Location-Aware Distributed Virtual Disk Storage for OpenStack

Keiichi SHIMA <keiichi@iijlab.net>
IIJ Innovation Institute Inc.

DataCenter とソフトウェア開発ワークショップ

2014-11-20

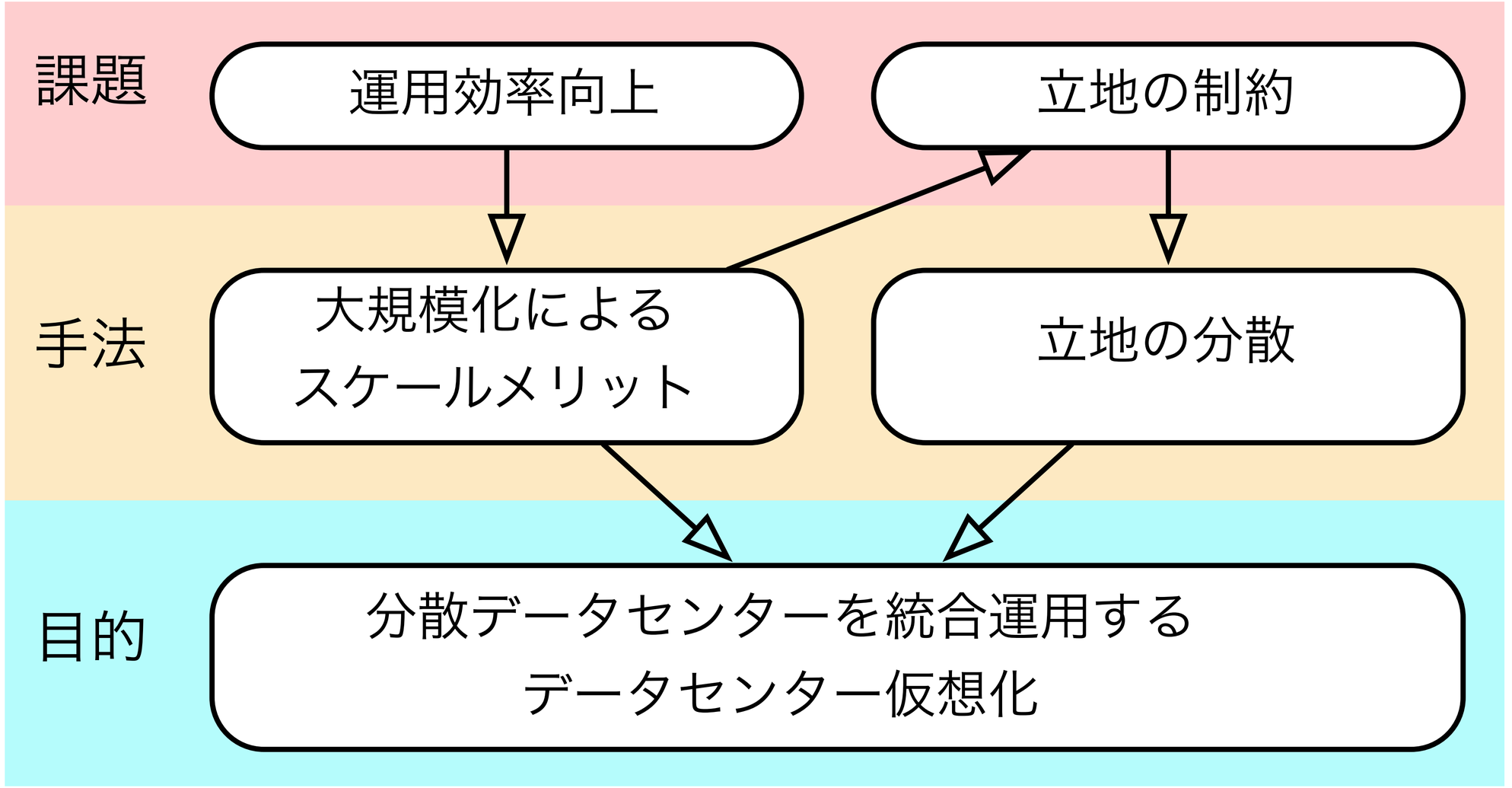
Background

- Widely spread virtualization deployment
- Integrated distributed datacenter
- Flexible resource management

Virtualization

- There is no service provider who doesn't use virtualization
- Many Internet services are now built on top of virtualized computers
- Easy to deploy, scale out when necessary, and scale down when shutting down services

分散統合データセンター



分散統合データセンターの課題

仮想資源の効率的な配置・再配置

CPU・メモリ
→ 利用可

ネットワーク
→ SDNなど

ストレージ
→ ???

Current Status

- Integrated Approaches
  - NFS or iSCSI storage devices
- Distributed Approaches
  - Ceph (RBD)
  - GlusterFS
  - Sheepdog

Current Status

- Integrated Approaches
  - NFS or iSCSI storage devices
- Distributed Approaches
  - Ceph (RBD)
  - GlusterFS
  - Sheepdog

Current Status

- Integrated solution
  - Single point of failure (typically)
  - Difficult to relocate resources

Current Status

- Distributed approaches
  - Difficult to adapt non-uniform environment
  - Administration issues (in the sense of a resource location control)

Idea is

- To use only good parts of the integrated and distributed solutions

Objectives

- Flexibility
- Redundancy
- Locality

Flexibility

- Flexible management of location
  - Place the data at the specified location
  - not at somewhere in a storage cluster
- Operation unit
  - Per virtual disk
- Flexibility in relocation timing
  - Try to prevent to disturb other traffic

Redundancy

- Replication of a virtual disk
  - Even more than two replicas
  - Dynamic operation of replication

Locality

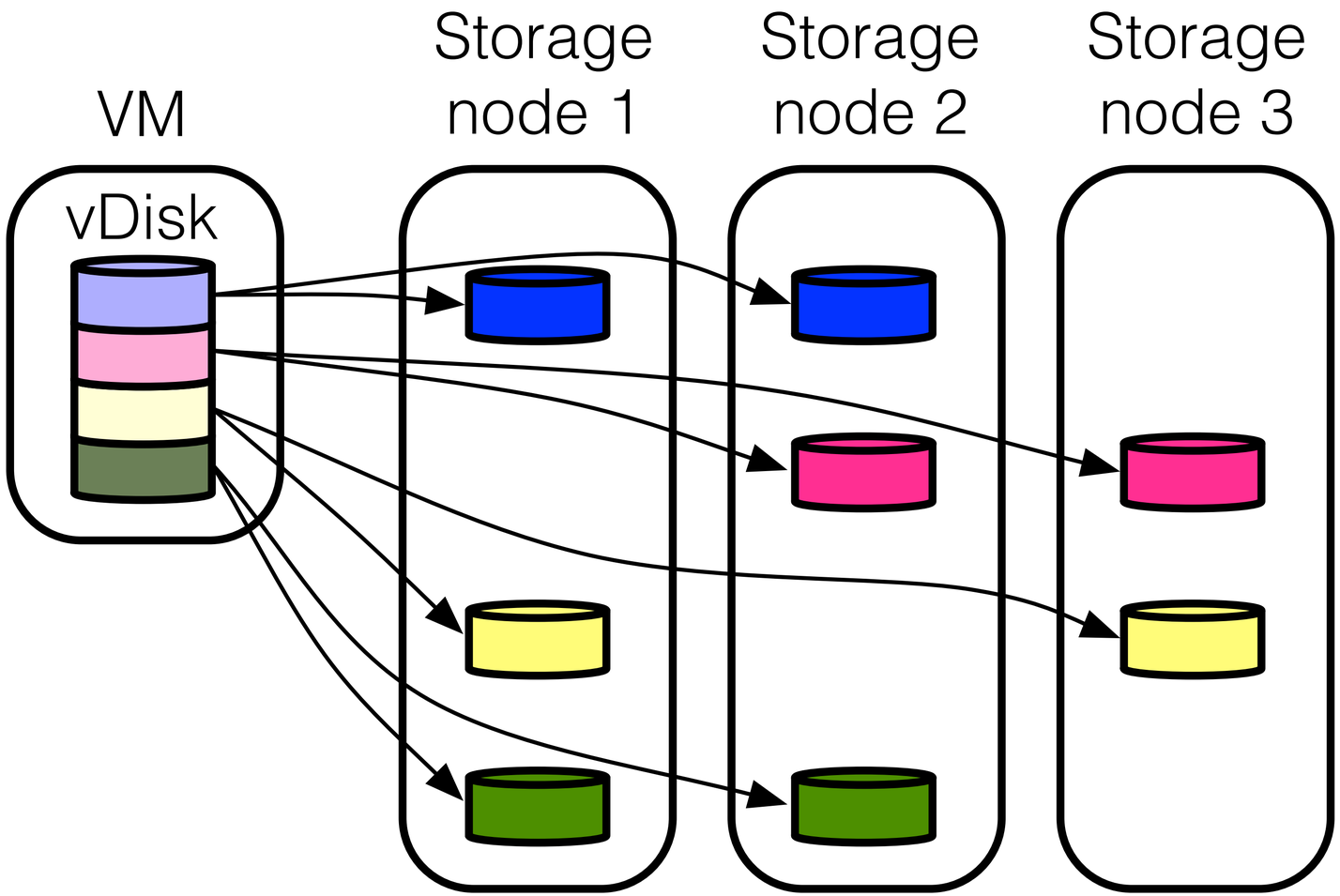
- Locate the actual data of a virtual disk as near as possible to the virtual machine using it

Operation Scenarios

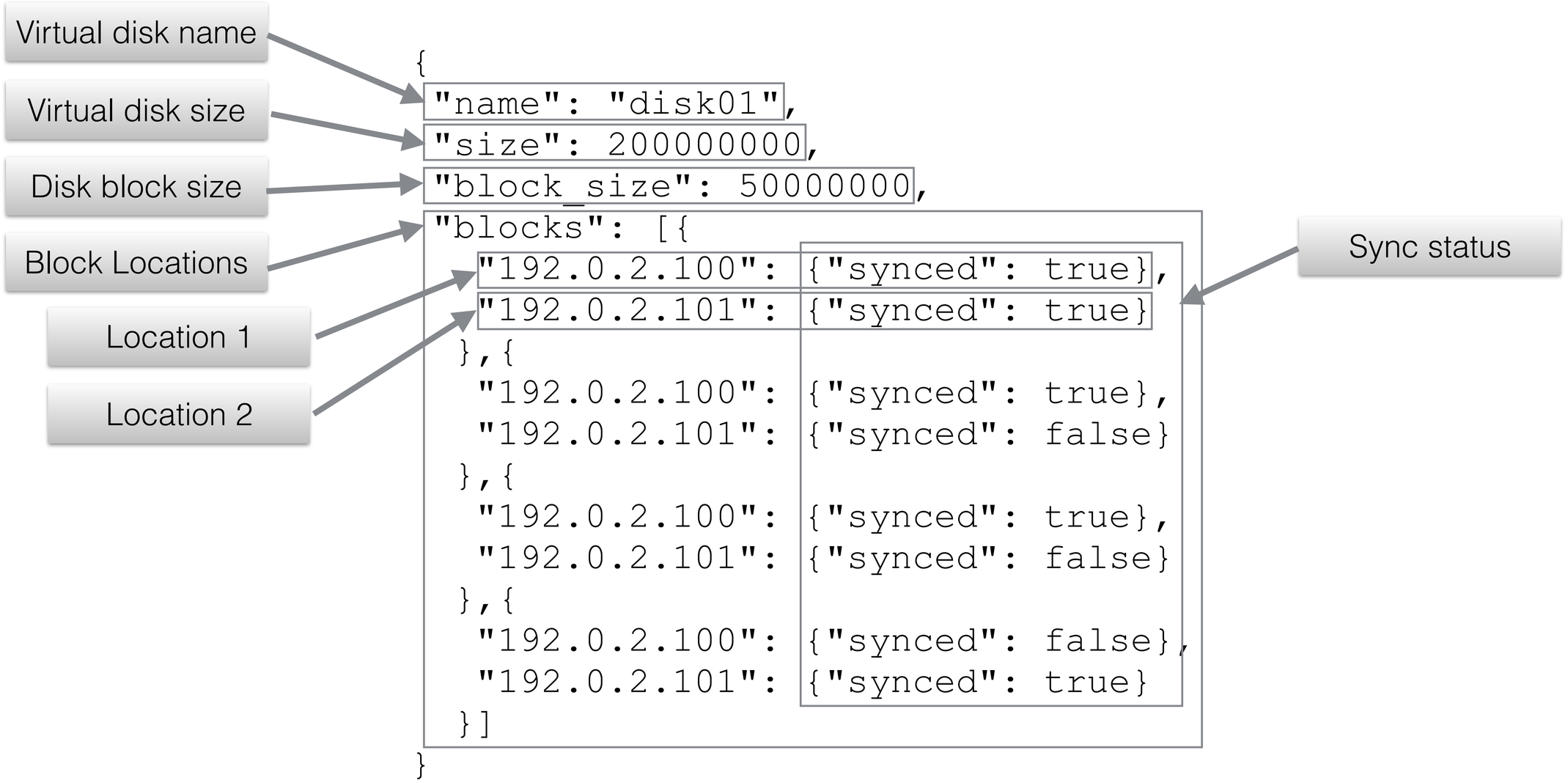
- Consolidation
  - Consolidate virtual machines to minimize the # of running hypervisors
- Move
  - Push virtual machines out from a hypervisor to upgrade the hypervisor
  - Move virtual machines from one DC to another DC, e.g. due to new DC launch, or discontinue
- Efficient resource usage
  - Use best performing pair of hypervisors and storage nodes

- A virtual storage device for virtual machines
  - Location control per block
  - Replication operation per block
  - No global lock required
    - Since a virtual storage device is owned by only one virtual machine at the same time

UKAI Virtual Disk Concept



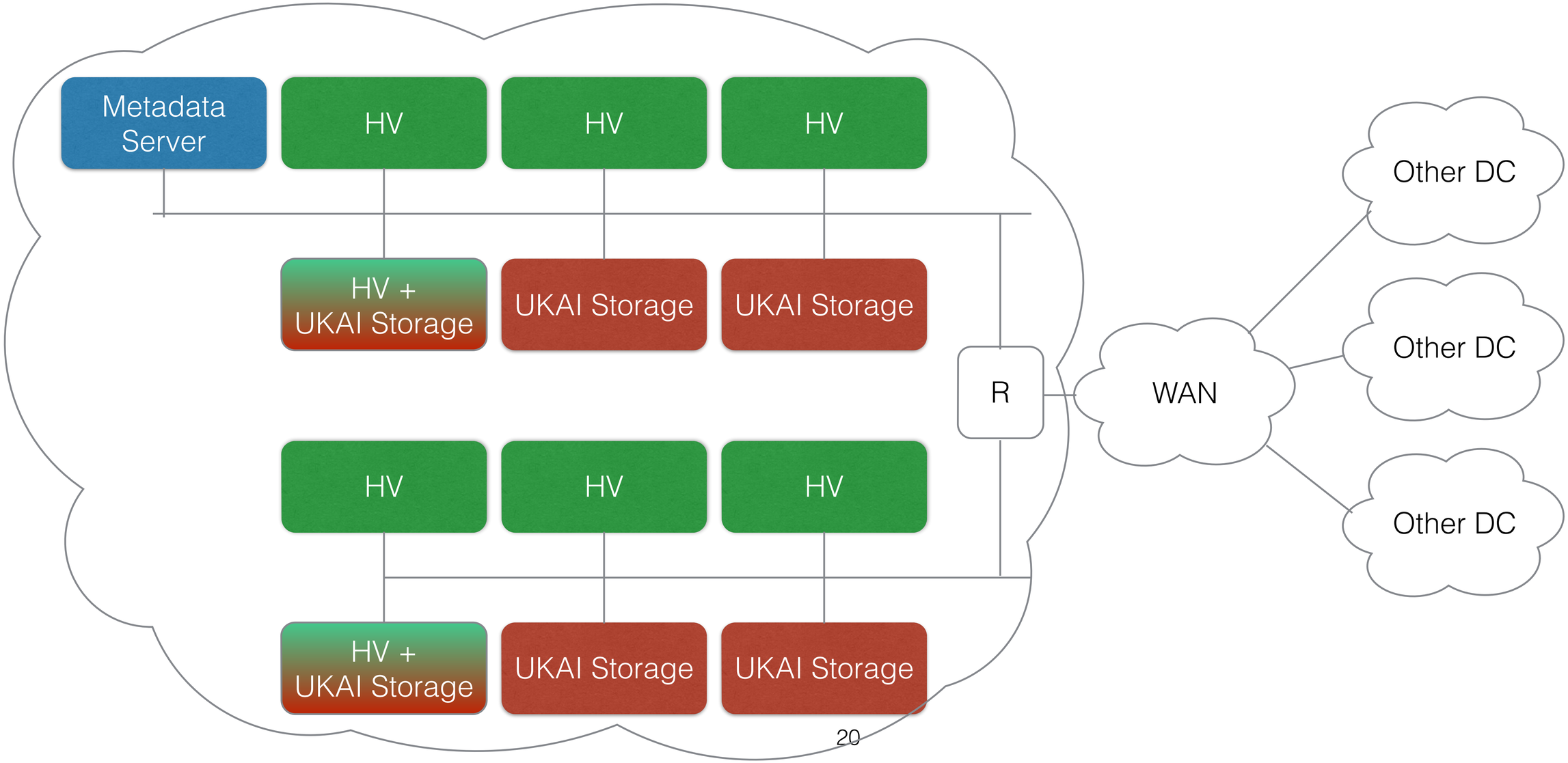
Metadata Structure



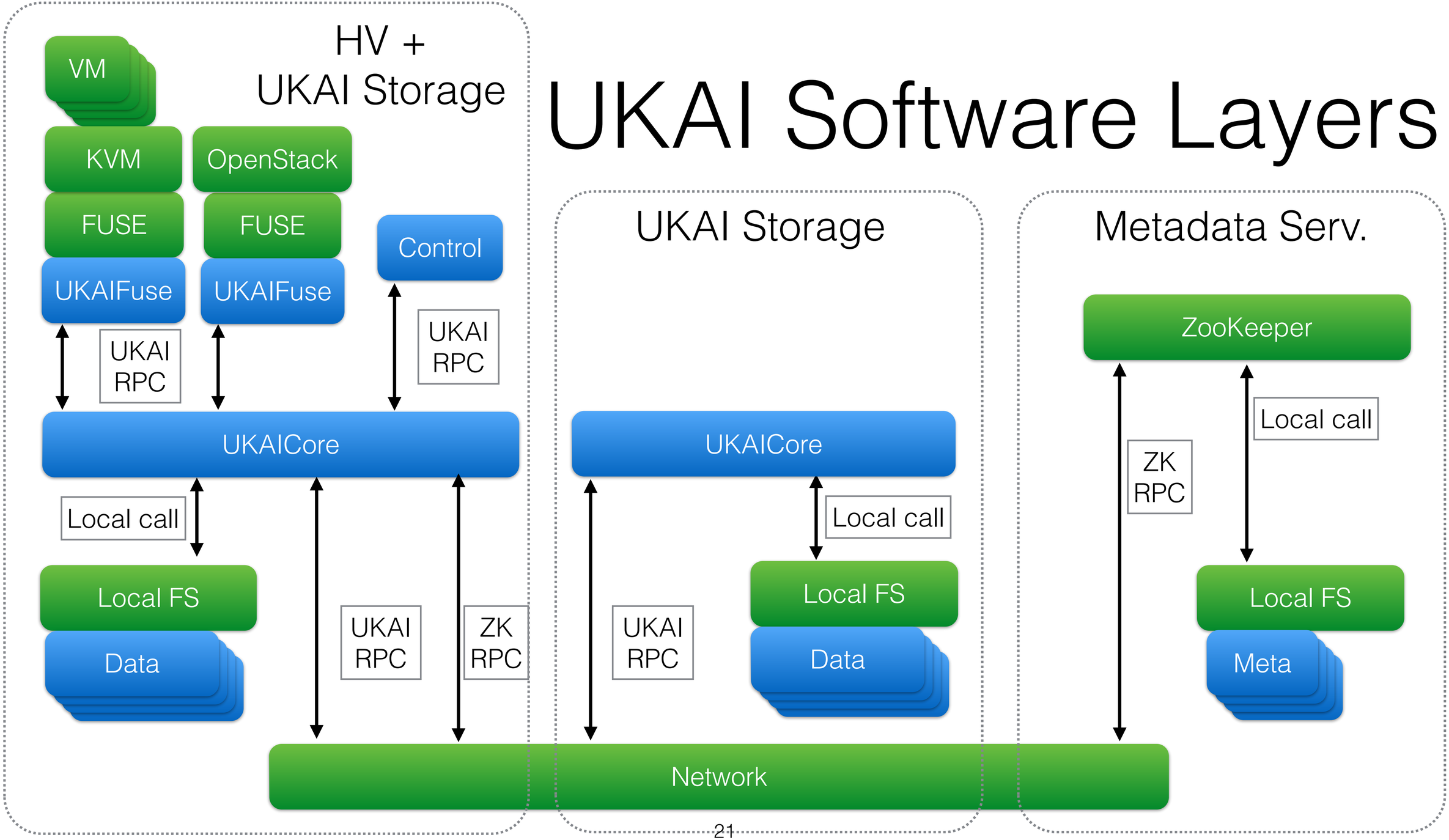
Metadata Server

- The server nodes keeping virtual data disk structure information
- Apache ZooKeeper is used

Typical Node Layout



UKAI Software Layers



Supported Disk Admin Ops

- Create a disk image / Destroy a disk image
  - Size and block size is configurable
- Add location / Delete location
  - Per block granularity
- Synchronize
  - Per block granularity

OpenStack Integration

- Compute (Nova)
  - A new virtual disk mount adapter module for UKAI is required
- Block storage (Cinder)
  - A new storage resource management module for UKAI is required
- Object storage (Swift)
  - Out of scope of this project

Demo

1. CinderによるUKAIボリュームの作成
2. GlanceからUKAIボリュームへCirrOSイメージをダウンロード
3. CirrOSボリュームからインスタンスブート
4. CirrOSボリュームを別のストレージノードへ移動

Hurdles

- Performance
  - At this moment, worse than NFS
    - Possible reasons: FUSE, RPC, no-caching
- Management facility
  - 100% manual operation is not always required
  - Need some support tools for node selection based on the predefined hypervisor/storage locations

Summary

- We need a new storage backend system to support distributed and integrated virtual infrastructure operation
- UKAI: a location-aware distributed storage system for virtual machines
  - Flexible location management, redundancy, and locality are provided
- OpenStack integration is possible
- Need more development efforts to get better performance

Availability

- UKAI
  - https://github.com/keiichishima/ukai
- OpenStack support
  - The UKAI branch in https://github.com/keiichishima/nova
  - The UKAI branch in https://github.com/keiichishima/cinder

Questions?

- UKAI
  - https://github.com/keiichishima/ukai
- OpenStack support
  - The UKAI branch in https://github.com/keiichishima/nova
  - The UKAI branch in https://github.com/keiichishima/cinder