Storage Area Network Design Overview Using Brocade DCX 8510 Backbone Switches

East Carolina University

Paola Stone Martinez

April, 2015

Abstract

The design of a Storage Area Networks is a very complex process. There are different factors that have to be considered before setting it up. The engineers have to consider what the expectations from the users are and which topology could be more adequate for the company. Performance and availability are some characteristics mainly expected from users. Based on the company history, it is necessary to decide if a Core-Edge or Mesh topology would work better for the environment.

Information about Storage Area Network is presented in the first part of this paper; why it is used and its benefits are explained. Concepts of fabric, zoning, and different port types are introduced as well. Then, a brief description of the Brocade DCX 8510 switch is presented including key features and management tools. This is followed by an explanation of how switches are interconnected to form a fabric, and the different available topologies. Finally, the study of a fabric using Brocade DCX 8510 switches is presented.

Storage Area Networks

The quick evolution of technology has been an important factor in the Information Technology (IT) field as most of the daily used systems produce a lot of information that needs to be saved. Companies have also realized that most employees do not need to have their own physical computer and have adopted the use of different virtualization tools. The success of this depends on information and data availability for people to complete their jobs. Because of all these changes, different storage companies have had to improve the performance of their equipment in order to satisfy their users' needs, and the design of storage network architecture quickly has become a popular procedure adopted by them. Network Attached Storage (NAS) and Storage Area Network (SAN) are the two main technologies that have been added to the data infrastructure of most companies as a way to centralize and secure their data storage resources. NAS is a low cost file-level storage that is easy to manage. It can be used for mass file storage and VMware (NFS). SAN is a flexible bloc-level storage that offers raw access to data blocks. It can be used for databases, exchange, VMware, and servers boot (EMC Corporation, 2013, p. 4). This paper presents information related to SAN, basic concepts and considerations when using Brocade DCX 8510 switches.

> Cameron states that:
>
> The implementation of a storage area network allows system and storage administrators to ensure the consistent storage and retrieval of data on a network. The multipath nature of a SAN, with its characteristic one-to-many relationships between host and storage devices, provides unmatched configuration flexibility and availability, as well as the load-balancing and increased connectivity essential to the creation of a scalable network design of new technologies have (as cited in Walker & O'Sullivan, 2009, p. 1).

According to EMC Corporation (2013), the benefits of SAN are (p.8):

- High availability and manageability
- Application performance
- Fast scalability
- Better backup, replication, and recovery options
- Storage consolidation optimizes total cost of ownership

The main purpose of SANs is to interconnect storage devices to servers clients or hosts (Walker, 2005, p. 1). In order to set up this architecture, companies need to invest in switches and fiber channel which are the main tools to build a SAN (Goda & Kitsuregawa, 2012, p. 6).Even though “fiber” and “fibre” are used indistinctly, it is necessary to know their difference and when to use them. The term “fiber” refers to the physical media; the word “fibre” refers to the Fibre Channel protocol and standards.

The switch used to connect all the devices is known as fabric. A fabric is the virtual space where all storage devices communicate with each other and its main function is to receive frames and send it to its target. SAN switches ports are auto-configurable, they can detect what type of devices every ports is connected to and the configuration on the port is set accordingly. The ports will acquire any of the following port types (EMC Corporation, 2012, p.8, 27):

- N or Node Port: transmits and receives Fibre Channel data frames in a Switches Fabric environment.
- F or Fabric Port: Port on a switch that has an N_port plugged into it.
- NL or Node Loop Port: transmits and receives Fibre Channel frames in an arbitrated loop (private) environment.
- FL or Fabric Loop Port: Allows a fabric to communicate with an arbitrated loop (public)
- E or Expansion Port: Port on a switch that has another switch plugged into it
- G or Generic Port: Port on a switch that defects the attached port type and auto-configures to match that port type.

Another important aspect of the fabric is the World Wide Name (WWN) which is a 64-bit address used by Fibre Channel to locate all elements in the network. It is assigned by the different vendors to Host Bus Adapters (HBA) in hosts, front end ports in storage arrays, or switch ports (EMC Corporation, 2012, p.30).

Once all connections are made, all devices are logging in, and the fabric is set up, it is necessary to complete the process of zoning. Zoning is the way in which we specify the fabric what ports are supposed to be communicating, sending and receiving frames between each other. There are two different methods in which zones can be created:

1. WWN zoning: The members of the zone are specified by using their WWN. In this case the connections can be moved to different ports within the fabric and it will not affect the transmission between the nodes
2. Port zoning: The members of the zone are specified by giving the switch port to which the device is connected. In this type of zone, the communication between the two devices will be interrupted if the cable is moved to a different port in the fabric.

All zones are contained in a zone set, and a fabric can only have one active zone set at the time.

Brocade DCX 8510 Backbone

The selection of the switch used to design a SAN is directly related to the number of connections needed from hosts and storage devices as well as the expected performance. In an environment where the number of clients is in the hundreds and storage availability depends on a SAN, using the Brocade DCX 8510 switches could be an assertive decision. EMC Corporation (2013) affirms that using large directors like the DCX lowers the price per port as the port density rises; they also reduce data center operational costs because of a smaller footprint and easier to manage (p. 10).

Brocade (2014) affirms that the DCX switch is suitable for environments where Reliability, Availability, and Serviceability (RAS), performance, and scalability requirements are needed. Reliability of data is accomplished though the use of Error Detection and Fault Isolation (EDFI), monitoring of field-replaceable units (FRUs), and dual control processors. Availability is offered with the redundancy on blades and FRUs; it is also achieved with the use of Fabric Shortest Path First (FSPF) rerouting in case of link failure. Serviceability is achieved mainly with the modular design and hot swappable components, it also offers status LEDs and background health check.

Key features of the switch are (Brocade, 2014, p. 11):

- High density SAN configuration: up to 512 external ports
- High performance port blades
- Flexibility: supports blades running different speeds and self-configuring ports

- High Availability: all parts are redundant
- Fibre Channel over IP (FCIP) and Fibre Channel over Ethernet (FCoE) capable
- Constant monitoring

Brocade switches and ports can be managed by using the command line, but they also count with Web Tools for a more user friendly interface that has three main tabs: Switch view, Port Admin, and Name Server (Brocade, 2013, p. 17). Brocade also provides Connectrix Manager Converged Network Edition (CMCNE) which is a web based client that allows users to provision, monitor, report and troubleshoot different switches belonging to the same or different fabrics. CMCNE is run on a Windows host that accesses the switches (EMC Corporation, 2015, p. 2).

Inter-Switch Link and Inter-Chassis Link

When more than one switch is needed for the configuration of a SAN, it is necessary to interconnect all the switches in order for them to act as a one network. These connections are called Inter-Switch Link (ISL) and Inter-Chassis Link (ICL).

An ISL is the connection between two switches using one of the ports for the available ports for devices. Depending on the number of switches, ISLs can consume a lot of ports reducing the number of available ports in the fabric. ISLs are usually configured as trunks. A trunk is configured when two or more links are able to behave as a one single link. Trunking helps to improve the performance on a fabric by distributing the frames route through out all available links instead of restricting it to a specific one. A trunk will automatically be formed on ISL connections as long as they have the same port settings and are connected to the same port group in the switch. On DCX 8510 switches, each blade counts with forty eight ports, and six port groups: 0-7, 8-15, 24-31, 32-39, and 40-47. A port group can contain different trunks. Trunking capability on the switches depends on the license available to the switch (Smith, Hultman, and Kloepping, 2015, p. 26).

An ICL is a high-performance connection between switches. The ports for ICL connections are located in a switch blade separated from the front-end ones intended for the fabric devices, they used dedicated backplane ports which maximizes the connectivity between the switches. On the DCX 8510 one ICL link is the combination of four 16Gbps links, totalizing the throughput of each cable to 64Gbps. They can reach 2Tbps on ICL bandwidth when using all 16 ports available (EMC Corporation, 2013, p. 18).

Topology Designs

A SAN can be formed by a single switch or multiple switches. There are different ways of classifying the topologies according to how many switches are in the fabric and how the switches are interconnected. According to the number of switches in the fabric, SAN topologies can be Simple Fibre Channel if there are less than four switches interconnected or Complex Fibre Channel SAN if there are four or more interconnected switches (Smith, Hultman, and Kloepping, 2015, p. 16).

While the interconnection in a Simple Fabric is very simple, a Complex Fibre Channel fabric set up needs a little bit more planning on selecting how the switches are going to be connected to each other. There are two main ways in which the can be connected:

1. Core – Edge topology

   There are two different models of this topology: two-tier and three-tier. In the two-tier, there is one edge tier and one core tier; hosts are connected to the edge switches while the storage arrays are connected to the core switches. In the three-tier, there are two edge tiers and one central core; all host are connected to one tier while all storage is connected to the other edge tier. The core is used just for connection between the switches (EMC Corporation, 2013, p.18).

   In this set up the switches are divided in core switches (usually two for redundancy) and edge switches. The core switches are connected to all edge switches, but there is no connection between the cores or between the edges. The data usually travels through the cores.
   When using ICLs the limit of the number of switches depends on the Fabric OS firmware version: for Brocade FOS v7.0.0x the maximum number of switches is six and for Brocade FOS v7.0.1x (and Later), it is ten (Brocade, 2014, p.11).

2. Mesh topology

   It can be a full mesh or partial mesh. In the full mesh, all switches are connected to all the other switches in the fabric. In a partial mesh, switches are connected jus to some switches. In this set up all switches are connected to every switch in the fabric. The number of switches that can be interconnected depends on the Fabric OS: for Brocade FOS v7.0.0x the maximum number of switches is three and for Brocade FOS v7.0.1x (and Later), it is nine (Brocade, 2014, p.11).

In some cases, both main topologies are mixed with a resulting set up known as Compound Core-Edge and Mesh topology.

The advantages and disadvantages of the both topologies are directly related to what is expected from the fabric. Some users' applications might need high performance while some others can need availability. Ideally, to plan the SAN design it would be necessary to know the destination and paths in which all data would travel. Since this information is usually not available when completing a SAN design most efforts are put towards keeping critical connections on core switches and having redundant connections from hosts to their storage devices. SAN designers also consider the number of hops that a host will have when zoned to a specific storage device and tried to keep it as low as it can be. The topology in which most SANs are based is on the Core-Edge mainly because of the port count needed to set up Mesh topologies when using ISLs; it can lower the number of available ports for devices connection drastically.

It is also necessary that SAN designers consider that the fabric could expand in the future; deciding where the new switches and connections could be added will avoid having to re-design and move connections that will affect the production time of the setup.

SAN Design Study and Monitoring Case

In the last year I have been working on resolving performance issues in a fabric reporting frames discard in different environments. The fabric characteristics are as follows:

- It was initially set up as a Core-Edge topology with six Brocade DCX 8510 switches, two cores and four edges.
- Each switch has 384 ports available for hosts and storage devices connections, a total of 2304 ports in the fabric
- It is using ICLs to interconnect the switches
- A seventh DCX switch was added to the fabric using ISLs.
- A hundred more ports were also added using ISLs from a virtual switch.
- The two Core switches are also connected to three Brocade MP8000 Fibre Channel over Ethernet (FCoE) switches using ISLs.

In an effort to reduce the frame discards ISLs were added between the core switches and the edges, the fabric is now a not well planned Compound Core-Edge and Mesh topology which is constantly reporting errors. The fabric has been monitored for some time and the following has been found:

- ISLs are not connected to the same port groups; this makes it impossible for them to form a trunk. There is a lot of bottleneck reporting on the single ISLs.
- Storage devices as well as host are connected all over the switches; there is no logic to the way the connections have been made. It looks like most of them were connected as available ports were found.
- There are hosts zoned to storage devices that have been removed. The hosts are sending requests for non-existing volumes all the time and this is generating a lot of errors on these ports.
- Storage devices configurations have been modified, but the hosts' volumes data base has not been refreshed.
- There is a HBA synchronization problem in host with a specific OS.
- For testing purpose, a jammer has been plugged to one of the core switches. As its name states, the purpose of this device is to generate errors in the switch. It is probably not a good idea to have this device connected to one of the core switches.

Based on the information gathered from the monitoring, the following are some actions that need to be taken in order to reduce and possibly mitigate the errors that are affecting the fabric performance:

- ISLs are being moved to the same port groups in the switches to have them join and form a trunk. This will help to distribute the bandwidth of the links accordingly.
- The fabric has around 3000 active connections that are part of the active zone set. This is being inventoried to register what hosts are connected to what storage device. Once that is done, the plan is to start moving the connections whit in the fabric to reduce the number of hops. If possible, the connections will try to be isolated to the same switch.
- All zones are being checked and member existence confirmed.
- All hosts volumes data bases are being refreshed in order to eliminate the constant request sent from them about storage that is no longer available.
- Drivers are being upgraded in hosts reporting problems.
- Jammer connections are being registered and the possibility of moving them to a different physical switch is being contemplated.
- Zones with members connected to the FCoE switches are being studied for compatibility or performance issues when zoned to 8 or 16GB connections.

This fabric has been poorly managed since its set up, so fixing up the performance issues is going to require a lot of time and planning. There is still a lot of monitoring work to be done in order to figure out all the root causes; and since there are a great number of ports being utilize, finding a way to isolate the connections going to a specific storage device could be a step to take. Connectrix Manager has been of great help to track errors and it has also offer helpful graphics about the behavior of devices.

This is a great example of what can happen if the necessary SAN planning is not done before starting to use it. There is a lot of documentation with recommendations that need to be carefully studied by all people who is setting up SANs. The size of the SAN does not matter; the design should always assume that the environments are going to grow and required more connections added to the fabric.

References

Brocade. (2013). Web Tools Administrator's Guide. Retrieved from https://support.emc.com/docu50173_Brocade-Fabric-OS-7.2.0-Web-Tools-Administrator's-Guide.pdf?language=en_US

Brocade. (2014). Brocade DCX Backbone. Hardware Reference Manual. Retrieved from https://support.emc.com/docu10597_Brocade-DCX-Backbone-Hardware-Reference-Manual.pdf?language=en_US

Brocade Communication Systems. (2014). Brocade SAN Scalability Guidelines: Brocade Fabric OD v7.0.x. Version 7.2, update 1. Retrieved from http://www.brocade.com/downloads/documents/matrices/scalability-matrix-fos-v7-mx.pdf

EMC Corporation. 2012. SAN Storage Protocols (Revision Number: MR-1CP-SYMCFG.5874.2.0). USA: Author

EMC Corporation. 2013. Connectrix Fundamentals (Revision Number: MR-1WP-CONFD.7694.9.0). USA: Author

EMC Corporation. Release Notes. (2015). EMC Connectrix Manager Converged network Edition. Retrieved from https://support.emc.com/docu55806_Connectrix-Manager-Converged-Network-Edition-12.3.1-Release-Notes.pdf?language=en_US

Goda, K., & Kitsuregawa, M. (2012). The History of Storage Systems. doi: 10.1109/JPROC.2012.2189787

Walker, C., & O'Sullivan, M. (2009). Core-Edge design of storage area networks—A Single-edge formulation with problem-specific cuts. doi: 10.1016/j.cor.2009.07.005

Walker, C., O'Sullivan, M., & Thompson, T. (2010). A mixed-integer approach to Core-Edge design of storage area networks. doi:10.1016/j.cor.2005.11.009

Smith, E., Hultman, R.., & Kloepping, D. (2015). Fiber Channel SAN Topologies Version 4.0. Retrieved from https://support.emc.com/docu33923_TechBook:-Fibre-Channel-SAN-Topologies.pdf?language=en_US