# 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:**

**HOW TO**

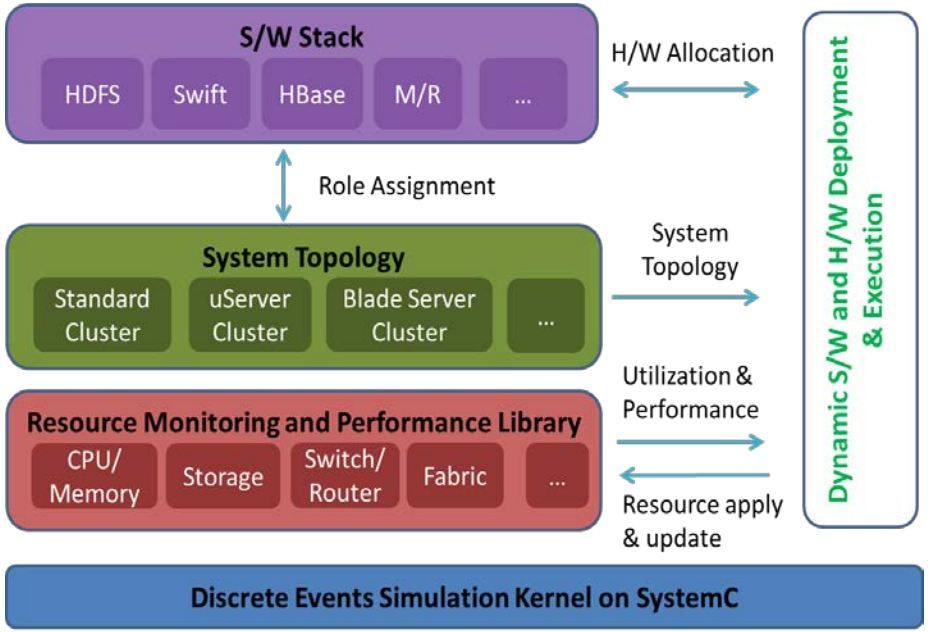
- ✓ ***Plan storage capacity?***
- ✓ ***Not over provisioning?***
- ✓ ***Minimize cost?***
- ✓ ***Meet the SLA/SLO requirements?***
- ✓ ***Predict system performance?***

# Complex Cluster Architecture

Storage
Network Topology

Processor
Memory

# Simulation Architecture

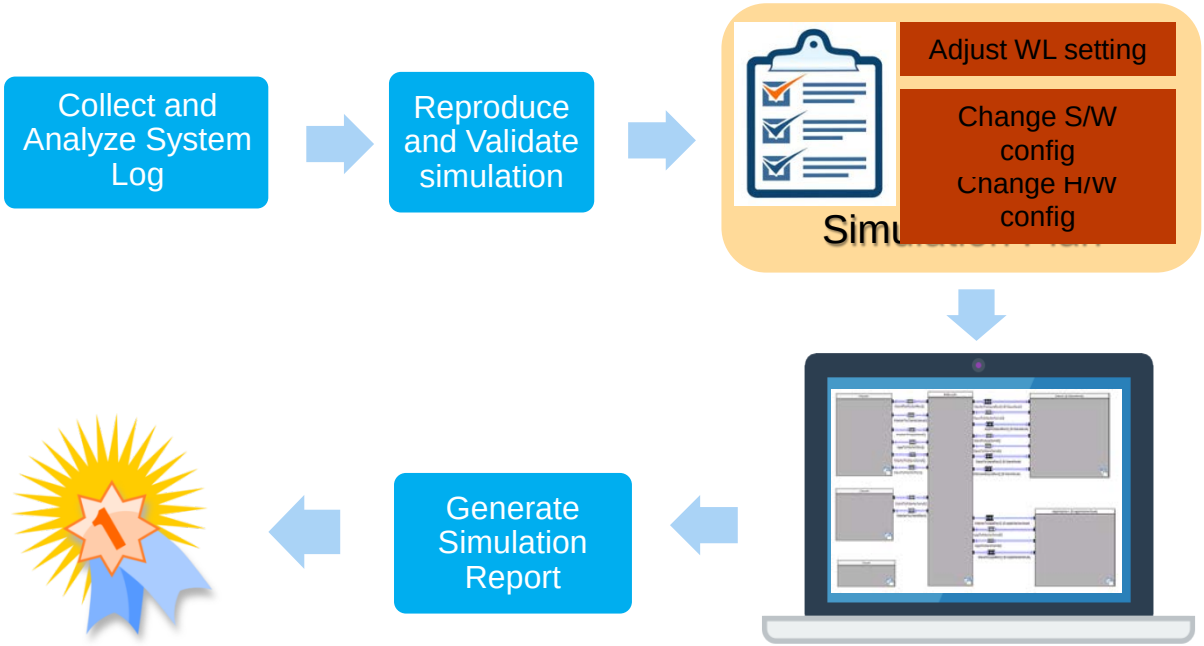



## What-If Analysis for

- 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*"**

# Simulation Approach

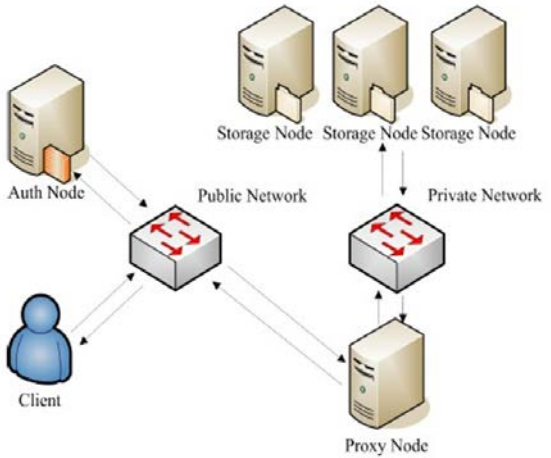


# Agenda

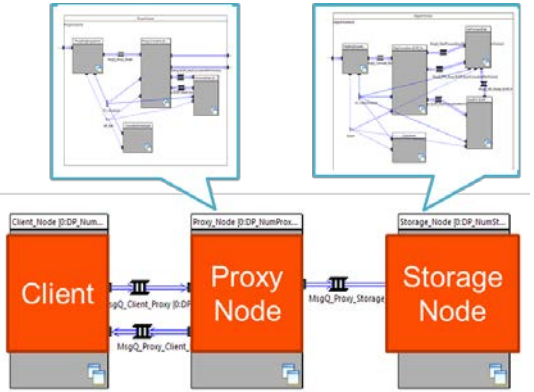
- Design challenges
- Cloud Storage system Swift modeling overview
- Use Case study

# Swift Simulator

- Logic View of a Swift Cluster



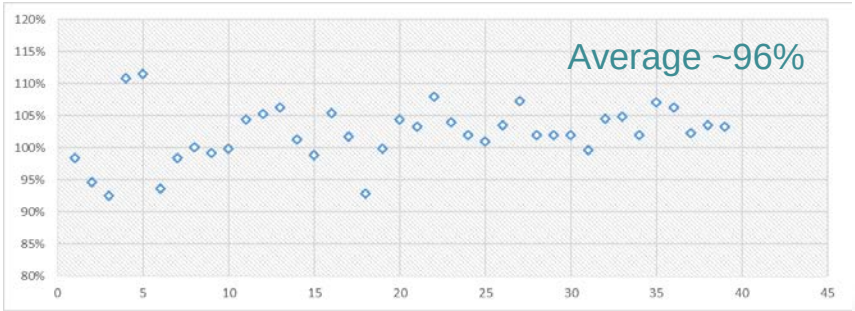

- Frontend model of Simulation

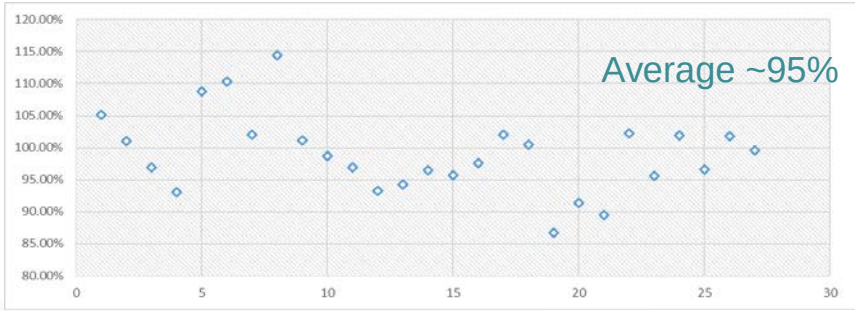



Backend model: System Topology (Star, CLOS… ), Ring parsing, Node mapping, Coroutine scheduling, Perf Lib …

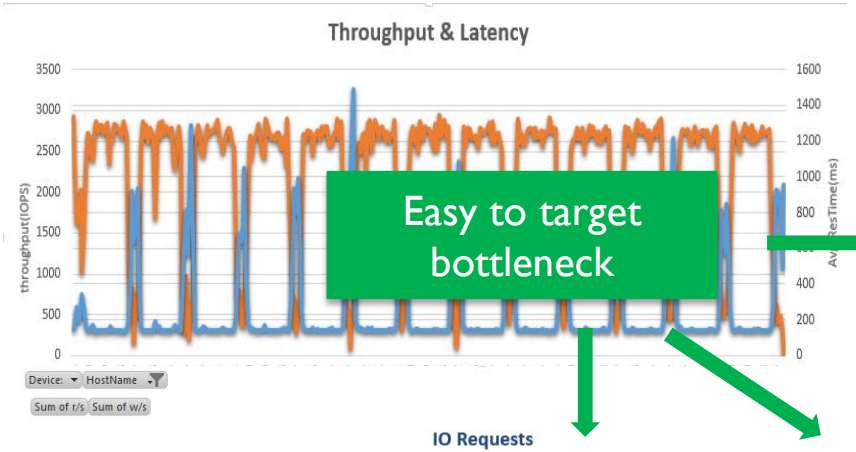
# Simulation Accuracy

- Hardware Validation
  - STAR, Fat-Tree, CLOS
  - 1GbE, 10GbE, 25GbE, 50GbE
  - HDD, SSD, NVMe
  - Node count
  - …
- Software Validation
  - Operation Type (PUT, GET, MIX)
  - Object Size (16KB ~ 1024MB)
  - Proxy worker#, Object worker#
  - Concurrency#
  - Write Barrier
  - …

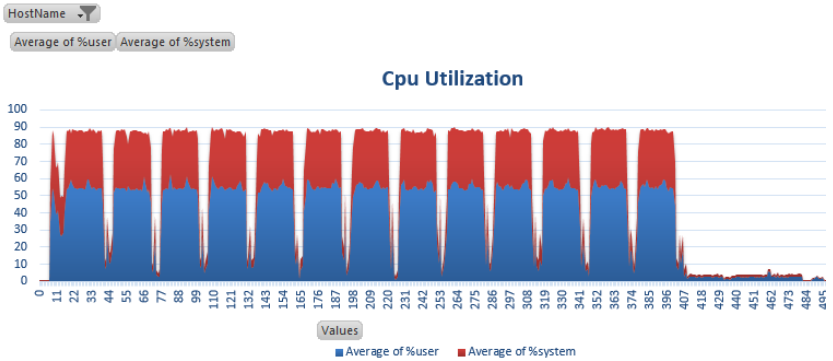


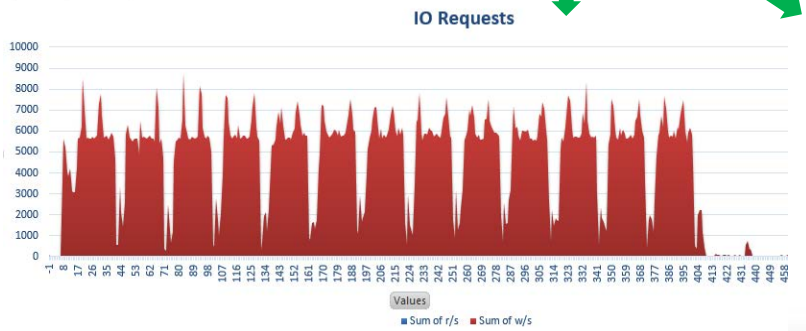



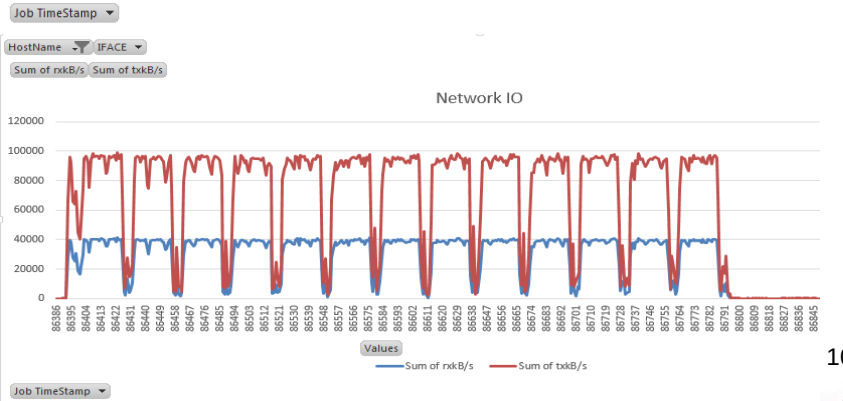
# Output: Workload and H/W metrics




Easy to target bottleneck



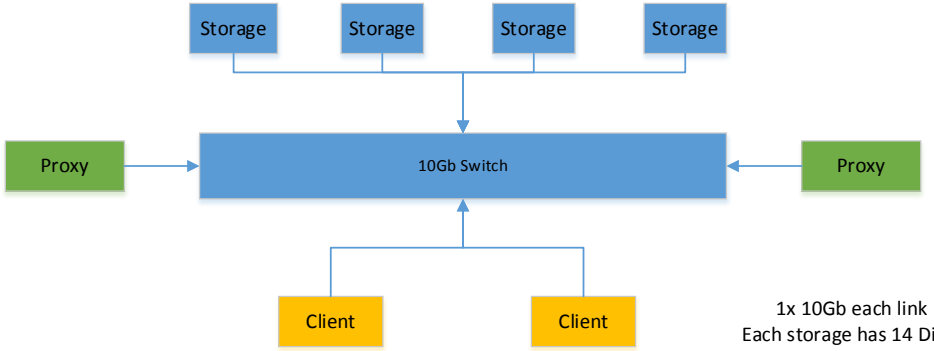
# Agenda

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

# Case Study: Optimize one cloud storage system

- Current: A media storage system for image thumbnail and audio.
- Design Goal:  Achieve to 2X IOPS and bandwidth for both small objects and large objects in read, write and mixed scenarios.

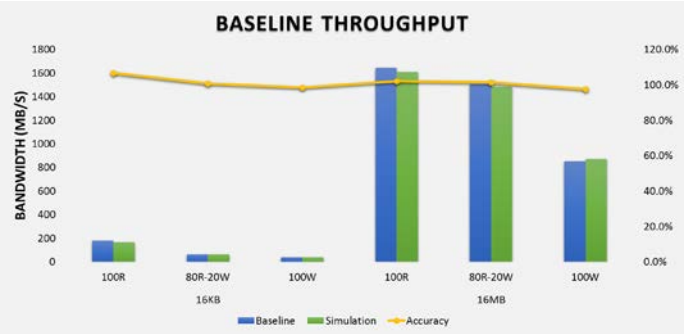
# Cluster Configuration

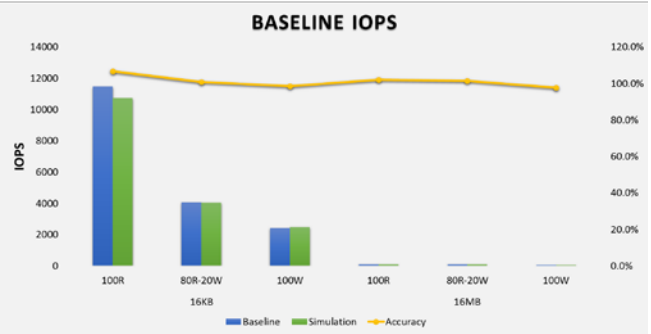



1x 10Gb each link
Each storage has 14 Disks

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

# Validation against baseline setup

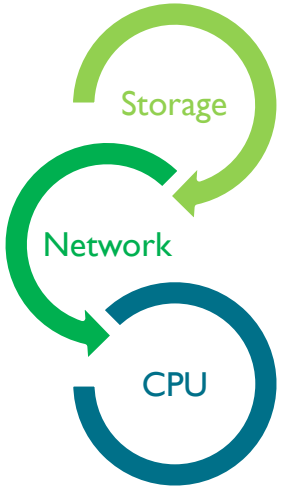


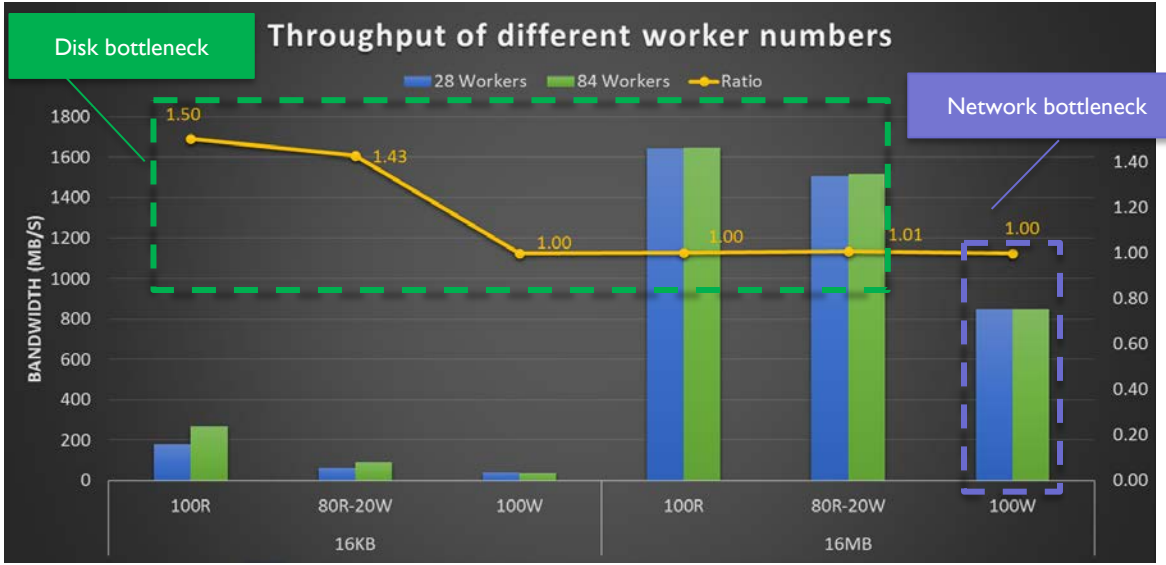

***Higher than 95% average accuracy***
***Ready to use simulator to predict performance***

# Design options

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

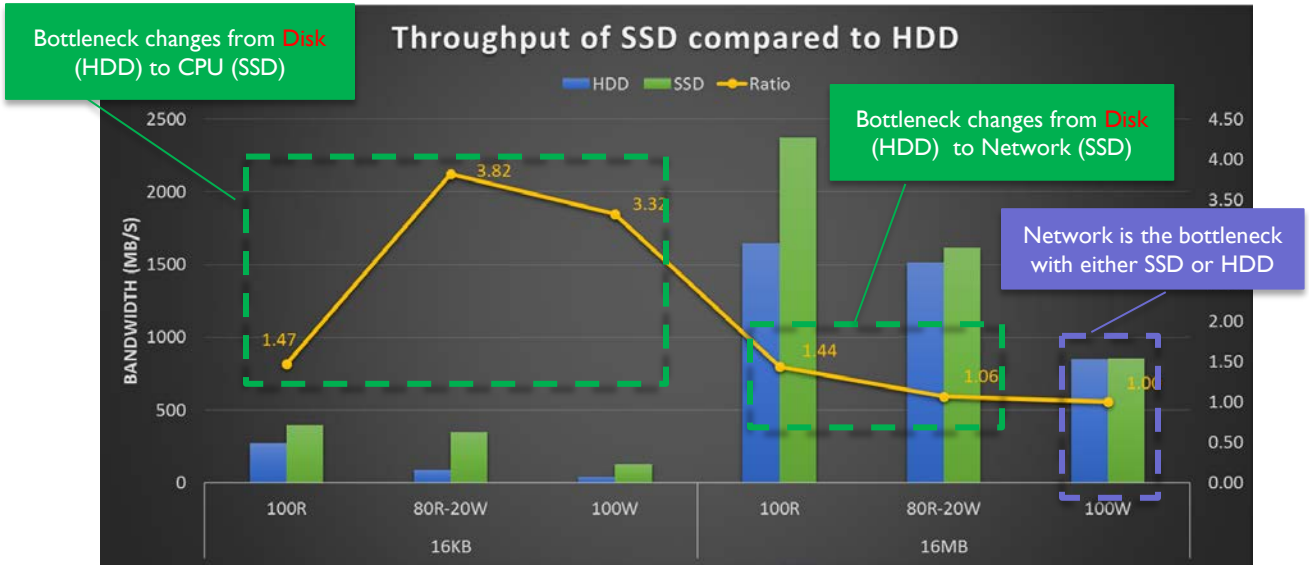


# Software optimization



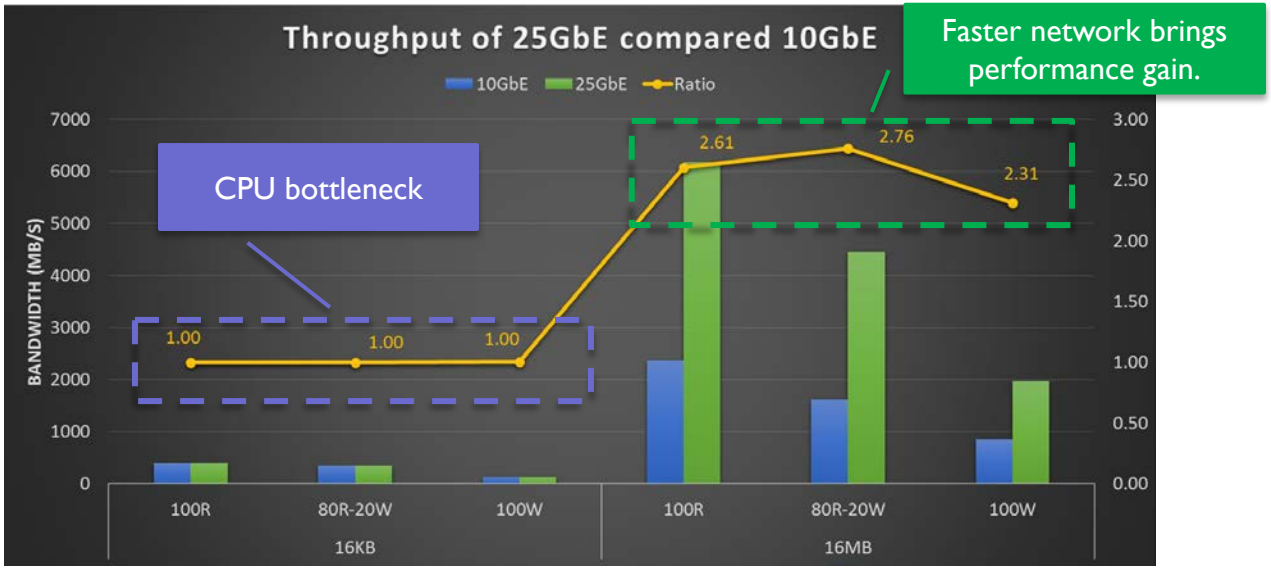

***Up to 1.5x improvement with software optimization***
***Set worker count for proxy and storage to 2 or 3 times the core count***

# Scale up: HDD -> SSD




***Expect SSDs to improve performance over HDD based storage***

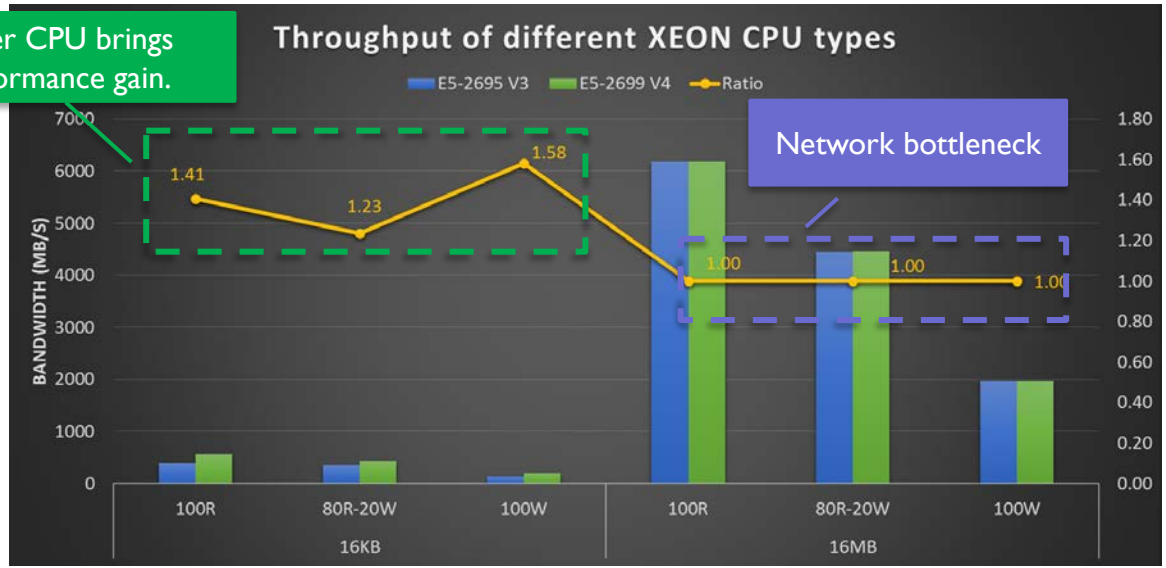
# Scale up: 10GbE -> 25GbE



# Scale up: XEON E5-2695 v3 -> E5-2699 v4

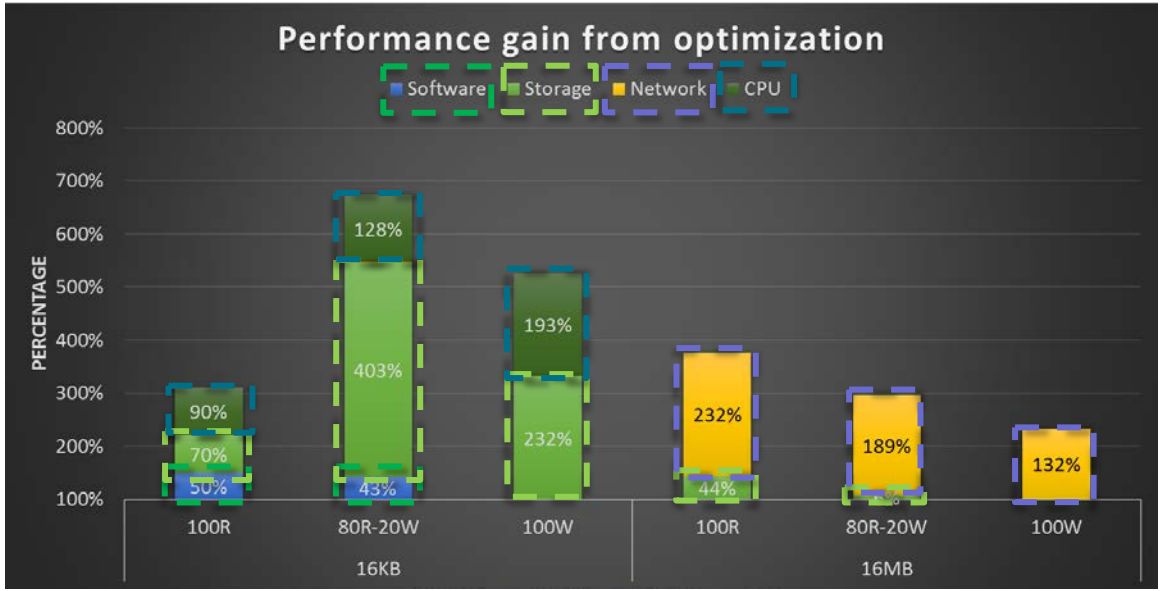
***Frequency: 2.3GHz vs. 2.2GHz Core count: 14 vs. 22***

Faster CPU brings performance gain.



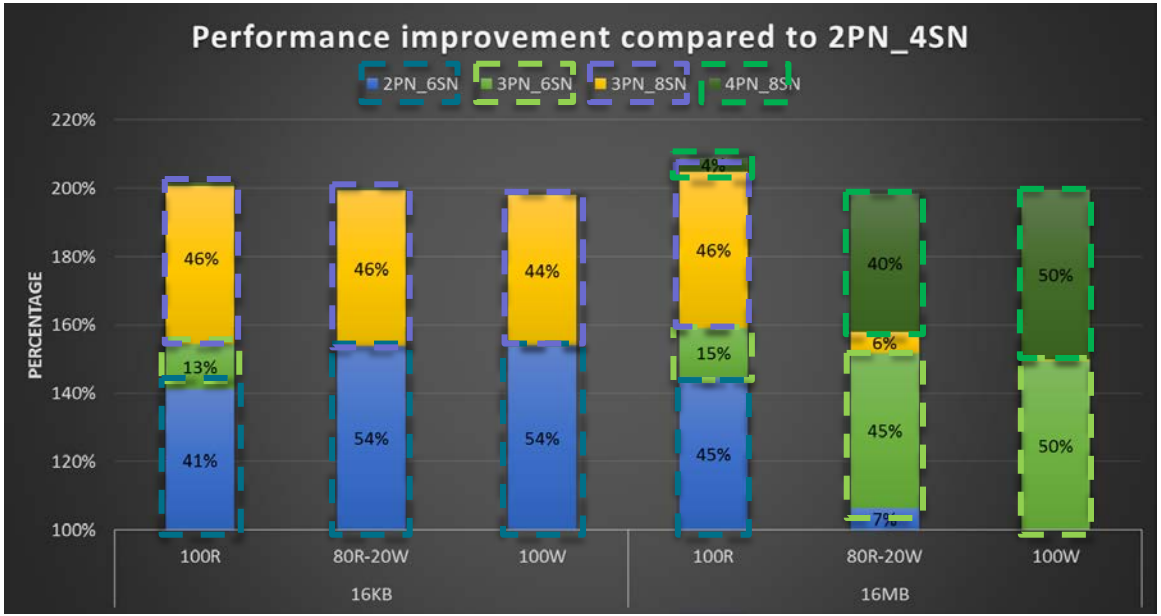

***Up to 1.58x improvement with CPU optimization***

# Scale-up optimization summary

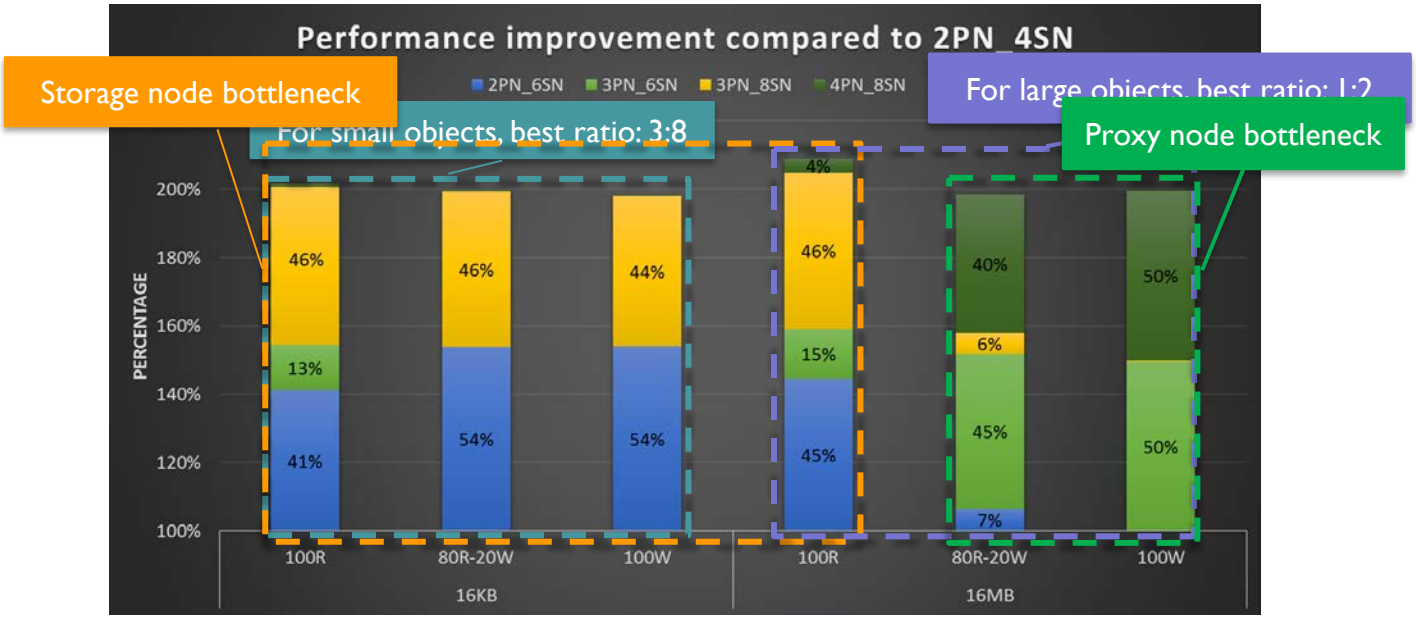



***Up to 6.73x improvement with scale-up optimization***

# Scale out optimization



# Scale out optimization

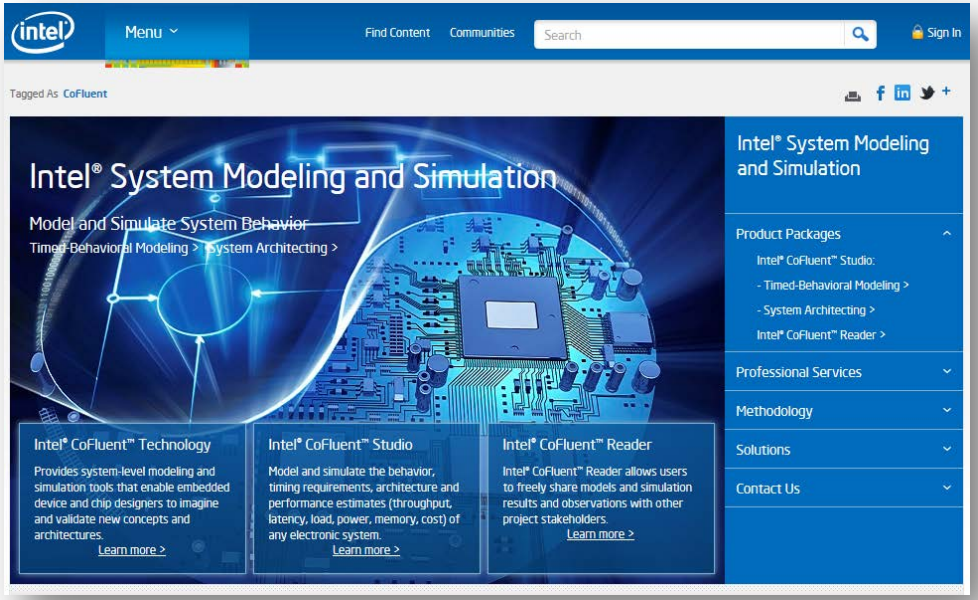



***Up to 2x improvement with scale-out optimization***
***Better scale up firstly, then scale out***

# Optimization analysis

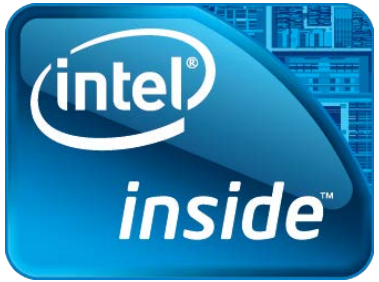
- Set worker count for proxy and storage to 2 or 3 times the core count.
- SSD improves performance over HDD.
- 10Gb is okay for small objects. Need to update network for large objects.
- CPU efficiency is high for small objects.
- Better scale up firstly, then scale out.
- Properly configure ratio of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

# cofluent.intel.com



# Backup

# What's Inside


INTEL® COFLUENT™ TECHNOLOGY FOR BIG DATA

Solutions for big data cluster Simulation, Planning and Optimization

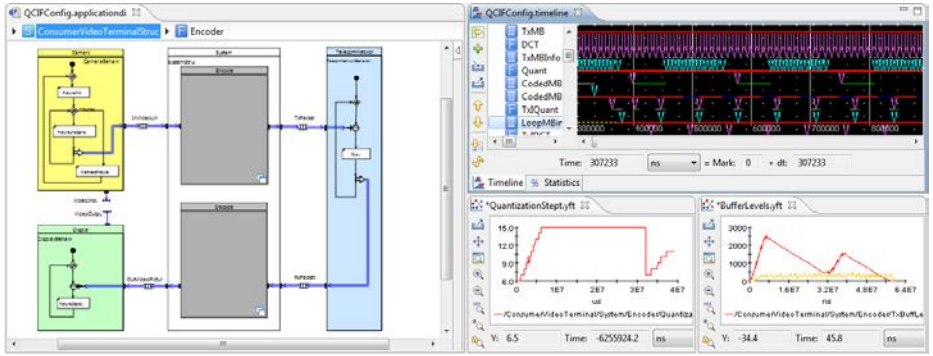
# Intel® CoFluent™ Technology for Big Data

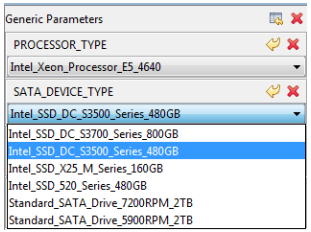
**FASTER CLUSTER DEPLOYMENT:**
Explore deployment options and meet performance goals

**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



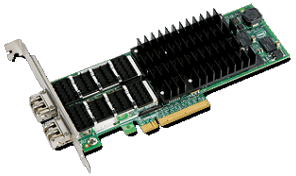

Enables fast “What if?” analysis with a virtual system

# Hardware Coverage

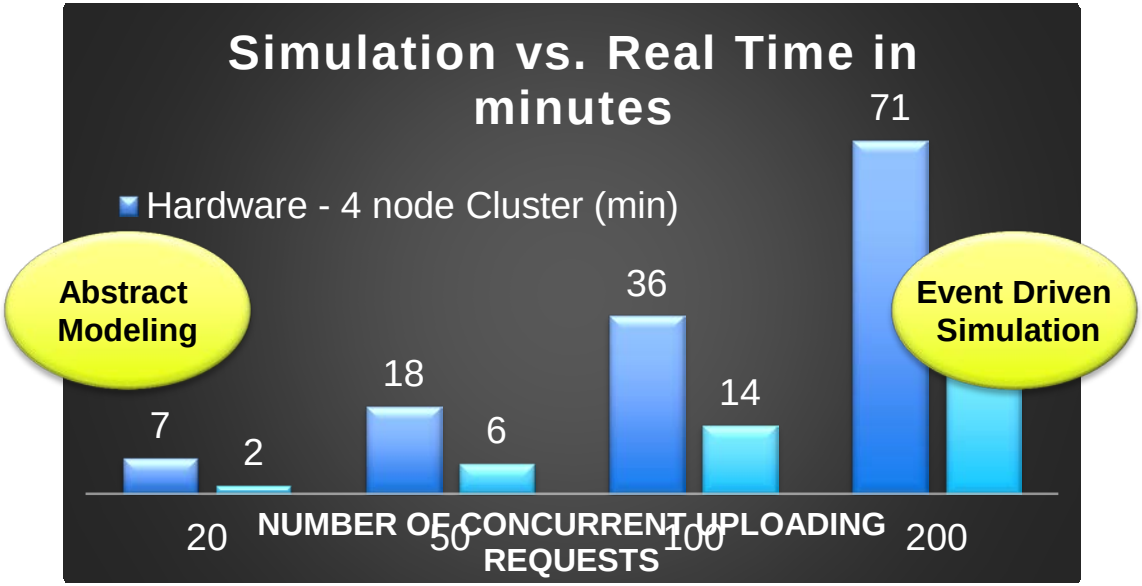
Validated: 700 Nodes


Rack Scale
Architecture





# Fast Simulation



# Host machine to run simulations

# 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 | | | | | |
|---|---|---|---|---|---|
| Class | Type | IOPS | | Throughput | |
| Storage Device | capacity | RND 4KB 100%R | RND 4KB 100%W | SEQ 64KB 100%W | SEQ 64KB 100%R |
| SSD S3700 | 1TBGB | 47,806 | 59,415 | 403MB/s | 492MB/s |
| 7200 RPM HDD | 1TB | 390 | 410 | 180MB/s | 180MB/s |