SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

Cloud Mobility and Data Movement

August 7, 2018

Today's Presenters

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

Alex McDonald
NetApp

Michelle Tidwell
IBM

Eric Lakin
University of Michigan

SNIA Legal Notice

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

- The material contained in this presentation 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.
- 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.

SNIA-At-A-Glance

170
industry leading organizations

2,500
active contributing members

50,000
IT end users & storage pros worldwide

Learn more: **snia.org/technical** **@SNIA**

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

What We Do

Educate vendors and users on cloud storage, data services and orchestration

Support & promote business models and architectures: OpenStack, Software Defined Storage, Kubernetes, Object Storage

Understand Hyperscaler requirements Incorporate them into standards and programs

Collaborate with other industry associations

Agenda

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

- The Value of Data: Why is Data Mobility Important?
- Key Challenges & Considerations
- Use Cases and Solutions for Multi-Cloud Data Mobility
- Gaining Business Insights
- Technologies for Data Mobility in Multi-Cloud Environments
- Storage at University of Michigan

Why is Data Mobility Important?

- Find it
- "Refine" it
- Recover it
- Process it
- Distribute it

Data Mobility: Unlocking the Value of Data

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

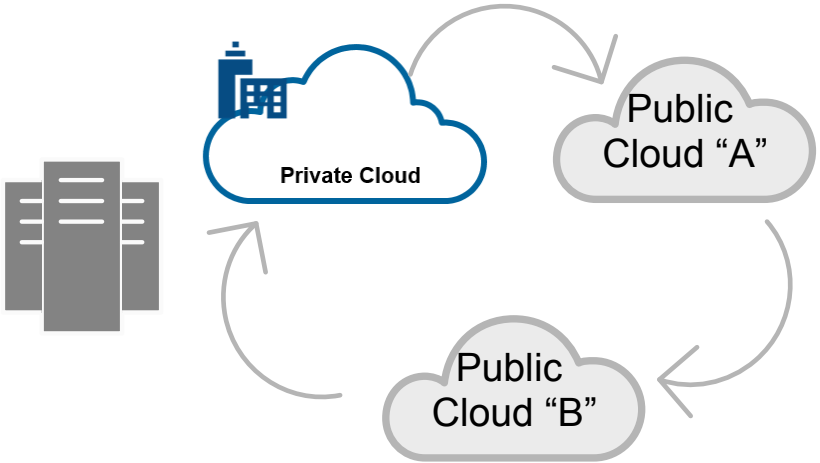
Data Mobility

- Modern & agile
- Multi-cloud enabled
- Secure
- Flexible – for both on premises private clouds and public clouds
- Appropriate for the workload.

Data Insights

- Data is the life blood of today's businesses
- With agile and flexible data mobility, extend the value of your data
- Leverage public and private clouds to gain insights through analytics, services and AI.

Key Considerations and Challenges



- Cold vs. hot data
- Block, file and object – The type of data brings a variety of data mobility solutions.
- Workload performance considerations
- Bridging incompatible storage infrastructure between on-premises and public cloud
- Consistent data management regardless of location
- Protecting data as part of data mobility
- Data mobility effects on production workloads – storage-based vs. server-based replication techniques
- What about high availability in the cloud?

Use Cases and Solutions

Archive: Cold vs. Hot Data

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

Cold Data

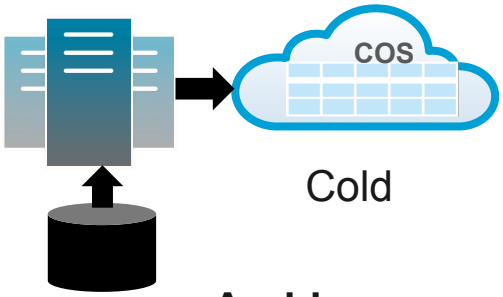
- Appropriate for object storage repositories in private or public clouds
- Write once, read never (or seldom)
- Longer Recovery times for Block Workloads (days, weeks)
- Purpose-built solutions from storage or cloud vendors can bridge between on-premises infrastructures and cloud archive storage
- Typical Interfaces Include: S3, CDMI and other RestFul API interfaces.

Hot Data – Primary Data

- Workload dependent and typically performance oriented
- Block or file protocols (typically)
- Near Zero RPO/RTO considerations
- Data management consistency on premises and on the cloud
- Data mobility options vary dependent on distance, bandwidth, and data reuse scenarios.
- Typical interfaces include: NFS, SMB, iSCSI, NVMe-oF

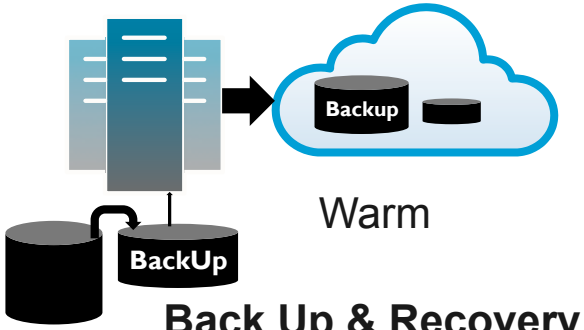
Data Protection & DR

- A range of data protection solutions with different RPO / RTO



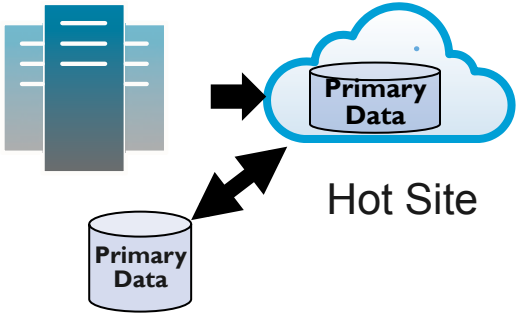
Cold

Archive



Warm

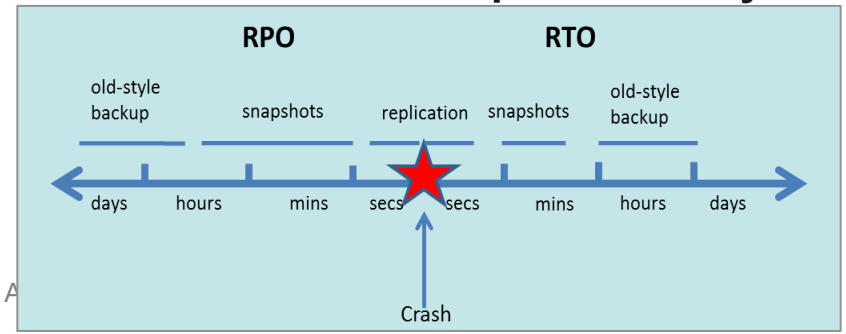
Back Up & Recovery



Hot Site

Real Time Replication

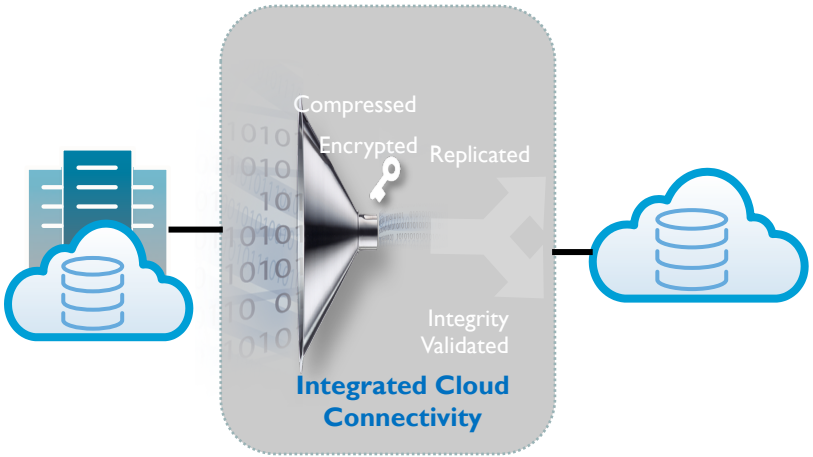
Zero or Near Zero RPO/RTO



- Data mobility in case of a disaster
- Physical AND virtualized apps
- Consistent management for on-premises and the cloud

Data Placement & Optimization

AI can bring new capabilities to data mobility for multi -cloud



- Application Response Time and IOPs Driven
  - Move to higher performance flash drives dynamically in public or private cloud with AI
  - Optimize provisioning of disk and flash
  - Examples: IBM Spectrum Virtualize and IBM Spectrum Virtualize for Public Cloud with EasyTier, NetApp OnCommand
- Lower Cost: Move Inactive data to Object Storage
  - As data ages, snapshot and replicate data to object storage over IP networks – built in
  - Examples: IBM Spectrum Scale with Transparent Cloud Tiering, NetApp OnCommand, Dell EMC Isilon CloudPools

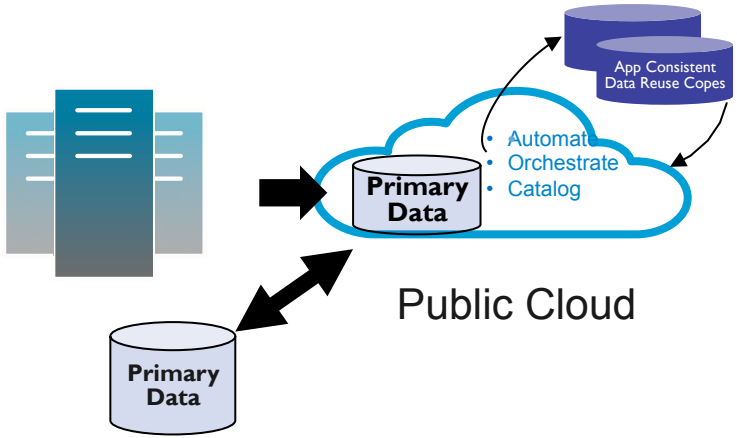
Other Related Use Cases

- Workload mobility
- Edge Compute & IoT
  - Bringing compute to storage
  - Consolidating data
  - Bringing back data
- Data protection & security
- Networking considerations
- Containers

Gaining Efficiency and Business Insights

Dev / Test

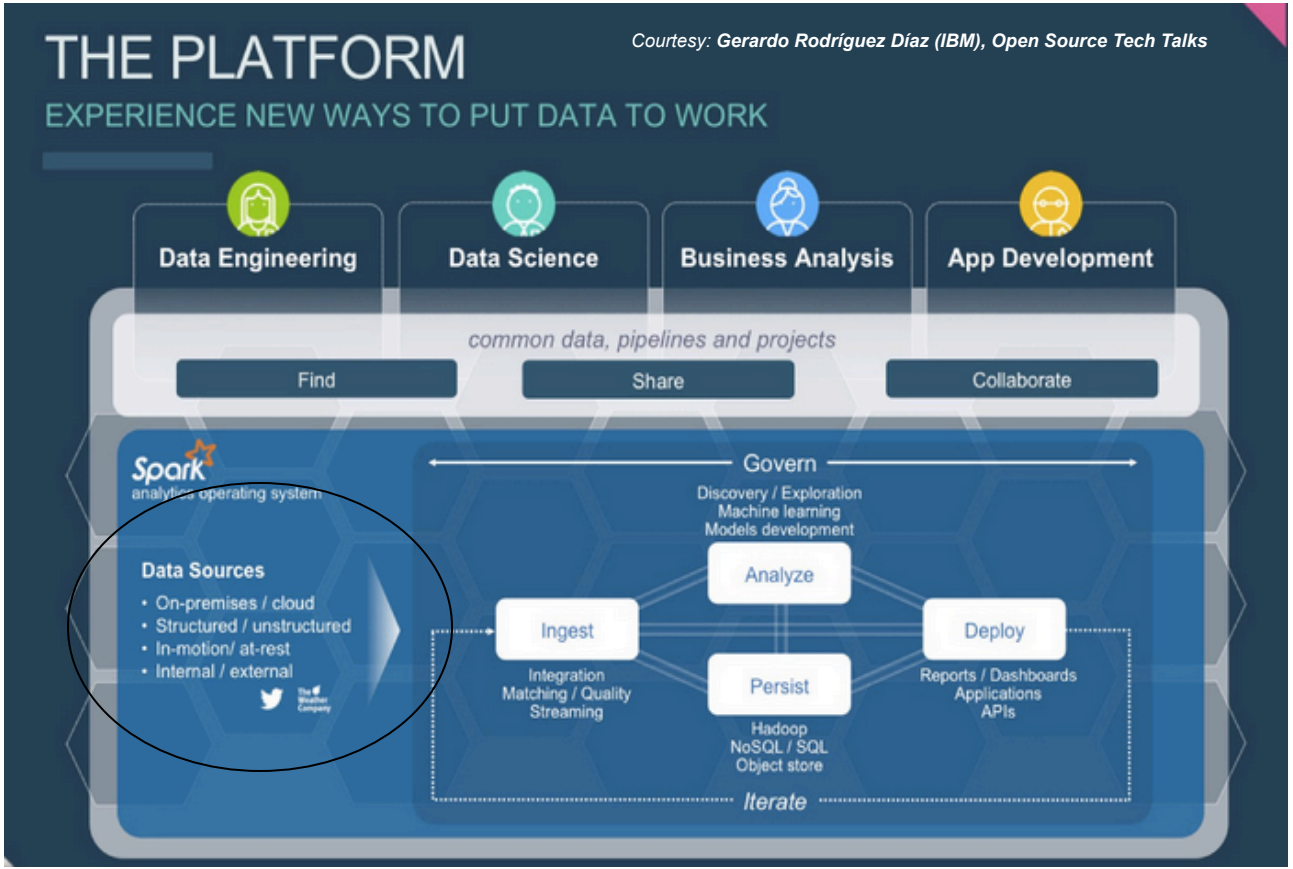
- You've used the appropriate data mobility technology to put your data in the right place, now what?
- What if you could make copies and reuse them for automated use in DevOps workflows?
- What if you could do that without having to ask IT to provision more storage?
  - Leverage public cloud infrastructure
  - With storage-based snapshots
  - Speed time to deliver for applications
  - Reduce bugs by developing and testing against the most recent data



DATA MOBILITY: Create synchronous IP-based replication of primary data to public cloud

DATA REUSE: Dev / Test against the most recent data for applications or VMs

Analytics Example: IBM Watson Data Platform



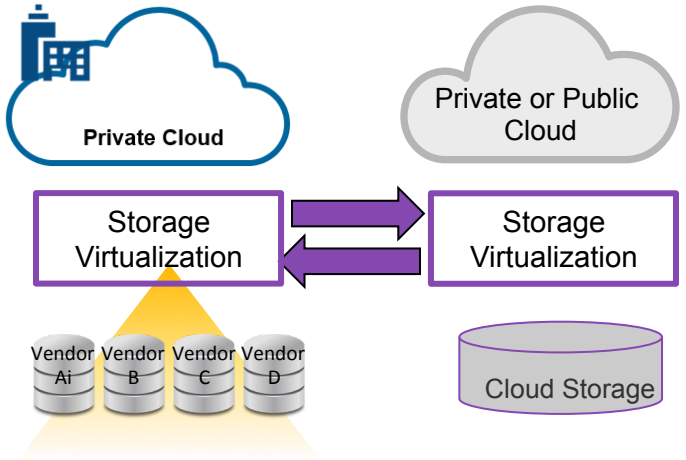
- Mobility of data is critical for scenarios such as Analytics.
- Here's an example from IBM – The Watson Data Platform.
- IBM Watson Data Platform is the first cloud-based data and analytics platform to support cognitive business.
- Data Sources can be On Premises or in the Cloud, Structured or Unstructured
- Data Ingest occurs through REST APIs, SMB, CIFs and other Raw API methods.
- Data is moved to Cloud Object Storage to create a Data Catalog for Analytics and Collaboration.

Technologies for Data Mobility in Multi-Cloud

Storage Virtualization

- Data mobility in multi-cloud environments means moving data between disparate infrastructures
- Create ***any to any*** data mobility solutions with storage virtualization
- Look for solutions with single-pane-of-glass management capability and built-in optimization
- Data replication capabilities across disparate storage infrastructure for IP/Ethernet and FC protocols
- This enables data mobility between disparate multi-cloud environments at the storage array level
- Further optimize underlying infrastructure for cost, performance, and data reuse scenarios

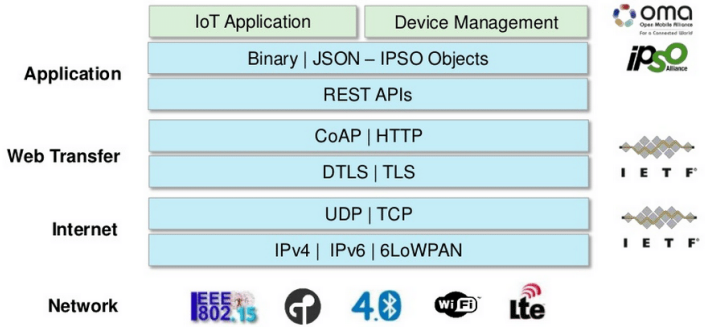
- Example: Storage virtualization and optimization with data mobility, such as IBM Spectrum Virtualize



Protocols and APIs

- Traditional methods of access
  - Block-based; SCSI over Fibre Channel and Ethernet
  - File-based; NFS and SMB
- Newer protocols designed for the cloud
  - SNIA's open Cloud Data Management Interface (CDMI)
  - Amazon's proprietary S3
  - Other object protocols, mainly key/value systems
- IoT (Internet of Things)
  - https://www.postscapes.com/internet-of-things-protocols/

Remember the I in IoT!



Credit: Simon Ford - Director of IoT Platforms ARM

Storage at University of Michigan

University of Michigan

- 19 schools and colleges with a common football team
- One of the top 5 employers in the state of Michigan
- Ann Arbor, Flint, and Dearborn campuses
- #1 public research University in the United States
- $106 million in annual industry sponsored research
- $1.5 billion in annual research expenditures

Storage at Michigan

- Does not include Michigan Medicine (our health system)
- On-premises and cloud-based offerings for units, faculty, and staff
- 10PB of NAS delivering CIFS and NFS to campus (MiStorage)
- 1PB of SAN for back-end systems
- 250TB of AFS for faculty shares and web hosting
- Support for HIPAA and export-controlled data
- 7 full-time storage engineers, 1.6 PB per FTE

MiStorage

- 20% annual growth
- Rate-based service by GB
- Rates collected are used to refresh hardware after depreciation
- Annual growth has been historically funded via capital request

The Problems

- No more capex for growth after FY16; Challenged by leadership to solve growth with operational funds
- Organization is averse to leasing equipment
- Storage vendor cloud support was immature
- HIPAA[1], ITAR[2], & EAR[3] scattered across the environment
- Fear, Uncertainty, and Doubt (FUD) regarding cloud storage

[1]HIPAA – Health Information Portability and Accountability Act (patient and health information)
[2]ITAR – International Traffic in Arms Regulations (research funded by Department of Defense)
[3]EAR – Export Administration Regulation (regulated commercial data)

The Decision

- Cloud Integration
  - Storage system support for native cloud integration – October ’17
  - BAA signed with primary cloud provider – November ’17
  - Testing and small-scale data movement (~20TB) – January ‘18 to present
  - Approval secured for HIPAA, ITAR, & EAR data – April ‘18
  - Pricing agreement and capacity commitment with cloud provider – May ‘18
  - Go-live with tiering and data movement – July ‘18

The Plan

- Integrate on-premise storage with a primary cloud provider
- Files not accessed > 180 days are moved to the cloud
  - CIFS and NFSv4 shares to government-approved cloud storage
  - NFSv3 shares to standard cloud storage (no sensitive data is allowed on NFSv3)
  - HSM model with stubs – transparent to users
- Goal – 4PB in the cloud over 5 years
- Shrink on-premise footprint while expanding cloud footprint
  - Reduce data center costs
  - Take advantage of cloud elasticity
  - Absorb 5 years of growth without increasing FTE count

Expected Challenges

- Financial
  - Is cloud storage really cheaper? That's TBD.
  - Our billing system isn't designed to include cloud costs.
- Political
  - Some units are resistant to data going in the cloud. Some have good reasons (DMP restrictions), others don't.
- Technological
  - How will it scale as we approach a multi-PB cloud footprint?
  - What is our data liberation plan? (We don't have one)

About Assumptions

- Don’t assume…
  - that you know where all of the sensitive data is stored
  - that your storage users know where their sensitive data is stored
  - that no response from your customers implies consent to a cloud strategy
- Do assume…
  - that your plans will be delayed as you discover more locations with sensitive data
  - that more data than you anticipated will be sensitive in nature
  - that vendors may exaggerate their capabilities

Lessons Learned

- Not all storage vendors integrate with cloud vendors in the same ways
  - Features and capabilities very widely, such as reporting and support for quotas
- Understand that sensitive data may exist in places it should not
- Communicate early and often with storage users
  - To many, cloud storage is an unknown wilderness fraught with vulnerabilities and weaknesses. It may take time to get people on-board with a cloud-integration strategy

After This Webcast

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

- Please rate this webcast and provide us with feedback
- This webcast and a PDF of the slides will be posted to the SNIA Cloud Storage Technologies Initiative website and available on-demand at https://www.snia.org/forum/csti/knowledge/webcasts
- A full Q&A from this webcast will be posted to the SNIA Cloud blog: www.sniacloud.com/
- Follow us on Twitter @SNIACloud

Thank You