Unstructured Data in the Cloud with ECS

Mikhail Vladimirøv Senior Sales Engineer

GLOBAL SPONSORS





How to build modern digital Archive (WEB 3.0)

Five reasons to choose ECS

Agenda

Some Definitions

Market Factors Driving Adoption

Trends in Archiving

DellEMC ECS Storage

Summary

Some Definitions

Latin archīum or archīvum, which is the romanized form of the Greek ἀρχεῖον (arkheion), "public records, town-hall, residence, or office of chief magistrates",itself from ἀρχή (arkhē), amongst others "magistracy, office, government, which comes from the verb ἄρχω (arkhō), "to begin, rule, govern".



Backup

Short-term Insurance Policy

Kept For Weeks Or Months - Overwritten On A Regular Basis

Secondary Copy - Operational Recovery

Altered State - Restore Required

Disaster Recovery / Business Continuance



Archive

Long-term Insurance Policy

VS

Kept For Years, Decades, or Forever

Primary Copy - Information Retrieval, Compliance, and eDiscovery

Native State - Immediately Available

WORM Used For Compliance Enforcement



Types Of Archiving

In-place (Passive Archive) Archive		e via Backup Active Archive			Long Term Preservation Archive			
Primary Storage		Secondary Storage		Long-term Storage				

Access times and dollar per gigabyte

In-Place Archive

- Do Nothing
- Wastes IT resources
- Management issues
- Regulatory challenges

Archive via Backup

- Inexpensive
- May not be indexed
- In altered state
- Might be offsite
- Regulatory challenges

Active Archive

- Fast access to data
- Non-intrusive
- Data can be retained for years
- Regulatory compliant
- In-place data analytics
- Often file based

Long Term Archive

- Data resides for decades
- Fast growing data sets
- In-place analytics
- Long tail data monetization
- Regulatory compliant
- Often Object/Cloud

What Is Driving Archiving Adoption?



80%

Because Most Of This Data Is Static

60-80%

OF DATA IN OPERATIONAL APPLICATIONS IS *INACTIVE*

INACTIVE (FINAL FORM)

ACTIVE

(IN PROGRESS)



Meaning The Probability Of Access And Value Decreases Quickly

The value of the data and likelihood of access goes down over time

But a litigation or

instantly increase

the value of the data

other event can

۲

- High availability and • fast access can be critical



The Result: Unprecedented Storage Costs

- Primary Storage is quickly consumed driving more purchases
- Backups can no longer be accomplished in window driving more capex and opex
- Power, cooling and other facility costs increase
- Human resources costs increase
- Value of data is marginalized
- Companies may put themselves at regulatory risk

The Solution: Intelligent Archiving



The Vicious Storage Cycle

Intelligent Archiving Breaks the Cycle



Reclaim expensive existing primary storage



Reduce on-going primary storage acquisition costs



Get static data out of the recurring backup process



Reduce backup storage acquisition costs



Reduce management and operating costs



Analyze data in place on cost effective storage

Three Trends In Archiving

- Data is coming from more sources that ever before
 - Archives must be designed for flexibility
- The era of big data analytics increased value of archived data
 - The archive should be capable of in-place data analytics
- Archives are moving to object clouds to deal with massive data sets
 - The archive data should be geo-distributed/geo-accessible

ECS Supports The Trends

Solving Archive Challenges At Geo-scale



- Modern Hyper-scale Cloud Architecture
 Scales from Petabytes to Exabytes
- Archive data from all sources
 - S3, Swift, Atmos and Centera CAS Object APIs
 - HDFS compatible with Cloudera, Hortonworks, Pivotal etc.
- Break down barriers with Geo-scale data access
- Innovation to enable scalability, efficiency and serviceability!



ECS Solutions For Archive

EMC Elastic Cloud Storage (ECS) - Hardware



U-Series

Available in multiple capacities within a rack **COTS** – x86 servers, JBOD DAS, 10GbE connectivity, SATA/SAS Disks Max 60 disks per DAE per node

D-Series

Denser Model Minimum of eight x86 servers Max 98 disks per DAE per node. Hyper-scale Aggressive seeking lowest \$/GB

ECS Software Certified on Dell Servers



Object Access

Namespace / Bucket|Container

- Namespace can span multiple instances of physical hardware, and data management functions like data replication and data distribution at object-level granularityy.
- Instead of organizing files in a directory hierarchy, object storage systems store files in a flat organization of containers/buckets.



File storage Data stored as 'files' in hierarchically nested 'folders'—ideal for active documents



Object storage Data stored as 'objects' in scalable 'buckets'—ideal for unstructured big data, analytics and archiving

ECS Data

SYSTEM META-DATA

- Identifiers and descriptors
- Encryption keys in encrypted format
- Internal flags
- Location information
- Timestamps
- Configuration/tenancy information



CUSTOMER META-DATA

Client=DellEMC

Event=DellEMC Forum

• ID=123

Access Protocols

- Representational State Transfer Protocol APIs
 - Comibnation of HTTP methods
 - Amazon S3
 - Openstack Swift
 - Atmos
 - Content Adress Storage (CAS)

ECS Supports A Wide Variety Of Workloads



Scale Effortlessly - Store Efficiently - Access Globally

Data Archives - Complete Cloud Storage Platform



- Lower cost than public cloud
- Unmatched combination of storage efficiency and data access
- Anywhere read/write access with strong consistency simplifies finding and using archived assets
- No single points of failure increases availability and performance
- Universal accessibility eliminates storage silos and inefficient archiving processes
- Comprehensive data types satisfy the broadest range of application needs

ECS Storage engine unique capabilities



Multi-Protocol Support

Access the same data from any access method

- CAS: SDK v3.1.54 or later support for upgrade
- Swift: byte range update within an object
- Retentions
- Keystone integration: drop-in replacement for OpenStack Swift



Global namespace
 with global locking

 Primary file system with native Ambari 2.2 integration

- Byte range updates
- Retentions
- Metadata search
 extension

CloudBoost to ECS

ENABLING LONG TERM RETENTION FOR BACKUP DATA



Compliance ready



Features

- Retention policy management
 - Retention enforcement
 - Data immutability
 - Advanced Rétention Management
- Access Locks
 - Lock/unlock user or bucket

Key Benefits

- Meet storage requirements for
 - ✓ SEC 17a-4(f)
 - ✓ CFTC 1.31(b)-(c)
 - ✓ NF Z 42-013

Geo Capabilities High durability with a low overhead



Storage Efficiency High performance and low overhead for both small and large objects



Metadata search

Uploading from client								
File Browse	30.jpg							
Bucket Name	•							
Retention to app	Retention to apply to protect the picture in days							
File size 668	File size 668565							
Image Width	3264							
Image Height	2448							
GPS Latitude	35.6532944444444							
GPS Longitude	139.76273055555555							
Datetime 20	15:11:01 09:52:31							
Upload file								

Save objects with metadata



Search objects using GPS coordinates, image resolution, ... /?query=x-amz-meta-image-gps-latitude>50&...

R Show B Delet



Native NFS support



Features

- Native NFS v3 capability
- ✓ Rich ACLs
- ✓ Global namespace
- Global locking
- Multi-protocol access object, NFS and HDFS

Key Benefits

- Ingest data in native format
- Requires no change on the application level, accelerating the move to an object platform



0 9	SERF WIN201	(R2AD paristal	- ECS CIFS Gatery	ey.			
	tiome Statu	8					
0 E	Add ICS Modify Drive ICS Dree	Delete ECS Delete Contriputation	Reporting Cloud	Force Paul	e Matual Doos in ad Ducament Explorer ECS Drive		
Drive	Description	Host	Status	Connectio	Samar Saminer	27017569	
€E		ECS Paris	Active	Connected	And the second		
Date/1	Time Me	ssage					

ECS Drive	Parameters [WIN2012R2AD.	paris.lab]					
Drive Settings Logging							
ECS Drive:	FA.		Adv	anced Ontions			X
Host:	ECS Paris		C\DecempOntel Et	AC ECE Colores Data E	Drivel	1	_
Bucket:	geodrive1	Local Cache: Configurations build of the second sec					
		Delay after file is closed before upload (Modify Wait Time)	ding it to the cloud:			Seconds	×
		How long to wait before replacing los (Access Wait Time Upload)	cal files with a stub afte	er upload:	2	Seconds	v
Protect files with checksum on all uploads		How long to wait before replacing local files with a stub after restore: (Access Wait Time Restore)				Minutes	v
Track Folder Security Write all files and folders to cloud i	in lower case	Access Wait Time if free d	fisk space is below:	50 68	2	Minutes	v
Preserve folder case ECS Options Translate ECS/Windows Permi	istions	Access Wait Time if free d	fisk space is below:	5 68	1	Minutes	v
	USN Journal Fil						
		How long to wait before checking the (Update Wait Time)	e cloud for updated co	intent:	1	Minutes	×
		Maximum time to wait for a file to clo (Open Wait Time)	ose before uploading to	to the cloud:	1	Minutes	v
				0	ĸ		Cancel

Features

- 🗸 S3 API
- ✓ Caching
- Multipart upload and download
- Retention & versioning
- ACL translation
- Client side load balancing

Key Benefits

- Ingest data in native format
- Requires no change on the application level, accelerating the move to an object platform

Five reasons to choose ECS

- Simple
- Powerful
- Efficient
- Scalable
- Support for various data types across the enterprise
- EMC ECS cloud storage platform combines the cost advantages of commodity infrastructure with high reliability, availability and serviceability

Thank you!

