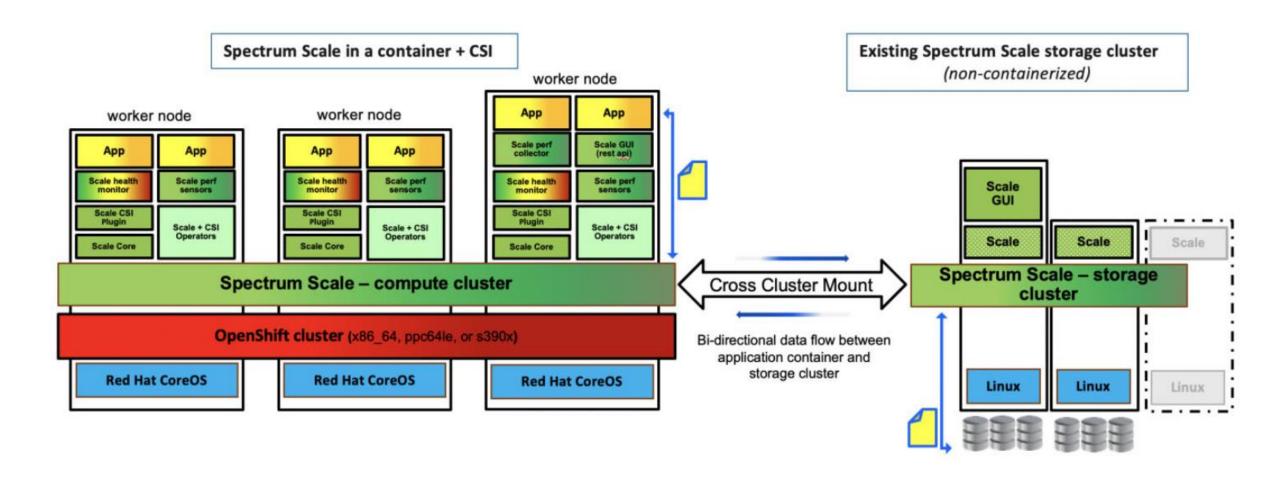




通过在容器中创建Spectrum Scale集群,可以在RedHat OpenShift集群中部署集群文件系统。并使用远程挂载附加文件系统,Spectrum Scale解决方案提供了持久性数据存储,应用程序可以通过使用持久性卷(PV)的Spectrum Scale容器存储接口(CSI)驱动程序来访问这些持久性数据存储。

Spectrum Scale in a contianer







目录

- 1 K8S CSI基本概念
- 2 K1 Power容器云CSI介绍
- 3 GPFS CSI vs NFS测试

测试场景-硬件&软件



✓ 硬件:

Server: E950 PowerVM

Storage: Flash840

NetWork: 10Gb sr-iov

✓ 软件版本:

OS: redhat7.6 ppc64le

K8S:1.18.16

GPFS:5.0.5-1

CSI:2.1.0

IOR:3.0.1

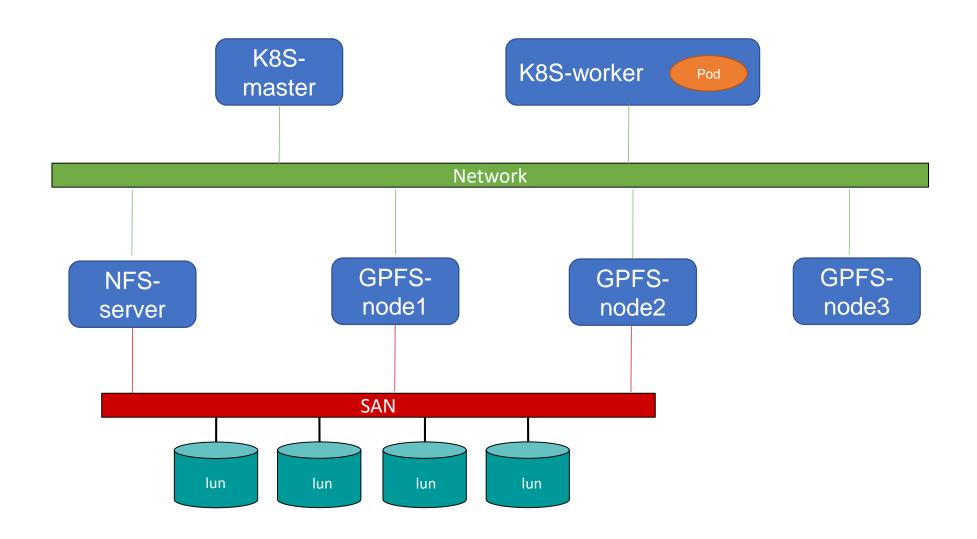




Node	Role
GPFS-node1	NSD Server
GPFS-node2	NSD Server
GPFS-node3	GUI,CES node
K8S-master	k8s master
K8S-work1	k8s worker,NSD client
NFS-server	nfs server

测试拓扑图





测试结果





fio -direct=1 -iodepth=32 -rw=randrw -ioengine=libaio -bs=4k -size=64G -numjobs=1 -runtime=1000 -group_reporting \ -filename=/mnt/nfs/fiotest -name=Rand_Write_Testing





请扫描二维码加入 官方微信公众号

谢谢

公司官方网址: https://www.inspurpower.com

浪潮官网IPS专区: https://www.inspur.com/lcjtww/2315499/2315503/2433653/index.html

合作伙伴网址: https://www.inspurpower.com/cop/

IPS微知营销工具平台: https://idisk.inspur.com/kw/indexKw (需IPS内部账号)

浪潮商用机器企业云创新中心: http://ipscloud.talkwithtrend.com/

IPS 2019年客户大会在线: http://server.51cto.com/act/201905/ips2019khdh

IPS 2019年先行者峰会: http://server.51cto.com/act/201905/ips