oVirt storage system and IBM's activity

Shu Ming
shuming@cn.ibm.com
IBM CSTL KVM technical leader

Topics today

- oVirt storage subsystem target
- Basic concept of oVirt storage subsystem
- Storage subsystem inside
- Management application view of oVirt storage subsystem
- What's new in oVirt 3.2
- oVirt storage community work items
- IBM oVirt storage work items

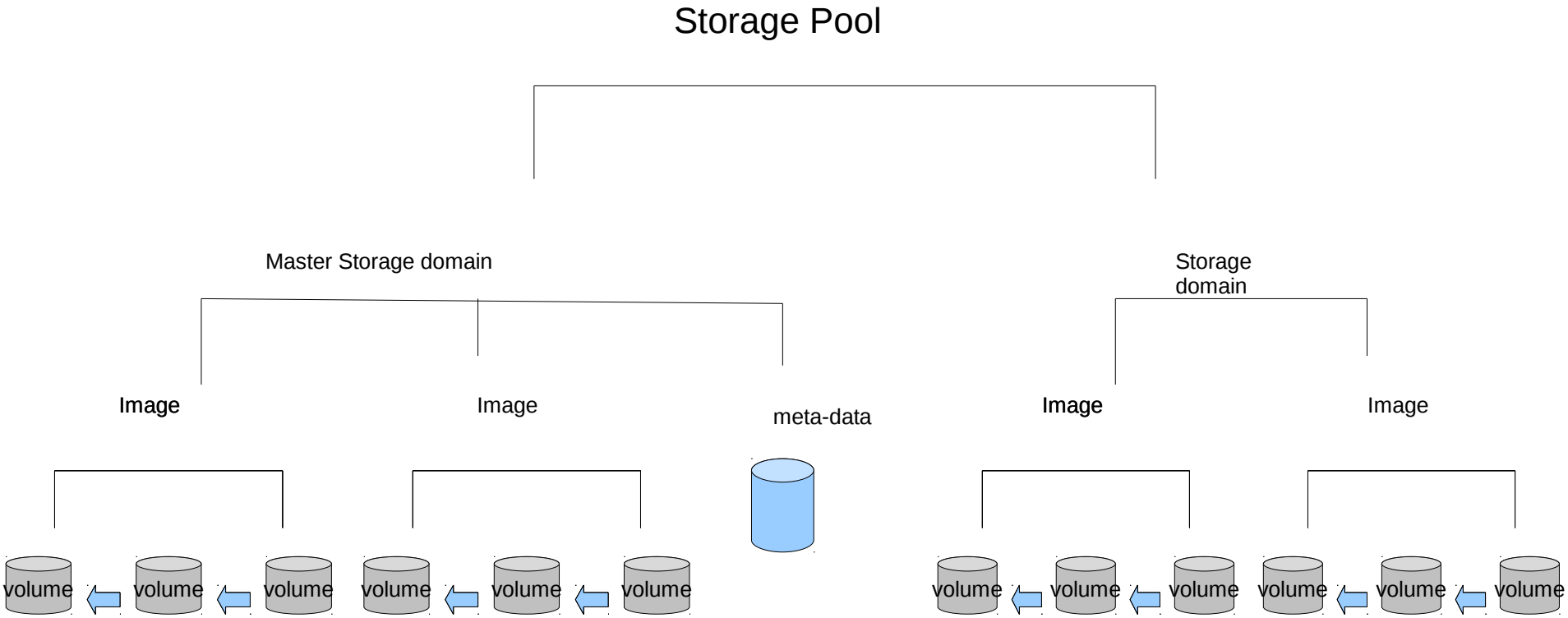
oVirt storage system target

- Tens of thousands of virtual disk images are hosted
- Virtual disk images can be shared by oVirt nodes
- The virtual disk images be performance and disk space efficient
- Both file based and block based Image support
- Multiple storage device support and ready to be extended to new device

Basic concept of oVirt storage subsystem

- Storage Pool
  - It is an abstract name for the nodes in the oVirt clusters and the sharing storage attached to the clusters
  - Rich types: NFS, iSCSI, FC, localFS, PosixFS
- Storage Domain
  - Dividing storage pool into different domains
- Image
  - It can be mapped to a disk to the VM with snapshots or be floating in oVirt
- Volume
- UIUID for Storage Pool, Domain, Image and Volume
- VM template and ISO

Basic concept of oVirt storage subsystem



Basic concept of oVirt storage subsystem

Ovirt engine

Management network

Cluster A

Cluster B

KVM host

KVM host

KVM host

KVM host

Storage pool master

NFS

ISCSI,FC

Shared Storage Pool

Metadata

Physical volume

Physical volume

Physical volume

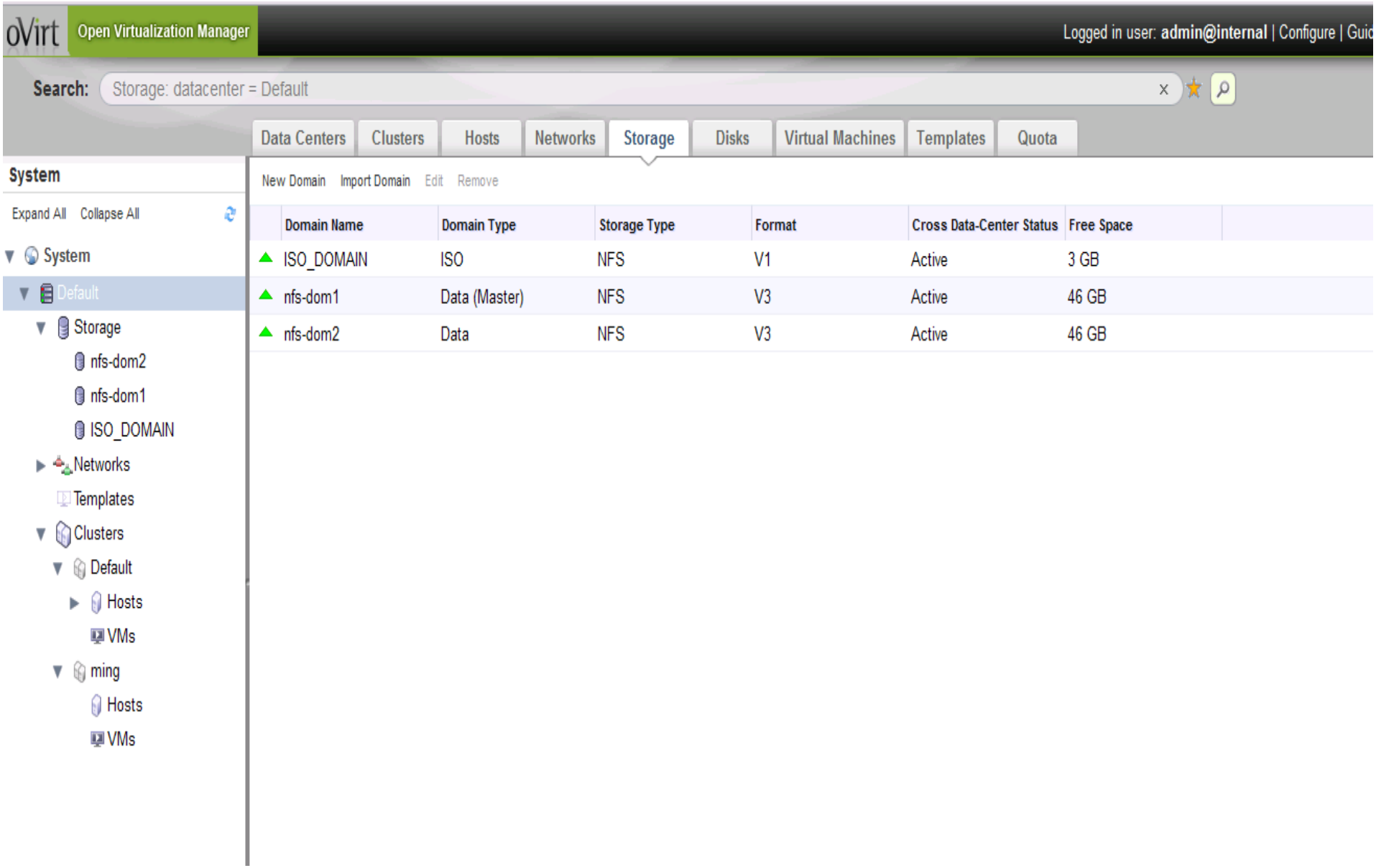
Storage subsystem inside I

- Developed by Python object oriented language
- Storage pool type
- The storage pool meta-data synchronization
- One master storage domain for a storage pool
- Storage pool master election

Storage subsystem inside II

- The lease of images are protected by Sanlock
- Block based images are built on top of Linux LVM systems and have no performance penalty from the file-system on the storage hardware
- File based images provide a cheap and flexible way to build virtual disk images for the Vms
- Direct Lun support to VMs

Management application view of storage subsystem



Management application view of storage subsystem I

- Storage subsystem is not a independent service to oVirt engine
- Storage pool meta data modification is centralized by one master node and other nodes can read the meta data
- The storage system can support synchronous and asynchronous tasks
- Asynchronous tasks can be persisted and recoverable

Management application view of Storage subsystem II

- XMLRPC or jason RPC APIs to management applications
- Administrator role per storage domain

Management application view of Storage subsystem III

- Storage live snapshot of the VM disks
- Preview of the VM snapshot
- VM snapshot merging
- Thin provision volumes
- Making template from existing VM and clone a new VM from the existing template

What's new in oVirt 3.2

- Live Storage migration was supported
- Live snapshot was supported from oVirt 3.1 and became stable in oVirt 3.2
- Support has been added for storage domain live upgrade

oVirt storage community work dashboard I

- Make the storage system an standalone image service like other cloud storage system
- Live backup of VM disks and restore from these backup images
- Storage pool removing and multiple storage domains
- Storage operation offloading to hardware

oVirt storage community work dashboard II

- Other shared file system integration
- New Storage API

IBM oVirt storage activity

- Decoupling the storage system APIs from other oVirt node level APIs
- oVirt storage system integration into Openstack
- oVirt storage system integration into IBM's product
- Gluster FS storage domain project followup
  - Ovirt 3.3 integration, transport type setting

IBM oVirt storage activity

- Allow creating ISO domain on other file-based storage except NFS
- Live backup and restore participation

- Thank You and Questions