



AWS Summits 2014

HPC on AWS for HDD Development

Hiroshi Kobayashi, Dev./Lab. IT Systems
HGST Japan, Ltd.
Jul 17, 2014



Index

- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- Data Collaboration
- Flexibility
- What's Next...
- Summary

Index

- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- Data Collaboration
- Flexibility
- What's Next...
- Summary

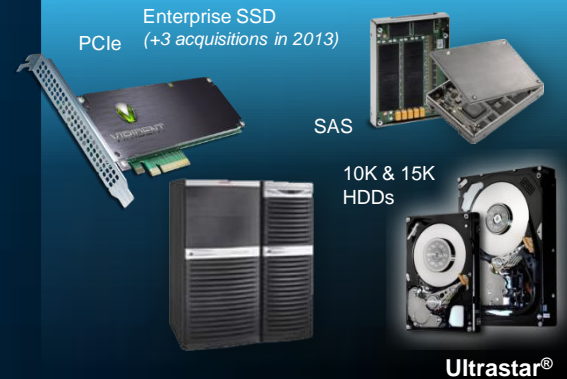
HGST, a Western Digital company



- Founded in 2003 through the combination of the hard drive businesses of IBM, the inventor of the hard drive, and Hitachi, Ltd (“Hitachi”)
- Acquired by Western Digital in 2012
- More than 4,200 active worldwide patents
- Headquartered in San Jose, California
- Approximately 41,000 employees worldwide
- Develops innovative, advanced hard disk drives, enterprise-class solid state drives, external storage solutions and services
- Delivers intelligent storage devices that tightly integrate hardware and software to maximize solution performance

Cloud & Datacenter

Performance Enterprise



Capacity Enterprise



Zero to Cloud << 12 Months

April 2013



By Mid 2014:

- ✓ Cloud eMail – Microsoft Office365
- ✓ Cloud eMail archiving/eDiscovery
- ✓ External Single SignOn (off VPN)
- ✓ Cloud File/Collaboration – BOX
- ✓ Salesforce.com
 - ✓ Integrated to save files in BOX
- ✓ Cloud–High Performance Computing (HPC) on Amazon's AWS
- ✓ Cloud – Big Data Platform on AWS
 - ✓ Extract insights from manufacturing data
- ✓ Cloud - Data mart and provisioning service using AWS Red Shift
- ✓ Cloud – HR Transformation via Workday Deployment



Index

- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- Data Collaboration
- Flexibility
- What's Next...
- Summary

Why choose AWS for HPC?

- Background

- 18 months ago, HPC implementation project was started. Project team investigated several cloud HPC services except for AWS. But did not match HGST's requirement.
- CIO Steve Phillpott recommended AWS for HPC. He had much experience of HPC on AWS at life-science industry.
- Through several Proof of Concept projects, began to understand Pros/Cons of On-premise and Cloud HPC.

- Key factors are...

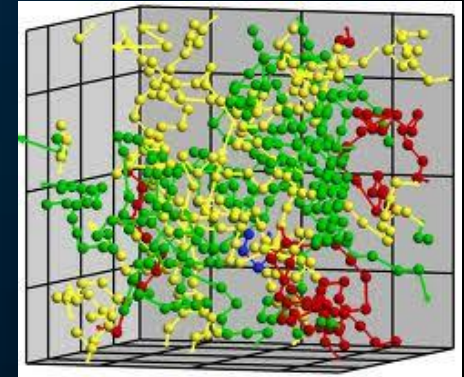
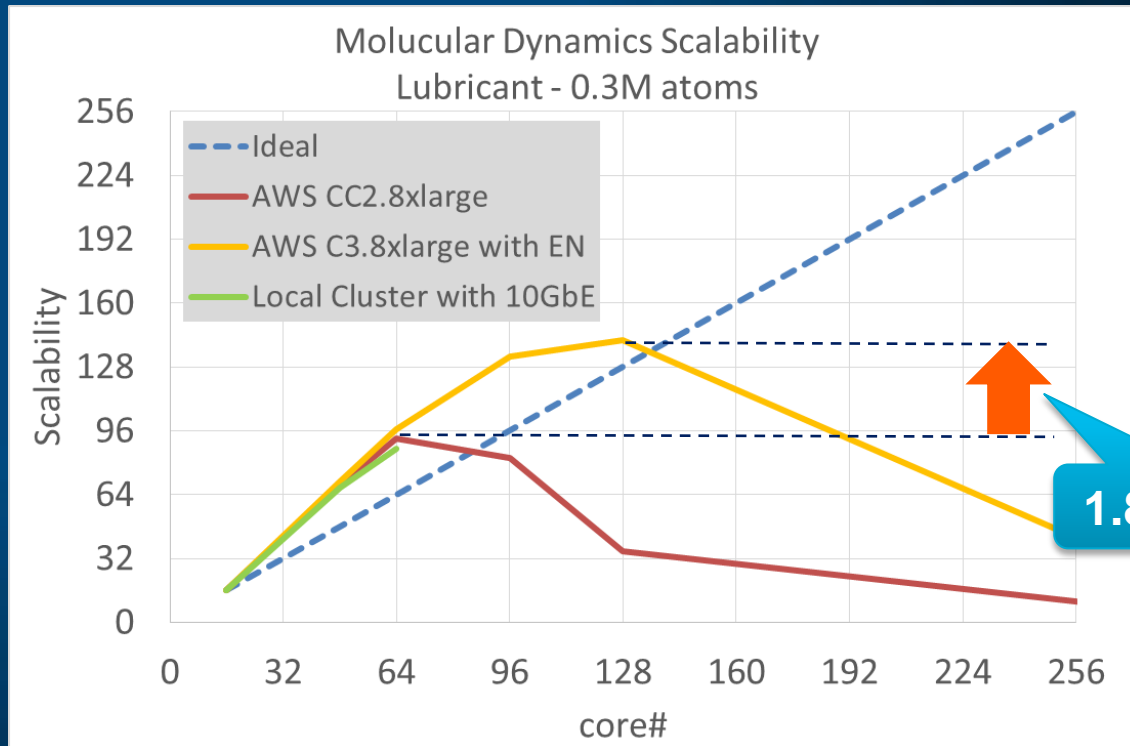
- Scalability, Data transfer, Remote Visualization
- Security, Commercial Application, Cost
- and more....

- ✓ What kind of benefit can HPC on AWS provide?

Index

- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- Data Collaboration
- Flexibility
- What's Next...
- Summary

Molecular Dynamics

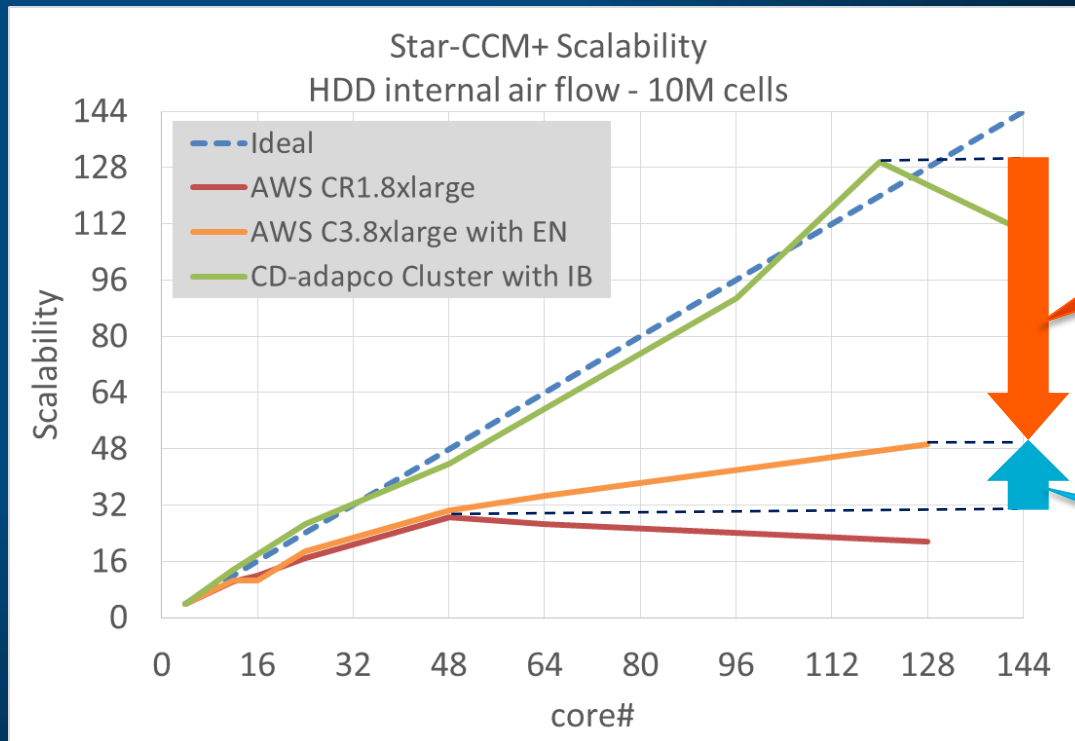


1.88x faster

- CC2 & C3 cluster have equivalent scalability of HGST local cluster with 10GbE around 64cores
- C3 provide significant improvement to the scalability
- C3 is **1.88x faster** than CC2

※1 EN = Enhanced Networking
 ※2 placement group enable
 ※3 evaluated by elapse time

CFD “Star-CCM+”



1.70x slower

1.81x faster

- CD-adapco Japan provided the scalability data on their cluster.
- C3 provide significant improvement to the scalability
- C3 is **1.81x faster** than CR1
- Still behind to physical cluster with InfiniBand

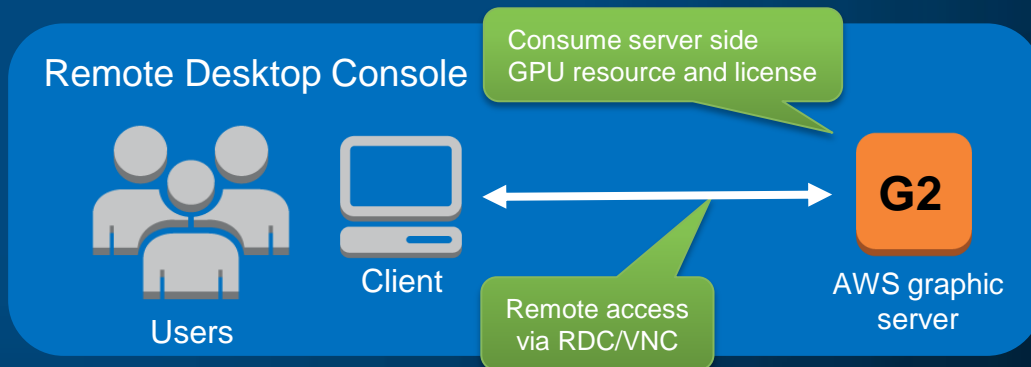
※1 EN = Enhanced Networking
※2 placement group enable
※3 evaluated by elapse time
※4 only 200steps

Index

- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- Data Collaboration
- Flexibility
- What's Next...
- Summary

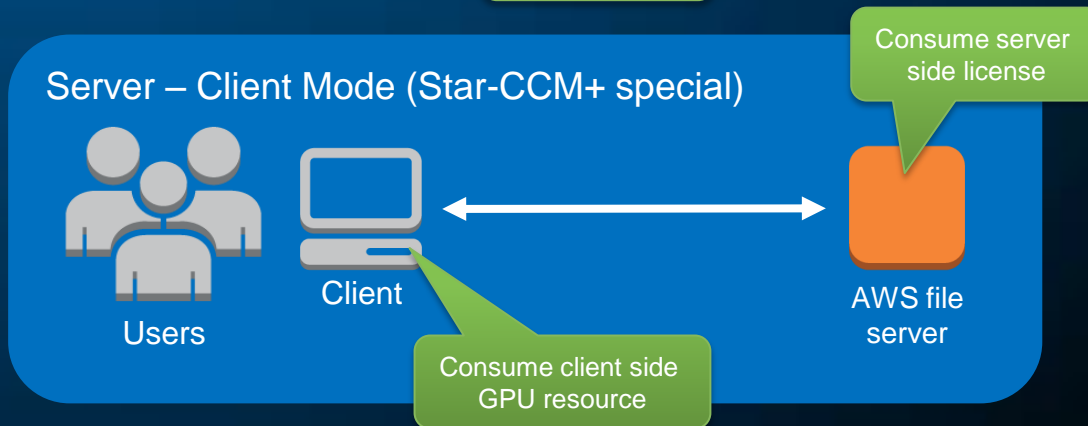
Remote Visualization

- Result data is too huge to download
- Transferring huge data is NOT a option
- Require Remote Visualization for huge result data



Not good performance...

- ✓ Slower response
- ✓ Slower rendering



Great performance!!!

- ✓ Almost same performance as local workstation with high-end graphic card

Index

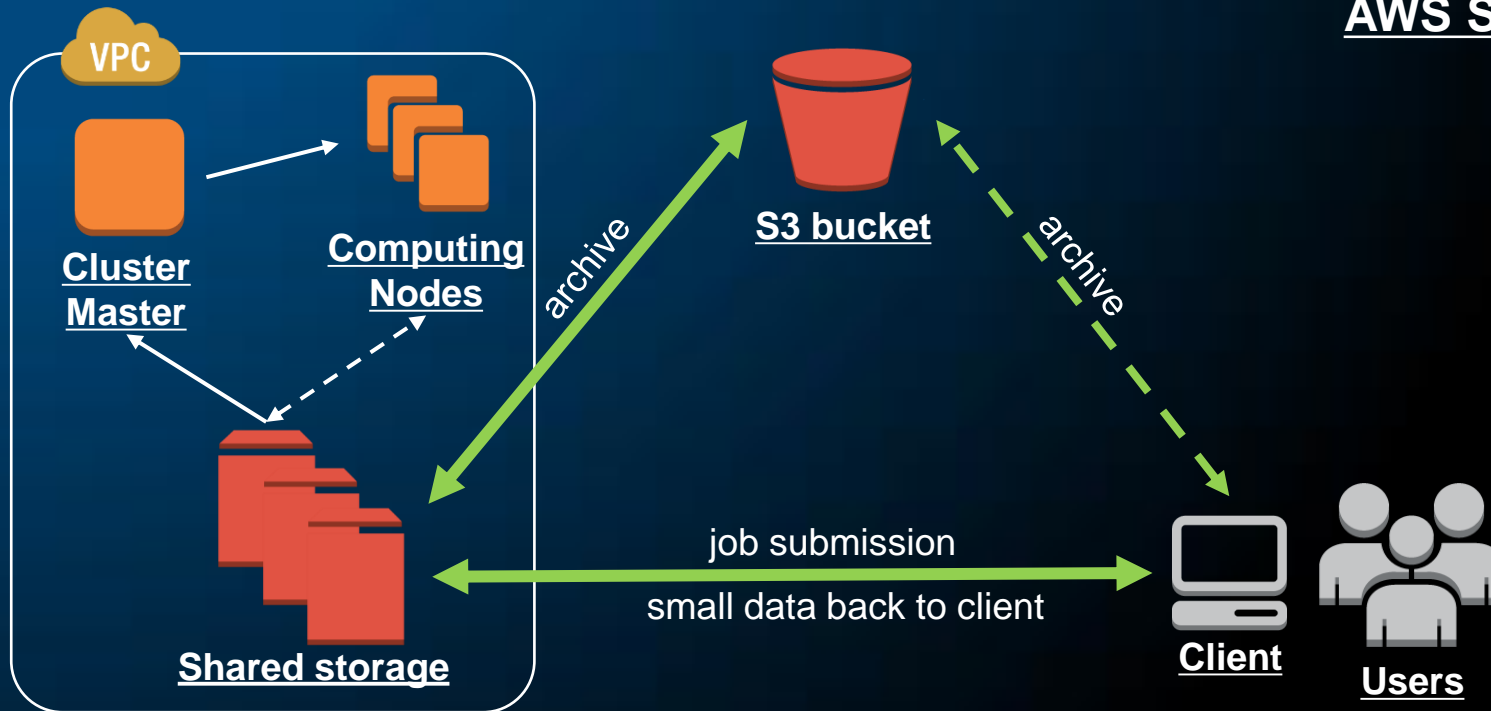
- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- **Data Collaboration**
- Flexibility
- What's Next...
- Summary

Data collaboration

- Transferring huge data is NOT a option
- Even 48TB of hs1.8xlarge may not be sufficient for long term / huge data repository
- High cost for re-computing of large scale model



AWS S3



Performance Summary

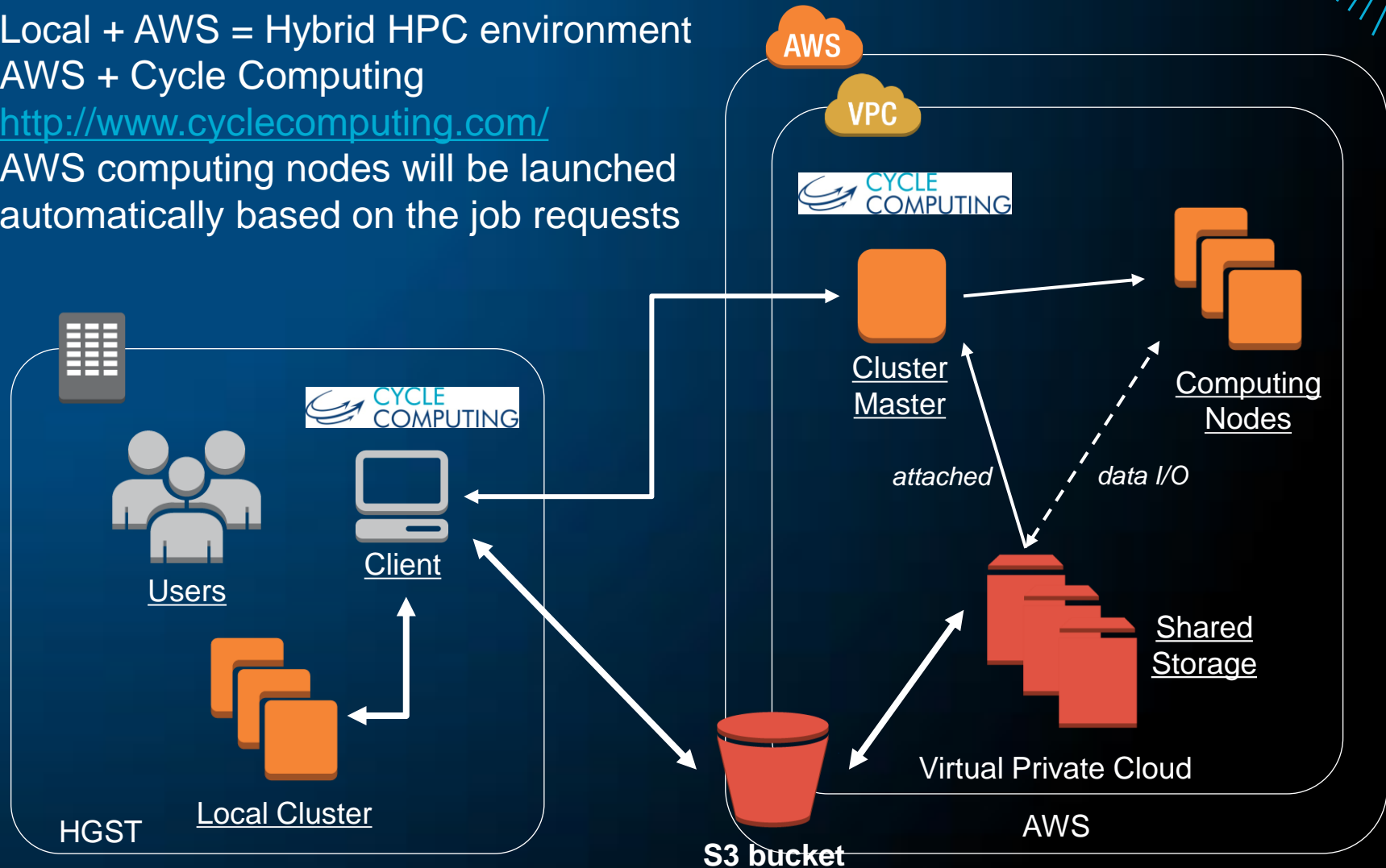
- Computing Performance
 - C3.8xlarge improved scalability dramatically
 - Higher scalability is better
- Remote Visualization
 - Star-CCM+ is ready
 - Other application are NOT ready
- Data Collaboration
 - No need to struggle with the storage capacity and durability
- AWS can support whole process of simulation works!!!

Index

- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- Data Collaboration
- **Flexibility**
- What's Next...
- Summary

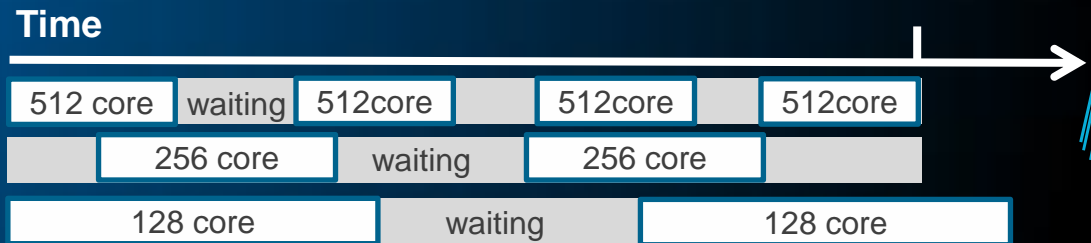
Hybrid HPC Architecture

- Local + AWS = Hybrid HPC environment
- AWS + Cycle Computing
<http://www.cyclecomputing.com/>
- AWS computing nodes will be launched automatically based on the job requests

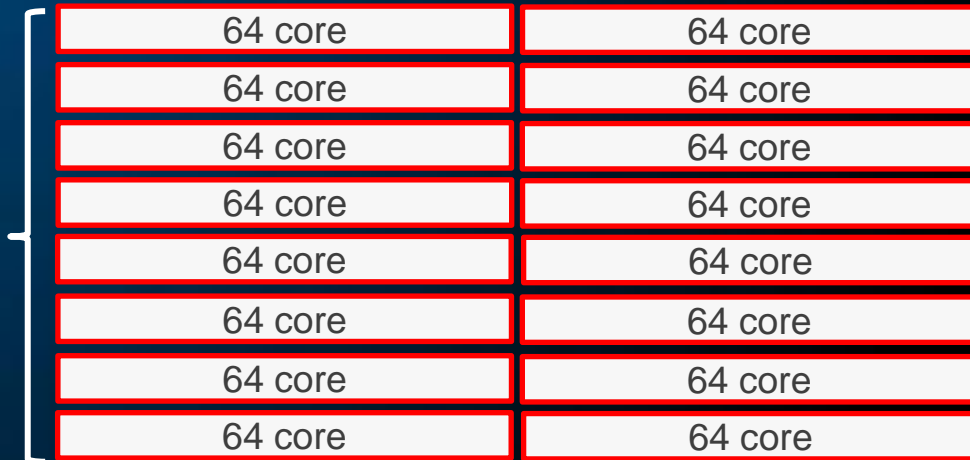


Shape Compute To Match Work To Be Done

Before:
Shared Cluster Computer

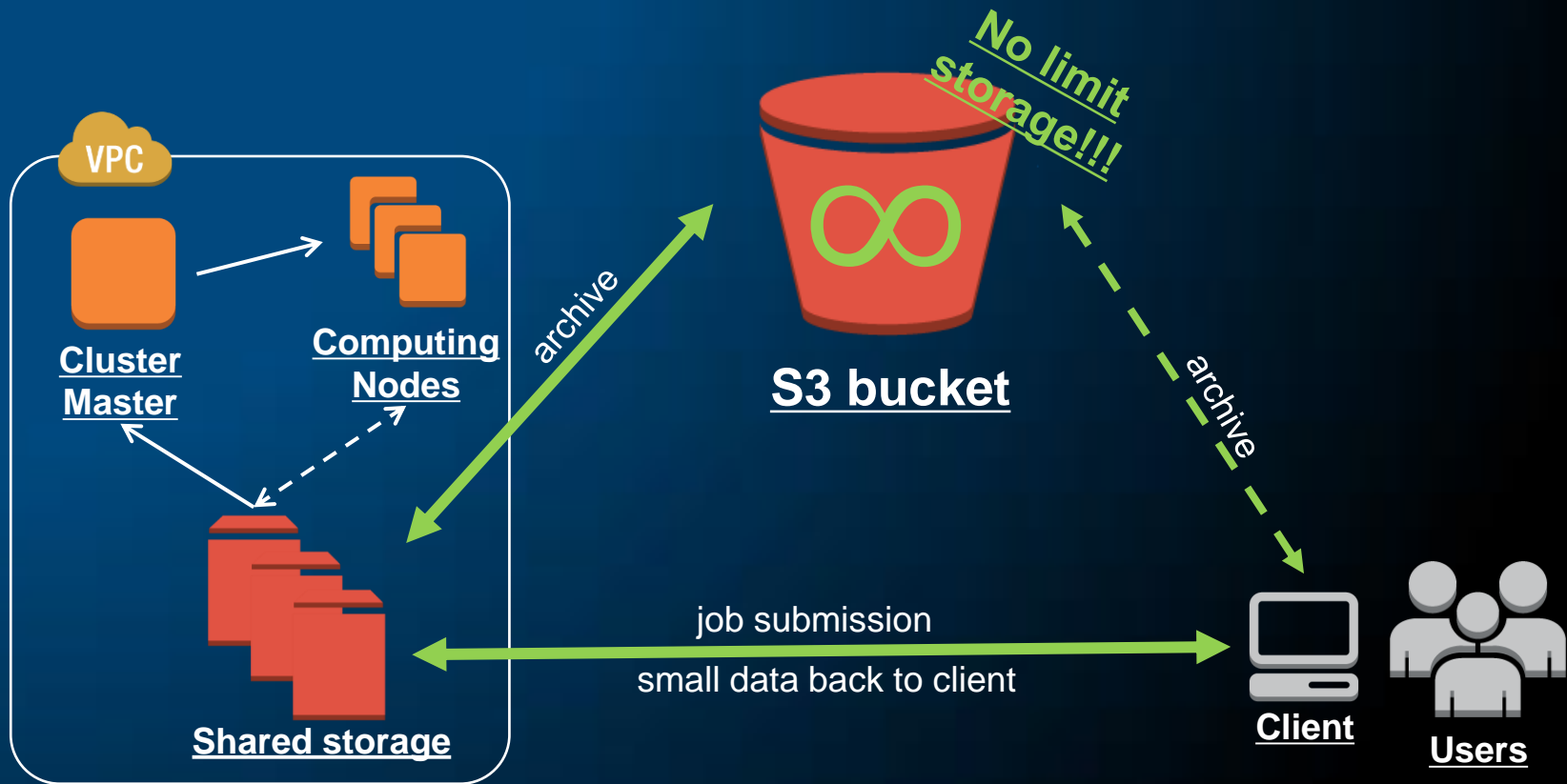


Today:
AWS EC2 CC2 Cluster
(Max Total 512 core)



All Jobs Run In Parallel on AWS → **1.67x** Throughput Improvement

Shape Storage To Match Work To Be Done



No need to struggle with the storage capacity and durability!!!

Index

- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- Data Collaboration
- Flexibility
- What's Next...
- Summary

What's next for AWS HPC...

- Computing Performance

- More scalability, like InfiniBand
- More resource capacity, especially c3.8xlarge



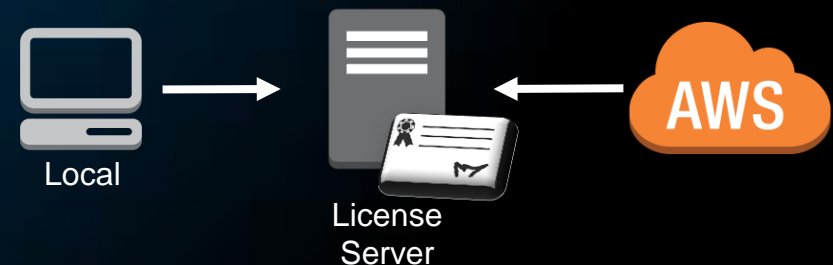
- Remote Visualization

- Higher performance than RDC-TCP/IP
- AppStream? PC over IP®?



- Commercial Application License

- End User License Agreement (EULA)
- Hybrid License Server
- Consumption Based License?



Index

- HGST
- Why choose AWS for HPC?
- Computing Performance
- Remote Visualization
- Data Collaboration
- Flexibility
- What's Next...
- Summary

Summary

- **At this moment, HPC on AWS is NOT perfect**
 - Scalability, Resource Capacity, Remote Visualization
- **HPC on AWS has extremely high flexibility**
 - Hybrid HPC, Shape Compute/Storage To Match Work To Be Done
- **Flexibility will drive to responding to the changing business model**
- Benefit of HPC on AWS should be verified with each applications based on its characteristic
- Required collaboration with application vendors



AWS Summits 2014

Tech Deep Dive TC-04

Thank You

