Deploying and Optimizing for Cloud Storage Systems using Swift Simulator

Gen Xu
Intel Corporation
Agenda

- Design challenges
- Cloud Storage system Swift modeling overview
- Use Case study
Cloud Storage System Design Challenges

Our customers are commonly addressing these challenges:

- Plan storage capacity?
- Not over provisioning?
- Meet the SLA/SLO requirements?
- Minimize cost?
- Predict system performance?
Complex Cluster Architecture

Storage
Network Topology

Processor
Memory
Simulation Architecture

What-If Analysis for

- S/W stack optimization
- Predict perf on varies node, network and disk configuration
- Explore against users number and cluster size

Details @ ICPP-2014 paper “Simulating Big Data Clusters for System Planning, Evaluation and Optimization”
Agenda

- Design challenges
- Cloud Storage system Swift modeling overview
- Use Case study
Swift Simulator

- Logic View of a Swift Cluster
- Frontend model of Simulation

Backend model: System Topology (Star, CLOS…), Ring parsing, Node mapping, Coroutine scheduling, Perf Lib...
Simulation Accuracy

- **Hardware Validation**
  - STAR, Fat-Tree, CLOS
  - 1GbE, 10GbE, 25GbE, 50GbE
  - HDD, SSD, NVMe
  - Node count
  - …

  Average ~96%

- **Software Validation**
  - Operation Type (PUT, GET, MIX)
  - Object Size (16KB ~ 1024MB)
  - Proxy worker#, Object worker#
  - Concurrency#
  - Write Barrier
  - …

  Average ~95%
Simulation Approach

Collect and Analyze System Log

Reproduce and Validate Simulation

Simulation Plan

- Adjust WL setting
- Change S/W config
- Change H/W config

Generate Simulation Report

- Reproduce and Validate simulation

9
Output: Workload and H/W metrics

- Throughput
- Latency
- Resource usage
  - CPU
  - Network
  - Disk I/O
  - ...

Easy to target bottleneck
Agenda

- Design challenges
- Cloud Storage system Swift modeling overview
- Use Case study
Case Study: Optimize one cloud storage system

Object type | Object size
--- | ---
Thumbnail | 16KB
Audio | 16MB

Design Goal: Achieve to 2X IOPS and 2X bandwidth for both small objects and large objects in read, write and mixed scenarios.
Design options

- Software optimization
- H/W scale up: Storage, Network, CPU
- H/W scale out: add more nodes
Validation against baseline setup

Higher than 95% average accuracy
Ready to use simulator to predict performance
Software optimization

Up to 1.5x improvement with software optimization
Set worker count for proxy and storage to 2 or 3 times the core count
Scale up: HDD -> SSD

Storage Type Impact

Bottleneck changes from Disk (HDD) to CPU (SSD)

- 100R
- 80R-20W
- 100W

16KB

Bottleneck changes from Disk (HDD) to Network (SSD)

- 100R
- 80R-20W
- 100W

16MB

Expect SSDs to improve performance over HDD based storage

Network is the bottleneck with either SSD or HDD

Expect SSDs to improve performance over HDD based storage
Scale up: 10GbE -> 25GbE

Faster network brings performance gain.
Scale up: XEON E5-2695 v3 -> E5-2699 v4

Frequency: Core count: 14 vs. 22

Faster CPU brings performance gain.

Network bottleneck

Up to 1.58x improvement with CPU optimization
Scale-up optimization summary

Performance gain from optimization

Up to 6.73x improvement with scale-up optimization
Scale out optimizing

For small objects, best ratio: 3:8
For large objects, best ratio: 1:2
Storage node bottleneck
Proxy node bottleneck

Up to 2x improvement with scale-out optimization
Better scale up firstly, then scale out
Optimization analysis

- Set worker count for proxy and storage to 2 or 3 times the core count.
- SSD improves performance over HDD.
- 10Gb is okay for small objects. Need to update network for large objects.
- CPU efficiency is high for small objects.
- Better scale up firstly, then scale out.
- Properly configure ratio of proxy to storage server.
Backup
What’s Inside

INTEL® COFLUENT™ TECHNOLOGY FOR BIG DATA
Solutions for big data cluster Simulation, Planning and Optimization
Intel® CoFluent™ Technology for Big Data

**FASTER CLUSTER DEPLOYMENT:**
Explore deployment options and meet performance goals

**OPTIMIZE CLUSTERS:**
Find performance bottlenecks and optimize software operation

**SCALE UP WITH CONFIDENCE:**
Simulate to determine the minimum cost to meet your future demand
Intel® CoFluent™ Studio Based Simulation

Enables fast “What if?” analysis with a virtual system
Hardware Coverage

Validated: 700 Nodes

Rack Scale Architecture
Fast Simulation

Simulation vs. Real Time in minutes

- Hardware - 4 node Cluster (min)
- Simulation Speed - Lenovo T420 (min)

Abstract Modeling

Event Driven Simulation

NUMBER OF CONCURRENT UPLOADING REQUESTS

2016 Data Storage Innovation Conference. © Intel Corporation. All Rights Reserved.
Host machine to run simulations

- Laptop
- Desktop computer
- Server
- Racks (not recommended)
Swift Simulation Input

- Workload Parameters
  - Concurrency
  - Request Size
- S/W Settings
  - Role setting (proxy, storage) for each node
  - Object ring
  - Proxy/object worker number
  - Object size
- H/W Settings:
  - Cluster size
  - System Components (CPU, Disks, Memory, Network)
  - Network topology
# Storage Optimizing: HDD VS SSD

## Storage Devices Assumed Performance

<table>
<thead>
<tr>
<th>Class</th>
<th>Type</th>
<th>IOPS</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Device</td>
<td></td>
<td>RND 4KB 100%R</td>
<td>SEQ. 64KB 100%W</td>
</tr>
<tr>
<td>SSD S3700</td>
<td>1TBGB</td>
<td>47,806</td>
<td>403MB/s</td>
</tr>
<tr>
<td>7200 RPM HDD</td>
<td>1TB</td>
<td>390</td>
<td>180MB/s</td>
</tr>
</tbody>
</table>
Heterogeneous Storage