Storage Grid using iSCSI

Felix Xavier
CloudByte Inc.

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

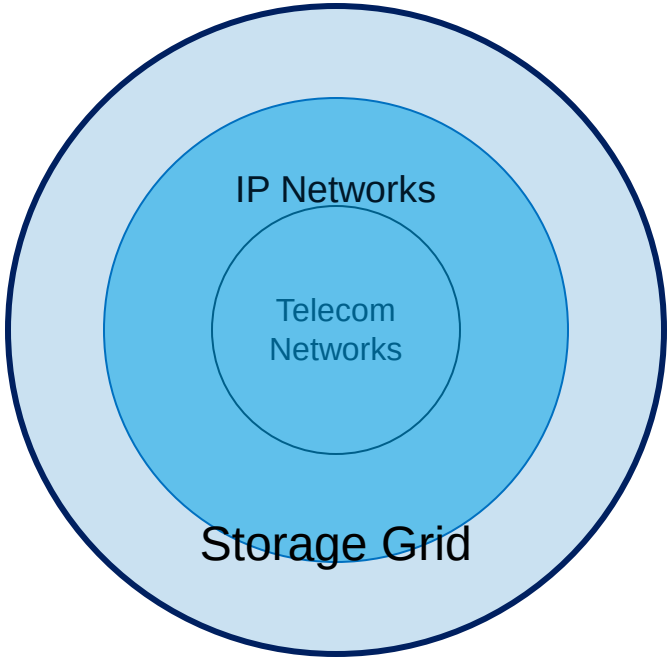
Abstract

- Advent of cloud brought a new requirement on storage: the storage nodes in the cloud have to communicate with each other and bring the hot data near the application across data centres
- The communication must be standard-based
- This session describes iSCSI protocol to achieve the inter-storage node communication for enterprise-grade cloud; this may not be applicable to object storage

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

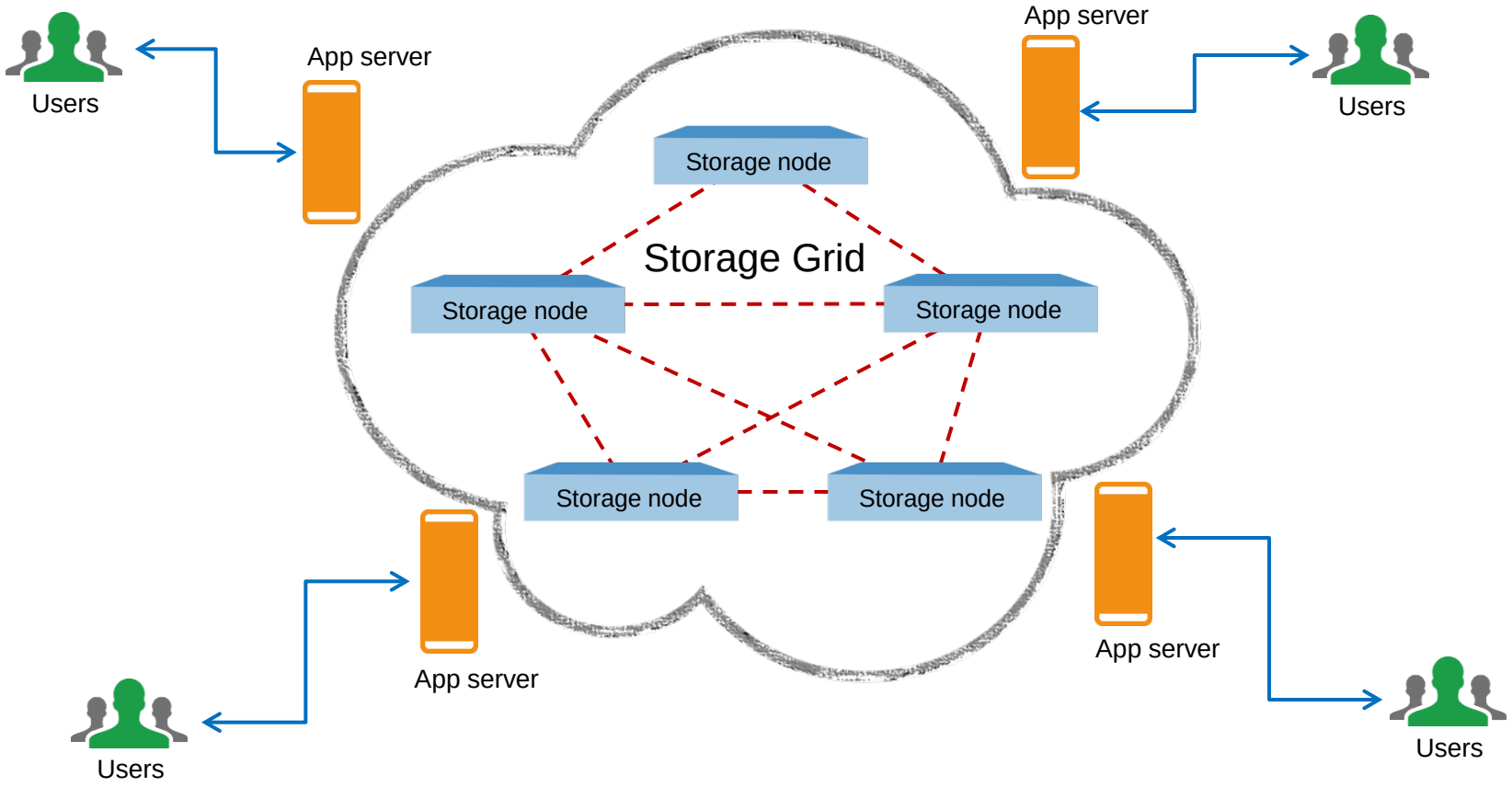
IT Evolution: The New Storage Core

- Telecom Networks
  - The first phase of electronic communication
  - Telex and fax operated on this network
- IP Networks
  - When the internet evolved, the IP networks became the core
  - www, FTP, and email ran on this network
  - Fusion between telecom and IP networks
- Storage Grid
  - Now Cloud is emerging where storage forms the new core
  - Web apps already use this grid while enterprise apps are moving towards it
  - Fusion between IP networks and storage



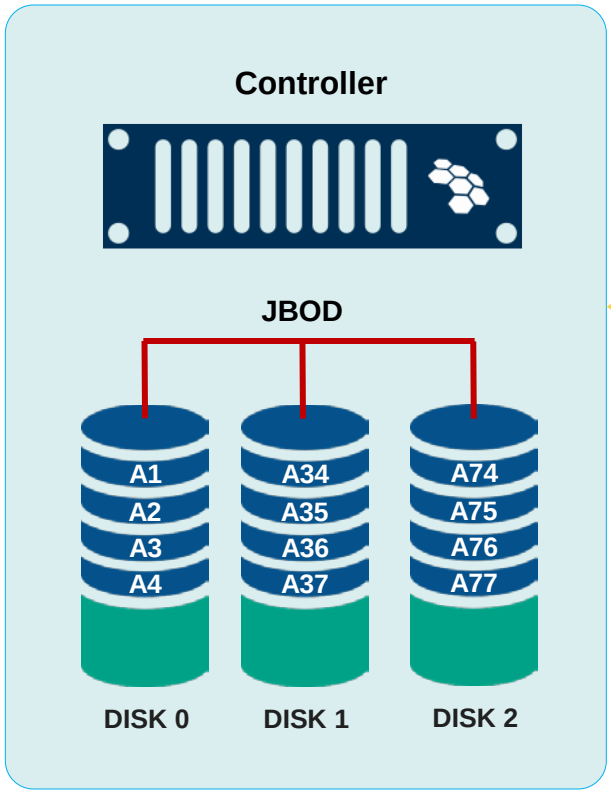
Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

The New Core: A Broader View



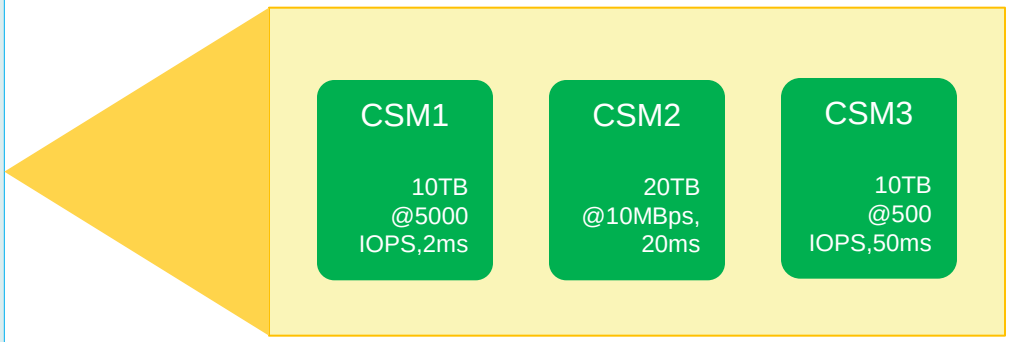
Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

The Intelligent Storage Node



Storage Node

Controller Software Architecture

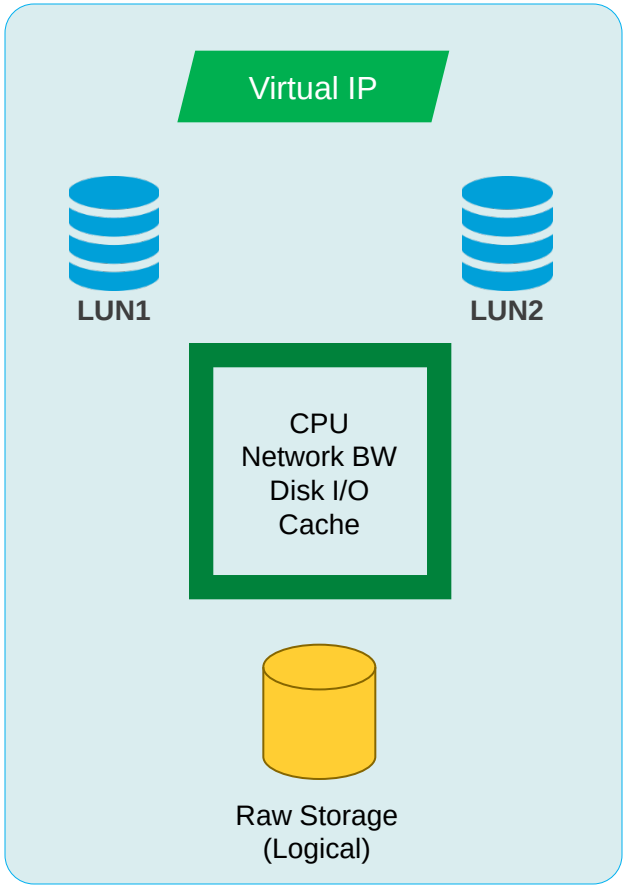


CSM – Cloud Storage Machine

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

Cloud Storage Machine (CSM) Architecture

- CSM abstracts the hardware characteristics into the software
- Each CSM has dynamically allocated hardware resources in terms of
  - CPU
  - Network bandwidth
  - Disk I/O
  - Cache
- Each CSM can host one or more storage volumes
- CSMs freely move across storage nodes

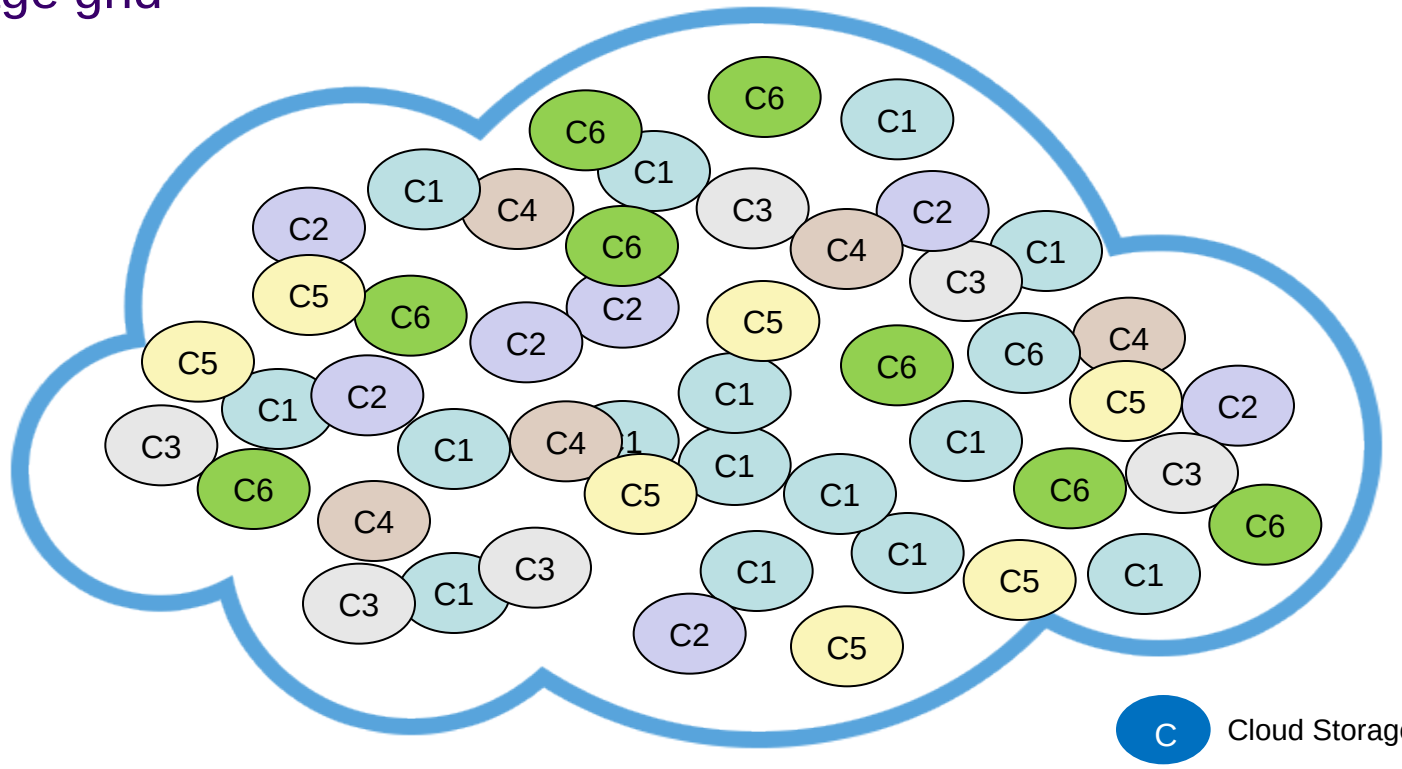


Cloud Storage Machine

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

Global Namespace for CSMs

- CSM is the fundamental block in the storage grid
- One CSM can have multiple instances in the grid, depending on the access patterns from the app servers
- CSM name can resolve to the closest instance of the CSM in the storage grid



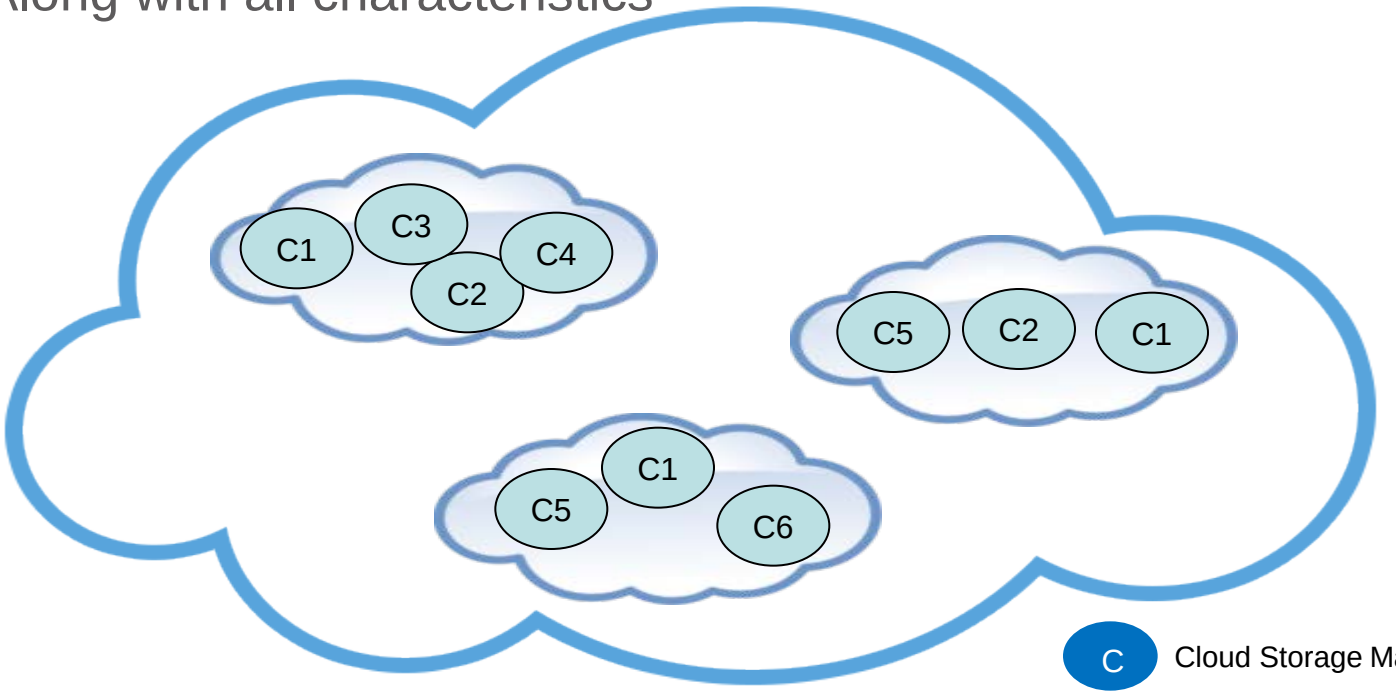
C Cloud Storage Machine

Storage Grid Using iSCSI

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

CSM – Migration

- CSM can completely migrate from one storage node to another without app disruption
  - Within the datacenter
  - Across datacenters
  - Along with all characteristics



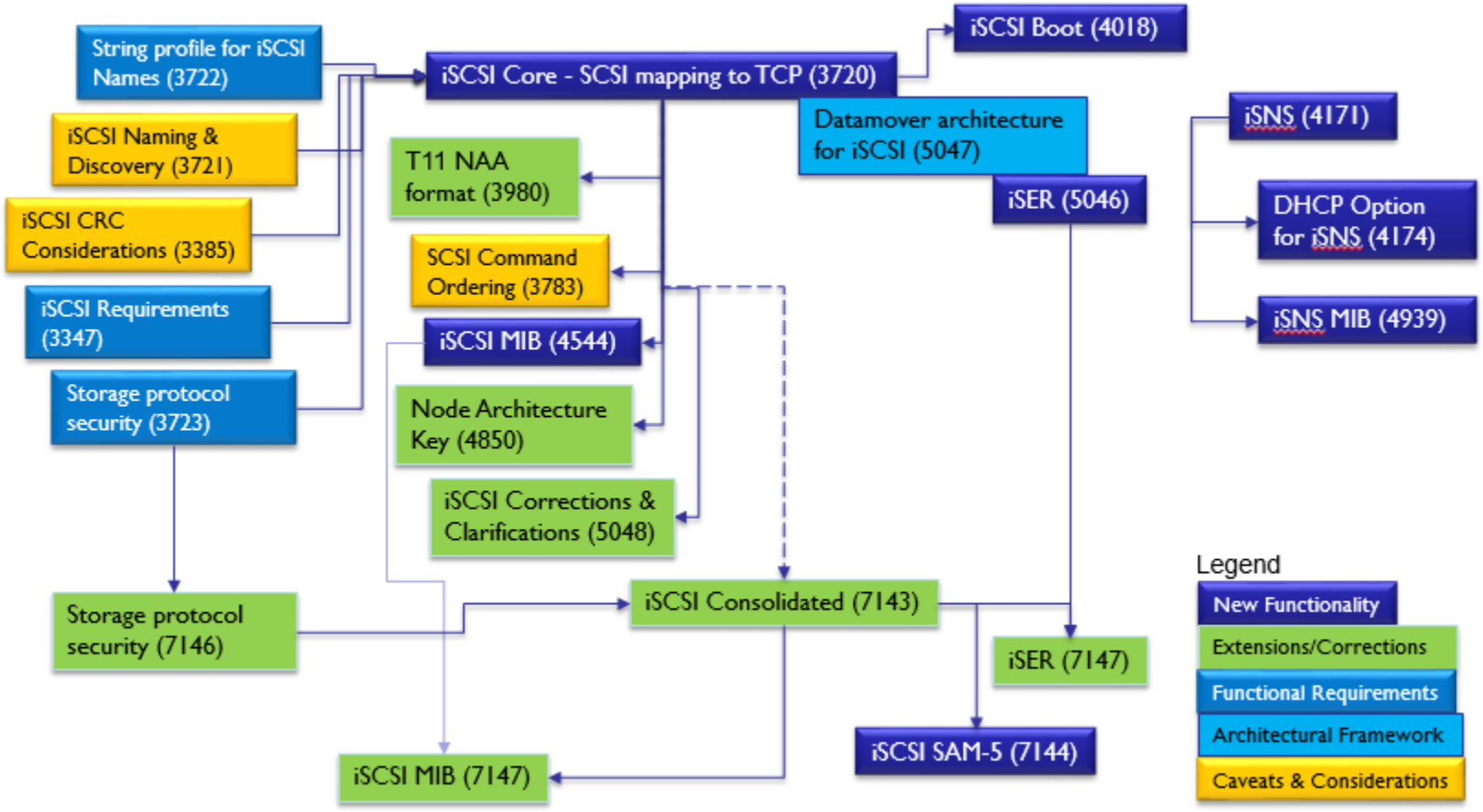
Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

iSCSI Overview

- A SCSI transport protocol that operates over TCP/IP
  - Encapsulates SCSI CDBs (operational commands, for example READ or WRITE) and data into TCP/IP byte streams
  - Allows IP hosts to access IP-based SCSI targets
- Broad industry support
  - Initiator support from server vendors
  - Native iSCSI storage arrays
  - Moving away from Fiber Channel
- Standards status
  - RFC 3720 - iSCSI Core, SCSI mapping to TCP
  - Collection of RFCs describing iSCSI

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

ISCSI Spec Landscape



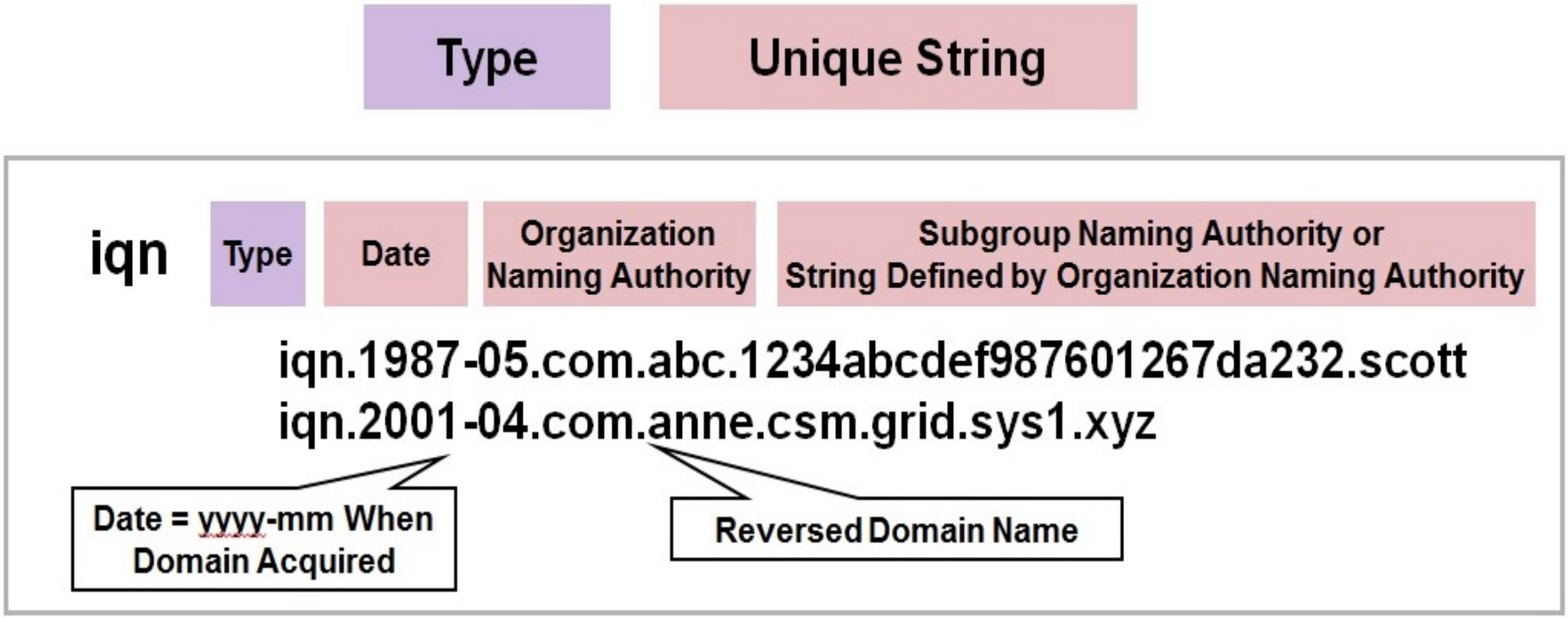
Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

Storage Networking over iSCSI

- iSCSI provides solution to carry storage traffic within IP
- Uses TCP, a reliable transport for delivery
- Applicable to local data center and long-haul applications
- Typical format

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

iSCSI Name Structure



Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

iSCSI Message Types

- Initiator to Target
  - NOP-out
  - SCSI Command
  - SCSI Task Management Command
  - Login Command
  - Text Command
  - SCSI Data-Out
  - Logout Command
- Target to Initiator
  - NOP-IN
  - SCSI Response
  - SCSI Task Management Response
  - Login Response
  - Text Response
  - SCSI Data-In
  - Logout Response
  - Ready to Transfer
  - Async Event

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

iSCSI Message Types – Extensions to support storage grid

- Read lock - sent to all other CSM instances on read request to lock start block to end block
- Write lock - sent to all other CSM instances on write request to lock start block to end block
- Advertise new instance – broadcast to other instances
- Elect master – sent with load index when the master is down, instance with the lowest index is chosen as master.
- In-sync – declare to initiator that new instances is ready
- Create instance/response – request and response to create new CSM instances
- Init transfer – initiate the first time data transfer to create new instance
- Update transfer – initiate subsequent data transfers

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

Management Workflow

Define CSM

- iqn name – iqn.2104.04.com.abc.2324809jhsdafs.xfx
- DNS name – csm1.xyz.com
- Capacity - 500GB
- Performance - 1200 IOPS , 5ms

Define number of CSM instance and location

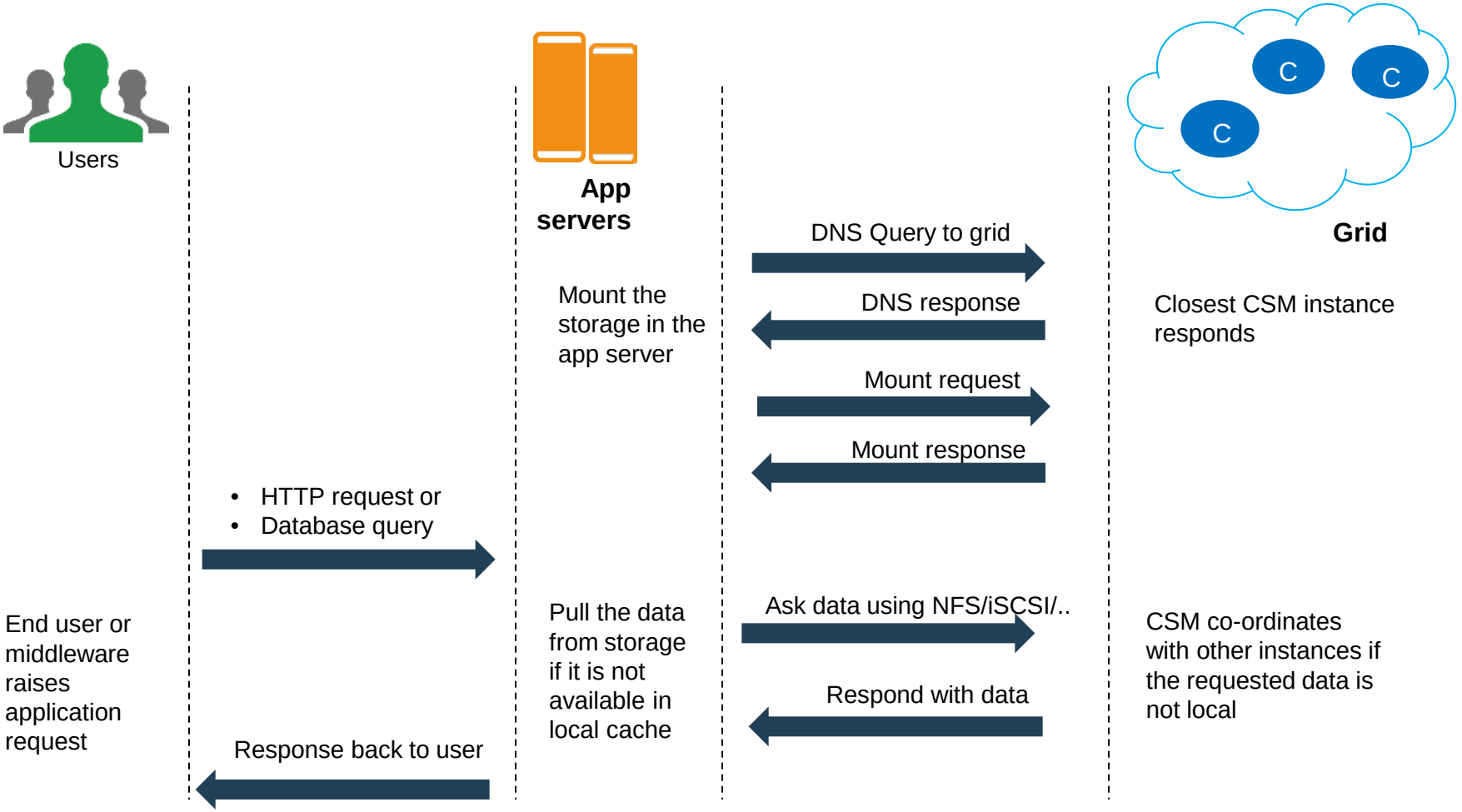
- Physical storage nodes hosting these instance for now
- These instances can move around
- Data center location

Define Master CSM instance

- Maintains house keeping information about the other nodes
- Constantly maintains keep alive to track the instances entering and exiting

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

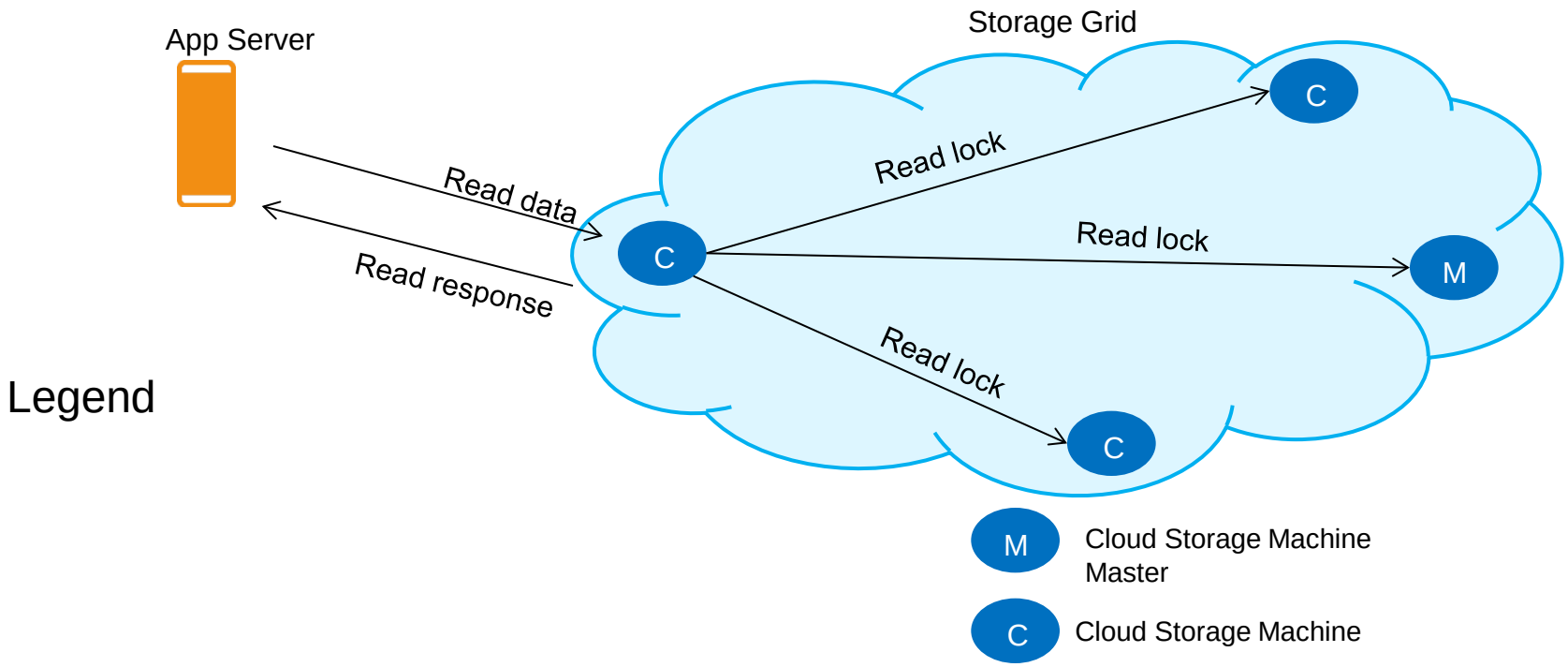
Data Flow



C Cloud Storage Machine

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

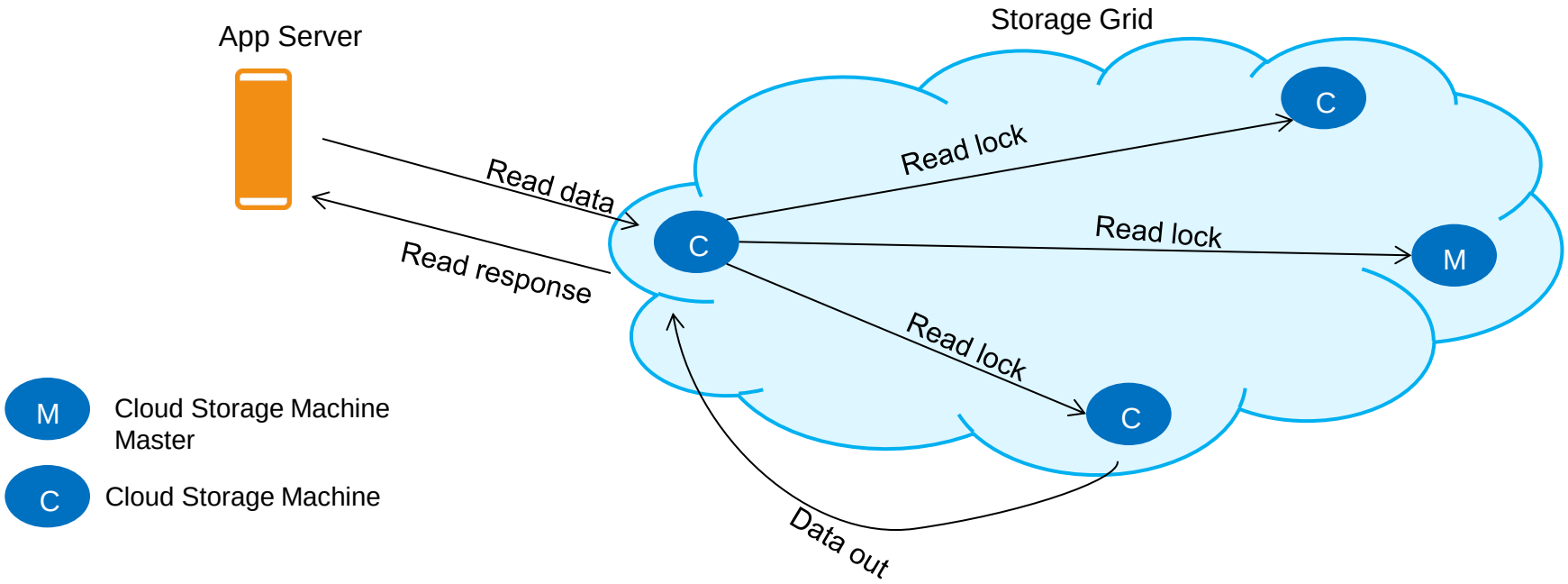
Read Data at Same Instances



Step 1. App server issues the data request to the connected CSM instance.
Step 2. CSM issues read lock to its peer instance for the group of blocks.
Step 3. On success of read lock, returns data to the application.

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

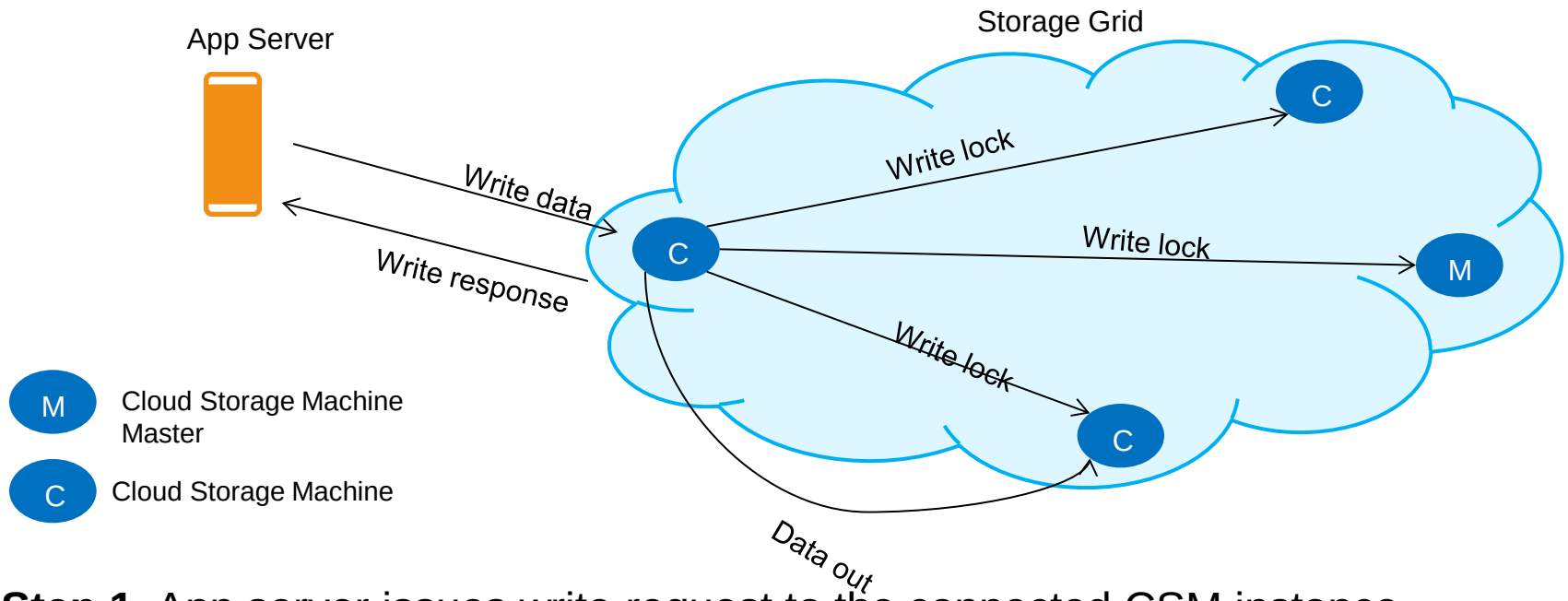
Read Data from Other Instances



Step 1. App server issues data request to the connected CSM instance.
Step 2.CSM issues read lock to its peer instance for the group of blocks.
Step 3. When someone has the latest data than this instance, fetches the data from there.
Step 4. Returns data to the application.

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

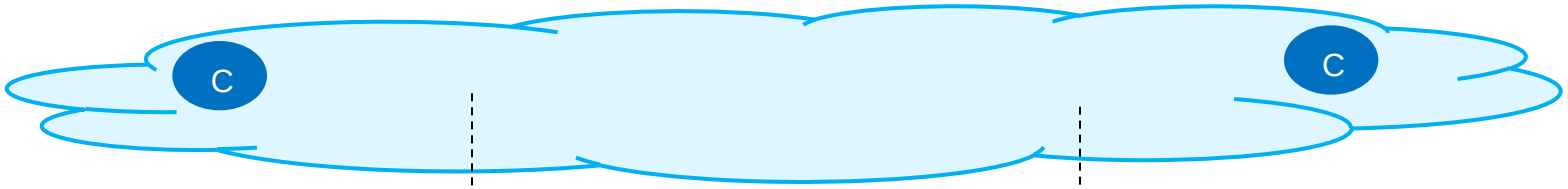
Write data



Step 1. App server issues write request to the connected CSM instance.
Step 2. CSM issues write lock to its peer instance for the group of blocks.
Step 3. Returns the acknowledgement to the app server after writing locally.
Step 4. Writes back to one of the instance immediately and to others upon read request.

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

Create New CSM Instance



Decide on the remote storage node where the new instance has to be created

Create instance

Create instance response

Allocate storage space and performance for this instance

Take the baseline snapshot and send it across the data to the newly created instance

Start initial transfer

Initial transfer complete

If there is any application traffic during this time, this node acts as the proxy to the source node.

Keep sending the incremental data post initial update

Update

Update

Announce in-sync when the last few updates are small enough; beyond this point, source stops sending updates.

In-sync

C Cloud Storage Machine

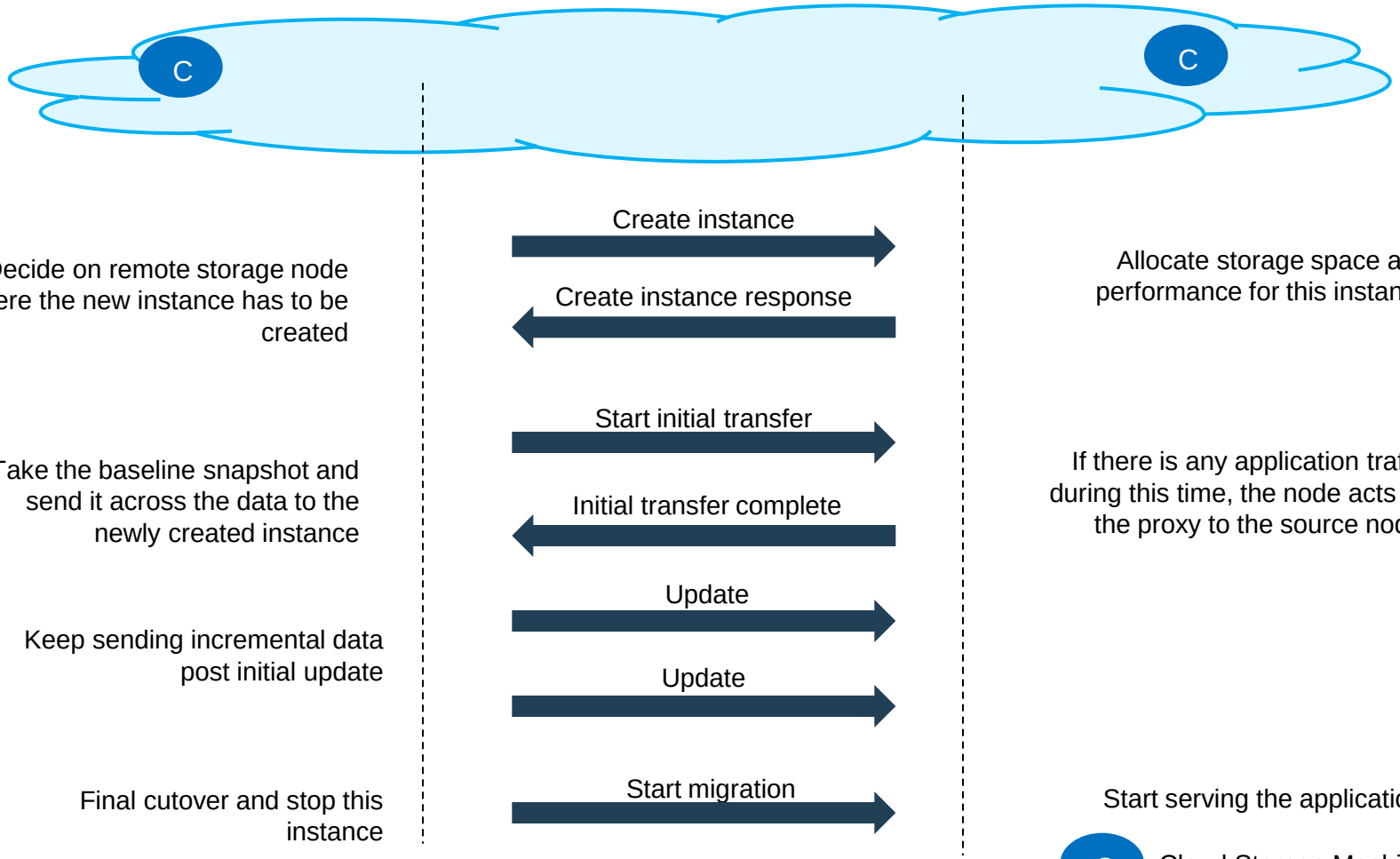
Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

Failure Handling

- Master CSM keeps sending keepalives to all the CSM instances
  - Maintain the state all the instance
- If Master CSM fails, other instances elect the new master based on traffic density
  - The highest traffic density instance within the surviving instance becomes the master
- App server reconnects with the surviving instance whenever there is an instance failure
- CSM instance re-entering into the system should first get into in-sync state

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

CSM Migration with One Instance



Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

Summary

- For cloud to be real for enterprise apps, hardware characteristics of storage need to be abstracted to software
- Abstracted storage must be available across the data center to get the same level of benefits of web apps today
- iSCSI is the standard supported by most of the storage arrays. Therefore, iSCSI can be used for inter-storage node communication

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

References

- https://www.ietf.org/rfc/rfc3720.txt

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.

Attribution & Feedback

The SNIA Education Committee thanks the following individuals for their contributions to this Tutorial.

Authorship History

Name/Date of Original Author here:
Felix Xavier

Updates:
Felix Xavier/February 2015
Name/Date
Name/Date

Additional Contributors

Umasankar Mukkara
Nadeem Kattangere

Please send any questions or comments regarding this SNIA Tutorial to ***tracktutorials@snia.org***

Approved SNIA Tutorial © 2015 Storage Networking Industry Association. All Rights Reserved.