SDC

STORAGE DEVELOPER CONFERENCE

SNIA SANTA CLARA, 2017

How we evaluate Storage Performance and do Capacity Planning for Telecom Cloud

Kei Kusunoki, Junji Arakawa

NTT Communications

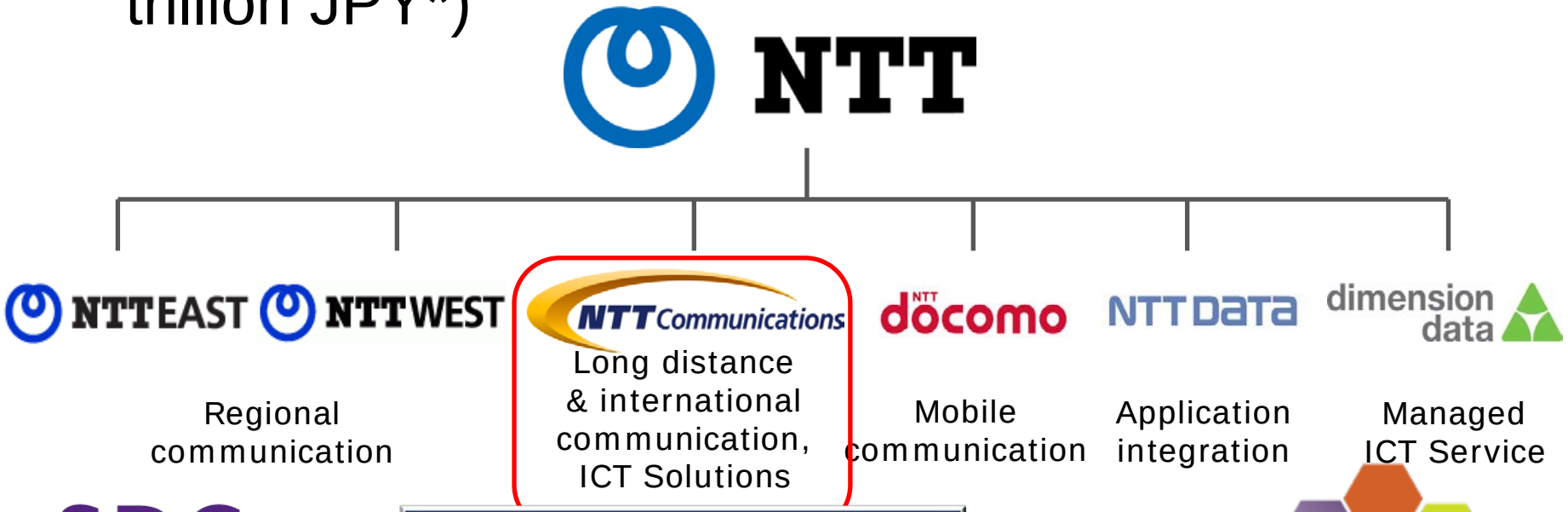
Agenda

- Our Public IaaS Overview and Required Storage Function
- Storage Evaluation
- Capacity Planning

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

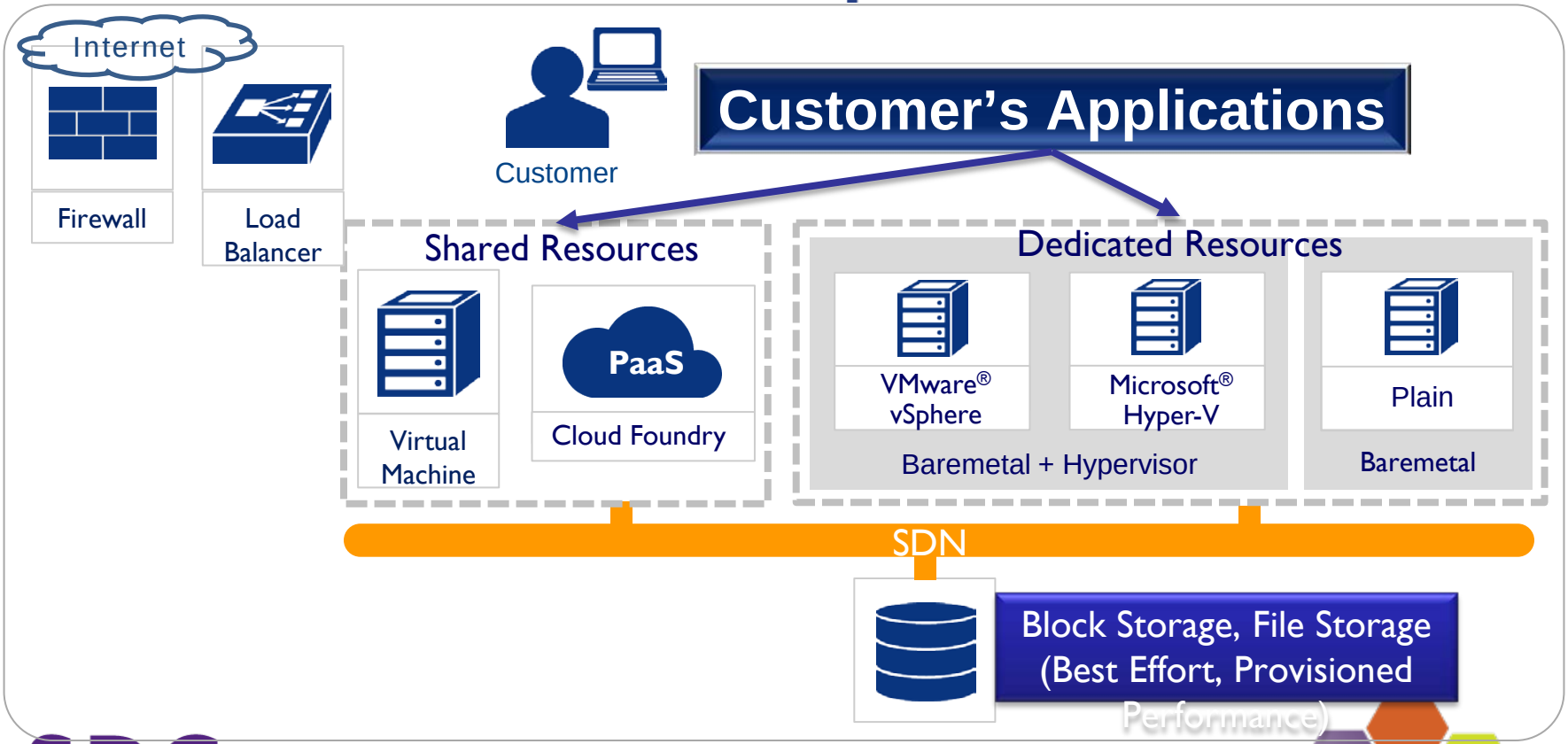
NTT Group overview

- One of the world's largest ICT companies (Consolidated revenue of approximately 11.4 trillion JPY*)



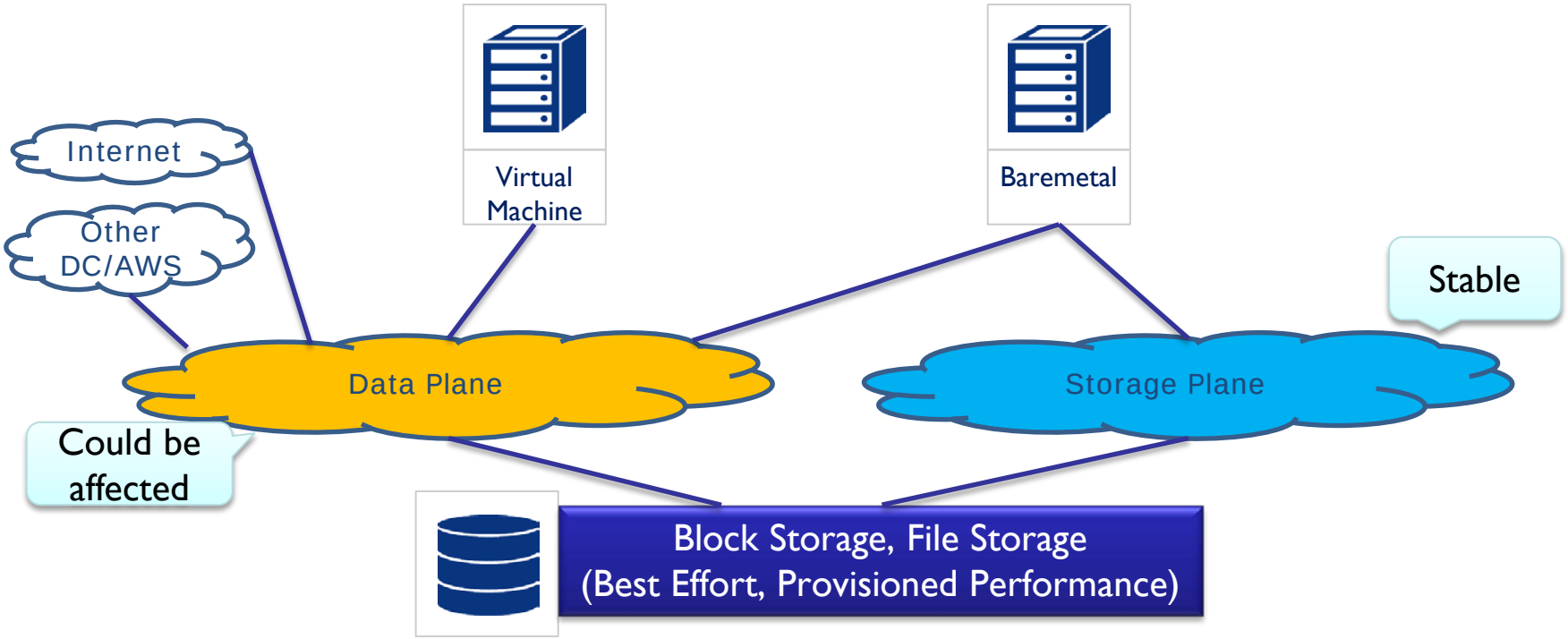
All Rights Reserved.

Our Cloud Service Enterprise Cloud



2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Storage Network



2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Our Storage Service

- Block Storage
  - Provisioned IOPS: 2IOPS/GB, 4IOPS/GB
- File Storage (NFSv3)
  - Best Effort
  - Provisioned Throughput (Up to 400MB/s)

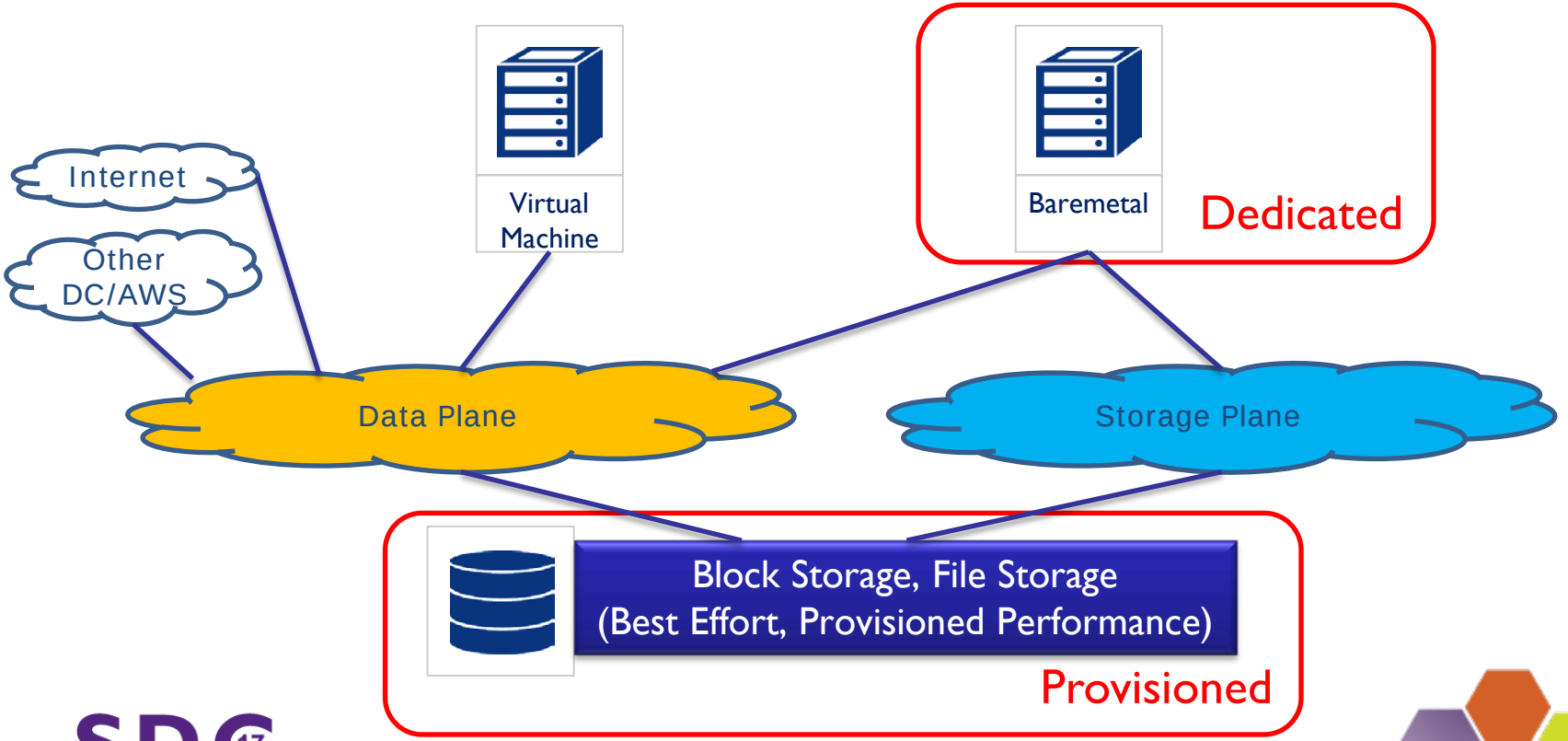
2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


How to provide “Provisioned” Storage Service in “shared” cloud infrastructure?

- Can Customer use provisioned storage performance at any time/situation?
  - Server -> Dedicated Baremetal
  - Network -> Best Effort
  - Storage -> Provisioned Storage
- We’ve not guaranteed end-to-end storage performance.

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

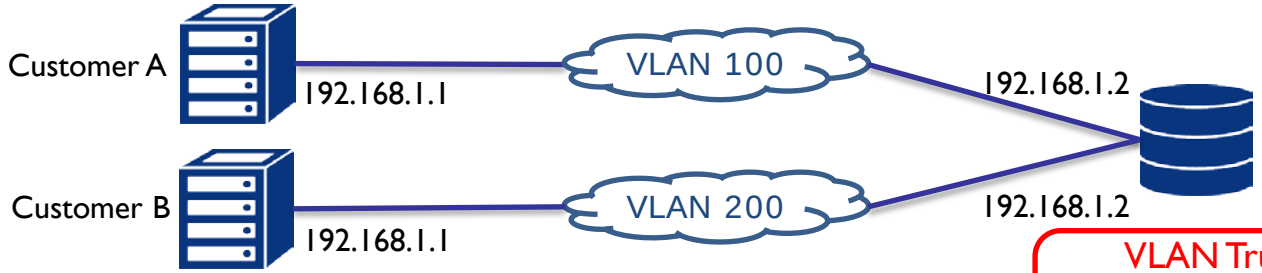

How to provide “Provisioned” Storage Service in “shared” cloud infrastructure?



2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Required Storage Functions for multi-tenant Provisioned Storage Service

- User tenant Isolation
  - Customer should not access another customer's storage
    - Lv1 Defense by CHAP authentication
    - Lv2 Prohibition by Config
    - Lv3 Invisible to all
- Network multi-tenancy



VLAN Trunk Tag support
Duplicated IP on each VLAN
(Virtual Routing and Forwarding)

2017 Storage Developer Conference. © NTT Communications. All Rights

Required Storage Technology for multi-tenant Provisioned Storage Service

- Storage Performance QoS
  - Capped at IOPS AND Throughput
  - Nice to Have: Minimum Performance, Burst Mode in low season
- Dedup Count Width
  - Which Width Storage can distinguish as duplication?
    - Lv1 Tenant/Volume
    - Lv2 Storage Array
    - Lv3 Whole Cluster

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Why we need such multi-tenancy/QoS for Storage itself?

- Other Solutions?
  - Storage Gateway
    - Bottle-neck, Higher cost, Complicated Data Path etc.
  - Routing Control on SDN
    - Bottle-neck, Need further SDN controller development
    - Network is more expensive than Storage

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Capacity Saving Function: Deduplication and Compression

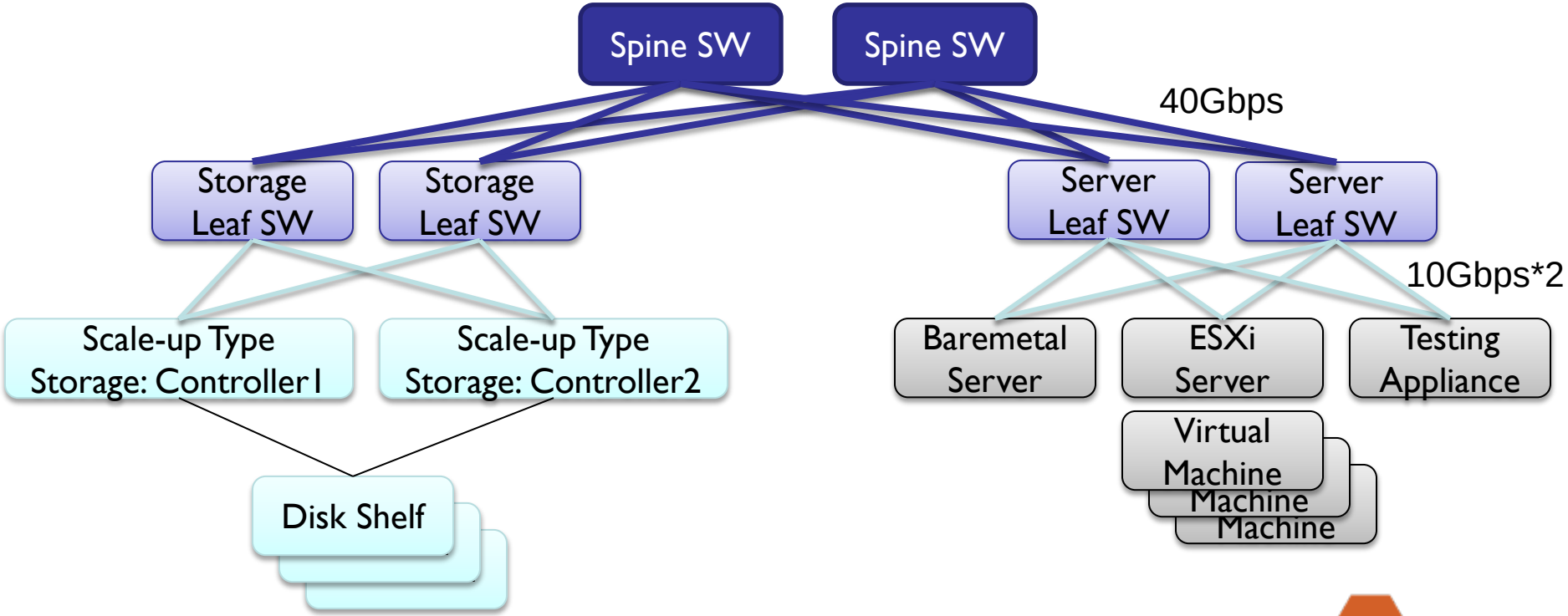
- x86CPU type Storage System
  - Trade-Off: Capacity or Performance
  - Some Storage use almost 100% CPU without Dedup/Comp.
- ASIC type Storage System
  - Relatively low performance impact

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Storage Evaluation

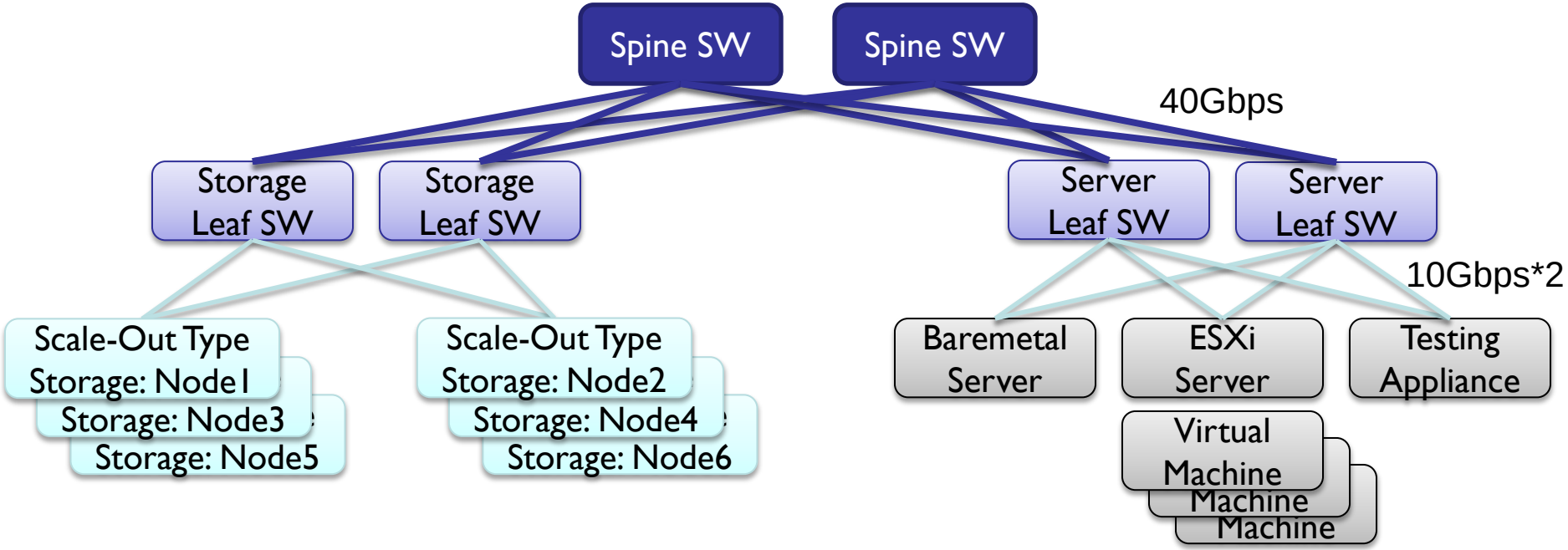
2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Testing Environment Overview



2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Testing Environment Overview



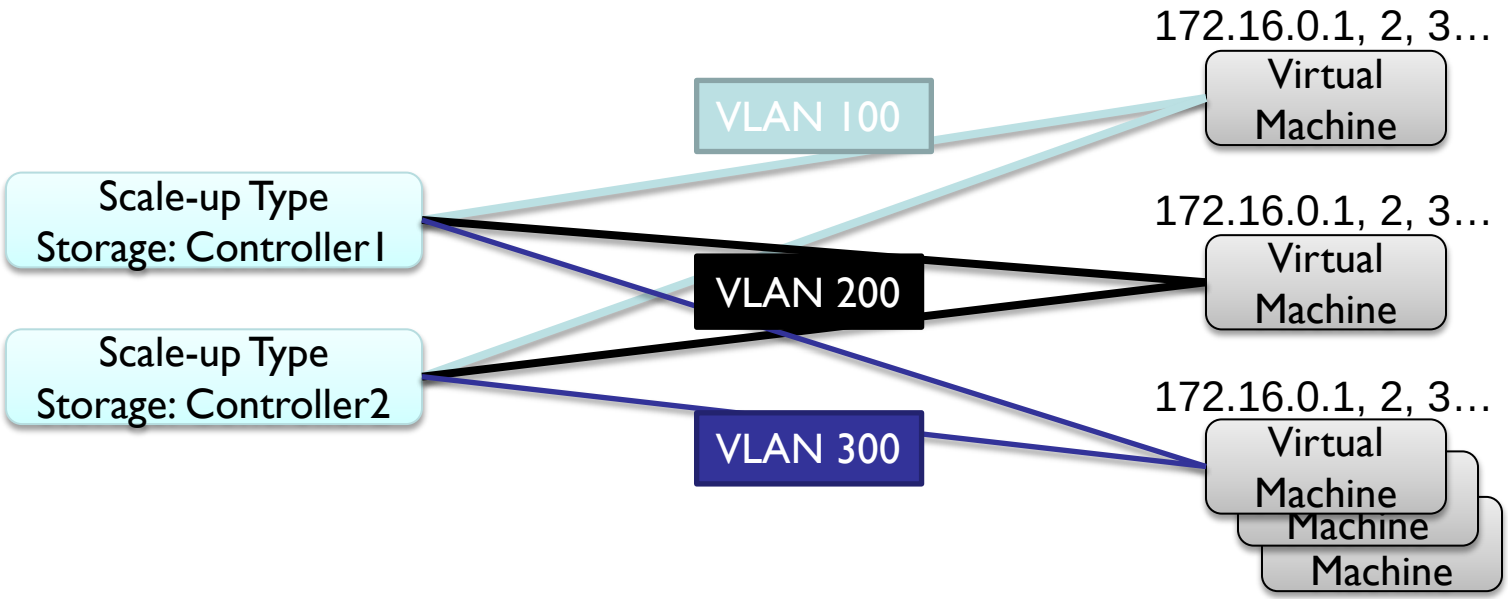
2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Storage Configuration

- Same as Production
  - RAID/Disk Group/Disk Shelf/Volume
  - VLAN/IP interfaces on the array/node
- Available capacity should be worst case
  - Ex. 20%, 10%
  - Flash Array generally needs free space for GC

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Network Example



2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Workload Client

- General Linux on Baremetal/ESXi VM
  - Each client has different IP/VLAN
  - Oracle Vdbench
- Testing Appliance
  - LoadDynamix
  - Exhaustive benchmark for some parameters
    - Asynchronous number, meta operation on NFS, etc.

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Workload Definition

- Can we expect real production's workload?
  - Public IaaS is not "Enterprise"
  - Nobody knows how customers use it.
    - DIY: App selection and OS/App configuration
    - Production measurement -> Quite Difficult
      - We cannot access customer' VM/Baremetal
      - Huge amount of mixed/tapped Storage Packet
      - It has non-negligible impact to capture analytics from the Storage

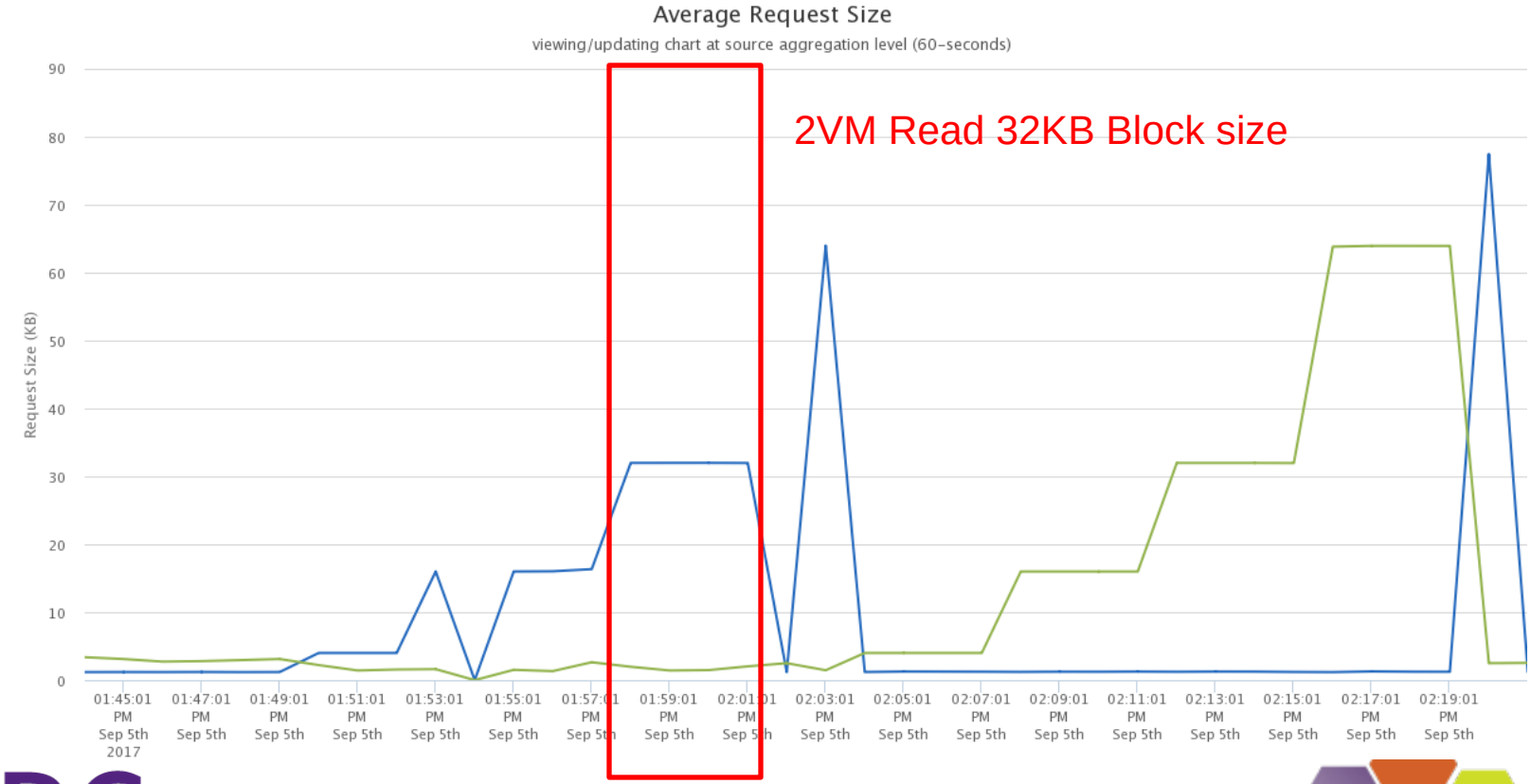
2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Is there any “typical” workload on IaaS?

- Does VMware iSCSI initiator aggregate storage access on same Datastore?
  - Ex. Two VMs write 8KB block to each vmdk file on same Datastore
- How about NFS?

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


We see no aggregated access…



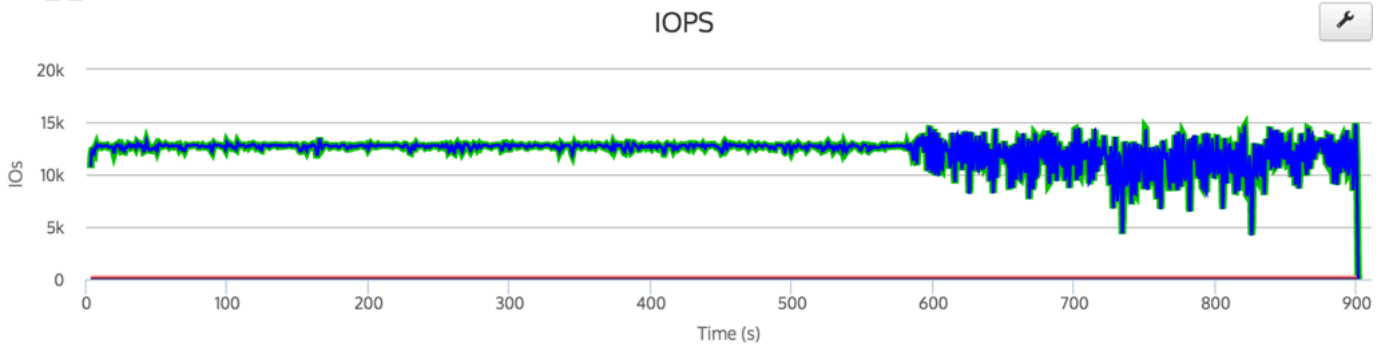
2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Workload Definition

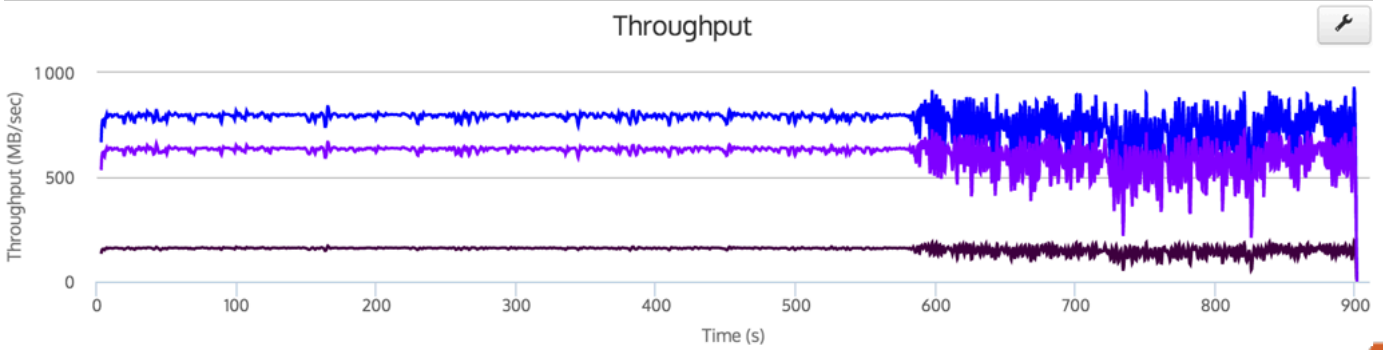
- Our service description
  - Block size 16KB, Random Access 100%, Read/Write 80/20%
- Evaluation Items
  - Max Performance with defined workload
  - R/W=50/50, 100/0, 0/100
    - Access is 100% random
    - Short and long evaluation time

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Example: Cache Overflow or GC impact on long time test



64KB Block, 64 Client /w 64VLAN



2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Testing with Normal/Failure case

- Storage should be evaluated not only with Normal Case, but also Failure case
- Supposed Failures
  - One component failure
  - Over two failures depends on Service SLA.
    - Ex. HDD is fragile. Two HDD down should be assumed.

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

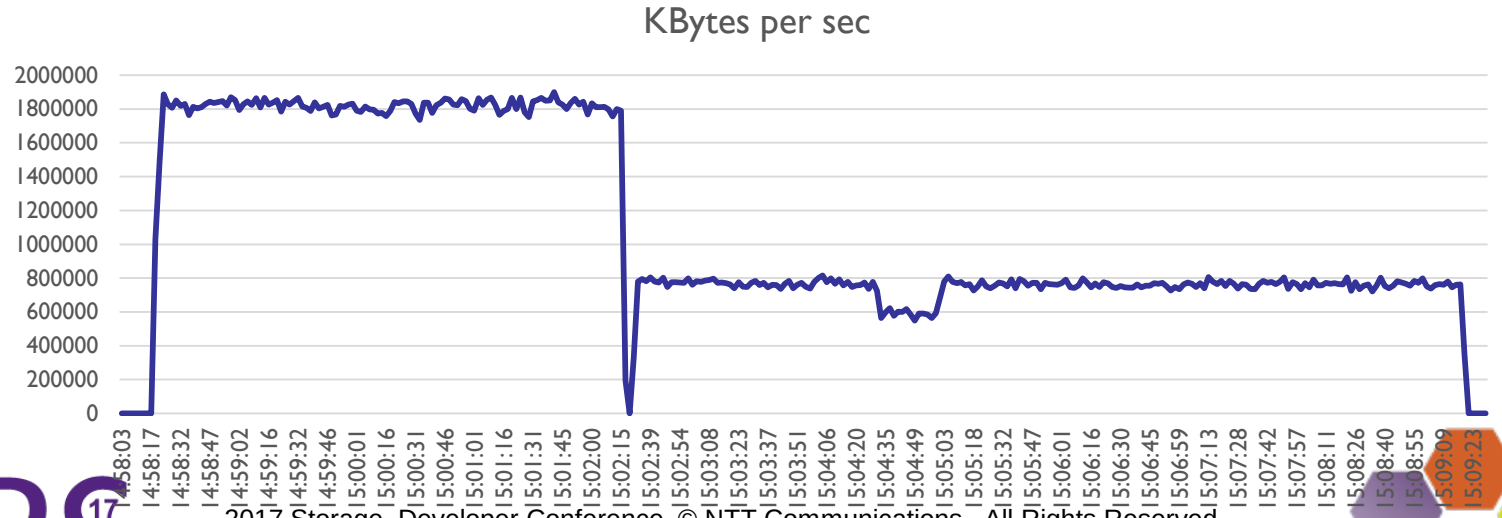

One Failure Pattern

- Scale Up Type Storage
  - Controller Failure
    - Sudden down and manual Take-over
  - SAS Card/Cable Failure
  - We've not supposed the loss of whole disk-shelf
    - Redundant Power has no performance impact
    - Other components in Disk-shelf is simple and robust

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Example: ¼ Controller Failure

- Storage Product A has 4 controllers. One node failure is supposed to cause ¼ performance degradation. But…



2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

One Failure Pattern

- Scale Out Type Storage
  - Node down (Sudden/Manual)
  - We’ve not assumed SAS/SATA card failure
    - It’s equivalent to the loss of all disks on the node (= Node down)

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


One Failure Pattern

- Common case in Scale-Up and Scale-Out
  - Disk Failure
  - Network Interface Card/Port Failure
  - Single Leaf Switch Failure

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Performance degradation pattern with Failure Case

- Momentary Outage of Entire Service
  - Controller/Node Failure during Take-over
  - Switch Failure during path changeover
- Performance degradation during short-time Failure
  - Controller/Node Failure during Take-over
  - RAID rebuild impact during failure and recovery
  - Garbage Collection on Flash Array

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Performance degradation pattern with Failure Case

- Performance degradation during long-time Failure
  - Controller/Node Failure after Take-over or during recovering process
    - Additional performance impact
  - RAID rebuild impact
    - High capacity SSD/HDD needs long rebuild time

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Capacity Planning “Examples”

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Capacity Planning from the results

- Based on the service policy
  - Tolerating some performance down with Failure
    - Normal performance would be used for the planning
  - Customer should not face any degradation for long time
    - Long-time Failure's perf would be used
  - No toleration even if Failure
    - Short-time Failure's perf would be user

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Items that could limit Capacity Planning other than Performance

- Even if the storage system has great performance, these items could limit the planning
  - Maximum Volume number
  - Maximum account container number
  - Maximum QoS configuration number
  - Logical network interface number (IP/VLAN/VRF)

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Example 1 (Scale Up)

- Storage Array X
  - Capacity RAW 10TB, Usable 8TB
  - IOPS 16K IOPS(Normal), 11K IOPS(/w short failure), 8K IOPS (/w long failure)
- 8000 IOPS / 8000GB = 1 IOPS/GB

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.

Example 1 (Scale Up)

- Unfortunately, only 64 VLANs can be configured
  - 8000GB/64VLAN=125GB/VLAN is much larger than 100GB
  - Our Storage menu starts from 100GB
- Recent storage have large SSD (ex. 15TB, 30TB etc)
  - Smaller rack space give us significantly lower cost
  - But above limitation prevents us to chose such large SSD.

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Example 2 (Scale Out)

- Storage cluster Y
  - 1 Node
    - Capacity RAW 10TB, Usable 8TB
    - IOPS 50K IOPS(Normal), 40K IOPS(/w short failure)
- 40,000 IOPS / 8000 GB = 5 IOPS/GB

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Example 2 (Scale Out)

- This product can have 512 accounts per cluster
  - 6Node: 8TB * 6 /512 = 93.75GB < 100GB
- We can’t have more nodes in one cluster
  - Less Dedup efficiency, redundancy, flexibility to deploy, etc.

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Summary

- We, carrier cloud provider need multi-tenancy and QoS to Storage.
- Using defined workload, we evaluate maximum storage performance with normal/failure case.
- Limitation number of multitenancy/QoS/Network could restrict Capacity Planning other than Storage performance

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.


Thank you and Q&A

2017 Storage Developer Conference. © NTT Communications. All Rights Reserved.