

# iSCSI Target Support for Ceph

**Lin Pei Feng**

pflin@suse.com

Ceph Day Shanghai 2015-10-18



# Ceph RADOS Block Device (RBD)

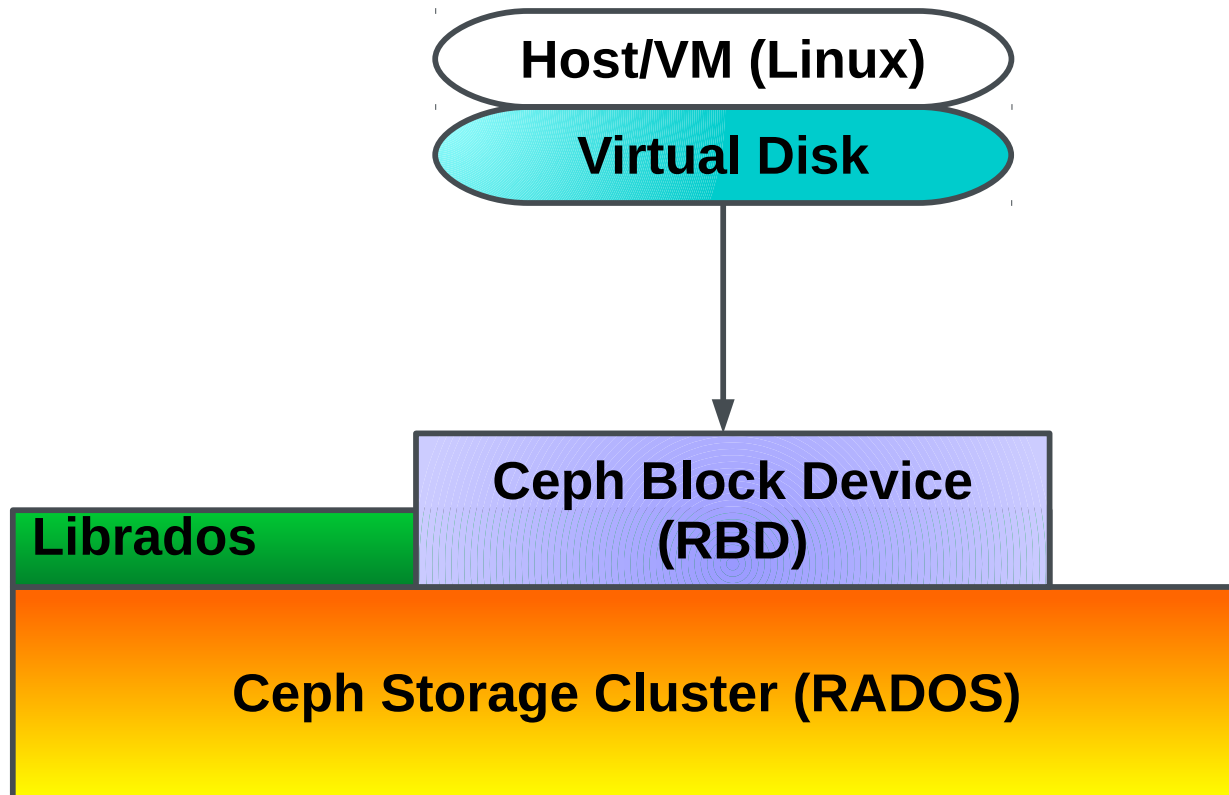
## Introduction

- Block device backed by RADOS objects
  - Objects replicated across Ceph OSDs
- Thin provisioned
- Online resize
- Snapshots and clones

# Ceph RADOS Block Device (RBD)

## Limitation

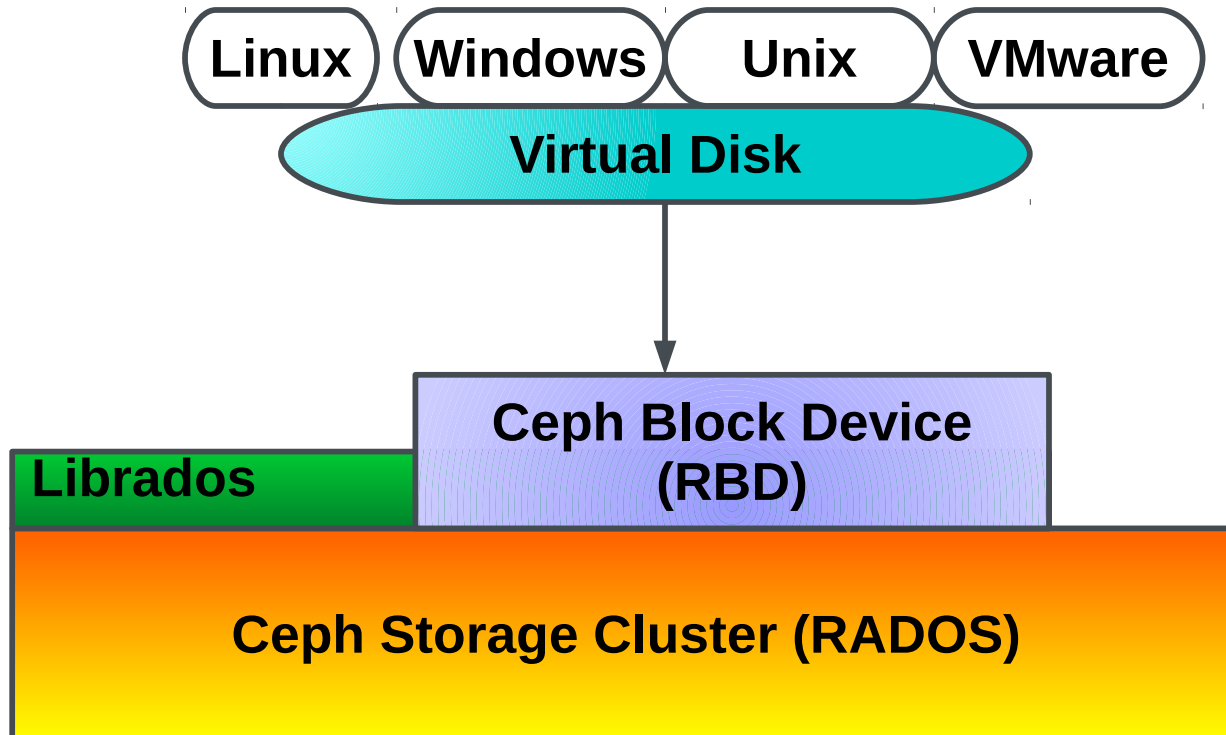
- Linux kernel or librbd clients
  - Usage restricted to a subset of operating systems and applications



# Ceph RADOS Block Device (RBD)

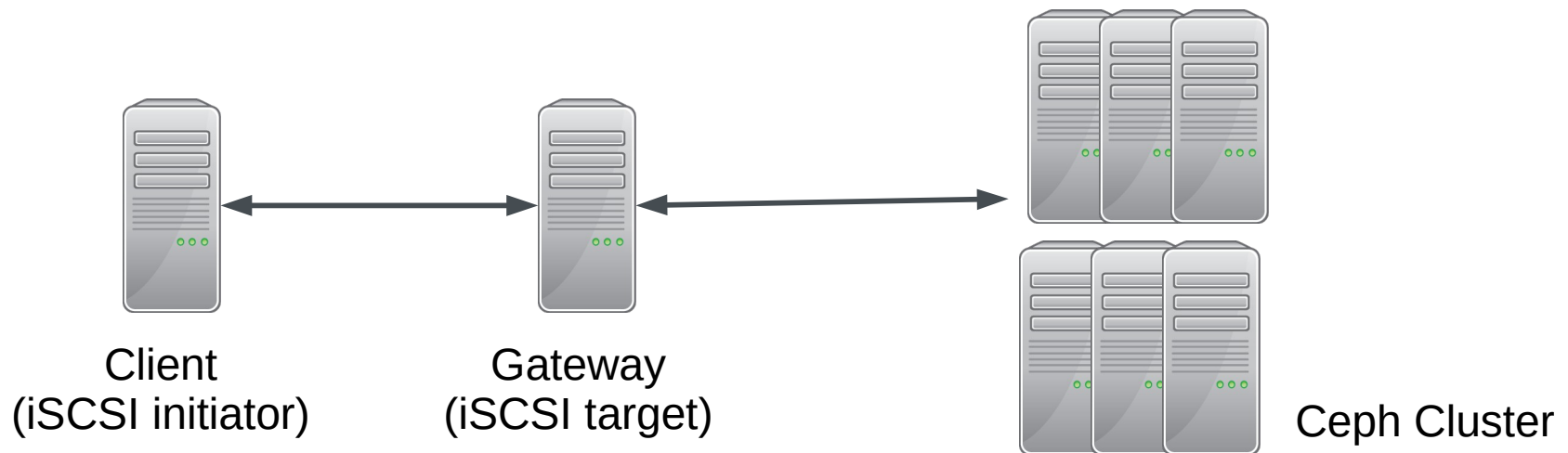
## Idea

- Heterogeneous OS access
- Easy to use



# iSCSI gateway for Ceph RBD

- Expose benefits of Ceph RBD to other systems
  - No requirement for Ceph aware applications or operating systems
- Standardised interface
  - Mature and trusted protocol (rfc3720)
- iSCSI initiator implementations are widespread
  - Provided with most modern operating systems

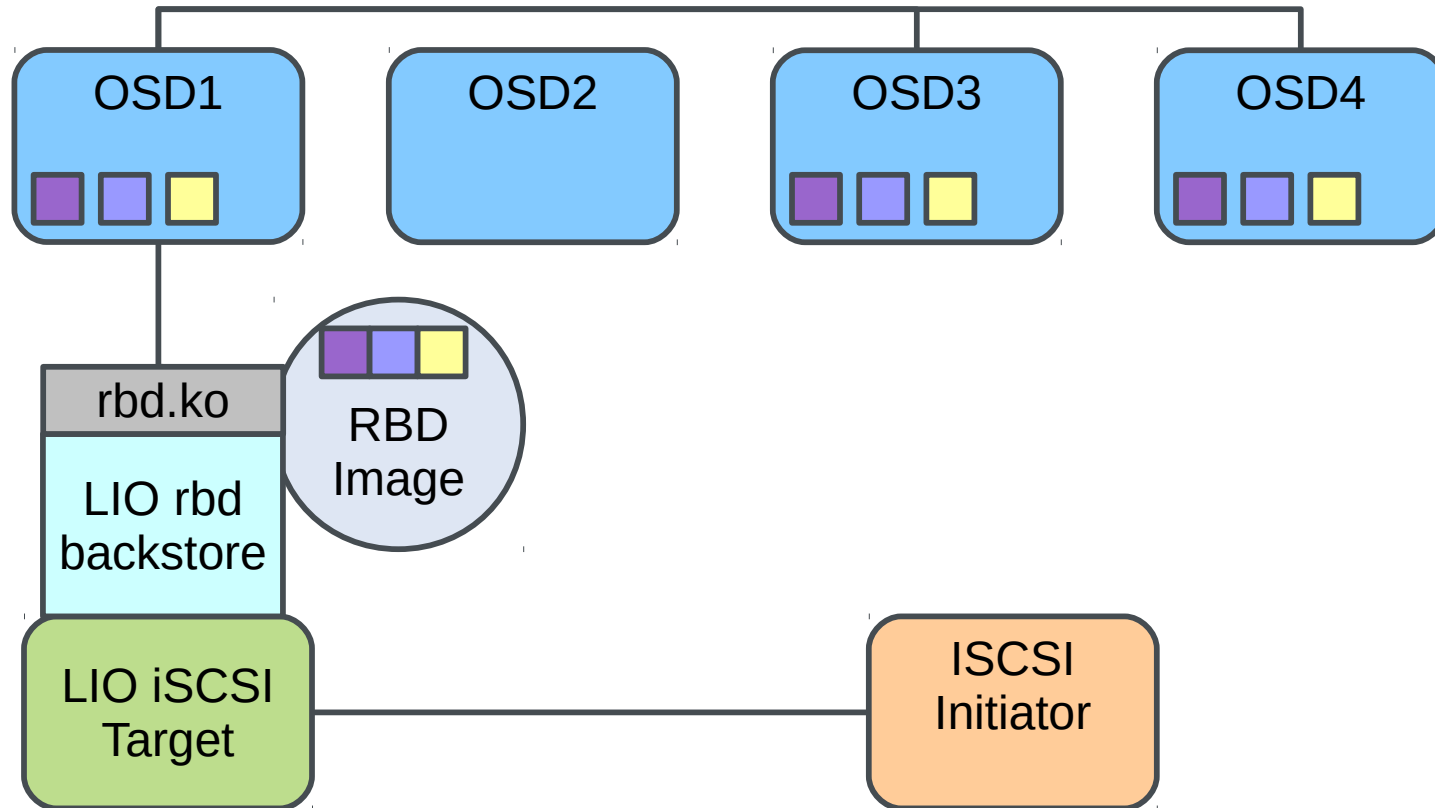


# iSCSI Target

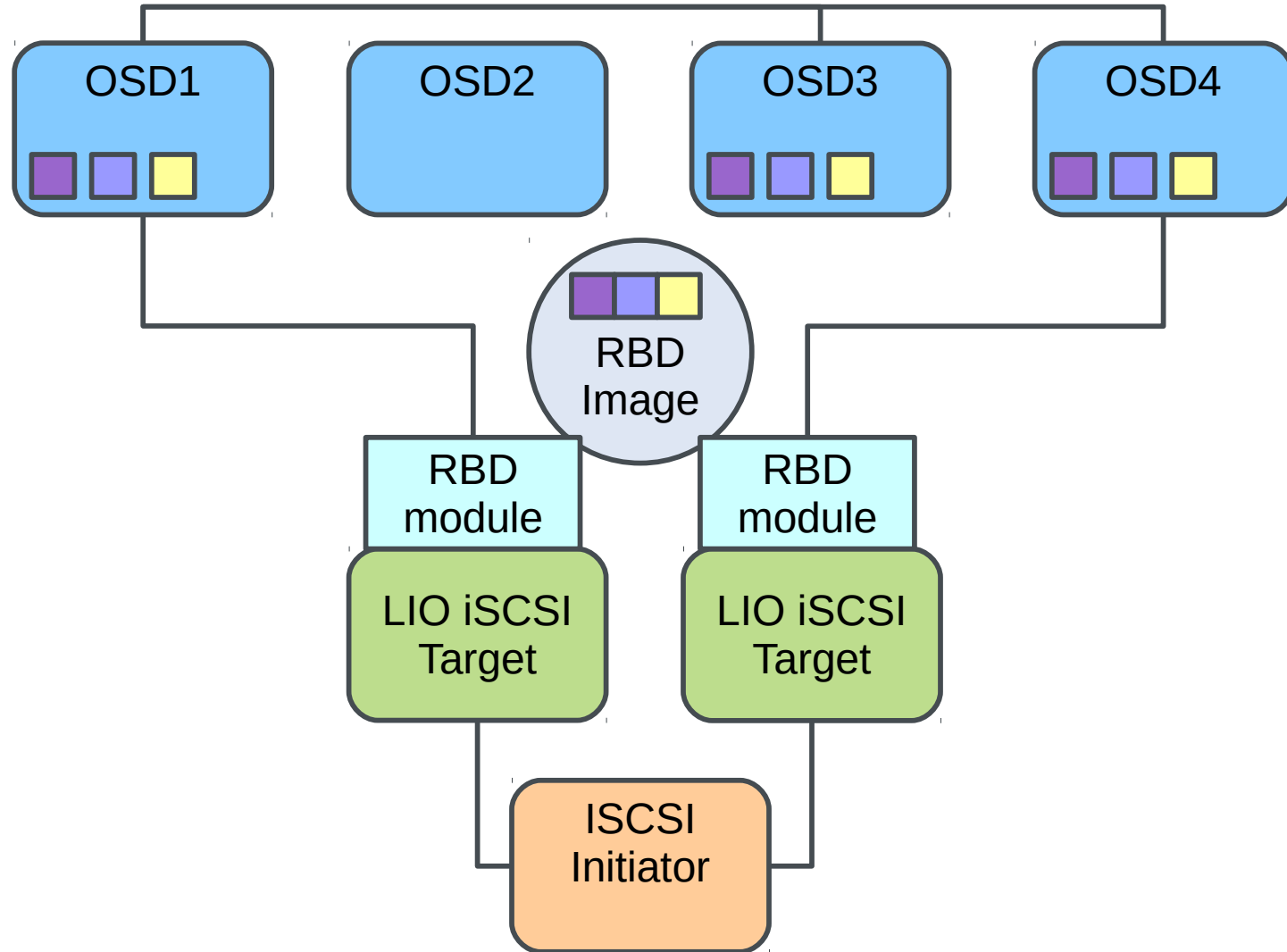
- LIO – Linux IO Target
- In kernel SCSI target implementation
  - Support for a number of SCSI transports
    - iSCSI, as well as iSER, FC and others
  - Pluggable storage backend
    - lblock, as well as fileio, pscsi and others
- Flexible configuration
  - targetcli utility

# LIO + RBD iSCSI gateway

## Overview



# LIO + RBD Cluster iSCSI gateway





# LIO + RBD Cluster iSCSI gateway

- LIO target configured with iSCSI transport fabric
- RBD backstore module
  - Translates SCSI IO into Ceph OSD requests
  - Special handling of operations that require exclusive device access
    - Atomic COMPARE AND WRITE
    - SCSI reservations
    - LUN reset
- Lrbd: Multi-node configuration utility
  - Applies iSCSI target configuration across multiple gateways via targetcli

# LIO + RBD iSCSI gateway

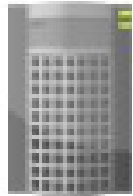
## Cluster support

- Allows for initiator access via redundant paths
  - iSCSI gateway node with multiple network adapters
    - Protection from network adapter failure
  - Multiple iSCSI gateways exporting same RBD image
    - Protection from entire gateway failure
- Initiator responsible for utilisation of redundant paths
  - Available paths advertised in iSCSI discovery exchange
  - May choose to round-robin IO or failover/failback

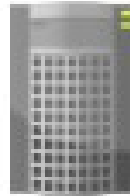
# iSCSI Initiators

- open-iscsi
  - Default iSCSI initiator shipped with SLES10 and later
  - Multipath supported in combination with dm-multipath
- Microsoft iSCSI initiator
  - Installed by default from Windows Server 2008 and later
  - Supports MPIO in recent versions
- VMWare ESX
  - Concurrent clustered filesystem (VMFS) access from multiple initiators

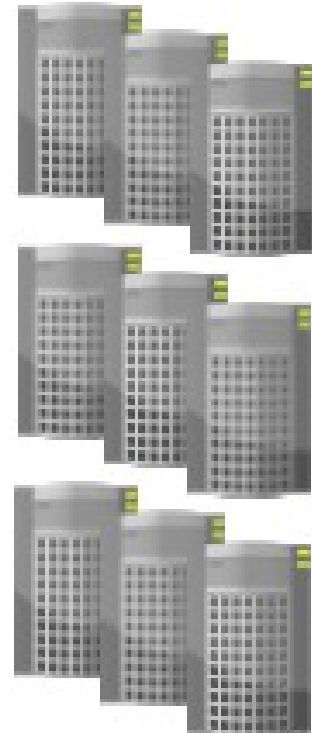
# A Simple iSCSI gateway for Ceph RBD



**Client**



**Gateway**  
**172.16.1.16**



**Ceph Cluster**

# Irbd configuration

```
"targets": [  
  {  
    "host": "igw1",  
    "target": "iqn.2003-01.org.linux-iscsi.generic.x86:sn.abcdefghijkl"  
  }  
],
```

```
"pools": [  
  {  
    "pool": "rbd",  
    "gateways": [  
      {  
        "host": "igw1",  
        "tpg": [  
          {  
            "portal": "portal1",  
            "image": "archive"  
          }  
        ]  
      }  
    ]  
  }  
]
```

```
"portals": [  
  {  
    "name": "portal1",  
    "addresses": [ "172.16.1.16" ]  
  }  
],
```

```
"auth": [  
  {  
    "host": "igw1",  
    "authentication": "tpg",  
    "tpg": {  
      "userid": "common1",  
      "password": "pass1"  
    }  
  }  
],
```

# Apply configuration

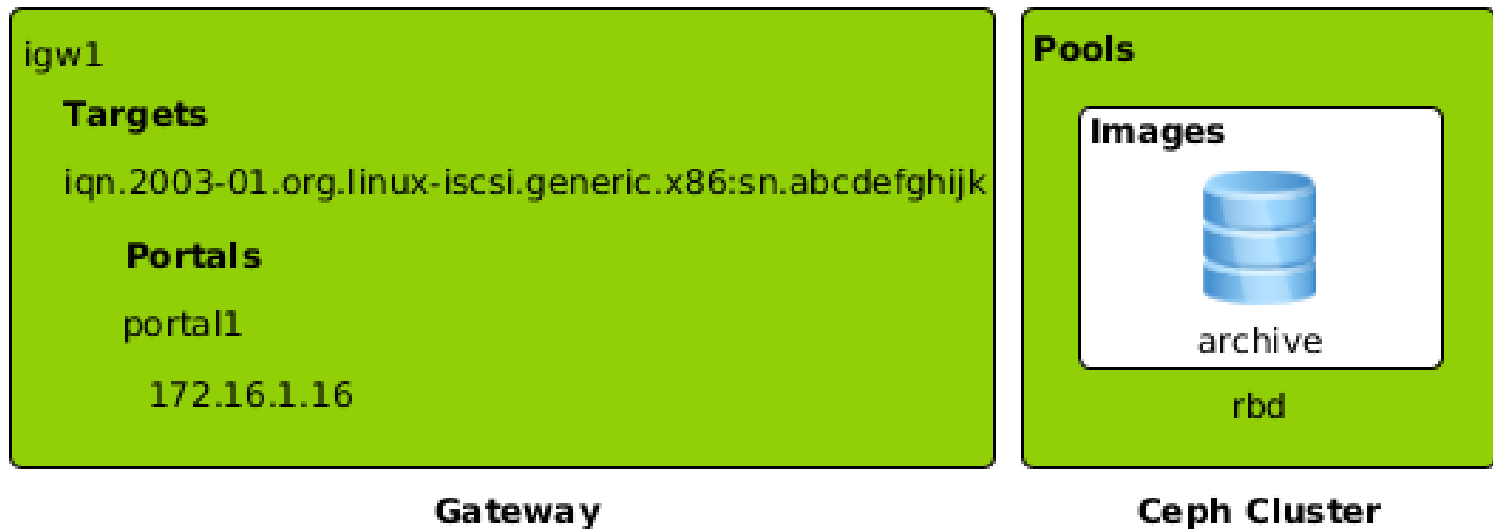
- # lrbd -f ~/simple.json

- # lrbd

- # targetcli ls

```
o- / ..... [...]
  o- backstores ..... [...]
    | o- fileio ..... [0 Storage Object]
    | o- iblock ..... [0 Storage Object]
    | o- pscsi ..... [0 Storage Object]
    | o- rbd ..... [1 Storage Object]
    | | o- archive ..... [/dev/rbd/rbd/archive activated]
    | o- rd_mcp ..... [0 Storage Object]
  o- ib_srpt ..... [0 Targets]
  o- iscsi ..... [1 Target]
    | o- iqn.2003-01.org.linux-iscsi.generic.x86:sn.abcdefghijkl ..... [1 TPG]
    |   o- tpg1 ..... [enabled]
    |     o- acls ..... [1 ACL]
    |       | o- iqn.1996-04.de.suse:01:e6ca28cc9f20 ..... [1 Mapped LUN]
    |         | o- mapped_lun0 ..... [lun0 (rw)]
    |       o- luns ..... [1 LUN]
    |         | o- lun0 ..... [rbd/archive (/dev/rbd/rbd/archive)]
    |       o- portals ..... [1 Portal]
    |         o- 172.16.1.16:3260 ..... [OK, iser disabled]
  o- loopback ..... [0 Targets]
  o- qla2xxx ..... [0 Targets]
  o- tcm_fc ..... [0 Targets]
  o- vhost ..... [0 Targets]
```

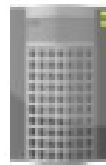
# Simple gateway for Ceph RBD



# Multiple iSCSI gateways for Ceph RBD



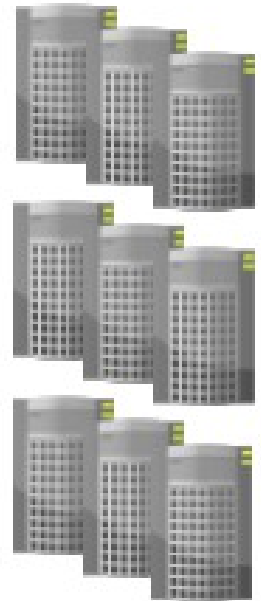
**Client**



**Gateway**  
**172.16.1.16**



**Gateway**  
**172.16.1.17**



**Ceph Cluster**



# Irbd configuration

```
"targets": [  
  {  
    "hosts": [  
      { "host": "igw1", "portal": "portal1" },  
      { "host": "igw2", "portal": "portal2" }  
    ],  
    "target": "iqn.2003-01.org.linux-iscsi.igw.x86:sn.redundant"  
  }  
],
```

```
"pools": [  
  {  
    "pool": "rbd",  
    "gateways": [  
      {  
        "target": "iqn.2003-01.org.linux-  
iscsi.igw.x86:sn.redundant",  
        "tpg": [  
          {  
            "image": "city"  
          }  
        ]  
      }  
    ]  
  }  
]
```

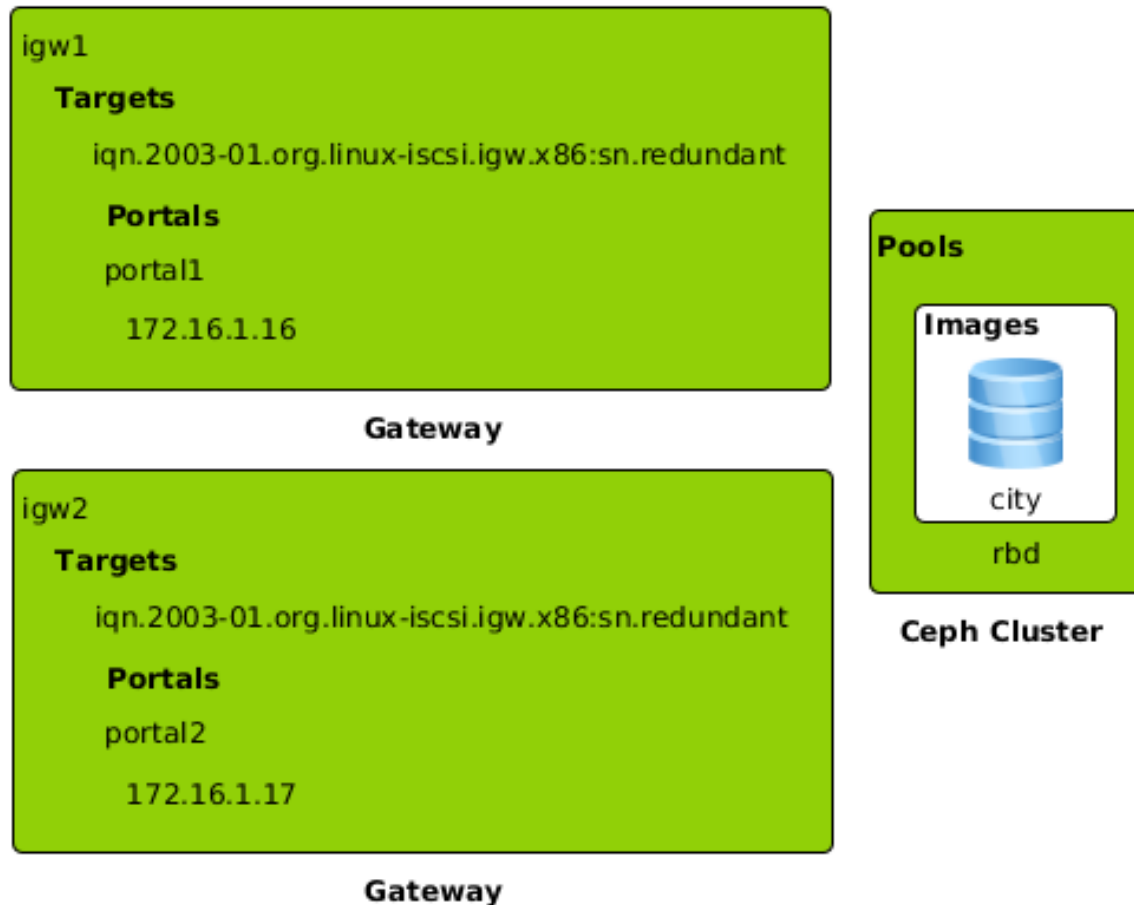
```
"portals": [  
  {  
    "name": "portal1",  
    "addresses": [ "172.16.1.16" ]  
  },  
  {  
    "name": "portal2",  
    "addresses": [ "172.16.1.17" ]  
  }  
],
```

```
"auth": [  
  {  
    "authentication": "none",  
    "target": "iqn.2003-01.org.linux-  
iscsi.igw.x86:sn.redundant"  
  }  
],
```

# Apply configuration

```
• # targetcli ls
o- / ..... [...]
  o- backstores ..... [...]
    | o- fileio ..... [0 Storage Object]
    | o- iblock ..... [0 Storage Object]
    | o- pscsi ..... [0 Storage Object]
    | o- rbd ..... [1 Storage Object]
    | | o- city ..... [/dev/rbd/rbd/city activated]
    | o- rd_mcp ..... [0 Storage Object]
  o- ib_srpt ..... [0 Targets]
  o- iscsi ..... [1 Target]
    | o- iqn.2003-01.org.linux-iscsi.igw.x86:sn.redundant ..... [2 TPGs]
      | o- tpg1 ..... [enabled]
        | o- acls ..... [0 ACLs]
        | o- luns ..... [1 LUN]
          | | o- lun0 ..... [rbd/city (/dev/rbd/rbd/city)]
          | o- portals ..... [1 Portal]
            | o- 172.16.1.16:3260 ..... [OK, user disabled]
      o- tpg2 ..... [disabled]
        o- acls ..... [0 ACLs]
        o- luns ..... [1 LUN]
          | o- lun0 ..... [rbd/city (/dev/rbd/rbd/city)]
          o- portals ..... [1 Portal]
            o- 172.16.1.17:3260 ..... [OK, user disabled]
  o- loopback ..... [0 Targets]
  o- qla2xxx ..... [0 Targets]
  o- tcm_fc ..... [0 Targets]
  o- vhost ..... [0 Targets]
```

# Multiple gateways for Ceph RBD



DEMO

# Questions?

## More to visit:

<https://github.com/SUSE/lrbd/wiki>

# Thank you.





**Corporate Headquarters**  
Maxfeldstrasse 5  
90409 Nuremberg  
Germany

+49 911 740 53 0 (Worldwide)  
[www.suse.com](http://www.suse.com)

Join us on:  
[www.opensuse.org](http://www.opensuse.org)



## **Unpublished Work of SUSE LLC. All Rights Reserved.**

This work is an unpublished work and contains confidential, proprietary and trade secret information of SUSE LLC. Access to this work is restricted to SUSE employees who have a need to know to perform tasks within the scope of their assignments. No part of this work may be practiced, performed, copied, distributed, revised, modified, translated, abridged, condensed, expanded, collected, or adapted without the prior written consent of SUSE. Any use or exploitation of this work without authorization could subject the perpetrator to criminal and civil liability.

## **General Disclaimer**

This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. SUSE makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for SUSE products remains at the sole discretion of SUSE. Further, SUSE reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All SUSE marks referenced in this presentation are trademarks or registered trademarks of Novell, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners.

