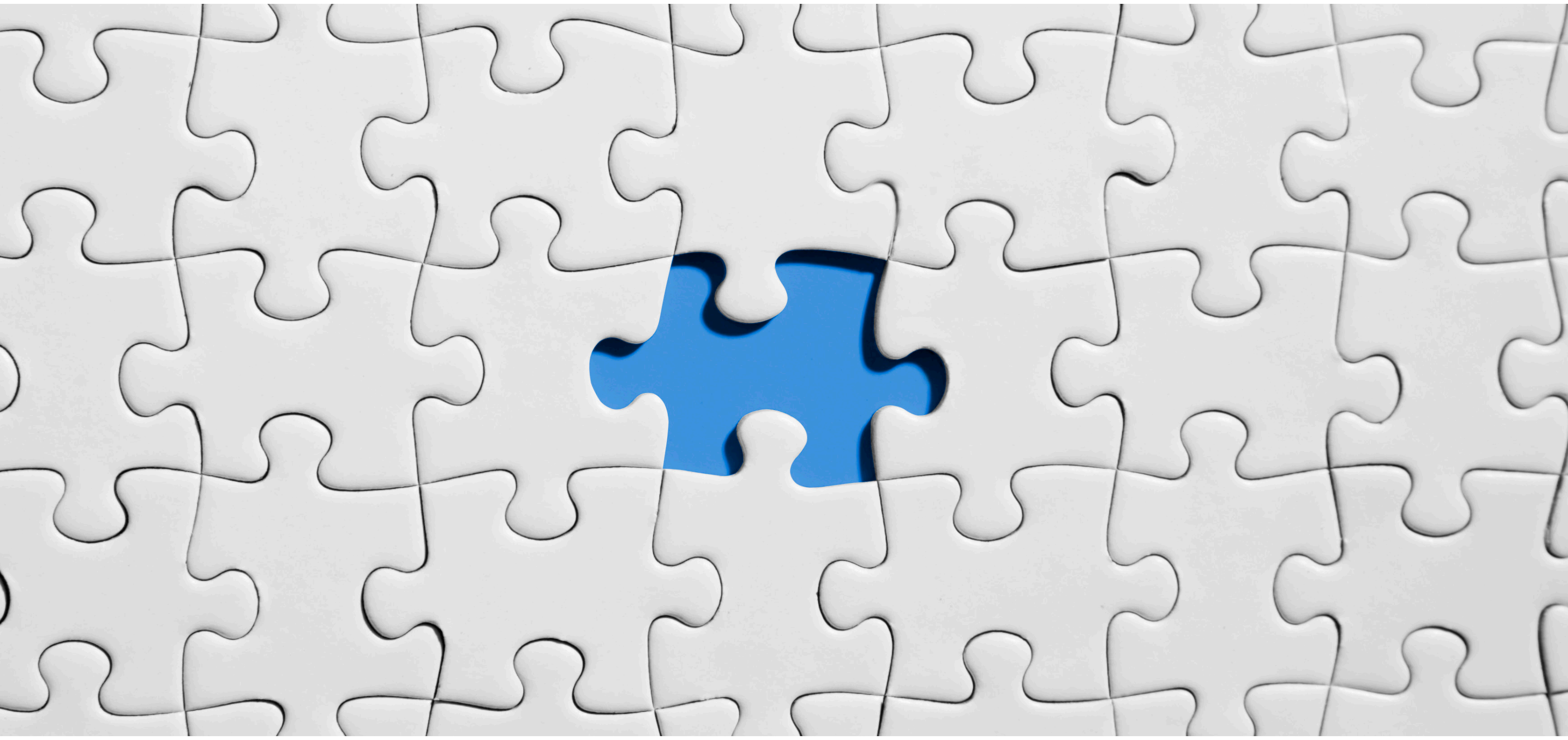
DELL EMC

Elastic Cloud Storage Sizing



Poorna Chandra Prashanth C

Senior Systems Engineer Analyst

Dell EMC – Presales Shared Services

Poorna.prashanth@dell.com

Knowledge Sharing Article

DELL EMC PROVEN PROFESSIONAL

Table of Contents

Executive Overview 3
Intended Audience 3
Introduction 3
Scope 4
ECS Appliance Specifications 4
Architecture Overview 10
ECS Sizing 12
Data Requirements for Sizing 13
Appliance Sizing Best Practices 15
Conclusion 17
Appendix 17

Executive Overview

This paper provides a high level overview of basic steps of Elastic Cloud Storage (ECS) appliance sizing using best practices. It also offers overview of the product architecture and hardware specifications that constitute an appliance.

Intended Audience

This paper is intended for Dell EMC sales and presales field personnel interested in gaining a high level understanding of the ECS architecture and basic methodology/process/procedures to size ECS Appliance to best meet customer requirements.

Introduction

While traditional SAN and NAS storage platforms were designed to meet the high performance requirements of critical applications, they were not designed to cater to cloud applications and cloud scalability. Elastic Cloud Storage appliance provides a complete software-defined object storage platform designed for today's cloud scale storage requirements.

ECS is a Third Platform Object-based Software-Defined Storage which is available as an appliance and is designed for traditional and next generation applications with high storage efficiency, resiliency and simplicity. ECS supports multi-protocol which can handle data from Internet of Things (IoT), modern applications and Big Data applications using native protocols. ECS also handles backup and archive applications.

Massive growth in structured and unstructured content is driving the need to store large files such as images and videos with universal accessibility (object, file and HDFS). ECS enables consolidation of multiple storage- and object-based archives into a single, efficient and globally accessible content repository. Due to unpredictable and unprecedented data growth, sizing ECS in real time is a challenge for any environment.

Scope

- Outlines ECS architecture and specifications on a high level
- Discusses basic procedures of ECS sizing from object based storage
- Understand the process to gather baseline information on the key user groups that have been identified as good candidates for ECS. The purpose of this step is to understand the performance characteristics and capacity requirement of the target users, i.e. Are the applications more CPU- or memory-intensive? Are there an excessive number of storage operations? What applications do they need? What type of network load is being generated by the end users' activities? What is the basic storage requirement? What native protocols are in place?
- Data collection of existing environment and interpretation of the collected data, results of which will be considered for ECS sizing
- Use internal tools or develop new tools to apply our findings to size ECS Appliance
- Best practices considered for ECS Sizing
    - How to determine number of nodes.
    - How to choose right object size.
    - Capacity to be considered.
    - Growth factors that may impact the overall sizing.

This paper will not discuss

- Replication
- Multisite configurations
- Geo Protection methodologies

Overall, this paper provides insight to ECS Sizing using Dell EMC best practices.

ECS Appliance Specifications

ECS appliance is a multi-purpose, software defined cloud storage platform that can be deployed on qualified industry-standard hardware or a turnkey storage appliance. The ECS Appliance Gen1 and G2 offer flexible configurations that can scale up to 6.2 PB in a single rack.

An ECS appliance consists of **nodes**, **switches** and **DAE's** with **drives**

Nodes: Two varieties; Gen1 and Gen2 with 2CPUs per node

Switch: Two types of switches; **Private Switch** and **Public Switch**

- **Private switch** is meant for management traffic. In C Series appliance would require second switch for appliance with >6 Nodes
- **Public switch** is meant for data traffic. In C Series appliance would require second switch for appliance with >6 Nodes

There are three appliance series:

1. U Series
2. D Series
3. C Series

U Series is commodity object storage with Servers and DAE's. Appliance includes the components shown in Table 1.

Component	Description
40U rack	Titan D racks that include: Single-phase PDUs with four power drops (two per side). The high availability configuration (HA) of four power drops is mandatory and any deviation requires that an RPQ be submitted and approved. Optional three-phase WYE or delta PDUs with two power drops (one per side) Front and rear doors Racking by EMC manufacturing
Private switch	One 1 GbE switch
Public switch	Two 10 GbE switches
Nodes	Intel-based unstructured server in four and eight-node configurations. Each server chassis contains four nodes (blades). Gen2 also has the option for five- and six-node configurations.
Disk array enclosure (DAE)	The U-Series disk array enclosure (DAE) drawers hold up to 60 3.5-inch disk drives. Features include: Gen1 hardware uses 6TB disks and Gen2 hardware uses 8TB disks Two 4-lane 6 Gb/s SAS connectors SAS bandwidth of 3500 MB/s SAS bandwidth of 3500 MB/s Drive service: hot swappable

Table 1: ECS Appliance: U-Series hardware components

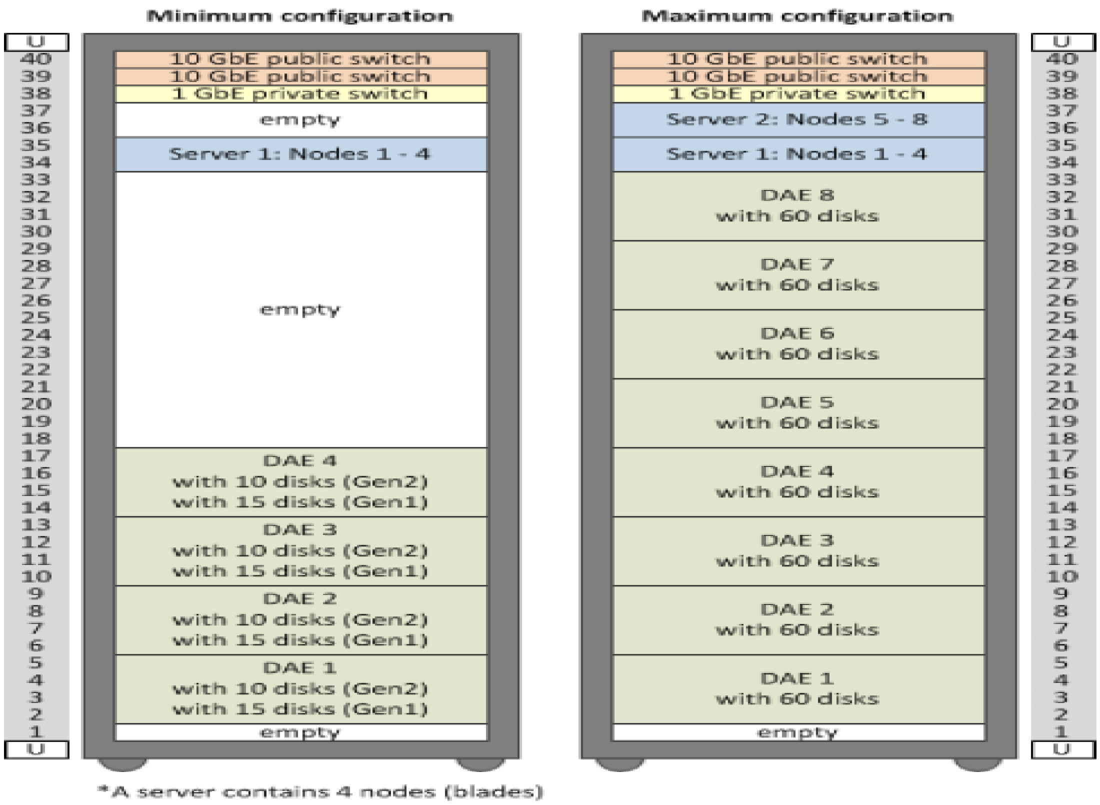


Figure 1: U-Series ECS Appliance minimum and maximum configurations

a) D Series is dense object storage with servers and DAE's

Component	Description
40U rack	Titan D racks that include: Single-phase PDUs with six power drops (three per side). The high availability configuration (HA) of six power drops is mandatory and any deviation requires that an RPQ be submitted and approved. Optional three-phase WYE or delta PDUs with two power drops (one per side) Front and rear doors Racking by EMC manufacturing
Private switch	One 1 GbE switch
Public switch	Two 10 GbE switches
Nodes	Intel-based unstructured server in eight-node configurations. Each server chassis contains four nodes.
Disk array enclosure (DAE)	The D-Series disk array enclosure (DAE) drawers hold up to 98 3.5-inch disk drives. Features include: 8TB disks Two 4-lane 12 Gb/s SAS 3.0 connectors SAS bandwidth of 5600 MB/s Drive service: cold service
Service tray	50-lb capacity service tray

Table 2 - ECS Appliance: D-Series hardware components

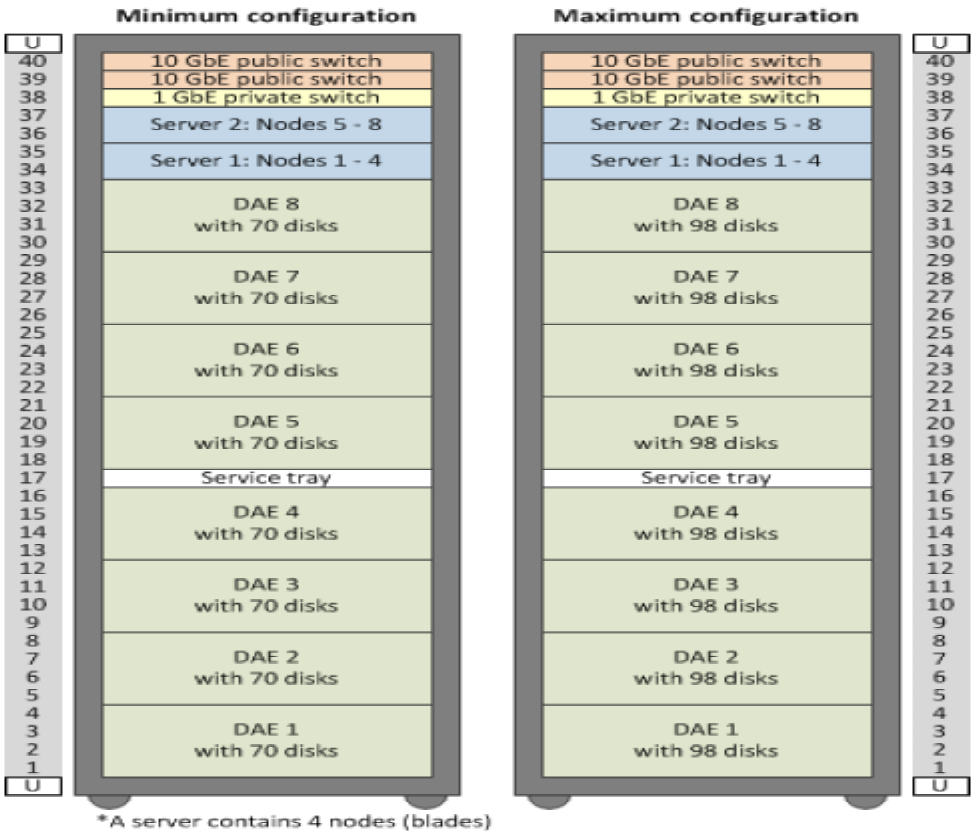


Figure 2 D-Series ECS Appliance minimum and maximum configurations

b) C Series is dense compute and storage solution of servers with integrated disks

Component	Description
40U rack	Titan D Compute racks that include: Two single-phase PDUs in a 2U configuration with two power drops. The high availability configuration (HA) of two power drops is mandatory and any deviation requires that an RPQ be submitted and approved. Optional two three-phase WYE or delta PDUs in a 2U onfiguration with two power drops Front and rear doors Racking by EMC manufacturing
Private switch	One or two 1 GbE switches. The second switch is required for configurations with more than six servers.
Public switch	Two or four 10 GbE switches. The third and fourth switches are required for configurations with more than six servers.
Nodes	Intel-based unstructured servers in eight through 48 node configurations. Each server chassis contains four nodes (blades).
Disks	TThe C-Series has 12 3.5-inch disk drives integrated with each server. Gen1 hardware uses 6TB disks. Gen2 hardware uses 8TB disks.

Table 3 ECS Appliance: C-Series hardware components

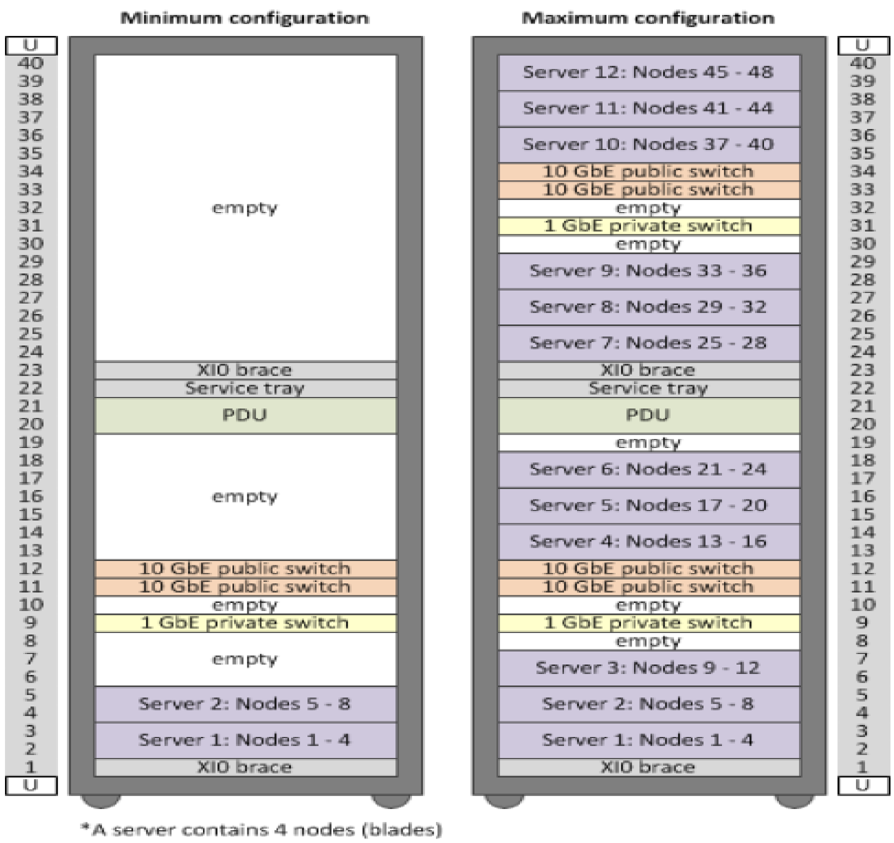


Figure 3 C-Series ECS Appliance minimum and maximum configurations

Table 5 describes U-Series, D-Series and C-Series Gen 2 configuration

- U-Series ECS Appliance is a commodity object storage solution
- D-Series ECS Appliance is a dense object storage solution using commodity hardware
- C-Series ECS Appliance is a dense compute solution using commodity hardware

U-Series (Gen2) configurations					
Model number	**Nodes**	**DAEs**	**Disks in DAE**	**Storage capacity**	**Switches**
U400 (minimum configuration)	4	4	10	320 TB	One Private Two Public
U400-E	5	5	10	400 TB	One Private Two Public
U480-E	6	6	10	480 TB	One Private Two Public
U400-T	8	8	10	640 TB	One Private Two Public
U2000	8	8	30	1.92 PB	One Private Two Public
U2800	8	8	45	2.88 PB	One Private Two Public
U4000 (maximum configuration)	8	8	60	3.84 PB	One Private Two Public
D-Series configurations					
D4500 (minimum configuration)	8	8	70	4.5 PB	One Private Two Public
D6200 (maximum configuration)	8	8	98	6.2 PB	One Private Two Public

Table 4 ECS Appliance: U-Series and D-Series Configurations

C-Series (Gen2) configurations			
Phoenix-12 Compute Servers	Nodes	Storage Capacity	Switches
2 (minimum configuration)	8	144 TB	One Private Two Public
3	12	216 TB	One Private Two Public
4	16	288 TB	One Private Two Public
5	20	360 TB	One Private Two Public
6	24	432 TB	One Private Two Public
7	28	504 TB	One Private Two Public
8	32	576 TB	One Private Two Public
9	36	648 TB	One Private Two Public
10	40	720 TB	One Private Two Public
11	44	792 TB	One Private Two Public
12 (maximum configuration)	48	864 TB	One Private Two Public

Table 5 ECS Appliance: C-Series Configurations

Appliance can be upgraded through HDD extension kits. Drives can be upgraded as per Table 6.

Disk upgrade kit Uses	Uses
10-Disk Upgrade	• Used to supplement other disk upgrade • kits to make up a valid configuration
40-Disk Upgrade	• Add 10 disks to each DAE in a four-node configuration • Add 5 disks to each DAE in a eight-node configuration • Populate a new DAE in a configuration with 40-disk DAEs
60-Disk Upgrade	• Add 10 disks to each DAE in a six-node configuration • Populate a new DAE in a configuration with 60-disk DAEs

Table 6 ECS Appliance: Disk Upgrade kit

Architecture Overview

ECS provides a software defined cloud storage solution that can be deployed on a set of qualified industry-standard hardware or a turnkey storage appliance. ECS is composed of:

1) **ECS Portal and Provisioning Services**
   Provides a Web-based portal that enables self-service, automation, reporting and management of ECS nodes. It also handles licensing, authentication, multi-tenancy and provisioning services.

2) **Data Services**
   Provides services, tools and APEs to support Object, HDFS and NFSv3
   Below table outlines Object APIs and protocol supported and protocols that interoperate

Protocols		Supported	Interoperability
Object	S3	Additional capabilities like Byte Range Updates and Rich ACLS	HDFS, NFS
	Atmos	Version 2.0	NFS (path-based objects only and not object ID style based)
	Swift	V2 APIs and Swift and Keystone v3 Authentication	HDFS, NFS
	CAS	SDK v3.1.544 or later	N/A
HDFS		Hadoop 2.7 compatibility	S3/NFS, Swift/NFS
NFS		NFSv3	S3/HDFS, Swift/HDFS , Atmos (path-based objects only and not object ID style based)

Table 7 - ECS Supported Data Services

3) **Storage Engine**
   Stores and retrieves data, manages transactions and protects and replicates data. ECS has a layered architecture with each layer acting independently. Each layer scales horizontally and across all nodes in the system and ensures high availability. All services in a storage engine are encapsulated within a Docker container and installed on each node, providing a distributed and shared service.

4) **Fabric**
   Provides clustering, health, software and configuration management. It also provides upgrade capabilities and alerting. Ensures that services are running and resources such as disks, containers, firewall and network are managed effectively. It tracks and reacts to device failures.

5) **Infrastructure**
   Uses SUSE Linux Enterprise Server 12 as the base operating system for the turnkey appliance.

6) **Hardware**
   Offers a turnkey appliance or qualified industry-standard hardware.

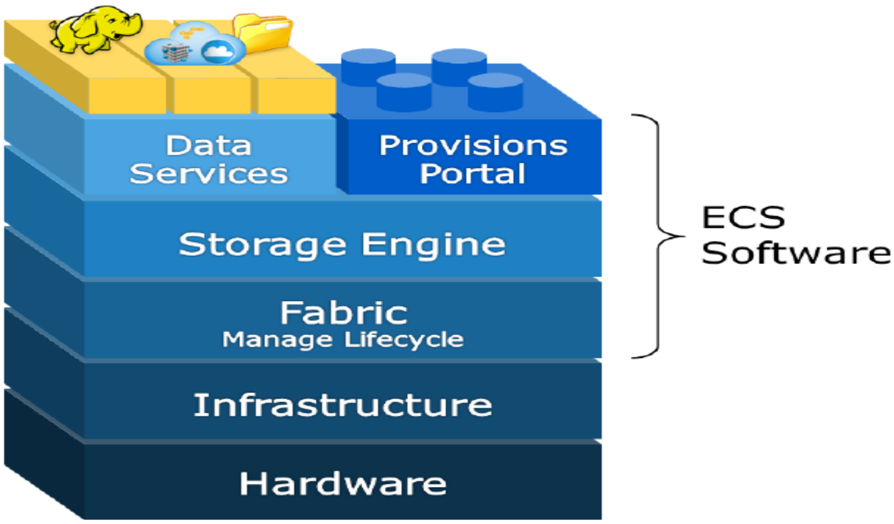


Figure 4 - ECS Architecture Overview

Data and Meta Data

Object is the smallest storage unity of an Object based storage system. Unlike files and blocks Object is the combination of data itself plus certain attributes that describes the data.

Summary of how data is stored is as follows:

Data: Actual content that is to be stored which includes images, videos, documents, etc.

Customer Meta Data: User-defined Meta data. This provides information about categorization of stored data.

System Meta Data: Attributes and Information related to the data.

All these data are stored as 128MB logical container of contiguous space – chunks– and each chunk can have data from different objects. ECS uses indexing to track all the parts of an object that may be spread across different chunks and nodes.

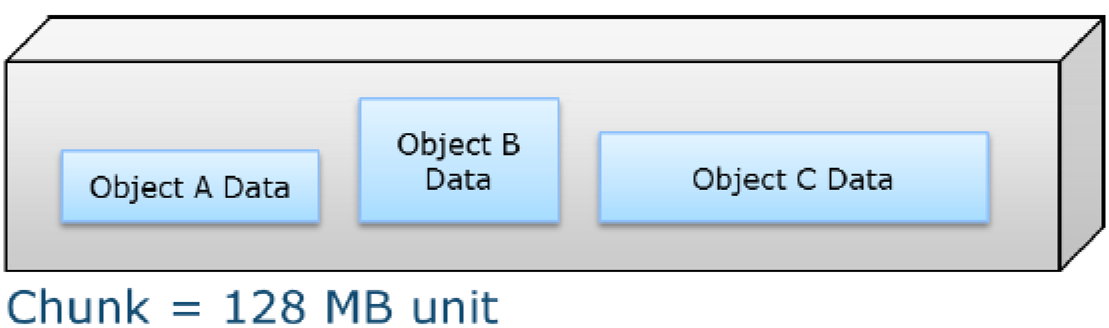


Figure 5 - Chunk (can store data from different objects)

These chunks are written in an append-only pattern and are triple mirrored to protect data during failures. Once the data is erasure coded, the mirrored copies are deleted. Erasure code is done in 12+4 or 10+2. Object is divided into 12/10 data fragments and 4/2 parity fragments which are spread across the nodes.

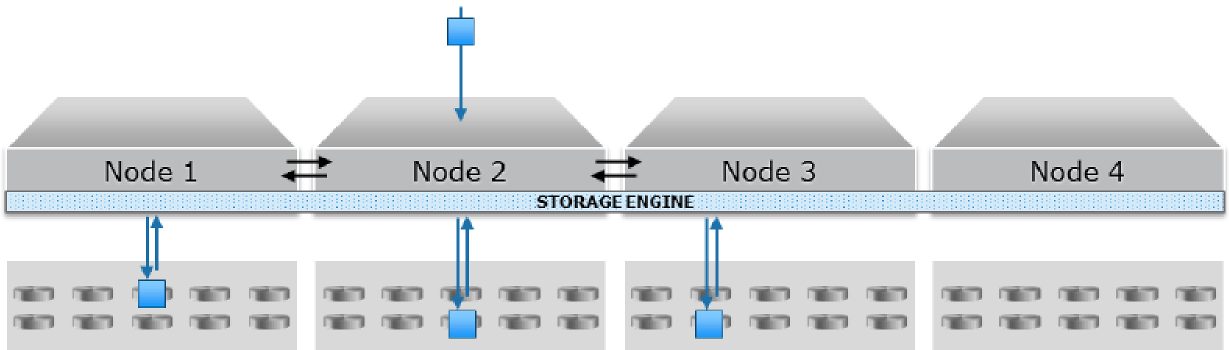


Figure 6 - Triple Mirroring of Chunks

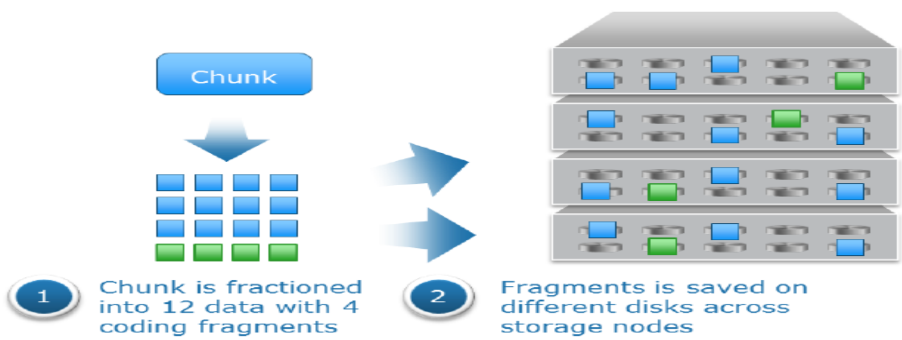


Figure 7 – Mirrored Chunks Replaced by Erasure Coding

ECS tolerates failures based on available number of nodes.

ECS Sizing

There is few design principles involved with sizing an ECS solution. We can follow these generic procedures that can be applied to any situations but there may be few exceptions with specific requirements

Preliminary requirement for sizing any solution is to understand the existing environment. We need to understand existing environment by asking questions and by extracting Metrics/information that is required to model the target solution. In the absence of data we can choose to model a solution with default assumptions with customer's consent.

Data Requirements for Sizing

Metrics/Information required for ECS solution modeling includes:

1) Capacity Requirement
2) Protection overhead
3) Object size
4) Thread Count
5) Growth requirements

1) **Capacity Requirement**
   Capacity of the existing environment can be extracted through various corporate tools. For most of storage products, **Mitrend scanner** can be run to get existing configuration information from which capacity details can be extracted. Capacity has to be consolidated from all servers and storage that are identified as candidates to be replaced by ECS.

   If you are doing Centera-to-ECS conversion, then Centera capacity can be obtained from Health Check Report.

2) **Protection Overhead**
   ECS achieves storage efficiency with data protection by using Erasure coding. As explained earlier, Storage Engine with 12/4 or 10/2 erasure coding scheme. The storage engine reconstructs an object from any of the 12/10 fragments. ECS requires a minimum of 4 nodes in a single site to achieve protection overhead of 1.33 with Active Archive with minimum of 4 nodes and Protection overhead of 1.2 can be achieved with Cold Archive with minimum of 8 nodes.

   Based on number of nodes, failures can tolerate Storage overhead as follows:

Number of Sites in Replication Group	Default Use Case (Erasure Code: 12+4)	Cold Archive Use Case (Erasure Code: 10+2)
1	1.33	1.2
2	2.67	2.4
3	2.00	1.8
4	1.77	1.6
5	1.67	1.5
6	1.60	1.44
7	1.55	1.40
8	1.52	1.37

Table 8 – Protection overhead based on Erasure Code

3) **Object Size**

Consider default object size of 128KB for any ECS Appliance modeling. If you have to consider existing object size for target modeling, Total Usable capacity/Object count should give you the average object size.

For any object storage-to-ECS conversions, average object size in the existing array can be obtained from the configuration information extracted from the health report.

For Centera, the formulae below can be used to calculate average Object count.

Object Size = Total Usable capacity *1024*1024*1024/Used Clip count

4) **Thread Count**

The number of users connected to the storage system requesting for writes and reads concurrently consume threads. The greater the size of the object, the fewer number of threads available. As the object size becomes larger – around >3MB – the number of threads available will get reduced.

As a rough estimate for sizing purposes, we use 150 threads per Node for end user workload. A full rack ECS solution will support 1200 threads (8nodes*150) regardless of node type. Based on the estimated load if the requirement is for 1800 threads, then 4 nodes have to be added to meet total thread count requirement.

5) **Growth Requirements**

Growth requirements should be understood based on the growth pattern monitored and recorded in the existing storage or this information can be collected from the customer. Adding buffer to meet growth requirements will avoid under sizing the solution. While extensive growth of data beyond predictions can be addressed through regular upgrades, the model for the solution should always be determined by considering additional headroom for future growth.

Appliance Sizing Best Practices

Best practices to be followed while modeling ECS.

Latest generation appliance should be considered for any new ECS solution modeling.

It is always wise to model ECS across two sites or at least 2 ECS solution where overall capacity will be distributed. This would ensure data protection against device or site failures.

A Storage pool must have a minimum of 4 nodes and at least 3 nodes should have more than 10% of headroom to allow incoming writes.

It is recommended that Load Balancer is used in front of ECS. In addition to segregating the load across nodes equally, a Load Balancer also provides high availability by channeling the traffic to healthy nodes. Configuration of a Load Balanced is based on type load balancer considered.

In addition to these, a few Generic steps involved in sizing ECS in a single VDC (Virtual Data Center) or cluster are illustrated below,

1) Calculate Overall Capacity
2) Derive number of spindles for overall capacity
3) Calculate Number of nodes based on spindle count
4) Validate determined number of nodes against performance
5) Redistribute spindles over adjusted Node count

1) Calculate Overall Capacity

Calculating overall capacity is the preliminary metric while modeling for ECS. There is other information to include while calculating overall capacity for the solution. This information should be collected from customer data and also asking related questions to the customer. A few of them are:

a) Actual data as determined by object size
b) Year-on-year growth
c) Retention Period
d) Protection overhead
e) Multisite requirement
f) Replication policies

Capacity can be extracted through Health Check report for Centera and Mitrend Scanner Report for other storage products. However without these reports we can still calculate capacity based on Object count and Object size using the formulae below.

Total Used Capacity = (Object Count * Object Size (n KB)) + (Object Count*Object Size (m KB))*Protection Overhead/1024*1024

2) **Derive number of spindles for the determined overall capacity based on the calculated overall capacity**
   We consider latest generation for target modeling, as of today all Gen 2 nodes carry 8TB Drives. Formatted capacity of 8TB Drive is 7.28 TB. Using the below formulae total number of spindles can be calculated to address required capacity

Number of spindles required = Total required capacity *Protection overhead* $(1+n)^m$ / Formatted Capacity of a Drive
n = growth factor in %
m – Number of year's growth has to be factored for

3) **Calculate Number of nodes for the spindle count**
   Node count can be calculated based on total drives equally distributed across all nodes. This procedure makes the selection of nodes difficult. As of today, we can distribute drives across nodes with half rack or full rack or over multiple racks.

   First rule is to distribute drives in groups of multiples of 10 in each node. For example if there are 64 drives required we will round it up to 70 drives and if each node has 10 drives, we will end up with 7 Nodes.

   All nodes can be expanded by ordering HDD drives expansions sets of 80 drives. Supported nodes in ECS can be scaled dynamically and disks can be added disruptively through HDD expansion kit. Overall ECS solution should be configured such a way that a single node failure should not eliminate more than 25% of available usable storage.

4) **Validate determined number of nodes against performance**
   In some cases, the performance component has to be considered for ECS Sizing. Metrics to consider to size for performance will be:
   a) Thread count
   b) Bandwidth
   c) Get/put Operational ratio
   d) Concurrency

   This can be validated through **ECS Performance Projection** tool for different values of each metric. The tool is available at **https://ecs-sizer.cfcp.isus.emc.com/**.
   Through this we can decide model and the final node count.

5) **Redistribute spindles over adjusted Node count**
   Once the count of nodes is determined based on the performance requirement, the total number of drives should be again distributed across nodes equally in a group of multiples of 10 in each node.

Conclusion

Elastic Cloud Storage is configurable technology and can address growing needs of archiving and retaining the data. Understanding customer environment and requirements and following standard methodology of sizing ECS Appliance would help Dell EMC Engineers position ECS Object Storage solution to best meet customer need.

Appendix

References:

1) ECS Architecture Guide.pdf
2) ECS Hardware and Cabling.pdf
3) ECS Storage Planning Guide.pdf
4) ecs-rfp-recommendations.pdf
5) Centera to ECS Migration
   http://www.emc.com/collateral/white-papers/h15115-centera-to-ecs-migration.pdf
6) h14071-ecs-architectural-guide-wp.pdf