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>
</tbody>
</table>

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 © December 2016 – August 2017 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners. Published in the USA [11/8/2018] [Technical Whitepaper] [H15718]

Dell believes the information in this document is accurate as of its publication date. The information is subject to change without notice.
Table of Contents

Revisions............................................................................................................................................................................. 2

Executive Summary ............................................................................................................................................................ 5

1 Introduction................................................................................................................................................................... 6
   1.1 Audience ............................................................................................................................................................. 6
   1.2 Scope.................................................................................................................................................................. 6

2 ECS Overview ............................................................................................................................................................... 7

3 ECS Network Overview ................................................................................................................................................ 8
   3.1.1 Traffic Types ....................................................................................................................................................... 8

4 ECS Network Hardware ............................................................................................................................................... 9
   4.1 Supported Switches ............................................................................................................................................ 9
   4.1.1 10 GbE Switches – Production Data .................................................................................................................. 9
   4.1.2 1 GbE Switch – Internal Private Management ................................................................................................. 10
   4.2 Customer Provided Network Switches ............................................................................................................. 12

5 ECS Network Configurations ...................................................................................................................................... 13
   5.1 Production Network .......................................................................................................................................... 13
   5.1.1 Top of Rack Data Switch Configurations and Node Links ............................................................................... 15
   5.1.2 Customer Uplinks Configurations ..................................................................................................................... 17
   5.1.3 Network Configuration Custom Requests ........................................................................................................ 20
   5.2 Internal Private Network (Nile Area Network) ................................................................................................... 22
   5.2.1 NAN Topologies ................................................................................................................................................ 22
   5.2.2 Segment LAN ................................................................................................................................................... 24
   5.2.3 Cluster LAN ...................................................................................................................................................... 24
   5.2.4 RMM Access from Customer Network (Optional) ............................................................................................ 25

6 ECS Network Separation ........................................................................................................................................... 30
   6.1.1 Network Separation Configurations .................................................................................................................. 31
   6.1.2 ECS Switch Configurations for Network Separation ........................................................................................ 35

7 ECS Network Performance ........................................................................................................................................ 42

8 Tools ........................................................................................................................................................................... 44
   8.1 ECS Designer and Planning Guide .................................................................................................................. 44
   8.2 ECS Portal ........................................................................................................................................................ 44
   8.3 Dell EMC Secure Remote Services (ESRS) .................................................................................................... 44
   8.4 Linux or HAL Tools ........................................................................................................................................... 45
Executive Summary

ECS™ is 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 types of deployment, a network infrastructure is required for the interconnection between the nodes and connection to customer environments for object storage access. This paper delves into ECS networking topologies in both internal and external configurations and describes some of the best practices.
1 Introduction
This whitepaper provides details on ECS™ networking. Specifics on ECS network hardware, network configurations, and network separation are discussed. It will also describe some ECS networking best practices.

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’s environment. Networking best practices, monitoring and some troubleshooting will be described.

1.2 Scope
This whitepaper explains ECS network configurations and topologies and provides some best practices. It does not cover ECS network installation and administration. Refer to ECS Product Documentation for more information on ECS installation and administration.

Updates to this document are done periodically and coincides usually with new features and functionality changes. To get the latest version of this document, please download from this link.
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. ECS can be deployed as an appliance or a software-only solution. The main use cases for ECS include: archival, global content repository, storage for “Internet of Things”, video surveillance, data lake foundation and modern applications.

ECS software and hardware components work in concert for un-paralleled object and file access. It can be viewed as a set of layered components consisting of the following:

- **ECS Portal and Provisioning Services** – provides a 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** – offers a turnkey appliance or qualified industry standard hardware composed of x86 nodes with internal disks or attached to disk array enclosures with disks, and top of rack switches.

For more in-depth architecture of ECS, refer to the [ECS Architecture and Overview](#) whitepaper.
3 ECS Network Overview

ECS network infrastructure consists of top of rack switches allowing for the following types of network connections:

- **Production Network** – connection between the customer network and ECS providing data.
- **Internal Private Network** – also known internally as “Nile Area Network” and is mainly for management of nodes and switches within the rack and across racks.

The top of rack switches include 10 GbE network switches for data and internal communication between the nodes and a 1 GbE network switch for remote management, console, and install manager (PXE booting) which enables rack management and cluster wide management and provisioning. From these switches, uplink connections are presented to customer’s production environment for storage access and management of ECS. The networking configurations for ECS are recommended to be redundant and highly available.

3.1.1 Traffic Types

Understanding the traffic types or patterns within ECS and the customer environment is useful for architecting the network physical and logical layout and configuration for ECS. The main types of communication or traffic that flow in and out as well as across the nodes on the 10 GbE network include:

- **Data Traffic** – customer data and I/O requests
- **Management Traffic** – provisioning and/or querying ECS via the portal and/or ECS Rest Management APIs as well as network services traffic such DNS, AD, and NTP.
- **Inter-node Traffic** – messages are sent between nodes to process I/O requests depending on owner of data and inter-node checks.
- **Replication Traffic** – 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 an ECS multi-site or geo-replicated deployment, all the above traffic will also go across the WAN.

For the 1 GbE network which is under Dell EMC control is entirely for node and switch maintenance and thus traffic types include:

- **Segment Maintenance management traffic** – traffic associated with administration, installation or setup of nodes and switches within rack.
- **Cluster Maintenance management traffic** - traffic associated with administration, installation or setup of nodes across racks within a site.
4 ECS Network Hardware

ECS supports Arista switches for the appliance and certified industry standard hardware for the 10 GbE switch. For the 1 GbE switch supported models are Arista and Cisco for the appliance and only Arista for ECS Software only on certified industry standard hardware offering. Customer provided switches can be used; however, customers would need to submit a Request for Product Qualification (RPQ) by contacting Dell EMC Global Business Service (GBS).

4.1 Supported Switches

There are several Arista models supported and the descriptions on ports and connections in this section shows examples of some of the models. For more detailed switch descriptions based on models and what is currently supported, please refer to the ECS Hardware and Cabling Guide.

4.1.1 10 GbE Switches – Production Data

Two 10 GbE, 24-port or 52-port Arista switches are used for data transfer to and from customer applications as well as for internal node-to-node communications. These switches are connected to the ECS nodes in the same rack. The switches employ the Multi-Chassis Link Aggregation (MLAG) feature, logically linking the switches and enabling active-active paths between the nodes and customer applications. Figure 1 shows an example view of the data switches and the port assignments for a 52 port Arista switch.

Figure 1 – Top of Rack Data Switches and Port Assignments

Each node has two 10 GbE ports, appearing to the outside world as one port via NIC bonding. Each 10 GbE port connects to one port in the top 10 GbE switch and one port in the bottom 10 GbE switch as pictured in Figure 2 below. These public ports on the ECS nodes get their IP addresses from the customer’s network, either statically or via a DHCP server. There are 8 SFP+ uplinks to customer network where 1 uplink is required at the minimum and 8 maximum per switch. The management ports on the data switches are connected to the 1 GbE management switch labeled 1 GbE management in figure below. As mentioned, both of the top of rack data switches are MLAG’ed together for redundancy and resiliency in case one of the switches fails.
4.1.2 1 GbE Switch – Internal Private Management

The 52-port 1 GbE Arista switch is used by ECS for node management and out-of-band management communication between the customer’s network and the Remote Management Module (RMM) ports of the individual nodes. The main purpose of this switch is for remote management and console, install manager (PXE booting), and enables rack management and cluster wide management and provisioning. Figure 3 shows a front view of an Arista 1 GbE 52 ports switch and port assignments.

Figure 3 – Example of 1 GbE Arista Management Switch and Port Assignments

In addition to Arista, there is now support for Cisco 52 port 1 GbE switch for management. This switch is meant to support customers who have strict Cisco only requirements. It is available only for new racks and is...
not supported to replace Arista management switches in existing racks. The configuration files will be pre-loaded in manufacturing and will still be under control of Dell EMC personnel. ECS 3.0 (with patches) is the minimum to support the Cisco management switch. Figure 4 illustrates the front view of a Cisco 1 GbE management switch and port assignments.

Figure 4 - Example of Cisco Management Switch and Port Assignments

The 1 GbE management port in a node connects to an appropriate port in the 1 GbE management switch and has a private address of 192.168.219.X. Each node also has a connection between its RMM port and a port in the 1 GbE switch, which in turn can have access to a customer’s network to provide out-of-band management of the nodes. To enable access for the RMM ports to the customer’s network, Ports 51 and/or 52 in the management switch are linked to the customer’s network directly. The RMM port is used by Dell EMC field service personnel for maintenance, troubleshooting and installation. Figure 5 depicts the network cabling for the 1 GbE management switch with the nodes and the 10 GbE switches.

Figure 5 - Example of 1 GbE Management Network Cabling for a Four Node ECS Single Rack
Best Practice

- 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 thru 24 can cause installation issues.
- RMM Connections are optional and best practice is to ask customer requirements for these connections. Refer to RMM Connection Section in this whitepaper for details.

You can expand an ECS rack by connecting ports 51 and 52 of management switch on a new rack to an existing rack’s ports 51 and 52 management switch. Topologies for multi-rack setup are described in the Internal Private Network section of this whitepaper.

4.2 Customer Provided Network Switches

The flexibility of ECS allows for variations of network hardware and configurations, however, a Request for Product Qualification (RPQ) would need to be submitted. An RPQ is a request for approval or review of non-standard configuration. An RPQ allows for Dell EMC product teams to review, offer guidance, and determine if ECS can properly function with the proposed configuration. As part of the RPQ, customers may need to submit physical and logical network diagrams which include information such as number of uplinks, upstream switch models, cabling, IP addresses, Virtual Local Area Networks (VLANs), Multi-Chassis Link Aggregation Group (MLAG), etc. Some popular network hardware that has gone thru the RPQ process includes the Cisco 3548P or 9372PX. For customer provided switches, configuration and support are the responsibility of the customer. Dell-EMC assistance is completely advisory for customer provided switches.

An RPQ cannot be submitted to replace the supported management switch for an ECS Appliance since these are solely for administration, installation, diagnosing and management of ECS nodes and switches. The supported management switch would need to remain under control of Dell EMC personnel. For ECS Software only solution, customer provided management and data switches can be implemented, however, an RPQ would also need to be submitted.

Best Practice

- Use 10 GbE switches for optimal performance.
- For high availability use two switches for ECS.
- Have dedicated switches for ECS and do not use “shared ports” on customer core network.
5 ECS Network Configurations

The previous section described the network hardware used for ECS, port connections and gives examples of the physical connections between nodes and the switches within a single rack. This section will explore the “production network” and ECS internal management network referred to as the “Nile Area Network” or internal network. 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 customer’s network and the two top of rack ECS 10 GbE 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 6 for a single rack.

Figure 6 - Production Network Traffic Flow from Customer Network to ECS Switches and Nodes in a Single Rack

For multi-rack, inter-node traffic will flow "north to south" and over to the customer network and to the other ECS racks as shown in Figure 7.
Customers can choose to add an aggregation switch external to the ECS racks so inter-node traffic will not flow thru their main network switch as shown in Figure 8.

Figure 7 - Production Network Traffic Flow from Customer Network to ECS Switches and Nodes in a Multi-Rack

Figure 8 - Traffic Flow with Aggregation Switch for Multi-Rack
Network connections in the production network as a best practice should be designed for high availability, resiliency and optimal network performance.

5.1.1 Top of Rack Data Switch Configurations and Node Links

Link Aggregation Group (LAG) combines or aggregates multiple network ports logically resulting in higher bandwidth, resiliency and redundancy in the data path. The ECS top of rack 10 GbE data switches utilizes the Multi-Chassis Link Aggregation Group (MLAG). MLAG is a type of LAG where the end ports are on separate chassis and are logically connected to appear and act as one large switch. The benefits of using MLAG include increased network bandwidth and redundancy. Each node within ECS has two network interfaces cards (NICs) bonded together using Linux bonding driver. The node bonds the two NICs into a single bonding interface and each NIC is connected to one port on top switch and to one port on the bottom switch. With MLAG switch pair and NIC bonding on the nodes, it forms an active-active and redundant connection between the ECS nodes and switches. Figure 9 is an example of one node bonded with the MLAG switch pair.

Figure 9 - Example of Node and 10 GbE Data Switches Connectivity

Figure 10 displays snippets of the basic configuration files for the 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 nodes ports are not yet configured as LAG, this setting allows for PXE booting from 10 GbE port of the nodes and forwarding of traffic.
Another snippet of the basic configuration files is shown in Figure 11 and illustrates 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.

The data switches are pre-configured on the ECS supported Arista switches. The configuration files for the data switches are located on each node in directory `/usr/share/emc-arista-firmware/config/ecs`. 

![10GbE (Bottom)](image)

<table>
<thead>
<tr>
<th>!10GbE (Bottom)</th>
</tr>
</thead>
<tbody>
<tr>
<td>interface Ethernet9</td>
</tr>
<tr>
<td>description MLAG group 1</td>
</tr>
<tr>
<td>channel-group 1 mode active</td>
</tr>
<tr>
<td>lacp port-priority 1</td>
</tr>
<tr>
<td>!</td>
</tr>
<tr>
<td>interface Port-Channel1</td>
</tr>
<tr>
<td>description Nile Node01 (Data) MLAG 1</td>
</tr>
<tr>
<td>port-channel lacp fallback</td>
</tr>
<tr>
<td>port-channel lacp fallback timeout 1</td>
</tr>
<tr>
<td>mlag 1</td>
</tr>
<tr>
<td>spanning-tree portfast</td>
</tr>
<tr>
<td>spanning-tree bpduguard enable</td>
</tr>
</tbody>
</table>

![10GbE (Top)](image)

<table>
<thead>
<tr>
<th>!10GbE (Top)</th>
</tr>
</thead>
<tbody>
<tr>
<td>interface Ethernet9</td>
</tr>
<tr>
<td>description MLAG group 1</td>
</tr>
<tr>
<td>channel-group 1 mode active</td>
</tr>
<tr>
<td>lacp port-priority 2</td>
</tr>
<tr>
<td>!</td>
</tr>
<tr>
<td>interface Port-Channel1</td>
</tr>
<tr>
<td>description Nile Node01 (Data) MLAG 1</td>
</tr>
<tr>
<td>port-channel lacp fallback</td>
</tr>
<tr>
<td>port-channel lacp fallback timeout 1</td>
</tr>
<tr>
<td>mlag 1</td>
</tr>
<tr>
<td>spanning-tree portfast</td>
</tr>
<tr>
<td>spanning-tree bpduguard enable</td>
</tr>
</tbody>
</table>

![Figure 11 - Snippet of Data Switch Configuration Files Showing the MLAG Peer Link Definitions Between the Switches.](image)
5.1.2 Customer Uplinks Configurations

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.

The first eight ports on each data switches are available to connect up to the customer network, providing 16 total ports total per rack. 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 in Figure 12. 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 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.

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 Cisco VPC or Arista MLAG. Customers are required to provide the
necessary connection information to establish communication to the nodes in the rack. Figures 13-17 illustrate some of the possible link connections and best practices.

Figure 13 is a single link connected to a single ECS data switch. In this setup, there is no redundancy and is only meant for non-production environments.

Figure 13 – Single Customer Switch with Single Link to one Data Switch for Non-Production Environments

A single switch with multiple links to each data switch is illustrated in Figure 14. Customers would need to create a port channel using LACP in active or passive mode. There should be an even number of links and should be split evenly between the data switches for proper and efficient network load balancing to the ECS nodes.

Figure 14 - Single Customer Switch with Multiple Links

Figure 15 is a customer example of a two port LAG for an Arista and Cisco single switch with multiple links.

Figure 15 - Example of a two port LAG for Customer’s Arista and Cisco switches

!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 16 and 17 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 17. 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 the customer would need to create port channels using LACP in active or passive mode on all switches participating in the multi-switch LAG. In Figure 18 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 in order to create a single port channel group.
Best Practice

- 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 other switch will only become active if one switch goes down.
- Balance 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.

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 network isolation between traffic types for security purposes. In these types of scenarios, an RPQ would need to be submitted to Dell EMC for review and approval. As an example, in Figure 19 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.
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. Figure 20 shows a sample script of how to change the VLAN membership. Both data switches would need to have the same VLAN configuration.

Figure 20 - Sample Script to Modify VLAN Membership on 10 GbE Data Switches

```
# 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-channel1-12
switchport access vlan 10

copy running-config startup-config
```

**Best Practice**

- Submit an RPQ for network configurations requiring modification to the basic default configuration of ECS switches
5.2 Internal Private Network (Nile Area 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. As previously mentioned, ports 51 and 52 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** – uplink ports 51 or 52 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 master node when a command like "setrackinfo" is done. 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 ports 51 and/or 52. 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 21 shows a simple topology linearly connecting the segments via ports 51 and 52 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.

Figure 21 - Linear or Daisy Chain Topology

Another way to connect the segments is in a ring topology as illustrated in Figure 22. 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.
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 10 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 23 would need to be added and would be an extra cost; however, it is the most reliable among the NAN topologies.

**Best Practice**

- 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 10 GbE switches within a rack to a LAN identified as VLAN 2. This consists of ports 1 thru 24 and ports 49 and 50 on the management and 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 data switch (bottom)
- **192.168.219.253** – reserved for the data switch (top)

Figure 24 identifies the ports associated with the Segment LAN (VLAN 2 untagged).

Figure 24 - VLAN Membership of Segment LAN

![Port 1-24 Management](image)

Best Practice

- 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 together to create a single Cluster LAN for administration and access to the entire cluster. Ports 51 and 52, also referred to as the cluster interconnect ports, on management switch can be connected to another management switch ports 51 and/or 52. In addition to the cluster interconnect ports, the blue network ports 1-24, the RMM ports 25-48 are members of the Cluster LAN. All members will tag their IP traffic with VLAN ID 4 as shown in Figure 25 and communicate via the
IPv4 link local subnet. During install of ECS software, all nodes in the rack will be assigned a unique color number. The color number acts as the segment ID and will be 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 \textbf{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.

![Figure 25 - VLAN Membership of Cluster LAN](image)

**Best Practice**

- 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)

The RMM ports 25-48 provide the out-of-band and remote management of the nodes via ports 51 or 52 on the management switch as depicted in Figure 26. 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.

![Figure 26 - VLAN Membership of RMM](image)
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 are able to 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 in Figure 27 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 to traverse through the customer network, then the switchport is configured as trunk and customer switch would need to be configured to forward VLAN 4 as tagged traffic to only the ports connected 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; however, an RPQ would need to be submitted.

**Figure 27 - Snippet of the Arista Switch Configuration for Ports 51 and 52**

```
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 out-of-band and remote RMM access to customers. In the linear topology, the rack segments are connected together end to end using port 51 and 52 as shown in Figure 28 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.
A customer switch with ECS segments in a ring topology is shown in Figure 29. 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 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.

In a star topology as depicted in Figure 30, port 51 on 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.
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 Figure 31.

![Diagram of Customer Switch with ECS Segments in a Star Topology]

Figure 30 - Customer Switch with ECS Segments in a Star Topology

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.

```
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
```
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.

<table>
<thead>
<tr>
<th>Best Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Providing RMM to customer is optional. Ask customer if it is absolutely required before setting up.</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>
6 ECS Network Separation

In ECS 2.2.1 and later, ECS network separation was introduced for segregating different types of network traffic for security, granular metering, and performance isolation. The main types of traffic that can be separated include:

- **Management Traffic** – traffic related to provisioning and administering via the ECS Portal and traffic from the operating system such as DNS, NTP, and ESRS.
- **Replication Traffic** – traffic between nodes in a replication group.
- **Data Traffic** – 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. Network separation is currently only supported for new ECS installations.

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

Table 1 - Services Associated with Logical Network.

<table>
<thead>
<tr>
<th>Services</th>
<th>Logical Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECS Portal, Provisioning, metering and management API, ssh</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, XOR</td>
<td>CAS only Data network (public.data1)</td>
</tr>
<tr>
<td>Replication data, XOR</td>
<td>Replication Network (public.repl)</td>
</tr>
<tr>
<td>ESRS (Dell EMC Secure Remote Services)</td>
<td>Based on the network that the ESRS Gateway is attached (public.data or public.mgmt)</td>
</tr>
<tr>
<td>DNS, NTP, AD, SMTP</td>
<td>Management network (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. For physical network separation,
customers would need to submit a Request for Product Qualification (RPQ) by contacting Dell EMC Global Business Service (GBS).

6.1.1 Network Separation Configurations

In addition to the default network configuration, network can be partially separated or all separated. Figure 32 illustrates examples of network separation configurations. For example,

- **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.
- **All Separated** – each traffic type is on its own VLAN in addition to the default “public” VLAN

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 has to match the customer’s VLAN ID.

Figure 32 – 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. So, planning for network separation requires decisions on how traffic should be segregated in VLANs, the static IP addresses, subnet, and gateway information needs to be determined. After network separation has been completed, you will observe that a virtual interface will be created for the VLANs and the interface configuration files will be of the form ifcfg-public.{vlanID} as shown in Figure 33 where the different traffic types are encased in different shapes.
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 in Figure 34.

The Hardware Abstraction Layer (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 in Figure 35. 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 who would create a JavaScript Object Notation (JSON file) with IPs and interface names and supplies this information to the object container. Figure 36 is an output from Fabric Command Line (fcli) showing the format of the JSON structure.

Figure 36 - JSON structure from the FCLI Command

```
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 Figure 37 in which the object layer can utilize to separate ECS network traffic.

Figure 37 - Contents of /host/data/network.json file.

```
{
  "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 packets take through the network. In general, the path packets come in will be the same path going out. Based on the "ip rules", local node originating the packet looks at the IP and first look 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.

In ECS 3.1, new features have been added for network separation which includes:

- Management of static routes using the “setrackinfo” command – in previous releases, static routes were set manually using interface specific route files; however, these can now be set using “setrackinfo” command.
- Ability to configure a second data network for CAS traffic only – another separate data network (public.data1) can be setup for CAS access.

Refer to ECS Networks and Node IP Change document from Solve on how to plan for network separation and instructions on how to configure.

### 6.1.2 ECS Switch Configurations 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. As previously discussed, physical separation would require an RPQ.

#### 6.1.2.1 Standard (Default)

The default settings for the standard data switches use the configuration files that are bundled with ECS. So 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 38 provide an example of a default ECS network setup with customer switches.

<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>

Table 2 – Standard Default Switch Configuration
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 39, data and replication traffic is 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 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>
Both data switch configurations files would need to be modified to handle the VLANs in above example. Figure 40 defines 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.

Figure 40 – Example Snippet of Single Domain Switch Settings with Two VLANs for 10 GbE Data Switches

```plaintext
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 thru 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 41 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>All Traffic Tagged</td>
</tr>
<tr>
<td>REPL</td>
<td>20</td>
<td>Yes (OS Level)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 41- 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 Figure 42, 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.

Figure 42 - Example Snippet of Single Domain with Two VLANs and Public in a VLAN Settings for 10 GbE Data Switches

```
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**

An RPQ needs to be submitted for physical separation of ECS traffic. For physical separation, an example setup may include multiple domains on the customer network defined for each type of traffic. Example of the setup and details are defined in Table 5 and illustrated in Figure 43. 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 MLAG together.

<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>

Table 5  An Example of Physical Separation Configuration

Figure 43- An Example of Physical Separation Setup

Figure 44 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”
For situations where customers would want the public network on a VLAN, Table 6 and Figure 45 provides example details of the configuration. In this case, all traffic are 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>
Best Practice

- Network separation is optional, it is important to ask WHY customer wants it to determine fit and best configuration.
- Keep the management traffic within the public to reduce the number of static routes.
- Although it is allowed to have only default gateway in public, it is recommended to have at least one of the traffic types to be in the public network.
- Do not use virtual IP/Secondary addresses for network isolation.
- Submit an RPQ for physical separation.

Figure 45 - Example Snippet of Arista Data Switches Settings of Physical Separation where 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
```
ECS Network Performance

Each customer uplink port on the supported ECS data switches has 10 Gigabit/s of bandwidth. The ports on both switches are LACP/MLAG and when all 16 ports are utilized it can provide 160 Gigabit/s of bandwidth. Network performance is one of the key factors 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 thru 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, it may not be sufficient to handle the performance necessary for all traffic specifically in multi-rack and single site deployment or when one switch fails. As discussed, inter-node traffic in a single site multi-rack deployment would need to traverse thru 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. So, 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 MLAG together, 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 the customer’s workload, deployment, current network infrastructure, requirements and expected performance is fundamental in architecting ECS network connections to customer. Some things to inquire about or 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 (VLANs, specific switch requirements, isolation of traffic required, known throughput or latency requirements, etc).

Network performance is only one aspect of overall ECS performance. The software and hardware stack both contribute as well. For overall ECS performance refer to the following whitepapers listed below (some of the links maybe internal only):

- ECS Performance Whitepaper
- ECS – Single Site CAS API Performance Analysis
- Hortonworks HDP with ECS-EMC ETD Validation Brief
**Best Practice**

- A minimum of 4 uplinks per rack (2 links per switch) is recommended to maintain optimal performance in case one of the switches fail
- Use sufficient uplinks to meet customer performance requirement.
- Get a good understanding of customer workloads, requirements, deployment, current network infrastructure and expected performance.
8 Tools
Several tools are available for planning, troubleshooting and monitoring of ECS networking. For instance, ECS Designer is a tool that is useful to help in planning for adding ECS on the customer network. There are other ECS specific tools and Linux commands for network monitoring and troubleshooting. This section will describe these tools.

8.1 ECS Designer and Planning Guide
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. It is designed to handle up to 20 racks with eight sites. It allows for ESRS and topology information to be created.

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

8.2 ECS Portal
Network related traffic metrics are reported on ECS portal. There are several portal screens or pages available to get 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 provides 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 ECS Administration Guide.

8.3 Dell EMC Secure Remote Services (ESRS)
Dell EMC Secure Remote Services (ESRS) provides a secure two-way connection between customer owned Dell EMC equipment and EMC customer service. It provides faster problem resolution with proactive remote monitoring and repair. ESRS traffic goes thru the ECS production network and not the RMM access ports on the ECS internal private network. ESRS offers a better customer experience by streamlining the identification, troubleshooting and resolution of customer issues. The configuration files and notification delivered via ESRS is useful in providing information to engineering and development teams on systems installed at customers’ sites.

For more information on ESRS, refer to the Enablement Center for EMC Secure Remote Services (ESRS) at this site: https://www.emc.com/auth/rpage/service-enablement-center/ec-for-esrs.htm
8.4 Linux or HAL Tools

ECS software runs on Linux Operating system that acts as the infrastructure layer. There are Linux tools that can be utilized to validate or get information on ECS network configurations. Some Linux tools useful for this include: ifconfig, netstat, and route. Also useful are the HAL tools such as "getrackinfo". Below are screenshots of some commands and sample output.

For instance, to validate if network separation configuration is working, run the "netstat" command and filter the processes that is part of the object-main container. A truncated output of “netstat” is in Figure 46 showing 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.

Figure 46 - Example 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

admin@memphis-pansy:/opt/emc/caspian/fabric/agent> sudo netstat -nap | grep nginx | grep tcp

tcp  0  0  10.10.20.55:80  0.0.0.0:*  LISTEN  68579/nginx.conf
tcp  0  0  127.0.0.1:4443  0.0.0.0:*  LISTEN  68579/nginx.conf
tcp  0  0  10.10.20.55:4443 0.0.0.0:*  LISTEN  68579/nginx.conf
tcp  0  0  10.10.20.55:443  0.0.0.0:*  LISTEN  68579/nginx.conf
```

Another tool that can also validate setup of network separation involves using the "domulti wicked ifstatus public.<traffic type>" command to check the state of the network interfaces. The state of each interface should be “up”. Figure 47 provides a snippet output of command to check the “public.data” interface.
Some of the HAL tools were covered in the Network Separation section, however, here is an output of “getrackinfo –a” in Figure 48 that lists the IP addresses, RMM MAC, and Public MAC across nodes within an ECS rack.
**Figure 48 - Example Output of "getrackinfo -a" Showing Network Interface Information.**

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

<table>
<thead>
<tr>
<th>Node private</th>
<th>Node Id</th>
<th>Status</th>
<th>Mac</th>
<th>Ip Address</th>
<th>Mac</th>
<th>Ip Address</th>
<th>Node Name</th>
</tr>
</thead>
</table>

Best Practice

- 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 ESRS for streamlining of issues to Dell EMC support team.
Network Services

In order to be able to deploy ECS, certain external network services need to be reachable by the ECS system which includes:

- **Authentication Providers** – users (system admin, namespace admin and object users) can be authenticated using Active Directory, LDAP or Keystone
- **DNS Server** – Domain Name server or forwarder
- **NTP Server** – 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** – only if assigning IP addresses via DHCP
- **Load Balancer** - (optional but highly recommended) distributes loads across all data services 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](#) whitepaper provides additional information on these network services. Also available are whitepapers that exemplify on how to deploy ECS with vendor specific load balancers:

- [ECS with HAProxy Load Balancer Deployment Reference Guide](#)
- [ECS with NGINX (OpenResty) Deployment Reference Guide](#)
- [ECS with F5 Deployment Reference Guide](#)
- [ECS with KEMP LoadMaster Deployment Reference Guide](#)
10 Conclusions

ECS supports specific network hardware and configurations in addition to customer variations and requirements via an RPQ. 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.
A References

- IEEE 802.1Q VLAN Tutorial
  - http://www.microhowto.info/tutorials/802_1q.html
- IEEE Spanning Tree overview
- MLAG overview (Refer to Chapter 10)
  - https://www.arista.com/assets/data/pdf/AristaMLAG_tn.pdf
- Link aggregation and LACP overview
  - https://www.thomas-krenn.com/en/wiki/Link_Aggregation_and_LACP_basics
- Request for Product Qualification
  - https://inside.dell.com/docs/DOC-190638

ECS Product Documentation

- ECS product documentation at support site or the community links:
  - https://support.emc.com/products/37254_ECS-Appliance-/Documentation/
  - https://community.emc.com/docs/DOC-53956
- ECS Technical Whitepapers
  - ECS Overview and Architecture
  - ECS Performance Whitepaper
  - ECS General Best Practices
- Load Balancers
  - ECS with HAProxy Load Balancer Deployment Reference Guide
  - ECS with NGINX (OpenResty) Deployment Reference Guide
  - ECS with F5 Deployment Reference Guide