Abstract
This white paper describes networking and related best practices for ECS, the Dell EMC software-defined cloud-scale object storage platform.

March 2020
Revisions

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 2016</td>
<td>Initial release</td>
</tr>
<tr>
<td>August 2017</td>
<td>Updated based on ECS 3.1</td>
</tr>
<tr>
<td>February 2019</td>
<td>Add Gen3 details</td>
</tr>
<tr>
<td>March 2020</td>
<td>Updated based on ECS Special Feature Configuration Support</td>
</tr>
</tbody>
</table>

Acknowledgments

This paper was produced by the following member of the Technical Marketing Engineering and Solution Architects team in the Unstructured Data Storage division of Dell Technologies:

Author: Jarvis Zhu, Chris Kraft

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2016–2020 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners. [3/30/2020] [Best Practices] [H15718.6]
# Table of contents

- Revisions ........................................................................................................... 2
- Acknowledgements ............................................................................................. 2
- Executive summary .............................................................................................. 4
  1 Introduction ....................................................................................................... 5
    1.1 Audience ...................................................................................................... 5
    1.2 Scope ........................................................................................................... 5
  2 ECS Overview .................................................................................................... 6
  3 ECS Network Overview ..................................................................................... 7
    3.1 Traffic Types ............................................................................................... 7
  4 ECS Network Hardware ..................................................................................... 8
    4.1.1 Public Switches ....................................................................................... 8
    4.1.2 Private Switches ..................................................................................... 9
    4.1.3 Customer-provided Switches ................................................................. 10
  5 ECS Network Configurations ........................................................................... 11
    5.1 Production Network .................................................................................. 11
      5.1.1 ToR Public Switch Configuration and Node Connectivity .................. 13
      5.1.2 Customer Uplinks Configurations ....................................................... 15
      5.1.3 Network Configuration Custom Requests ........................................ 18
    5.2 Internal Private Network ........................................................................... 19
      5.2.1 NAN Topologies ................................................................................ 19
      5.2.2 Segment LAN ................................................................................... 21
      5.2.3 Cluster LAN ...................................................................................... 21
      5.2.4 RMM Access from Customer Network (Optional) ............................ 22
  6 ECS Network Separation .................................................................................. 26
    6.1.1 Network Separation Configurations .................................................... 27
    6.1.2 ECS Switch Configurations for Network Separation ......................... 31
  7 ECS Network Performance .............................................................................. 37
  8 Tools .................................................................................................................. 38
    8.1 ECS Portal .................................................................................................. 38
    8.2 ECS Designer and Planning Guide ............................................................. 38
    8.3 Secure Remote Services ............................................................................. 38
    8.4 Linux or HAL Tools .................................................................................... 38
  9 Network Services ............................................................................................... 41
  10 Conclusions ..................................................................................................... 42
  A Technical support and resources ....................................................................... 43
ECS is the Dell EMC cloud-scale, object storage platform for traditional, archival, and next-generation workloads. It provides geo-distributed and multi-protocol (Object, HDFS, and NFS) access to data. In an ECS deployment, a turn-key appliance or industry standard hardware can be utilized to form the hardware infrastructure. In either type of deployment, a network infrastructure is required for the interconnection between the nodes and customer environments for object storage access. This paper describes ECS networking and some best practices to consider.
1 Introduction

This white paper provides details on ECS networking. ECS network hardware, network configurations, and network separation are discussed. It will also describe some ECS networking best practices. This paper should be used as an adjunct to the following Dell EMC ECS documentation:

- ECS Hardware Guide - for Gen1 and Gen2 hardware
- ECS EX-Series Hardware Guide
- Networks Guide for D- and U- Series (Gen 1 and Gen 2) Hardware
- Networks Guide for EX300 and EX3000 (EX-Series) Hardware

1.1 Audience

This document is targeted for Dell EMC field personnel and customers interested in understanding ECS networking infrastructure and the role networking plays within ECS as well as how ECS connects to customer environments.

1.2 Scope

This white paper explains ECS network configurations and topologies and provides some best practices. It does not cover ECS network installation and administration. Refer to official Dell EMC ECS product documentation for information on ECS installation and administration.

Updates to this document are done periodically and often coincides with new features and functionality changes.
ECS overview

ECS features a software-defined architecture that promotes scalability, reliability and availability. ECS was built as a completely distributed storage system to provide data access, protection and geo-replication. The main use cases for ECS include storage for modern applications and as secondary storage to free up primary storage of infrequently used data while also keeping it reasonably accessible.

ECS software and hardware components work in concert for un-paralleled object and file access. The software layers are presented in Figure 1 along with the underlying infrastructure and hardware layers. It can be viewed as a set of layered components consisting of the following:

- **ECS Portal and Provisioning Services** - provides an API, CLI and Web-based portal that allows self-service, automation, reporting and management of ECS nodes. It also handles licensing, authentication, multi-tenancy, and provisioning services.
- **Data Services** - provides services, tools and APIs to support Object, and HDFS and NFSv3.
- **Storage Engine** - responsible for storing and retrieving data, managing transactions and protecting and replicating data.
- **Fabric** - provides clustering, health, software and configuration management as well as upgrade capabilities and alerting.
- **Infrastructure** - uses SUSE Linux Enterprise Server 12 as the base operating system for the turnkey appliance or qualified Linux operating systems for industry standard hardware configuration.
- **Hardware** - industry standard hardware composed of x86 nodes with internal disks or attached to disk array enclosures with disks, and top-of-rack (ToR) switches.

For more in-depth architecture of ECS, refer to the *ECS Architecture and Overview* white paper.
3 ECS network overview

ECS network infrastructure consists of a set of ToR switches which allow for the following two types of network connections:

- **Public Network** - connection between the customer production network and ECS.
- **Private Network** - for management of nodes and switches within and across racks.

The ToR switches are dedicated to either the public (production) network, or to the private, internal-to-ECS-only network. For the public network a pair of 10/25 GbE network switches are used which service data and internal communication between the nodes. For the private network, depending on the hardware generation, either a single 1 GbE switch for Gen1 or Gen2 (Gen1/2) is used, or a pair of 25 GbE switches (Gen3) are used. The private network is used for remote management, console access and PXE booting which enables rack management and cluster-side management and provisioning. From these set of switches, uplink connections are presented to the customer production environment for storage access and management of ECS. The networking configurations for ECS are recommended to be redundant and highly available.

3.1 Traffic types

Understanding the traffic types within ECS and the customer environment is useful for architecting the network physical and logical layout and configuration for ECS.

The public network carries the following types of traffic:

- **Data** - customer data and I/O requests.
- **Management** - provisioning and/or querying ECS via the portal and/or ECS Rest Management APIs as well as network services traffic such as DNS, AD and NTP.
- **Inter-node** - messages are sent between nodes to process I/O requests depending on owner of data and inter-node checks.
- **Replication** - data replicated to other nodes within a replication group.

In a single site single-rack deployment, inter-node traffic stays within the ECS rack switches; whereas in single site multi-rack deployment, inter-node traffic traverses from one rack set of switches up to the customer switch and to the other rack switches to process requests. In a multi-site or geo-replicated deployment, all the above traffic will also go across the WAN.

The private network, which is under Dell EMC control, is entirely for node and switch maintenance and thus traffic types include:

- **Segment Maintenance Management** - traffic associated with administration, installation or setup of nodes and switches within rack.
- **Cluster Maintenance Management** - traffic associated with administration, installation or setup of nodes across racks within a site.
ECS network hardware

Each ECS appliance rack contains three or four switches. Gen1/2 appliances have three switches, two for the public network and one for the private network. Gen3 has two public switches and two private switches.

Switch details, including model numbers, along with designated switch port usage and network cabling information, can be found in the ECS Hardware Guide for Gen1/2 appliances and the ECS EX-Series Hardware Guide for Gen3 appliances.

4.1.1 Public switches

Public (a.k.a. production or front-end) switches are used for data transfer to and from customer applications as well as for internal node-to-node communication. These switches connect to the ECS nodes in the same rack. For Gen1/2 appliances, two 10 GbE, 24-port or 52-port Arista switches are used. For Gen3 appliances, two 10/25 GbE (EX300) or two 25 GbE (EX3000) 48-port Dell switches are used. To create a High Availability (HA) network for the nodes in the rack, the public switches work in tandem using LACP/MLAG, with the Arista switches in Gen1/2 appliances, and Virtual Link Trunking (VLT), with the Dell EMC switches in Gen3 appliances. This pairing is for redundancy and resiliency in case of switch failure.

Across all generations of hardware, Gen1-3, each ECS node has two Ethernet ports that directly connect to one of the ToR public switches. Due to NIC bonding, the individual connections of a node appear to the outside world as one. The nodes are assigned IP addresses from the customer’s network either statically or via a DHCP server. At a minimum one uplink between each ToR public switch in the ECS appliance to the customer network is required. The public switch management ports connect to the ToR private switch(es).

Best Practices

- For redundancy and to maintain a certain level of performance, have two uplinks per switch to customer switch, or, four uplinks per rack minimum.
- Use 25 GbE switches for optimal performance when using customer-provided public switches.
- Have dedicated switches for ECS and do not use “shared ports” on customer core network.
### 4.1.2 Private switches

Private switches are used by ECS for node management. For Gen1/2 appliances, the private switches also allow for out-of-band (OOB) management communication between customer networks and Remote Management Module (RMM) ports in individual ECS nodes. Gen1/2 appliances have a 52-port 1 GbE Arista switch, or a Cisco switch for organizations with strict Cisco only requirements. Gen3 appliances contain two 25 GbE 48-port Dell private switches identical in model to the public switches.

Gen3 does **not** allow for OOB management communication from customer networks.

The management ports in each node connect to the private switch(es). They use private addresses such as 192.168.219.x. Each Gen1/2 node also has a connection between its RMM port and the private switch, which in turn can have access to the customer network to provide OOB management of the nodes. Gen3 nodes also have a connection between their iDRAC (Integrated Dell Remote Controller) and one of the private switches, however, there is no customer-facing OOB management for Gen3 ECS nodes.

Dell EMC switches are required for the private network. Private switches cannot be customer-provided.

---

**Best Practices**

- When physically connecting nodes to the management switch, do so in an ordered and sequential fashion. For instance, node 1 should connect to port 1, node 2 to port 2 and so on. Connecting nodes to an arbitrary port between 1 through 24 can cause installation issues.
- RMM Connections are optional and best practice is to ask customer requirements for these connections.
4.1.3 **Customer-provided switches**

The flexibility of ECS allows for variations of network hardware and configurations which should meet the Dell EMC standards like [ECS Appliance- Special Feature Configuration Support](#). However, An RPQ is a request for approval or review of non-standard configuration.

Regarding customer provided switches, configuration and support are the responsibility of the customer. These switches should be dedicated to ECS and not shared with other applications. Dell EMC assistance is completely advisory for customer provided switches.

[ECS Appliance- Special Feature Configuration Support](#) is an internal-only authenticated reference, customers need ask Presales/Sales help if they want to read this reference

An RPQ cannot be submitted to replace the private network switches for an ECS appliance since these are solely for administration, installation, diagnosing and management of ECS nodes and switches. The private network switches need to remain under control of Dell EMC personnel.
5  **ECS network configurations**

The previous section briefly described the switches and related networks used ECS appliances. This section will explore further the public production network, and the private ECS internal management network referred to as the *Nile Area Network*, or NAN for short. Design considerations and best practices in both production and internal networks are discussed to offer guidance for network architects.

5.1  **Production network**

The production network involves the connections between the customer’s network and the two ToR ECS front-end, public data switches as well as the connections within the ECS rack. These connections act as the critical paths for in and out client requests and data (“north to south”), and inter-node traffic (“east to west”) for replication and processing requests as shown in Figure 4 for a single rack.

![Figure 4](image)

Figure 4  **Production Network Traffic Flow between Customer Network and ECS Hardware in a Single Rack**

For multi-rack, inter-node traffic flows “north to south” and over to the customer network and to the other ECS racks as shown in Figure 5 on the next page.
Production network configurations

Customers may choose to add an aggregation switch external to the ECS racks to prevent inter-node traffic from flowing through their main network switch infrastructure as shown in Figure 6.

Network connections in the production network as a best practice should be designed for high availability, resiliency and optimal network performance.
5.1.1 ToR public switch configuration and node connectivity

The public ToR switches work in tandem using LACP/MLAG, with the Arista switches in Gen1/2 appliances, Virtual Link Trunking (VLT), with the Dell EMC switches in Gen3 appliances, to create a HA network for the nodes in the rack. Similarly, if Cisco switches are used the vPC LAG protocol is used for HA. The aggregation of multiple ports results in higher bandwidth, resiliency and redundancy in the data path.

Each node in the rack is connected to both switches through two NICs which are aggregated together using a Linux bonding driver. The node is configured to bond the two NICs into a single LACP bonding interface also known as a "mode 4" bond. This bonding interface connects one port to each switch as demonstrated in the following images.

Figures 7 and 8 are examples of nodes bonded with a public switch pair.

![Figure 7](image1)

**Figure 7** Gen1/2 Node and Public Switch Connectivity via MLAG

![Figure 8](image2)

**Figure 8** Gen3 Node and Public Switch Connectivity via VLT

The terminal output on the next page displays snippets of the basic configuration files for the Gen1/2 Arista data switches for the ports associated with the nodes starting at port 9. As can be seen from these snippet examples, the physical port 9 on each switch are configured as an active Link Aggregation Control Protocol (LACP) MLAG. LACP is a protocol that would build LAGs dynamically by exchanging information in Link Aggregation Control Protocol Data Units (LACDUs) relating to the link aggregation between the LAG. LACP sends messages across each network link in the group to check if the link is still active resulting in faster error and failure detection. The port channels on each switch are MLAG-ed together and can be visible from the nodes. They are configured to allow for fast connectivity via the spanning tree portfast command. This command places the ports in forwarding state immediately as opposed to the transition states of listening, learning and then forwarding which can cause 15 seconds of delay. Port channels also are set to lACP fallback to allow all ports within the port-channel to fall back to individual switch ports. When the node’s ports are not yet configured as LAG, this setting allows for PXE booting from public switch ports of the nodes and forwarding of traffic.
Snippets of the basic configuration files are shown below and illustrate the MLAG peer link definition for the data switches. This sets up a Port Control Protocol connection between the two switches for MLAG communication. So, the peer address defined in `mlag configuration` points to the IP address of its peer on the other switch. MLAG peer link has several functions which include:

- Manage port channel groups, check status of each link and to update any Layer 2 protocol information for instance where the MAC address tables are updated to allow for quick forwarding.
- Pass traffic to correct link when traffic is sent to a non-MLAG destination.

Once the peer-link is up and there is bi-directional TCP connection between the peers, the MLAG peer relationship is established. Here is output of data switch configuration file showing the MLAG Peer Link definitions between the switches.

Similar Gen3-specific configuration examples, such as for `vlt-port-channel` and `channel-group`, are shown in the ECS Networks Guide for EX300 and EX3000 (EX-Series) Hardware documentation.

The data switches are pre-configured on ECS supported switches. The configuration files for the data switches are located on each node in directory. For example, Gen3 Dell EMC switch configuration files are located inside `/usr/share/emc-dell-firmware/config/ecs/`. 
5.1.2 Customer uplink configuration

Any networking device supporting Static Link Aggregation Group or IEEE 802.3ad Link Aggregation Control Protocol (LACP) can connect to the MLAG switch pair. With Static Link Aggregation, all settings are defined on all participating LAG components whereas LACP sends messages across each link in the group to check their state. An advantage of LACP over Static Link Aggregation is faster error or failure detection and handling.

Each Gen1/2 Arista public data switch has eight ports available to connect to the customer network, providing sixteen uplink ports per rack. Gen3 Dell EMC switches each have eight 10/25 GbE and four 100 GbE ports, providing either sixteen or eight uplink ports per rack. See the appropriate (Gen1/2 or Gen3) Networks Guide for ECS Hardware for complete details including switch configuration examples.

For Gen1/2 appliances, as with the ports used for the node, the eight uplink ports on each of the data switches are configured as a single LACP/MLAG interface as shown below in the code output. The port-channels are also configured to be in lacp fallback mode for customers who are unable to present LACP to the ECS rack. This mode will only be activated if no LACP is detected by the protocol. If there is no LACP discovered between the customer link and the ECS switches then the lowest active port will be activated and all other linked ports in the LAG will be disabled until a LAG is detected. At this point, there is no redundancy in the paths.

In addition, the data switches in Gen1/2 appliances are not configured to participate in the customer’s spanning tree topology. They are presented as edge or host devices since a single LAG for the eight ports in each switch is created. The spanning-tree bpdufilter enable setting in the configuration file filters all spanning tree bridge protocol data units (BPDUs) from the uplink ports. This setting separates the customer network from any ECS link or switches failure in the rack as well as address customer’s concerns relating to ECS interfering with their spanning tree topology by one of the ECS switches becoming root. It also simplifies the setup of the ECS switches in the customer network. Here is the output for Gen1/2 public switches basic configuration file showing the uplink definitions for port 1.

![10GbE (Bottom)](image)
```
!10GbE (Bottom)
interface Ethernet1
   description MLAG group 100
   channel-group 100 mode active
   lacp port-priority 1
!
interface Port-Channel100
   description Customer Uplink (MLAG group 100)
   port-channel lacp fallback
   port-channel lacp fallback
   timeout 1
   spanning-tree bpdufilter enable
   mlag 100
```

![10GbE (Top)](image)
```
!10GbE (Top)
interface Ethernet1
   description MLAG group 100
   channel-group 100 mode active
   lacp port-priority 2
!
interface Port-Channel100
   description Customer Uplink (MLAG group 100)
   port-channel lacp fallback
   port-channel lacp fallback
   timeout 1
   spanning-tree bpdufilter enable
   mlag 100
```

Connections from the customer network to the data switches can be linked in several different ways, for instance, as a single link, multi-link to a single switch using LACP or multi-link to multiple switches using a multiple switch LACP protocol like Arista MLAG, Dell EMC VLT or Cisco vPC. Customers are required to provide the necessary connection information to establish communication to the nodes in the rack. Figures 9-11 illustrate some of the possible link connections and best practices.
The code below shows an example of a two port LAG for an Arista and Cisco single switch with multiple links.

**Arista configuration**

```
interface Ethernet 1-2
channel-group 100 mode active
```

**Cisco configuration**

```
interface Ethernet1/1
channel-group 100 mode active
interface Ethernet1/2
channel-group 100 mode active
```

Figure 10 exhibits a multiple port uplink to multiple switches with a LAG configuration. A better approach would be to configure more than two links per ECS switch as presented in Figure 11. The links should be spread in a bowtie fashion (links on each customer switch should be distributed evenly between the data switches) for redundancy and optimal performance during failures or scheduled downtime.
In either of these configurations, both port channels will need to be connected using a multi-switch LAG protocol like Arista MLAG or Cisco virtual Port Channel (vPC) to connect to the ECS MLAG switch pair port channel. Also, customers need to create port channels using LACP in active or passive mode on all switches participating in the multi-switch LAG. In the output below are sample configurations for Arista and Cisco with multi-switch LAG protocols definitions. Note the vPC or MLAG numbers on each switch would need to match to create a single port channel group.

### Arista Configuration

**Switch A**
- interface Ethernet 1-2
- channel-group 100 mode active
- mlag 100

**Switch B**
- interface Ethernet 1-2
- channel-group 100 mode active
- mlag 100

### Cisco Configuration

**Switch A**
- interface Ethernet1/1
- channel-group 100 mode active
- interface Ethernet1/2
- channel-group 100 mode active
- vpc 100

**Switch B**
- interface Ethernet1/1
- channel-group 100 mode active
- interface Ethernet1/2
- channel-group 100 mode active
- interface port-channel 100
- vpc 100

### Best Practices

- For multiple links, setup LACP on customer switch. If LACP is not configured on customer switches to ECS switches, one of the data switches will have the active connection(s) and the port(s) connected to other data switch will be disabled until a LAG is configured. Connection(s) on the other switch will only become active if one switch goes down.
- Balance the number of uplinks from each switch for proper network load balancing to the ECS nodes.
- When using two customer switches, it is required to utilize multi-switch LAG protocols.
- For multi-rack environments, consider utilizing an Aggregation Switch to keep inter-node traffic separated from customer core network.

Gen1/2 switches are configured **not to participate** in the customer’s spanning tree topology. Gen3 switches are configured **to participate** in the customer’s spanning tree topology with Rapid Spanning Tree Protocol (rstp). For Gen3 switch configuration details see the *ECS Networks Guide for EX300 and EX3000 (EX-Series) Hardware* documentation.
5.1.3 **Network configuration custom requests**

Customers may have requirements needing modifications to ECS basic configuration files for the data switches. For instance, customers who require physical network isolation between traffic types for security purposes. As an example, in Figure 12 is an example of multiple ports uplinked to multiple domains. In this setup, it would need changes in the data switches basic configuration files to support two LAGs on the uplinks and to change VLAN membership for the LAGs.

![Figure 12: Multiple Ports Uplinked to Multiple Domains](image)

Another example is if the customer needs to configure the uplinks for a specific VLAN. The VLAN membership should only be changed if the customer requirement is to set the uplink ports to VLAN trunk mode. Only the port-channels for the uplink and nodes need to be changed to setup the VLAN. The output below shows a sample of how to change the VLAN membership. Both data switches would need to have the same VLAN configuration.

```
# Create new vlan
vlan 10
exit
# change to vlan trunk mode for uplink
interface port-channel100
switchport mode trunk
switchport trunk allowed vlan 10
# change vlan membership for access port to the nodes
interface port-channel11-12
switchport access vlan 10
```

copy running-config startup-config

**Best Practices**

- Refer to the *ECS Appliance- Special Feature Configuration Support* for the configuration changed
- Submit an RPQ for network configurations requiring modification if you cannot find the corresponding configuration in the *ECS Appliance- Special Feature Configuration Support* document.
ECS network configurations

5.2 Internal private network

The internal private network also known as the Nile Area Network (NAN) is mainly used for maintenance and management of the ECS nodes and switches within a rack and across racks. Ports on the management switch can be connected to another management switch on another rack, creating a NAN topology. From these connections, nodes from any rack or segment can communicate to any other node within the NAN. The management switch is split in different LANs to segregate the traffic to specific ports on the switch for segment only traffic, cluster traffic and customer traffic to RMM:

- **Segment LAN** - includes nodes and switches within a rack
- **Cluster LAN** - includes all nodes across all racks
- **RMM LAN** (Gen1/2 only) - uplink ports to customer LAN for RMM access from customer’s network

5.2.1 NAN topologies

The NAN is where all maintenance and management communications traverse within rack and across racks. A NAN database contains information such as IP addresses, MAC addresses, node name and ID on all nodes within the cluster. This database is locally stored on every node and is synchronously updated by the master node using the `setrackinfo` command. Information on all nodes and racks within the cluster can be retrieved by querying the NAN database. One command that queries the NAN database is `getrackinfo`.

The racks are connected via the management switches on designated ports. These connections allow nodes within the segments to communicate with each other. There are different ways to connect the racks or rack segments together. Each rack segment is specified a unique color during installation and thus identifying the racks within the cluster. The figures below depict some of the topologies and give some advantages and disadvantages of each NAN topology.

Figure 13 shows a simple topology linearly connecting the segments via ports of the management switches in a daisy-chain fashion. The disadvantage of this topology is that when one of the physical links breaks, there is no way to communicate to the segment(s) that has been disconnected from the rest of the segments. This in effect causes a “split-brain” issue in NAN and forms a less reliable network.

![Linear or Daisy Chain Topology](image-url)
Another way to connect the segments is in a ring topology as illustrated in Figure 14. The advantage of the ring topology over the linear is that two physical links would need to be broken to encounter the split-brain issue, proving to be more reliable.

![Ring Topology Diagram](image)

**Figure 14  Ring Topology**

For large installations, the split-brain issue in the ring or linear topologies could be problematic for the overall management of the nodes. A star topology is recommended for an ECS cluster where there are ten or more racks or customers wanting to reduce the issues that ring or linear topologies pose. In the star topology, an aggregation switch as shown in Figure 15, would need to be added and would be an extra cost; however, it is the most reliable among the NAN topologies.

![Star Topology Diagram](image)

**Figure 15  Star Topology**

**Best Practices**

- Do not use linear topology
- For large installations of ten or more ECS racks, a star topology is recommended for better failover.
### 5.2.2 Segment LAN

The Segment LAN logically connects nodes and switches within a rack to a LAN identified as VLAN 2. This consists of designated ports on the management switch(es) and are referred to as the “blue network”. All traffic is limited to members of this segment for ease of management and isolation from the customer network and other segments within the cluster. The Ethernet ports on the nodes are configured with a private IP address derived from the segment subnet and node ID number. Thus, the IP address is of the form 192.168.219.[NodeID]. The IPs are not routable and packets are untagged. These addresses are reused by all segments in the cluster. To avoid confusion, it is not recommended to use these IP addresses in the topology file required when installing the ECS software on the nodes. There are several IP addresses that are reserved for specific uses:

- **192.168.219.254** - reserved for the master node within the segment. Recall from the previous section that there is a master node designated to synchronize the updates to the NAN database.
- **192.168.219.251** - reserved for the management switch
- **192.168.219.252** - reserved for the public switch (bottom)
- **192.168.219.253** - reserved for the public switch (top)

Figure 16 identifies the ports associated with the Segment LAN (VLAN 2 untagged) in Gen1/2 appliances.

#### Best Practices

- Although ports 1-24 on management switch are available for any node and are on the same VLAN, they should be physically linked starting from port 1 and in order with no gaps between ports otherwise, there will be installation issues.
- For troubleshooting a suspect node, administer the node via the “blue network” or Segment LAN (i.e. connect laptop to port 24 or unused port) to not interfere configurations of other segments within the cluster.

### 5.2.3 Cluster LAN

Multiple segment LANs are logically connected to create a single Cluster LAN for administration and access to the entire cluster. Designated interconnect ports on management switches provide interconnectivity between management switches. In addition to the cluster interconnect ports, the designated blue network ports, and the RMM designated ports (Gen1/2 only) are members of the Cluster LAN. All members will tag their IP traffic with VLAN ID 4 as shown in Figure 17 and communicate via the IPv4 link local subnet. During software installation, all nodes in the rack are assigned a unique color number. The color number acts as the segment ID and is used together with the node ID to comprise the new cluster IP address for every node in the cluster. The IP addresses of the nodes in the cluster LAN will be in the form of 169.254.(SegmentID).[NodeID]. This unique IP address would be the recommended IP address to specify in the topology file for the nodes within the cluster.
VLAN Membership of Cluster LAN on Gen1/2 Management Switch

Best Practices

- ECS does not yet have support for IPv6, so do not enable IPv6 on these switches or send IPv6 packets.
- If troubleshooting a segment within the cluster, administer the segment via the Segment LAN to not affect the configuration of the entire cluster.
- Use the IP address in the Cluster LAN in the topology file to provide a unique IP for all nodes within the cluster.

5.2.4 RMM access from customer network (optional)

For Gen1/2 hardware only, designated RMM ports provide OOB and remote management of the nodes via ports on the management switch. RMM access from customer network is optional and it is recommended to determine specific requirement from customer. A relevant use of the RMM connection would be for ECS software only deployments where the hardware is managed and maintained by customers or when customers have a management station in which they would require RMM access to all hardware from a remote location for security reasons.

To allow for RMM connections from customer switch, Ports 51 and 52 on the management switch are configured in a hybrid mode allowing the ports to handle both tagged and untagged traffic. In this setup, the ports can be used for multiple purposes. The uplinks to the customer switch are on VLAN 6 and packets are untagged. A snippet of the Arista management switch basic configuration with added comments is illustrated by the code below and shows how ports 51 and 52 are configured as a hybrid port. For RMM traffic, the ports on management switch should be setup to assign untagged traffic coming into VLAN 6 and strip the tag from the VLAN 6 traffic on the way out. To allow for both cluster and customer traffic to travel on same physical link, it may require some modification on the customer switch or aggregation switch. The peer connection from the customer network must also be setup as an access port to allow untagged traffic in and out of the management switch if there is only RMM traffic going thru customer switch. If it is required that NAN and RMM traffic traverse through the customer network, then the switchport is configured as a trunk and the customer switch would need to be configured to forward VLAN 4 as tagged traffic to only the ports connected.
ECS network configurations

to management as well as add VLAN 4 in the forbidden list on the customer network. It is important that the NAN is a closed network. If the customer requires a different VLAN ID for the RMM customer uplink (VLAN 6), it is possible.

interface Ethernet51
   description Nile Area Network Uplink
   mtu 9212
   ! Assign untagged packets to VLAN 6 on the way in and untag packets from VLAN 6 on the way out.
   switchport trunk native vlan 6

   ! Only traffic from the list VLAN are allowed to forward
   switchport trunk allowed vlan 4,6 traffic

   ! Enable tagged and untagged traffic on the port
   switchport mode trunk

! interface Ethernet52
   description Nile Area Network Uplink
   mtu 9212
   switchport trunk native vlan 6
   switchport trunk allowed vlan 4,6
   switchport mode trunk

The following figures present the addition of customer switch into the different NAN topologies to allow for OOB and remote RMM access to customers. In the linear topology, the rack segments are connected end to end using port 51 and 52 as shown in Figure 19 and are on VLAN 4. Port 51 on management connects to the customer network on VLAN 6 and configured as an access port. As an access port on the customer switch, it will drop tagged traffic. For extra protection or if there are some limitations on the customer switch that does not allow for setting an access port, VLAN 4 can be added to the forbidden list on the customer switch to prevent leaking of cluster traffic (NAN traffic) to the customer network. As previously mentioned, the linear topology will cause a split-brain issue when one link fails and thus not an ideal topology.

![Customer Switch with ECS Segments in Linear Topology](image)

Figure 19  Customer Switch with ECS Segments in Linear Topology

A customer switch with ECS segments in a ring topology is shown in Figure 20. Port 51 of one management switch (Segment Purple) and port 52 of another management switch (Segment Blue) are connected to the customer network forming the ring topology. Ports 51 and 52 connect the cluster to the customer network on VLAN 6. In this setup, the customer uplink ports on the management switch would need to be configured to isolate VLAN 4 traffic to only between the management uplinks. It should also allow tagged and untagged traffic between the uplinks for VLAN 4 traffic and allow the untagged traffic to be forwarded up to the customer
network. The customer switch would also need to be configured to forward VLAN 4 as tagged traffic to only the ports connected to the ECS management switches.

![Customer Switch with ECS Segments in Ring Topology](image1)

**Figure 20**  Customer Switch with ECS Segments in Ring Topology

In a star topology as depicted in Figure 21, port 51 on the management switch is connected to the customer network on VLAN 6 (untagged) for RMM access to all the RMM ports in the Cluster LAN. The connection to the aggregation switch from the customer switch can also be connected for better fault tolerance of the RMM access. Ports 52 of the management switches on each segment are connected to the aggregation switch and are on VLAN 4.

![Customer Switch with ECS Segments in a Star Topology](image2)

**Figure 21**  Customer Switch with ECS Segments in a Star Topology
The settings for the aggregation switch will be similar to the settings of ports 51 and 52 on the management switch. The customer uplink on the aggregation switch must also be configured as an access port to prevent the leaking of VLAN 4 traffic to the customer network. Sample configuration for an aggregation switch is shown in the terminal output below.

```
vlan 4,6
interface Ethernet1-51
    description Nile Area Network Uplink
    mtu 9212
    switchport trunk native vlan 6
    switchport trunk allowed vlan 4,6
    switchport mode trunk
    exit
interface Ethernet52
    description Customer Uplink for RMM
    mtu 9212
    switchport access vlan 4
    exit
```

If the RMM ports are not configured for access from the customer switch, the baseboard management of each segment can still be accessed on the blue network. The baseboard management is available by way of the 1 GbE port on the blue network whose default IP address is 192.168.219.(NodeID+100). Thus, when a laptop is connected to one of the unused blue network ports, a browser can be used to get to the Integrated Baseboard Management Controller (BMC) Web Console of the node console via these IP addresses. The Intelligent Platform Management Interface (ipmi) tools also use these IP addresses for management of nodes. The IP addresses for the RMM ports on the customer network are obtained by default via DHCP or it can be set statically if desired.

Inter-rack connectivity is possible between Gen3 EX-series and Gen1/2 D- or U-series racks. Refer to the Networks Guide for EX300 and EX3000 (EX-Series) Hardware for complete details.

<table>
<thead>
<tr>
<th>Best Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Providing RMM to customer is optional and only available in Gen1/2 appliances.</td>
</tr>
<tr>
<td>• Determination should be made if customer access to RMM is absolutely required before configuration.</td>
</tr>
<tr>
<td>• Ensure that NAN traffic on VLAN 4 does not leak to customer network when adding RMM access to customer network.</td>
</tr>
<tr>
<td>• Use star topology for best failover protection for RMM access and large installations.</td>
</tr>
</tbody>
</table>


## ECS network separation

Network separation allows for the segregation of different types of network traffic for security, granular metering and performance isolation. The types of traffic that can be separated include:

- **Management** - traffic related to provisioning and administering via the ECS Portal and traffic from the operating system such as DNS, NTP and SRS.
- **Replication** - traffic between nodes in a replication group.
- **Data** - traffic associated with data.

There is a mode of operation called the “network separation mode”. When enabled during deployment, each node can be configured at the operating system level with up to three IP addresses or logical networks for each of the different types of traffic. This feature has been designed for flexibility by either creating three separate logical networks for management, replication and data, or combining them to either create two logical networks. For instance, management and replication traffic is in one logical network and data traffic in another logical network.

ECS implementation of network separation requires each network traffic type to be associated with specific services and ports. For instance, the portal services communicate via ports 80 or 443, so these ports and services will be tied to the management logical network. Table 1 below highlights the services fixed to a logical network. For a complete list of services and their associated ports, refer to the most recent version of the ECS Security Configuration Guide.

<table>
<thead>
<tr>
<th>Services</th>
<th>Logical Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECS Portal, provisioning, metering and management API, SSH, DNS, NTP, AD and SMTP</td>
<td>Management network (public.mgmt)</td>
</tr>
<tr>
<td>Data across NFS, Object and HDFS</td>
<td>Data network (public.data)</td>
</tr>
<tr>
<td>Replication data and XOR</td>
<td>CAS only data network (public.data2)</td>
</tr>
<tr>
<td>SRS (Dell EMC Secure Remote Services)</td>
<td>Replication network (public.repl)</td>
</tr>
<tr>
<td></td>
<td>Based on the network that the SRS Gateway is attached (public.data or public.mgmt)</td>
</tr>
</tbody>
</table>

Network separation is achievable logically using virtual IP addresses, using VLANs or physically using different cables. The command `setrackinfo` is used to configure the IP addresses and VLANs. Any switch-level or client-side VLAN configuration is the customer’s responsibility.
### 6.1.1 Network separation configurations

In addition to the default network configuration, network can be partially separated, or all separated, via the following:

- **Standard (default)** - all management, data and replication traffic in one VLAN referred to as public.
- **Partial (Dual)** - two VLANs where one VLAN is the default public which can have two traffic types and another VLAN for any traffic not defined in public VLAN.
- **Partial (Triple)** - one VLAN for public VLAN and two VLANs where one traffic type is placed in public VLAN and two different VLANS are defined for the other two traffic types not in public.

Network separation configures VLANs for specific networks and utilizes VLAN tagging at the operating system level. There is an option to use virtual or secondary IPs where no VLAN is required; however, it does not actually separate traffic but instead just provides another access point. For the public network, traffic can be tagged at the switch level. At a minimum, the default gateway is in the public network and all the other traffic can be in separate VLANs. If needed, the default public VLAN can also be part of the customer’s upstream VLAN and in this case, the VLAN ID for public must match the customer’s VLAN ID.

![Network Separation Configurations Diagram](image)

**Figure 22 Examples of Network Separation Configurations**

Network separation is conducted during ECS installation before the installation of Hardware Abstraction Layer (HAL) or in an existing ECS environment. It requires static IP addresses. Planning for network separation requires decisions on how traffic should be segregated in VLANs, the static IP addresses required, and subnet and gateway information needs to be determined. After network separation has been completed virtual interfaces are created for the VLANs and the interface configuration files will be of the form `ifcfg-public.(vlanID)`. For examples see the following terminal output:

```bash
admin@memphis-pansy:/etc/sysconfig/network> ls ifcfg-public*
ifcfg-public ifcfg-public.data ifcfg-public.mgmt ifcfg-public.repl
```
The operating system presents the interfaces with a managed name in the form of `public.{trafficType}` such as `public.mgmt`, `public.repl`, or `public.data` as can be observed by `ip addr` command output in the following code.

```
admin@memphis-pansy:/etc/sysconfig/network> ip addr | grep public
3: slave-0: <BROADCAST,MULTICAST,SLAVE,UP,LOWER_UP> mtu 1500 qdisc mq master public
5: slave-1: <BROADCAST,MULTICAST,SLAVE,UP,LOWER_UP> mtu 1500 qdisc mq master public
37: public: <BROADCAST,MULTICAST,MASTER,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP
inet 10.245.132.55/24 scope global public
39: public.mgmt@public: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP
inet 10.10.20.55/24 scope global public.mgmt
40: public.repl@public: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP
inet 10.10.30.55/24 scope global public.repl
41: public.data@public: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP
inet 10.10.10.55/24 scope global public.data
```
The HAL searches for these managed names based on the active_template.xml in /opt/emc/hal/etc. It finds those interfaces and presents those to the Fabric layer. Output of cs_hal list nics is shown in the code below. As can be seen from the output, the network traffic types are specified, tagged and used for the mapping.

```
admin@memphis-pansy:~> sudo -i cs_hal list nics
Nics:
    Name: public Type: Bonded
    SysPath: [/sys/devices/virtual/net/public]
    IfIndex : 37
    Pos : 16421
    Parents : ( slave-1, slave-0, public )
    Up and Running : 1
    Link detected : 1
    MAC : 00:1e:67:e3:19:82
    IPAddress : 10.245.132.55
    Netmask : 255.255.255.0
    Bond Info: Mode: 4 miimon: 100 Slaves: (slave-0, slave-1) OtherOptions:
    NetworkType: public
    Name: public.mgmt Type: Tagged
    SysPath: [/sys/devices/virtual/net/public.mgmt]
    IfIndex : 39
    Pos : 32807
    Parents : ( slave-1, slave-0, public, public.mgmt )
    Up and Running : 1
    Link detected : 1
    MAC : 00:1e:67:e3:19:82
    IPAddress : 10.10.20.55
    Netmask : 255.255.255.0
    Tag Info: VID: 2000 base dev: public
    NetworkType: mgmt
    Name: public.repl Type: Tagged
    SysPath: [/sys/devices/virtual/net/public.repl]
    IfIndex : 40
    Pos : 32808
    Parents : ( slave-1, slave-0, public, public.repl )
    Up and Running : 1
    Link detected : 1
    MAC : 00:1e:67:e3:19:82
    IPAddress : 10.10.30.55
    Netmask : 255.255.255.0
    Tag Info: VID: 3000 base dev: public
    NetworkType: repl
    Name: public.data Type: Tagged
    SysPath: [/sys/devices/virtual/net/public.data]
    IfIndex : 41
    Pos : 32809
    Parents : ( slave-1, slave-0, public, public.data )
    Up and Running : 1
    Link detected : 1
    MAC : 00:1e:67:e3:19:82
    IPAddress : 10.10.10.55
    Netmask : 255.255.255.0
    Tag Info: VID: 1000 base dev: public
    NetworkType: data
```
The HAL gives the above information to the Fabric layer which creates a JavaScript Object Notation (JSON file) with IP addresses and interface names and supplies this information to the object container. The code below is an output from Fabric Command Line (fcli) showing the format of the JSON structure.

```
admin@memphis-pansy:/opt/emc/caspian/fabric/cli> bin/fcli agent node.network
{
   "etag": 12,
   "network": {
      "mgmt_interface_name": "public.mgmt",
      "mgmt_ip": "10.10.20.55",
      "data_interface_name": "public.data",
      "data_ip": "10.10.10.55",
      "hostname": "memphis-pansy.ecs.lab.emc.com",
      "private_interface_name": "private.4",
      "private_ip": "169.254.78.17",
      "public_interface_name": "public",
      "public_ip": "10.245.132.55",
      "replication_interface_name": "public.repl",
      "replication_ip": "10.10.30.55"
   },
   "status": "OK"
}
```

The mapped content of this JSON structure is placed in object container in the file `/host/data/network.json` as shown in the terminal output below in which the object layer can utilize to separate ECS network traffic.

```
{
   "data_interface_name": "public.data",
   "data_ip": "10.10.10.55",
   "hostname": "memphis-pansy.ecs.lab.emc.com",
   "mgmt_interface_name": "public.mgmt",
   "mgmt_ip": "10.10.20.55",
   "private_interface_name": "private.4",
   "private_ip": "169.254.78.17",
   "public_interface_name": "public",
   "public_ip": "10.245.132.55",
   "replication_interface_name": "public.repl",
   "replication_ip": "10.10.30.55"
}
```

Network separation in ECS utilizes source-based routing to specify the route that packets take through the network. In general, the path that packets come in on will be the same path going out. Based on the `ip rules`, the local node originating the packet looks at the IP, and then first looks at local destination, and if it is not local, then looks at the next. Using source-based routing reduces static routes that need to be added.
6.1.2  ECS switch configuration for network separation

Depending on customer requirements, network separation may involve modification of the basic configuration files for the data switches. This section will explore examples of different network separation implementations in the switch level such as the default, single domain, single domain with public set as a VLAN, and physical separation.

6.1.2.1  Standard (default)

The default settings use configuration files that are bundled with ECS. In this scenario there is no VLAN and there is only the public network. Also, there is no tagged traffic in the uplink connection. All ports are running in access mode. Table 2 and Figure 23 provide an example of a default ECS network setup with customer switches.

Table 2  Standard Default Switch Configuration

<table>
<thead>
<tr>
<th>Interface</th>
<th>VLAN ID</th>
<th>Tagged</th>
<th>Uplink Connection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>None</td>
<td>No</td>
<td>MLAG:po100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No tagged traffic</td>
</tr>
</tbody>
</table>

Figure 23  Example of Standard Default Switch Setup
6.1.2.2 Single domain

In a single domain, a LACP switch or an LACP/MLAG switch pair are configured on the customer side to connect to the ECS MLAG switch pair. Network separation is achieved by specifying VLANs for the supported traffic types. In the example in Table 3 and Figure 24, data and replication traffic are segregated into two VLANs and the management stays in the public network. The traffic on the VLANs will be tagged at the operating system level with their VLAN ID which in this case is 10 for data and 20 for replication traffic. The management traffic on the public network is not tagged.

Table 3: An Example of Single Domain Switch Configuration

<table>
<thead>
<tr>
<th>Interface</th>
<th>VLAN ID</th>
<th>Tagged</th>
<th>Uplink Connection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>None</td>
<td>No</td>
<td>MLAG:po100</td>
</tr>
<tr>
<td>Data</td>
<td>10</td>
<td>Yes</td>
<td>All named traffic tagged</td>
</tr>
<tr>
<td>Repl</td>
<td>20</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Figure 24: An Example of a Single Domain Switch with Two VLANs

Both data switch configurations files would need to be modified to handle the VLANs in above example. The terminal output below shows how this can be specified for Arista switches. Things to note from the configuration file include:

- The switchport have been modified from access to trunk.
- VLANs 10 and 20 created to separate data and replication traffic are allowed. They also need to be created first.
- VLAN 1 corresponds to the public.
- Ports-channels are utilized it will supersede and ignore Ethernet level configurations.

This example shows a single domain switch settings with two VLANs for public switches.

```
vlan 10, 20
interface po1-12
switchport trunk native vlan 1
switchport mode trunk
switchport trunk allowed vlan 1,10,20

!For 7050S-52 and 7050SX-64, the last port channel is 24

interface po100
switchport mode trunk
switchport trunk allowed vlan 1, 10,20
```
6.1.2.3 Single domain and public VLAN

Customers may desire to have the public network in a VLAN and in this scenario, the traffic going through the public network will be tagged at the switch level and the other VLANs will be tagged at the operating system level. Table 4 and Figure 25 provides switch and configuration details for a single domain with public VLAN setup.

Table 4  Single Domain and Public VLAN Configuration Example

<table>
<thead>
<tr>
<th>Interface</th>
<th>VLAN ID</th>
<th>Tagged</th>
<th>Uplink Connection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>100</td>
<td>Yes (Switch)</td>
<td>MLAG:po100</td>
</tr>
<tr>
<td>Data</td>
<td>10</td>
<td>Yes (OS level)</td>
<td>Al traffic tagged</td>
</tr>
<tr>
<td>Repl</td>
<td>20</td>
<td>Yes (OS level)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 25  An Example of Single Domain Public VLAN Switch Setup

The settings within the configuration files of the data switches would need to be changed to include all the VLANs specified for network separation. As can be seen from the terminal output below, an update to the native VLAN is done to match the customer VLAN for public. In this example the public VLAN is identified as VLAN 100.

Here is an example code that shows single domain with two VLANs and public in a VLAN settings for public switches.

```bash
vlan 10, 20, 100
interface po1-12
switchport trunk native vlan 100
switchport mode trunk
switchport trunk allowed vlan 10,20,100
interface po100
switchport mode trunk
switchport trunk allowed vlan 10,20,100
```
6.1.2.4 Physical separation

For physical separation, an example setup may include multiple domains on the customer network defined for each type of traffic. An example of the setup and details are defined in Table 5 and illustrated in Figure 26. As can be observed from Table 5, the public network is not tagged and will be on port-channel 100, data traffic will be on VLAN 10, tagged and on port-channel 101 and replication traffic will be on VLAN 20, tagged and on port-channel 102. The three domains are not MLAGed together.

Table 5 An Example of Physical Separation Configuration

<table>
<thead>
<tr>
<th>Interface</th>
<th>VLAN ID</th>
<th>Tagged</th>
<th>Uplink Connection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>None</td>
<td>No</td>
<td>MLAG:po100</td>
</tr>
<tr>
<td>Data</td>
<td>10</td>
<td>Yes</td>
<td>MLAG:po101</td>
</tr>
<tr>
<td>Repl</td>
<td>20</td>
<td>Yes</td>
<td>MLAG:po102</td>
</tr>
</tbody>
</table>

Figure 26 An Example of Physical Separation Setup

The terminal output below shows what the settings would be on the data switches for this configuration on Arista switches. Port-channel 100 is set up to remove uplink ports 2-8, leaving only the first uplink for the public network. Port-channel 101 defines the settings for the data traffic and port-channel 102 is for the replication traffic where the corresponding VLANs are allowed and switchport is set to trunk. Connections to the data nodes are defined by interface po1-12.
The terminal output below shows code from the Gen1/2 Arista public switches settings for physical separation.

**!Uplink ports**

```plaintext
interface po100
  no interface eth2-8
```

**!Data Traffic**

```plaintext
interface Ethernet 2
  channel-group 101 mode active
interface port-channel 101
  switchport trunk allowed vlan 1,10
  description MLAG 101 - data
  mlag 101
```

**!Replication Traffic**

```plaintext
interface Ethernet 3
  channel-group 102 mode active
interface Ethernet 3
  switchport trunk allowed vlan 1,20
  description MLAG 102 - repl
  mlag 102
```

**!ECS Nodes**

```plaintext
interface po1-12
  switchport trunk native vlan 1
  switchport mode trunk
  switchport trunk allowed vlan 1,10,20
```

For situations where customers would want the public network on a VLAN, Table 6 and the subsequent terminal output provide example details of the configuration. In this case, all traffic is tagged and public is tagged with ID 100, data traffic tagged with 10 and replication tagged with 20. Uplink connections, port-channel 100 is setup as trunk and VLAN 10, 20, and 100 are allowed. The connections to the nodes defined in `interface po1-12` are also set accordingly.

**Table 6  Physical Separation with Public VLAN Example**

<table>
<thead>
<tr>
<th>Interface</th>
<th>VLAN ID</th>
<th>Tagged</th>
<th>Uplink Connection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>100</td>
<td>Yes (Switch)</td>
<td>MLAG:po100</td>
</tr>
<tr>
<td>Data</td>
<td>10</td>
<td>Yes</td>
<td>All traffic tagged</td>
</tr>
<tr>
<td>Repl</td>
<td>20</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
The code below shows code in Gen1/2 Arista public switches settings of physical separation with the public is on a VLAN.

!Uplink Ports
interface port-channel 100
    switchport trunk allowed vlan 10,20,100

!ECS Nodes
interface po1-12
    switchport trunk native vlan 100
    switchport mode trunk
    switchport trunk allowed vlan 10,20,100

<table>
<thead>
<tr>
<th>Best Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Network separation is optional. If used, it is important to determine fit and best configuration.</td>
</tr>
<tr>
<td>• Keep the management traffic within the public to reduce the number of static routes.</td>
</tr>
<tr>
<td>• Although it can have only default gateway in public, it is recommended to have at least one of the traffic types to be in the public network.</td>
</tr>
<tr>
<td>• Do not use virtual IP or secondary addresses for network isolation.</td>
</tr>
</tbody>
</table>
ECS network performance

Network performance is a major factor that can affect the ability of any cloud storage platform to serve data. When architecting or designing the customer network to connect with ECS data switches, there are some considerations to maintain optimal performance. Data, replication, inter-node traffic and management traffic (i.e. ECS Portal, Rest APIs and traffic to network services such as DNS, AD, etc.), flows through the data switches and thus, a reliable and highly available network is also important.

For production network, one uplink per switch to customer switch is required at the minimum. However, one per switch may not be enough to handle the performance necessary for all traffic specifically in multi-rack and single site deployment or when one switch fails. Inter-node traffic in a single site multi-rack deployment traverses through one rack, up to the customer network and down to the next rack of switches in addition to handling traffic associated with data, replication and management. It is recommended at the minimum, four uplinks per rack (two links per switch) for performance and high availability. Since both the data switches are peers, if link to either switch is broken, one of the other switches are available to handle the traffic.

Network Latency is one of the considerations in multi-site or geo-replicated environments. In a multi-site configuration, recommended maximum latency between two sites is 1000ms.

Understanding workload, deployment, current network infrastructure, requirements and expected performance is fundamental in architecting ECS network connections. Some additional areas to understand include:

- Multi-rack ECS deployment
- Multi-site or geo-replicated deployment
- Rate of data ingress, average size of objects, and expected throughput per location if applicable.
- Read/write ratio
- Customer network infrastructure such as VLANs, specific switch requirements, traffic isolation requirements, known throughput or latency requirements

Network performance is only one aspect of overall ECS performance. The software and hardware stack both contribute as well.

<table>
<thead>
<tr>
<th>Best Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>• A minimum of 4 uplinks per rack (2 links per switch) is recommended to maintain optimal performance in case one of the switches fail.</td>
</tr>
<tr>
<td>• Use enough uplinks to meet any performance requirements.</td>
</tr>
<tr>
<td>• Get a good understanding of workloads, requirements, deployment, current network infrastructure and expected performance.</td>
</tr>
</tbody>
</table>
Tools

This section describes tools available for planning, troubleshooting and monitoring of ECS networking.

8.1 ECS portal
The WebUI provides a view of network metrics within ECS. For instance, the average bandwidth of the network interfaces on the nodes can be viewed in the Node and Process Health page. The Traffic Metrics page provides read and write metrics at the site and individual node level. It shows the read and write latency in milliseconds, the read and write bandwidth in bytes/second and read and write transactions per second. The Geo-Replication monitor page shows information relating to geo-replication occurring between sites. For instance, the rates and chunks page provide the current read and write rates for geo-replication and the chunks broken by user data, metadata and XOR data pending for replication by replication group or remote site. ECS portal also provide a way to filter based on timeframe to get a historical view of traffic. Note that updates to any rate information in the ECS portal can take some time. For more information on the ECS Portal refer to the most recent ECS Administration Guide.

8.2 ECS designer and planning guide
The ECS Designer is a tool to assist in streamlining the planning and deployment of ECS. It integrates the ECS Configuration Guide with the internal validation process. The tool is in spreadsheet format and inputs are color coded to indicate which fields require customer information. The sheets are ordered in a workflow to guide the architects in the planning.

Contact your account team to obtain the latest copy of the ECS Designer.

Also available is an ECS Planning Guide that provides information on planning an ECS installation, site preparation, ECS installation readiness checklist and echoes the considerations discussed in this white paper.

8.3 Secure remote services
Secure Remote Services (SRS) provides secure two-way communication between customer equipment and EMC support. It leads to faster problem resolution with proactive remote monitoring and repair. SRS traffic goes through the ECS public network and not the RMM access ports on the ECS private network. SRS can enhance customer experience by streamlining the identification, troubleshooting and resolution of issues.

8.4 Linux or HAL tools
ECS software runs on a Linux operating system. Common Linux tools can be utilized to validate or get information on ECS network configurations. Some tools useful for this include: `ifconfig`, `netstat` and `route`. Also useful are the HAL tools such as `getrackinfo`. Below is some output as examples.

For instance, to validate if network separation configuration is working, running and filtering the `netstat` command for processes that are part of the `object-main` container. A truncated output of netstat below shows the open ports and processes using it such as the `georeceiver` used by `object-main` container to pass around the data and `nginx` directs requests for the user interfaces.
The following code shows an example of truncated output of `netstat` to validate network separation.

```
admin@memphis-pansy:/opt/emc/caspian/fabric/agent> sudo netstat -nap | grep georeceiver | head -n 3
```

```
tcp 0 0 10.10.10.55:9098 :::* LISTEN 40339/georeceiver
tcp 0 0 10.10.30.55:9094 :::* LISTEN 40339/georeceiver
tcp 0 0 10.10.30.55:9095 :::* LISTEN 40339/georeceiver
```

Another tool that can validate the setup of network separation is the `domulti wicked ifstatus public.<traffic type>` command which shows the state of the network interfaces. The state of each interface should be `up`. Here is the command being used to check the `public.data` interface.

```
admin@boston-pansy:~> domulti wicked ifstatus public.data
```

```
192.168.219.9
----------------------------------------
public.data     up
  link:     #14, state up, mtu 1500
  type:     vlan public[1000], hwaddr 00:1e:67:e3:1c:46
  config:   compat:suse:/etc/sysconfig/network/ifcfg-public.data
  leases:   ipv4 static granted
  addr:     ipv4 10.10.10.35/24 [static]
192.168.219.10
----------------------------------------
public.data     up
  link:     #13, state up, mtu 1500
  type:     vlan public[1000], hwaddr 00:1e:67:e3:28:72
  config:   compat:suse:/etc/sysconfig/network/ifcfg-public.data
  leases:   ipv4 static granted
  addr:     ipv4 10.10.10.36/24 [static]
192.168.219.11
----------------------------------------
public.data     up
  link:     #13, state up, mtu 1500
  type:     vlan public[1000], hwaddr 00:1e:67:e3:29:7e
  config:   compat:suse:/etc/sysconfig/network/ifcfg-public.data
  leases:   ipv4 static granted
  addr:     ipv4 10.10.10.37/24 [static]
192.168.219.12
----------------------------------------
public.data     up
  link:     #11, state up, mtu 1500
  type:     vlan public[1000], hwaddr 00:1e:67:e3:12:be
  config:   compat:suse:/etc/sysconfig/network/ifcfg-public.data
  leases:   ipv4 static granted
  addr:     ipv4 10.10.10.38/24 [static]
```
Some of the HAL tools were covered in the Network Separation section, however, here is an output of `getrackinfo -a` that lists the IP addresses, RMM MAC and Public MAC across nodes within an ECS rack.

```
admin@hop-u300-12-pub-01:~> getrackinfo -a
```

<table>
<thead>
<tr>
<th>Node IP Address</th>
<th>Id</th>
<th>Status</th>
<th>Public Mac</th>
<th>Private IP Address</th>
<th>RMM Mac</th>
<th>Public IP Address</th>
<th>Node Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.168.219.5</td>
<td>N/A</td>
<td>noLink</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>192.168.219.6</td>
<td>N/A</td>
<td>noLink</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>192.168.219.7</td>
<td>N/A</td>
<td>noLink</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Best Practices**

- Use ECS Designer to assist in the planning of ECS network with customer network.
- Use the ECS portal to monitor traffic and alerts.
- Setup and enable SRS for streamlining of issues to Dell EMC support team.
Network services

9 Network services

Certain external network services need to be reachable by the ECS system which includes:

- **Authentication Providers** - (optional) System and namespace administrative users can be authenticated using Active Directory and LDAP. Swift object users can be authenticated using Keystone. Authentication providers are not required for ECS. ECS has local user management built-in. Local users on ECS are not replicated between VDCs.
- **DNS Server** - (required) Domain Name server or forwarder.
- **NTP Server** - (required) Network Time Protocol server. Please refer to the NTP best practices for guidance on optimum configuration.
- **SMTP Server** - (optional) Simple Mail Transfer Protocol Server is used for sending alerts and reporting from the ECS rack.
- **DHCP server** - (optional) only necessary if assigning IP addresses via DHCP.
- **Load Balancer** - (optional but **HIGHLY** recommended) evenly distributes client/application load across all available ECS nodes.

Also, the data switch uplinks would need to reside in the same network or accessible by the ECS system. The ECS General Best Practices white paper provides additional information on these network services. Also available are white papers that exemplify on how to deploy ECS with vendor specific load balancers.
10 Conclusion

ECS supports specific network hardware and configurations in addition to customer variations and requirements. The switches utilized as part of the ECS hardware infrastructure provide the backbone for the ECS communication paths to the customer network, node to node communication as well as node and cluster wide management. It is best practice to architect ECS networking to be reliable, highly available and performant. There are tools to assist in planning, monitoring and diagnosing of the ECS network. Customers are encouraged to work closely with Dell EMC personnel to assist in providing the optimal ECS network configuration to meet their requirements.
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.

**ECS Appliance- Special Feature Configuration Support** is an internal-only authenticated document describes how to configure some non-default features of the Dell EMC ECS Gen1/Gen2/Gen3 appliance.