Best Practices

Dell EMC ECS: General Best Practices

Abstract

This document provides general best practices for the deployment, configuration, and use of the Dell EMC™ ECS™ platform.

March 2019

Revisions

Date	Description
March 2017	Initial release
January 2018	Minor updates to the Operations section
October 2018	Add link to new KEMP load balancer paper.
March 2019	Updated for ECS 3.3 release

Acknowledgements

This paper was produced by the Dell EMC Unstructured Technical Marketing Engineering and Solution Architects team. Please send comments, suggestions, or feedback to unstructured.tme.sa@emc.com.

Table of contents

Executive summary

Dell EMC™ ECS™ is a software-defined, cloud-scale, storage platform for traditional, archival, and next-generation workloads. It provides geo-distributed and multi-protocol (Object, HDFS, and NFS) access to data. With ECS, any organization can deliver scalable and simple public cloud services with the reliability and control of a private-cloud infrastructure.

This document provides general best practices for the deployment, configuration, and use of ECS.

1 Introduction

This document highlights general ECS best practices relating to physical deployment, networking requirements for external infrastructure services, provisioning, and application development when utilizing ECS APIs. It describes some of the common pitfalls associated with deployment and provisioning, and lists best practices to mitigate them.

1.1 Audience

This document is primarily intended for operations personnel such as storage administrators responsible for designing, deploying, and managing ECS. Application developers may also find the paper useful.

1.2 Scope

This document is intended to supplement and highlight some of the content in current ECS product documentation. Hence, this document does not cover installation, administration, and upgrade procedures for ECS. It is assumed that the reader already has an understanding and working knowledge of ECS and has familiarized themselves with available documentation for ECS. References to other documentation for further reading are provided.

2 Architecture overview

ECS is a strongly-consistent, indexed, object storage platform. It is a scalable solution providing secure multi-tenancy; and superior performance for both small and large objects. ECS was built as a completely distributed system following cloud principles. The ECS software running on commodity nodes forms the underlying cloud storage, providing protection, geo replication, and data access. The software was built with six design principles in mind:

- Layered services for horizontal scalability.
- Both the index and data use the same underlying storage mechanism.
- Good small and large object performance.
- Multiple protocol access — Object, HDFS, and File.
- Geo replication with lower storage and Wide Area Network (WAN) overhead.
- Global access — read and writes access from any site within a replication group.

Figure 1 illustrates the different layers of ECS. For additional information, please review the ECS Overview and Architecture white paper.

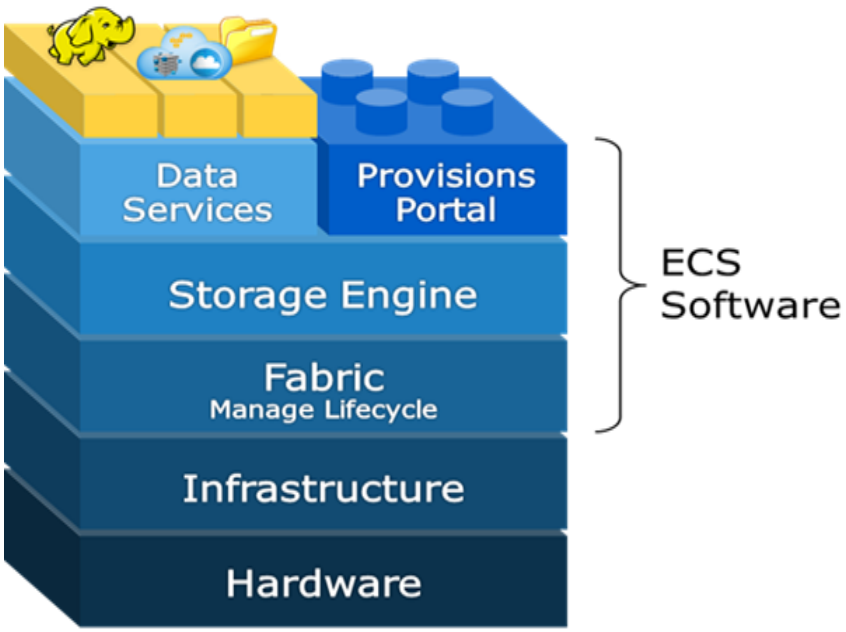

Figure 1 ECS layers

3 Physical deployment

Strategic planning is critical to the success of any ECS deployment. Some of the factors to consider during physical deployment relates to the following:

- Space and power
- Networking
- Single-site and Multi-site considerations

Working closely with Dell EMC personnel, reading thru the documentation, and utilizing tools available for planning are important in designing ECS.

3.1 Planning documentation and tools

Making assumptions relating to power, space, and infrastructure services such as firewall/network, ACL, DNS, NTP, etc. is a common pitfall and poses challenges for ECS installation. Thus, knowledge of requirements and existing infrastructure at customer site is important to mitigate this issue. There are documentation and tools available to help plan, prepare and design ECS to fit your requirements and eliminate the guess work.

Just to review, the following components illustrated in Figure 2 form the basis of an ECS deployment:

- **Site:** A unique physical location, for example, a data center in Arizona, USA. An ECS deployment consists of one or more sites.
- **Site ID:** Dell EMC assigns a unique identifier to each site. All hardware, software, and services are tied to individual site IDs.
- **Rack:** A rack consists of hardware that is physically located in a single data center floor tile space.
- **Node:** A node is basically a server in a rack. Racks generally consist of five or more nodes.
- **Cluster:** One or more racks of hardware physically connected at a single site. In general, each site has one cluster that is made up of one or more racks of hardware and federation is done between at most one cluster at each site. That is, it is possible to have two clusters at a single site, but, ECS is designed to federate geographically not locally. A cluster is also referred to as a Virtual Data Center (VDC).

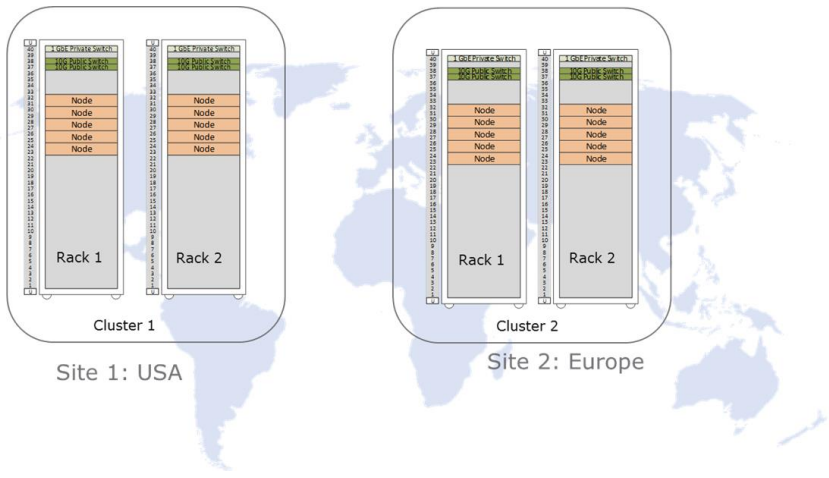

Figure 2 Physical ECS deployment

A VDC/site is built up of one or more racks where each rack requires a tile space on the data center floor. Racks communicate across the site's Local Area Network (LAN) via uplink network connections through a pair of 25GB switches. These switches may be purchased with the ECS as part of the solution, or may also be customer provided switches. In addition, ECS communicates privately over a closed backend for administrative tasks; no data travels over the backend network. The quantity of racks deployed at each site is primarily determined by storage and performance requirements. Floor space and plans for future growth are also considerations.

A multi-site deployment is built by federating two or more sites. ECS enables you to configure replication either within a single site or across multiple sites. This provides flexibility in solution design allowing for data segregation, protection against many types of failures, and global access.

After understanding the terminology and components, there are documentation and tools that can assist in the planning and deployment which include:

- Planning Guide: Along with a general overview of ECS and ECS data protection, the section on planning an ECS installation contains an ECS readiness checklist for infrastructure components and requirements that is essential.
- Site Preparation Guide: Regardless of whether an ECS appliance or customer rack is used, this document contains critical must-know information such as requirements for site floor load-bearing and power.
- Security Configuration Guide: A guide that provides an overview of settings, and configurations for secure operation.
- ECS Designer (Available internally and via Dell EMC Sales): An excel spreadsheet available to record and centralize required information.

Regarding the ECS Designer, all hardware and software and licensing are associated with a specific and unique site ID. It is critical site information be kept up-to-date and verified for accuracy from the earliest planning stages, through the ordering process, and all the way through provisioning, alerting, and remote access. Support issues are tied to site IDs as well.

Planning documentation and tools best practices:

- Make no assumptions; understand all requirements and existing infrastructure.
- Carefully review the planning and site preparation guides.
- Obtain and utilize the ECS Designer.
- Validate Site ID information is accurate and that all hardware, software, and licenses are associated with sites properly. Verify license for encryption is ordered correctly and received for each site.
- Account for growth and retention requirements when planning.
- Design deployment based on your High Availability and Disaster Recovery requirements.

3.2 Power and space

Power and space are important considerations when planning an install. Under specifying the power requirements can cause overload and overheating issues. Another example would be to not consider the total weight of the rack. A fully loaded eight node ECS appliance weighs over a ton. Due to the density of ECS hardware, ECS may have unique requirements such as custom rack size, depth, cable management and brackets which some locations may not be generally equipped with. Knowledge of the power and space requirements assists in alleviating issues and plan for future growth.

The documentation and tools referenced in previous section must be leveraged to make installation location(s) within the datacenter compatible with requirements. Adhering to the requirements outlined in the documentation assist facilities in supporting ECS. Best practices related to power and space are described as follows.

Power and space best practices:

- Customers who purchase ECS appliances but move the hardware to their own rack should plan for the disposal of the cabinet purchased with the appliance.
- When expanding ECS clusters, purchase nodes for existing racks to consolidate space, and purchase racks to allow for future consolidation.
- Consider reserving additional tiles for cluster growth.
- Allow extra time when purchasing hardware outside of a rack as the switches and nodes do not come preinstalled with operating systems and require additional inspection.
- Consult the most recent hardware specifications guide when ordering hardware for power requirements, dimensions and weight.

3.3 Networking

Three primary categories of switches (illustrated in Figure 3) and their inter-connectivity need consideration during deployment:

- **Customer Network**: ECS attaches to customer network allowing for multiprotocol access to data stored on ECS by applications and users over standard protocols including S3, Swift, and NFS.
- **ECS FE network**: Production traffic, UI and API management, replication, and data. 25 GbE switches are deployed as a pair in each rack and configured for high availability to allow for sub-second recovery time and for performance using Link Aggregation Control Protocol (LACP). Two optional Dell EMC S5148F 25 GbE 1U ethernet switches can be obtained for network connection or the customer can provide their own 25 GbE HA pair for the front-end pair. The two switches must be configured with LACP/MLAG to create a single LAG interface. Configuration should also include eight customer uplinks available per 25 GbE switch, for a total of 16 uplinks per rack. Each node in the rack is connected to both data switches via its two 10/25 GbE network interface cards (NICs) and are aggregated together using the Linux bonding driver. The node is configured to bond the two 25 GbE NICs into a single LACP bonded interface.

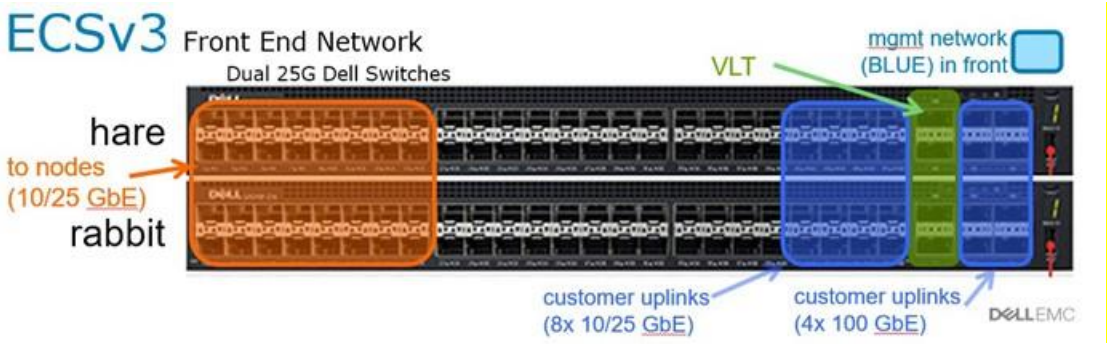

Figure 3 Front-end network switching in ECS deployment

- **ECS provided 25 GbE**: Remote Monitoring and Management (RMM), low-level node-to-node interconnectivity, and general admin traffic. ECS refers to the node-to-node network as the Nile Area Network (NAN).

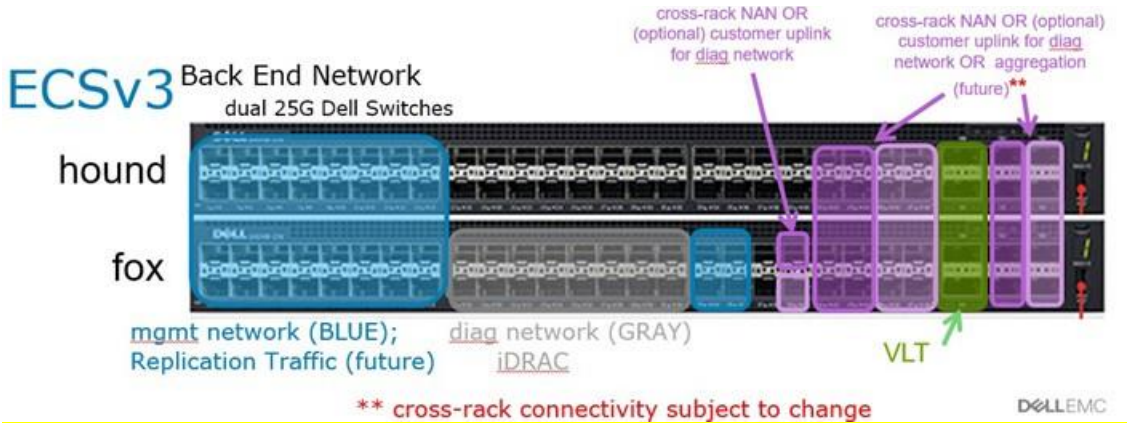

Figure 4 Back-end network switching in ECS deployment

The following documents should be consulted for ease in switch, switch port, and overall network planning:

- **ECS Designer** (Available via Dell EMC Sales): Absolutely critical document in the design and provisioning process, especially around switches and their related configuration, and guides users through important questions.
- ECS Hardware and Cabling Guide: Provides information on supported hardware configurations, upgrade paths, and rack cabling requirements.
- ECS Networking and Best Practices: A white paper that describes details of ECS networking and specifics on ECS network hardware, network configurations, and network separation.

Networking best practices:

- Use the ECS Designer throughout the design and deployment process. Record customer provided switch manufacturers, models, and firmware versions.
- Record ECS rack uplink information along with switch and port identifiers and cabling descriptions.
- Reserve the necessary number of ports on the customer's switch infrastructure.
- Understand the options for port channel configuration.
- Refer to the ECS Networking and Best Practices white paper.

4 Customer-provided infrastructure

An ECS deployment depends upon certain customer provided infrastructure requirements need to be reachable by the ECS system as shown in Figure 5. A list of required and optional components includes:

- **DNS Server**: Domain Name server or forwarder.
- **NTP Server:** Network Time Protocol server.
- **DHCP server**: Only required if assigning IP addresses via DHCP.
- **Authentication Providers**: Users (system admin, namespace admin and object users) can be authenticated using Active Directory or LDAP or Keystone.
- **SMTP Server**: (Optional) Simple Mail Transfer Protocol Server is used for sending reports from the ECS rack.
- **Load Balancer**: (Optional but highly recommended) evenly distributes loads across all nodes.

Best practices associated with these external services are described in this section.

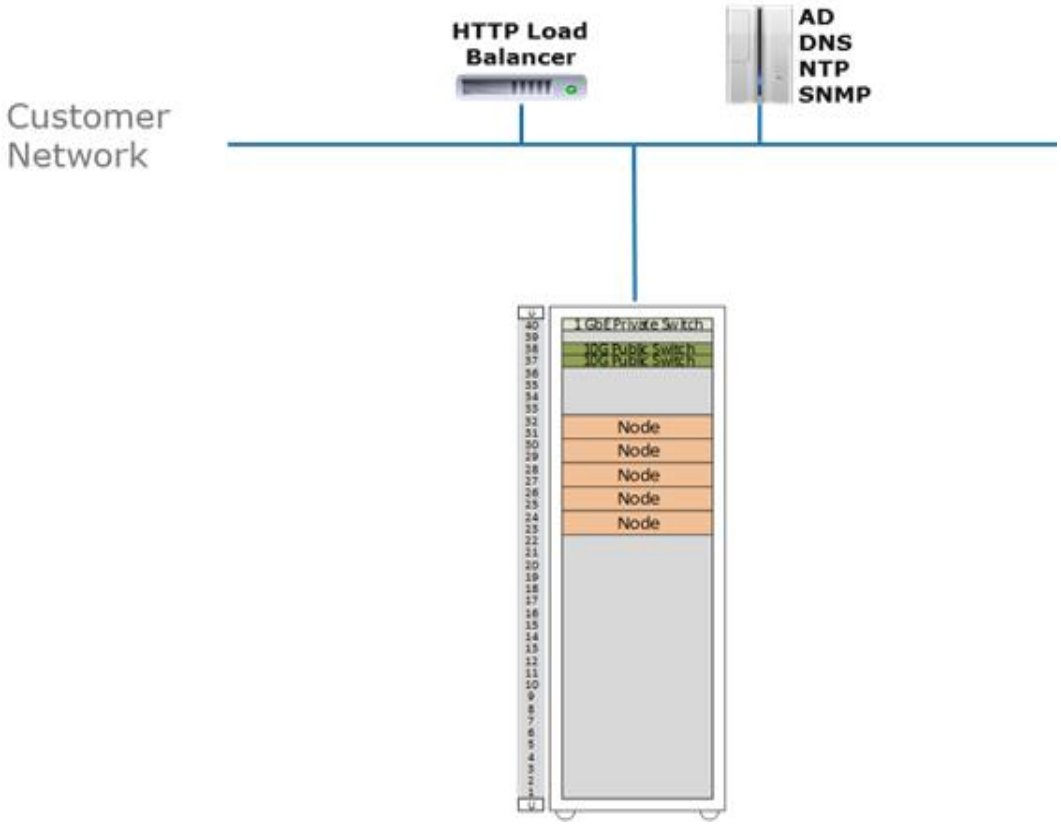

Figure 5 Customer-provided infrastructure

4.1 Domain Name System (DNS)

Each node in an ECS cluster requires both forward and reverse DNS entries as well as access to one or more domain name servers. There is potential for each workflow to require unique DNS entries (and IP and load balancer configuration). DNS administrators should be given ample time to meet with all necessary application and workflow engineers so that the naming requirements can be fully understood and deployed correctly.

DNS best practices:

- Use a minimum of two DNS servers for redundancy.
- Planning should include a record of site DNS server IP addresses, server names, and relevant search domains per site.
- Obtaining domain names and associating them with IP addresses can often take longer than expected. Be sure to engage all relevant groups as early in the design phase as possible.
- Work directly with application developers and workflow owners so that all required domain names can be obtained, properly recorded, and configured when needed.

4.2 Network Time Protocol (NTP)

Network Time Protocol (NTP) accessibility is essential for ECS to operate correctly. Precise time is necessary for consistent ECS for clock synchronization between nodes in ECS which insures clean log and journal entries for chunk timestamp values. Unstable NTP sources can result in data corruption, or excessive system journal activity. Multi-site ECS deployments should use common sources. Include NTP server IP addresses and names for each site in planning documents. Refer to industry NTP best practices for more information.

NTP best practices:

- Use either one or four NTP servers. Utilizing any number in between, like two or three may cause issues

4.3 IP addressing and Dynamic Host Configuration Protocol (DHCP)

At a minimum, one customer-provided IP address is required for each node. Use of local and/or global load balancers require additional IP addresses. If hosts will retrieve IP addresses from a DHCP server, record DHCP server IP addresses and names. In addition, if traffic separation is used more IP addresses may need to be reserved. Sufficient IP addresses or subnets need to be identified and reserved for deployment. If a separate layer 3 network team exists, planners should reach out to them early during the planning and design phase. They are instrumental in deciding which deployment model work best for each site and allocating and reserving of IP addresses or subnets.

DHCP is utilized for assigning IP addresses. Many customers choose to use static IP addresses, often with reservations in DHCP. For large scale-out environments however DHCP could be leveraged to avoid hard-coding a large number of addresses.

Putting DHCP in a DMZ is an often-common requirement for cloud-based storage which may not be part of the traditional model. Begin this conversation early to give ample time for all involved to plan accordingly.

DHCP best practices:

- If using DHCP, MAC addresses should be persistent so that nodes get the same IP addresses during reboot.
- When two or more VDCs are federated, Network Address Translation (NAT) cannot be used within unnamed public, named replication, and named management networks.

4.4 Load balancing

Load balancers are highly recommended in ECS deployment to evenly distribute data loads across all service nodes. Although customers are responsible for implementing and configuring their deployed load balancers, Dell EMC does provide recommendations and suggestions on how to configure some of them with ECS workflows. Load balancing needs should be examined at the workflow level. Each workflow may justify or rule out the use of load balancers. Use of load balancing is important for Dell EMC Atmos™ traffic, generally recommended for S3, can be used with NFS. They are not required for CAS since CAS workflows have load balancing built in to the client applications.

Both local and global load balancers are recommended where work flows justify their need. In addition to distributing the load across ECS nodes, a load balancer provides High Availability (HA) for the ECS cluster by routing traffic to healthy nodes. For each workflow that utilizes a load balancer, each load balancer's IP address and Fully Qualified Domain Name (FQDN) should be recorded in planning documents.

For multi-site deployments consider when load balancing should be implemented to provide a method to balance writes across sites to take advantage of ECS's XOR data reduction capability. As an example, in a three-site deployment using an archive use case; if the application is performing writes in only one site, ECS will not take advantage of its XOR capabilities despite having all three VDCs in a replication group. This can be corrected by using a load balancer to redirect traffic and balance writes across sites.

Several white papers are available that provides references on how to implement a load balancer with ECS:

- ECS with HAProxy
- ECS with NGINX (OpenResty)
- ECS with F5
- ECS with KEMP

Load balancing best practices:

- Great care should be taken to configure and sized load balancers correctly such that they do not reduce or provide a bottleneck to performance. For the load balancer to not hinder peak throughput (based on PUT/GET), check that the maximum transaction rate and bandwidth can pass through the load balancer.
- Deploy redundant load balancers (as per manufacturer's instructions) to eliminate single points of failure.
- Only utilize DNS Round Robin if you cannot implement Global DNS / Load Balancing as it is a better approach.
- For best performance, terminate SSL connections at load balancer(s), passing traffic unencrypted to the ECS nodes. This offloads the encryption from ECS to the load balancer. NOTE: For workflows carrying Personally Identifiable Information (PII), do NOT terminate SSL at the LB. This is important to prevent clear text transmission of PII between routers and ECS nodes.
- If SSL termination is required on ECS nodes itself, then use Layer 4 (TCP) to pass through the SSL traffic to ECS nodes for handling. The certificates would need to be installed on the ECS nodes and not on the load balancer.
- For NFS traffic, use only the high available functionality of the load balancer.
- When federating three or more ECS sites, employ a global load balancing mechanism to distribute load across sites to take advantage of ECS XOR storage efficiency. It is also important to optimize the local object read hit rate in a global deployment.
- Enable web monitoring of traffic.

4.5 Authentication providers

Many customers use local ECS authentication for management users. The management users then define all object users, generally one per application. For customers that highly leverage AD and/or LDAP, groups or users are assigned to management roles, as opposed to local user accounts. Some things to note when utilizing authentication providers include:

- **Active Directory (AD)**: An AD domain group can only be the namespace admin for one namespace. Generally, storage administrators create an AD group for each namespace and assigned AD users to that group. Namespace users can use the Web UI and only see things pertaining to their namespace.
- **Lightweight Directory Access Protocol (LDAP)**: LDAP users can be administrative users in ECS. LDAP groups are not used in ECS.
- **Local**: Local management users are not replicated between sites.

4.6 Simple Network Management Protocol (SNMP)

SNMP servers, also known as SNMP Agents, are optional. SNMP provide data about network managed device status and statistics to SNMP Network Management Station clients. ECS supports SNMP basic queries and SNMP traps.

SNMP best practices:

- For each SNMP server that will be used with an ECS deployment, plan for their IP addresses, names, ports, version and type of SNMP service used, and community name.

4.7 Firewalls

Certain ports need to be open for ECS traffic. Firewalls rules would need to be modified to open up the ports required for ECS traffic.

Firewall best practices:

- When firewalls are in use, reference the latest version of the ECS Security Configuration Guide for a complete list of ports to open and define rules in your firewall accordingly.

5 Provisioning

Once the physical hardware is installed and deployed and the external services configured and available, the next step is to provision VDCs, namespaces, replication groups, users, buckets, etc. to provide data access to ECS storage platform. Let's review some of the terminology associated with the components that can be provisioned (also illustrated in Figure 6):

- **Virtual Data Center** (VDC): A geographical location defined as a single ECS deployment within a site. Multiple VDCs can be federated and managed as a unit.
- **Storage Pool**: A storage pool can be thought of as a subset of nodes and its associated storage belonging to a VDC. An ECS node can belong to only one storage pool; a storage pool can have any number of nodes, the minimum recommended being five. A storage pool can be used as a tool for physically separating data belonging to different applications.
- **Replication Group**: Replication groups define where storage pool content is protected and locations from which data can be read or written. Local replication groups protect objects within the same VDC against disk, node, and rack failures. Global replication groups span multiple VDCs and protect objects against disk, node, rack, and site failures.
- **Namespace**: A namespace is a logical construct and is conceptually the same as a "tenant." The key characteristic of a namespace is that users from one namespace generally cannot access objects belonging to another namespace. Namespaces can represent a department within an organization or a group within a department.
- **Buckets**: Buckets are containers for object data and are created in a namespace to give applications access to data stored within ECS. In S3, these containers are called "buckets" and this term has been adopted by ECS. In Atmos, the equivalent of a bucket is a "subtenant," in Swift, the equivalent of a bucket is a "container," and for CAS, a bucket is a "CAS pool." Buckets are global resources in ECS. Where the replication group spans multiple sites, a bucket is similarly replicated across sites.

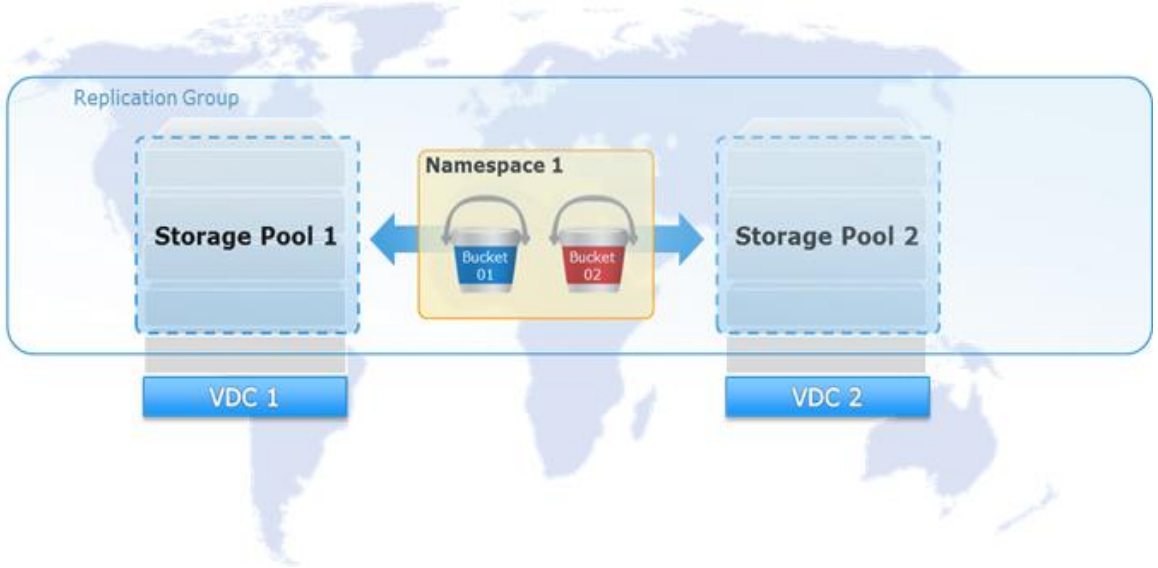

Figure 6 ECS components

There are several best practices and considerations when provisioning ECS which this section highlights. When provisioning, there are certain things that can only be set during creation time and once set cannot be modified. Table 1 provides a list of items that need to be decided prior to provisioning since they cannot be changed once set. The details in this table are re-iterated in each of the sub-sections below where applicable.

Table 1 ECS settings (uneditable once set during creation)

Level setting	Sub-level setting	Settings	Default
Storage Pool	Erasure Coding	10+2, 12+4	12+4
Replication Group	Replicate to All Sites	Disabled, Enabled	Disabled
Namespace	Name	User defined	N/A
	Server-side Encryption	Disabled, Enabled	Disabled
	Compliance	Disabled, Enabled	Disabled
Buckets	Name	User defined	N/A
	Namespace	User defined, associated with bucket	N/A
	Replication Group	User defined, associated with bucket	N/A
	Server-side Encryption	Disabled, Enabled	Disabled if not set in namespace level Enabled if set in namespace level
	File System	Disabled, Enabled	Disabled
	CAS	Disabled, Enabled	Disabled
	Metadata Search	Disabled, Enabled	Disabled

5.1 Naming conventions

Defining proper names for components is sometimes overlooked when provisioning and maybe problematic in some cases and at most inconvenient to change once set. Use DNS appropriate naming conventions for all ECS constructs - hosts, clusters, VDCs, storage pools, replication groups, namespaces and buckets. While some constructs may allow additional characters, such as underscore, limiting characters to those that are acceptable to DNS eliminates potential application-related conflicts that may arise when valid namespace or bucket names are in use that do not translate DNS. Use only the following characters:

- Lower case letters (a-z). Do not use upper case letters.
- Numbers (0-9).
- Hyphens. Avoid the use of underscores.

Naming conventions best practices:

- Use DNS appropriate naming conventions.
- Do not use personal or confidential information as names.

5.2 Storage pool

The first step in provisioning a site is creating a storage pool and assigning it nodes. Storage pools are logical constructs that contain physical nodes. They provide a means to physical separate data on a cluster, if required. Erasure coding (EC) is configured at the storage pool level during pool creation. The two EC options on ECS are 12+4 or 10+2 (aka cold storage). EC cannot be changed once created.

All cluster nodes can belong to a single storage pool. Implement the minimum number of storage pools required at each VDC. Storage pools along with their associated replication groups are integral in ECS indexing so keeping them to a minimum required minimizes unnecessary overhead.

Currently there are only two reasons to create additional storage pools within a VDC:

- Erasure coding is done at the storage pool level. Generally, only a maximum of two pools are required when both 12+4 and 10+2 EC is used.
- Physical separation of data. If data must be physically separated between nodes additional storage pools are required. Again, it is important to keep the number of storage pools to a minimum.

A storage pool must have a minimum of five nodes and must have three or more nodes with more than 10% free space data/object writes to continue. System metadata, user data and user metadata all coexist on the same disk infrastructure. Space is reserved so that ECS does not run out of space while persisting system metadata. Storage pool space considerations are also important when sites are replicated. Multi-site environments require sufficient space available to handle temporary and permanent site failures. When adding additional storage capacity to a site, expand other sites as needed to accommodate space requirements.

Storage pool best practices:

- Size storage pools to account for minimum free space needed to allow for writes.
- For multi-site, account the space needed in case of temporary site outage and permanent site removal.

5.3 Virtual Data Center (VDC)

A VDC identifies the nodes that are participating in an ECS instance. The first VDC must contain the nodes from the local ECS instance. Additional VDCs can then be configured identifying all the nodes in that remote ECS instance. Adding remote VDCs to a local ECS instance creates the federation of ECS instances. To create a replication group that includes storage pools from a remote VDC, that remote VDC must be federated with the local VDC.

Generally, a physical site has one VDC. Some organizations have multiple VDCs per site, for example, one for engineering and one for operations; and can be federated together for ease of management. However, it is not recommended to create replication groups consisting of VDCs that are all in one local site to make use of the ECS XOR feature for storage efficiency. This is not recommended primarily because in this scenario when a site is down, more than one VDC becomes unavailable.

Virtual Data Center best practices:

- Plan for redundancy and availability. Replication to VDC in different geographic locations increases data availability.
- VDC names must be unique.
- VDC names cannot be reused.

5.4 Replication group

Replication groups allows grouping of storage pools from different geographically located VDCs for replication of data between sites. Replication of data across sites has the following advantages:

- In case of site failure, data is accessible from surviving site(s) within the replication group.
- For three or more sites, ECS XOR feature provides better storage efficiency.

Similar to storage pools, the minimum number of replication groups should be created. This is because of the indexing overhead associated storage pool/replication group pairs. There is no reason to have two or more replication groups that do the same thing. That is, for example, two replication groups containing the same set of VDC storage pools are of no value and add additional unnecessary overhead.

The standard scenario is one replication group for local data (non-replicated), and one for replicated data that spans all VDCs. Organizations with more than two sites may consider more replication groups for times when data should only be replicated to a subset of all sites. Generally, one replication which spans all sites is sufficient. Compliance may dictate additional replication groups be created, for example, where data privacy or sovereignty laws prohibit shared data across specific borders.

When three or more sites are in replication group efficiencies in storage overhead can be gained. ECS can XOR chunks written at two sites at a third site. It is important to understand that in order to gain these efficiencies, new writes must occur at two or more sites. To balance the efficiency across all sites in a replication group, all sites must have relatively similar write workload. This benefit may not be appropriate for all workloads especially in scenarios where WAN latency creates unacceptable bottlenecks. However, there are tradeoffs when spreading data across sites. For instance, there is an additional latency for WAN lookups of objects not local to the VDC. Geo-caching does alleviate some of this; however, this latency can pose some issues for applications if data is not in cache.

Replication group best practices:

- Limit the number of replication groups to reduce indexing overhead.
- Replication groups cannot be deleted so it is critical they are planned for correctly.
- For three or more sites, distribute write requests across sites to take advantage of XOR feature benefits. However, be aware of the latency tradeoffs for WAN lookups of objects not in local cache.
- Federate all VDCs prior to attempting to create a replication group.

5.5 Namespace

Namespace provides a way to organize or group items for purposes of segregating the space for different uses or purposes. It allows for the multi-tenancy feature in ECS. Unlike storage pools and replication groups, many namespaces can be created. Some environments may do well with a single namespace. Here are a few reasons to create a namespace:

- One per business unit.
- One per application.
- One per reporting boundary. **Note**: Buckets can also be reported on.
- One per subscriber. It may make sense to have a namespace for each subscriber like for Internet Service Providers for example. This is how Dell EMC ECS Test Drive is configured. That is, a unique namespace is created for each user.
- As a workaround, namespace can be used to allow targeting buckets in specific replication groups for legacy applications. Some legacy applications cannot access a specific storage pool so it may be

necessary for these applications to use buckets that access storage pools via specific replication groups.

Namespace best practices:

- In a multi-tenant environment, create a namespace administrator for configuration based on tenant requirements.
- Understand the settings in the namespace that are set during creation time cannot be modified (refer to Table 10 above) once set and plan accordingly.
- For best performance, recommended to have less than 1000 buckets in namespace.

5.6 Bucket

Buckets are containers for object data. Buckets are created in a namespace to give applications access to data stored within ECS. Buckets are global resources in ECS. Where the replication group spans multiple sites, a bucket is similarly replicated across sites.

Bucket best practices:

- Use buckets for specific environment, workflow, or uses. For instance: **dev, test, finance, operations**, etc.
- In multi-site deployments, create buckets at the VDC site closest to the application accessing and updating the objects. There is overhead involved with checking the latest copy if the ownership of object is at a remote site.
- Bucket names must be unique within a namespace. Naming convention for buckets as mentioned should be followed preserving DNS best practices.

5.7 Users and roles

Users of ECS can be management users, object users, or both as pictured in Figure 7. Management users have access to the ECS via its web portal and thru the management API. Object users have access to the ECS object interfaces for S3, OpenStack Swift, Atmos, Dell EMC Centera™ CAS, HDFS, and NFS. An object user uses a native object interface (e.g., the standard S3 API) to perform data access operations such as reading, writing, or listing objects in buckets. They can also create or delete buckets. If a user requires access to both the ECS portal and the ECS object interfaces, that user must be created as both a management user and an object user. ECS does not know, for example, that a single individual named Bob is both a management user and also an object user. To ECS, management and object users are not correlated.

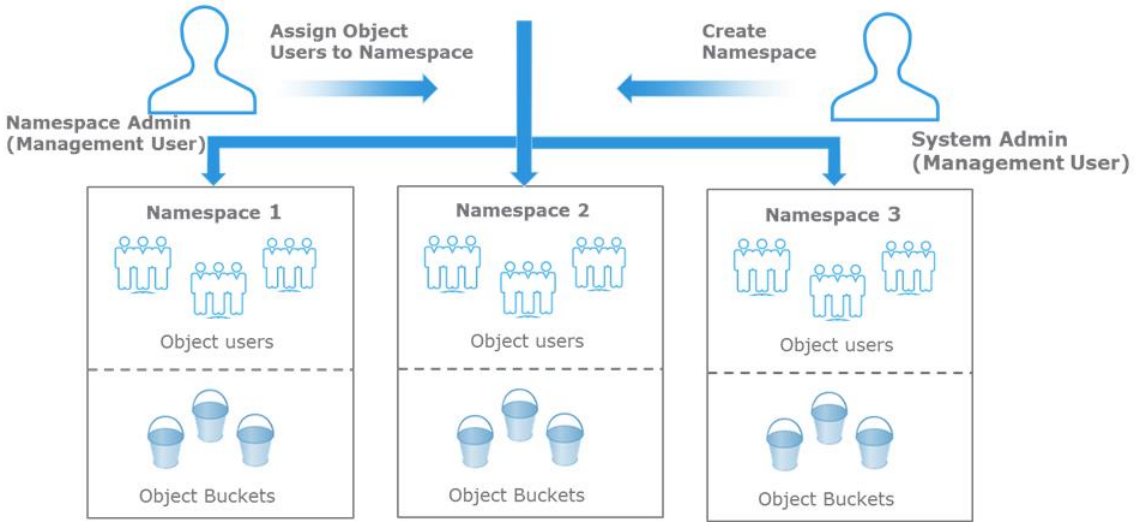

Figure 7 Types of ECS users

Users and roles best practices:

- When there are is a large group of users to be given access to the object store, leverage existing AD/LDAP infrastructure.
- A common pitfall to make names unique and consistent with AD names is to create local accounts using a domain-style. This implies that authentication is performed by AD or LDAP. However, in ECS authentication is done using secret keys. So, do not use domain-style names as local accounts that are not part of any domain to avoid confusion.
- Management users, whether local or domain based, are not replicated across geo-federated VDCs. This means all admin except Namespace admin must be created at each VDC that requires the account/role. Domain-based namespace admin accounts are excluded in this caveat because namespaces are global constructs and as such their associated admin are also global.
- Local management accounts are not replicated across sites, so a local user who is a Namespace Admin can only log in at the VDC at which the management user account was created. If you want the same username to exist at another VDC, the user must be created at the other VDC. As they are different accounts, changes to a same-named account at one VDC, such as a password change, are not propagated to the account with the same name at the other VDC.
- Namespace Admin can only be the administrator of a single namespace.
- The user scope setting must be made before the first object user is created. That is, once first object user is created in a VDC, the user scope setting cannot be changed. The default user scope setting is GLOBAL. If you intend to use ECS in a multi-tenant configuration and you want to ensure that tenants are not prevented from using names that are in use in another namespace, you should change this default configuration to NAMESPACE.

6 Security

In addition to assigning specific roles for certain access and control for users for security, additional measures must be taken to make ECS less vulnerable to unwarranted access, common user mistakes or security data breach. ECS provides several features to enable security of customer's data such as encryption, platform lockdown, retention, etc. Features available and best practices in protecting ECS are described in this section.

6.1 Protection from unwarranted access

ECS has features to protect against unwarranted access that include:

- **Platform Lockdown**: Disable SSH access to nodes.
- **Retention Policies**: Limiting the ability to change records or data under retention using retention policies, time-period and rules.
- **Audit Events**: Records change in the system configuration, tracks logins, and sudo commands run on node, bucket operations such as setting bucket permissions, and user operations such as set/delete password.

Protection from unwarranted access best practices:

- Immediately change the ECS default account password for admin on nodes and for root on ECS portal.
- Use individual user accounts for day-to-day administration as opposed to the ECS built-in account.
- Use the "Platform Lockdown" feature if there is requirement that ECS nodes should not be accessible via SSH.
- Set appropriate retention for objects to protect from accidental deletions.
- Use SSL for additional security.
- Monitor "unauthorized" access and modifications through audit events.

6.2 Data at Rest Encryption (D@RE)

ECS provides server-side encryption to protect data on disk. Key management is either done automatically or specified by user. Enabling encryption is done at the namespace level or bucket level, allowing customers to have level of control at what level to handle encryption. If in the namespace level, all buckets within namespace are encrypted unless at bucket creation time it is specifically disabled. If not enabled at namespace level, buckets can enable encryption individually at create time.

D@RE best practices:

- Be aware of the performance impact to workflows when using encryption.
- Avoid double encryption scenarios. For example, if encryption is in place with the use of Isilon Cloud Pools, don't use encryption on ECS as well.

7 Application development

ECS provides a set of REST APIs for customers to utilize for data access and management of ECS through their applications. There are few best practices and considerations when developing or customizing an application for ECS. These are highlighted in this section in categories as it relates to namespaces and buckets, objects, retention, extensions, security and data management.

ECS was designed predominately for archival, content repository, Internet of things, video surveillance, and modern applications. Thus, some things to consider when designing an application for ECS include:

- ECS was designed mainly for applications or use cases that don't require high IOPS or low latency, so expect response times > 100ms.
- ECS has a 99.99% success rate for transactions so handle failures accordingly by either utilizing the built-in retry mechanism in most software development kits (SDKs) or creating appropriate error handlers.
- Use an SDK for your programming language. No need to reinvent the wheel.
- Use ECS S3 if you want to take advantage of ECS features.
- Use AWS SDK if you want to maintain compatibility with AWS.
- Use the protocol that best fits your needs and skills, S3, Swift, or Atmos.

7.1 Namespaces and buckets

As mentioned previously, following proper DNS naming conventions for buckets and namespaces are important for compatibility and accessibility. These best practices are important to re-iterate when developing applications for ECS:

- Bucket names must be unique within a namespace.
- Namespace and bucket names should be DNS compatible since they can appear in a DNS record. For instance, no underscores. Use lower case letters a-z, numbers 0-9, and hyphens.
- Do not use personal or confidential information as names since users could probe existence due to error codes (e.g., 409 Conflict).
- Assign a prefix to each group with access to the system and use it on your highest level of granularity e.g., namespaces, buckets, and objects to avoid name conflicts and ease of management.
- Run regular admin level scans to identify non-conforming namespaces, buckets, or objects.

Although some drivers and applications emulate a directory structure (prefix+delimiter), here are some things to consider relating to "flat" structure layout of ECS object storage platform:

- S3 has no support for directory structures.
- Because of its flat structure nature, for best performance, it is recommended to have less than 1000 buckets in a namespace.
- Avoid overly repetitive and costly bucket listing operations (enumeration of all objects within a bucket). A bucket stores all objects "flat" or equal without hierarchy, so very little changes between bucket listings when new objects are added.
- Avoid listing all buckets in a namespace since this operation will fail during an ECS temporary site outage.
- Paginate bucket listings.
- Use separate buckets for your environments instead of lumping all objects in one bucket. For instance, create buckets based on function or use like "dev," "test," "staging," "production," etc. This makes it easier for developers and development operations to manage.

7.2 Objects

Unlike traditional filesystems that have limits, ECS is designed to handle trillions of objects. This section provides useful tips when handling small and large objects within your application. It also provides some information on ECS versioning and compression details and options.

7.2.1 Small objects

An object that is less than 100 KB is considered small. While ECS has a write unit of 128MB, the standard size of all ECS chunks, the small object threshold is where the work associated with writing, journaling, and erasure coding the data begins to outweigh the combined gains of network speed for small payloads and multi-threaded write operations. To compensate, ECS has a special internal mechanism called box-carting which helps performance for data writes of small objects. Box-carting aggregates multiple small data objects queued in memory and then write them in a single disk operation, up to 2 MB of data. This improves performance by reducing the number of roundtrips down to disk to process individual writes to storage. Although ECS has optimizations for small writes, if there is an option in your application to define a size then choose a larger size (e.g., 1 MB rather than 64 KB) or a value that aligns to the ECS internal buffer size of 2 MB for performance.

7.2.2 Large objects

A general consideration for working with large object (>100MB) read/write operations is performance. ECS provides certain API features to reduce the impact on performance for large objects, such as, multipart uploads and byte range reads. Some tips to alleviate some of the issues for large object access include:

- When working with large objects (> 100 MB), utilize the multipart upload feature. This allows pause and resume uploads for large objects. Since ECS internal buffer size is 2 MB, for size < 1 GB, use multiple of 2 MB (e.g., 8 MB). Since ECS chunk size is 128 MB, for size > 1 GB, use 128 MB part size.
- Performance throughput can be improved by parallelizing uploads within your application.
- Use Byte Range reads for parallel downloads of large objects.
- Use APIs that allows for easy upload and download, for instance:
  - In Java, use the TransferManager.
  - In .NET, use TransferUtility.

7.3 Versioning

Inherent to ECS design is versioning. If S3 versioning is enabled and an older version of the data is needed, it can be retrieved or restored to a previous version using the S3 REST API. This is useful when an application requires rollbacks to previous versions or if a “Recycle Bin” feature is implemented in the application.

7.4 Compression

ECS has a basic built-in compression mechanism. ECS will determine if data is already compressed, and if so, it will not re-compress data. So, if a more sophisticated compression or specific compression is required for your objects, consider using client-side compression. ECS Java SDK supports ZIP and LZMA.

7.5 Temporary Site Outage (TSO)

In a multi-site ECS deployment, ECS offers an access during outage (ADO) feature that would allow access to data when there is a temporary disconnect or site outage between two sites or a failure of one site due to a power failure or natural disaster. If access is required by an application in case of temporary site outage, it is best to enable ADO when creating the bucket. However, there some things to consider when enabling ADO:

- FS buckets (NFS/HDFS) are read-only during TSO.
- During rejoin, conflict resolution favors secondary site, though it is non-deterministic.
- Listing of some buckets may fail during a TSO.
- If possible, use a Global Load Balancer to handle failover so that requests are automatically directed to available site in case of failure.

7.6 Traffic management

Communication to ECS is via HTTP/HTTPs, hence, it is best practice to keep in mind the back and forth traffic or how to mitigate traffic issues within your application. Some tips to reduce traffic impact include:

- Use pre-signed URLs. ECS supports pre-signed URLs to enable users to access objects without needing credentials.
- Object update frequency should be low since object storage platforms are not designed for transactional workloads but ideally for static content such sensor data, images, videos.
- Only one application should write to each bucket. Other applications may read from them, but not write.
- Use the object copy operation instead of downloading and uploading the object again.
- Beware of the concurrent requests for the same object.
- If order needs to be guaranteed, use Conditional PUTs (ECS extensions).
- If there is no external load balancing in your ECS deployment, implement client-side load balancing to distribute load across ECS nodes for increased performance.
- Use **Range Reads** for listing objects. Align the range to your application and request only what is needed. There are **Markers**, **NextMarker**, and **MaxKeys** parameters available to paginate listings.

7.7 ECS extensions

ECS APIs have support for additional extensions not available in the standard S3 APIs. These features extend ECS capabilities and provide an advantage over other solutions. Extensions relating to metadata search, byte range upload and retention and expiration are covered in this section.

7.7.1 Metadata search

ECS provides a facility for metadata search of objects to improve performance of queries. ECS maintains an index of the objects in a bucket, based on their associated metadata, allowing S3 object clients to search for objects within buckets based on the indexed metadata using a rich query language. Search indexes can be up to thirty system and user metadata fields per bucket and are configured at the time of bucket creation through the ECS Portal, ECS Management REST API, or S3 REST API. Some considerations when developing applications utilizing the metadata search capability include:

- Supported operations include **<, >, <=, >=, =, !=, AND/OR**.
- Metadata search must be enabled during bucket creation. Also, fields and values must be specified at bucket creation time.

- Performance is lower for accessing object on buckets configured for metadata search so use the feature wisely and after careful consideration. The more indexes created the larger the performance hit.

7.7.2 Byte range extensions

Unlike AWS S3 in which objects are immutable, ECS REST APIs provides byte range extensions to update and read parts of an object. Some features that ECS provides as part of this extension include:

- Partial reads and updates within an object (which still maintains append-only behavior).
- Overwrite part of an object: Overwrite by providing only the starting offset in the data request.
- Atomic append to an object: Ability to atomically append data to the object without specifying and offset and the offset is returned in the response. This is useful for multi-client streams, e.g., syslog, sensor data.

7.7.3 Retention and expiration

Retention means you cannot update or delete the object until retention period ends. There are three ways to assign retention:

- At the bucket level (compatible with generic S3).
- At the policy defined at namespace and assigned to objects (e.g., email = 5 years, documents = 3 years).
- Explicit retention period at an object level.

When retention is defined in multiple places, the longest time wins. Objects are automatically deleted when the expiration time is reached. Also, if object-level retention period is assigned at the application level, do not use ECS to assign a retention period greater than the application retention period. This may lead to application errors.

As an extension to general retention, ECS supports write-once-read-many (WORM) for data ingested via NFS protocol. When buckets are file-enabled, ECS can accommodate WORM access behavior by providing an auto-commit function on data written to the bucket. It is a bucket level setting and is only available from the ECS bucket controls. The setting allows the administrator to define a time-interval delay period after which files are converted to read-only.

NFS WORM best practices:

- The auto-commit period should be as short as possible to minimize the chance that the file is modified while waiting for conversion to write-only. A 24-hour maximum auto-commit period should be observed, less if possible.

7.8 Security

Security is important to safeguard your credentials and data being transmitted over the internet. Here are some tips relating to security:

- Use TLS (HTTPS) to transport your data.
- Validate your server's certificate; otherwise application will be vulnerable to man-in-the-middle attacks.
- Revoke access to unused applications.

- Store your tokens securely.
- Grant as few permissions as possible, for instance, no need to grant read-write permissions if your application only needs to read data.
- Use client-side encryption for maximum protection and keep your master key secure. If you lose your master key, you lose your data.

ECS supports the use of external key servers to store top level KEKs (key encrypting keys). Customers may take advantage of the additional layer of security provided by HSM based key protection, and latest encryption technology, provided by specialized key management servers. In addition, data stored on ECS is protected against loss of the entire appliance by storing top level key information outside of the appliance.

8 Operations

Maintaining the health of ECS requires the use of tools such as the ECS portal to monitor overall system-level health and performance information, syslog, and SNMP. This section provides best practices for day-to-day operations for ECS administrators. It includes subsections on monitoring, EMC Secure Remote Services (SRS – formerly ESRS) and product alerts and updates.

8.1 Monitoring

These four methods are primarily used for monitoring ECS:

- **ECS Portal**: The dashboard on ECS portal will provide the first view into health of system. From here, one can drill down to major issues using the other monitoring panes provided by the ECS portal. Situations to watch in the dashboard which indicate whether to investigate further include:
  - Nodes and disks with a red X on it or yellow caution marks. If you see any of these, go to the Nodes and Process health pane to determine which disk and node is not working and investigate further using this view.
  - Critical alerts. Examine the Events pane to determine critical alerts and if further handling of situations needs to be done.
  - Capacity and if capacity is reaching the limits. The Metering Pane indicates which namespace or buckets are utilizing capacity. View Capacity Utilization to check if more disks need to be added.
  - Performance data. Determine if performance is expected for workload using the historical view.
  - Geo Monitoring. Look at failover progress to validate that failover is as expected.
- **Audit logs**: Audit logs record of change in the system configuration. Things to watch out for in audits include changes unauthorized modifications such as owner or ACL changes, quota changes or creation and deletion of buckets and users.
- **Event Notifications**: Types of event notification in ECS include:
  - SNMP: Information about network managed device status and statistics to SNMP network management clients.
  - Syslog: Provides a method for centralized storage and retrieval of system log messages.
- **ECS Service logs**: ECS Service Logs provide further viewing and diagnosing. These logs are available on each node and are accessible via SSH by the system administrator user. These are output logs collected for each of the services running on node for instance, authsvc, blobsvc, eventsvc, etc. These logs are designed more for ECS experts and engineering to further probe and diagnose possible issues. The location of these service logs are in /opt/emc/caspian/fabric/agent/services/object/main/log.

There are situations to watch based on type of issue. Here are a few to look at via the ECS Portal:

- **Hardware health:** View the node states and if marked “suspect,” check on connectivity of node and if “bad,” determine which hardware is bad and check Node and Process health.
- **Node and process health**: Look out for extreme unevenness in the CPU, memory and NIC bandwidth. Check load balancer or network path or node health for imbalances.
  - If there are a lot of process restarts, check hardware health and node and process health views to look for further clues on what is occurring in the system.

- **Performance metrics**: The Monitoring section in the ECS Portal has a transactions page which contains a Performance page.
  - Keep an eye on the two, primary, metrics for performance on ECS, Transactions per Second (TPS) and Bandwidth, and the rates of each are broken out between read and write operations.
  - Read the latest ECS Performance white paper, available to customers under NDA, to become familiar with why TPS and bandwidth are the preferred measures of an object storage system. In short, TPS is the appropriate metric to measure small object performance and bandwidth is used to measure large object performance.
  - Latency data, although provided, is generally not considered a major factor in Restful API object workloads on ECS. It may best be used to look for trends over time.
  - Look for performance metrics that are outside of the norm. For instance, if the read bandwidth rate is not within the normal range, check for any alerts and look at the overall health of the system. Knowing which rates are expected require historical reference and an understanding of the normal workload occurring on the system.

- **Request failure rates**: The Monitoring section in the ECS Portal has a transactions page which contains a Requests page.
  - Check the requests page periodically for unreasonably high recent transaction failures. When found look at the error codes shown and check hardware health or node and process health.
  - Each VDC name in these pages is a link used to breakdown the metrics on a per node basis. If one or more nodes present with unusually lower rates, investigate further. Traffic may be being serviced by the cluster in an unbalanced manner which is less than ideal from a performance standpoint.

- **Alerts**: Critical alerts are available for hardware and software failures such as bad disks and nodes, fabric alerts and quota alerts.
  - Filter the errors and watch for errors that are consistently happening in order to determine if there are errors that may give insight to potential issues.
  - Filter warnings to look for indicators that something is going to fail or go down.

Monitoring best practices:

- Keep an eye out for unevenness of CPU, memory, and network bandwidth between nodes.
- Become familiar with the performance of the system and the metrics that are expected over time so that if rates are out of the normal range investigation can be initiated.
- Do not let ECS get too full. Account for rebalancing time when expanding.
- Keep an eye out for a higher than normal number of failed requests and determine root cause.
- Regularly check events and audit logs.

8.2 Dell EMC Secure Remote Services (SRS)

SRS provides secure two-way connection between customer-owned Dell EMC equipment and Dell EMC customer service. It provides faster problem resolution with proactive remote monitoring and repair. Although use of SRS is optional, it is highly recommended and should be included during deployment planning. For each site the following contact information is required for ECS, customer support.emc.com credentials, port (if not default of 9443), SRS IP addresses, and SRS server names required for configuration.

For more information on SRS, refer to the Enablement Center for SRS.

8.3 Product alerts and updates

We recommend ECS administrators sign up to receive product updates and alerts. At support.emc.com, clicking on a user's 'Preferences' link opens the 'Account Settings and Preferences' page which contains a 'Subscription and Alerts' tab that allows users to manage their product update subscriptions. Similarly, in the 'Alerts' section on the same page, product advisories can be subscribed to. The minimum recommended subscriptions are for 'ECS Software' or 'ECS Appliance' and 'ECS Software with Encryption.' A search for ECS reveals all available subscription options.

Product alerts and updates best practice highlights:

- Sign up to receive ECS-related product updates and alerts.
- Review Release Notes for details on new features and known problems and limitations.

9 Conclusion

Most of the best practices outlined in this document are pulled from existing Dell EMC product documentation. We recommend that the reader adheres to the globally accepted best practice of reading and becoming familiar with all existing documentation on ECS. We encourage working closely with appropriate internal teams and Dell EMC personnel during the planning phase and refer to appropriate hardware specifications.

A Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Storage technical documents and videos provide expertise that helps to ensure customer success on Dell EMC storage platforms.

A.1 Related resources

Note: Links in this section may require login access to Dell EMC Support site or internal site.

ECS product documentation:

- ECS product documentation at support site
- ECS product documentation at community site

ECS technical assets:

- Planning Guide
- Site Preparation Guide
- Security Configuration Guide
- ECS Hardware and Cabling Guide
- ECS Specification Sheet
- ECS ISV Compatibility Guide

ECS APIs and SDKs:

- ECS Rest API Reference
- Dell EMC Data Services – Atmos and S3 SDK
- Data Access Guide

ECS technical white papers

- ECS Architecture Guide
- ECS General Best Practices (this paper)
- ECS Networking and Best Practices
- Load balancers:
  - ECS with HAProxy Load Balancer
  - ECS with NGINX (OpenResty)
  - ECS with F5
  - ECS with KEMP

Community:

- ECS Community
- ECS Test Drive
- Enablement Center for SRS

Third-party resources:

- Industry NTP Best Practices