Use of Fibe-Channel Full Fabric and Storage Area Network Technology. Are you ready for the Challenge?

Ian D Norman, GRADINE Technical Director
UK Ministry of Defence, Washington DC
Rod Roderique, Principal Systems Engineer
DoD Special Detachment #1, Washington DC
Phone: +1-301-688-9977, e-mail: k3qii@erols.com

Presented at the THIC Meeting at the Sheraton Barcelona
Annapolis MD 21401-3094

May 8, 2001

THIC Technology Conference

8-9 May 2001

"Using Fibre-Channel Full Fabric and Storage Area Network Technology. Are You Ready for the Challenge?"

Presentation by:

Ian D. Norman, GRADINE Technical Director

United Kingdom Ministry of Defence, Washington D.C.

Rod Roderique, Principal Systems Engineer

Department of Defense, Special Detachment #1, Washington D.C.

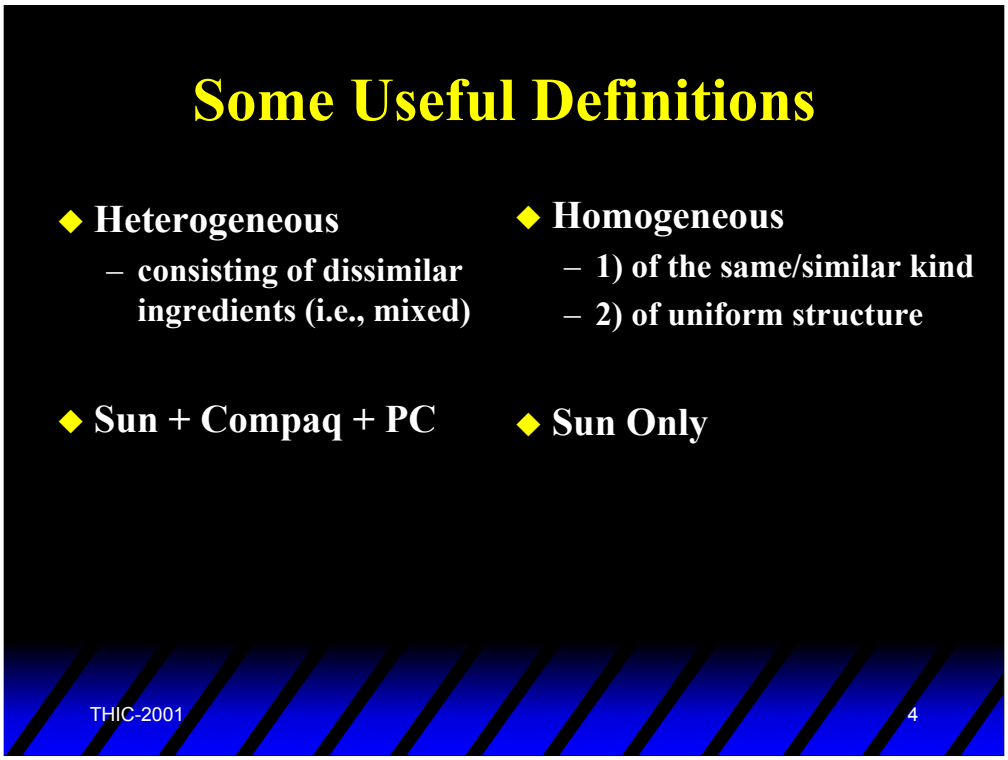
Good Afternoon,

It is my pleasure to speak with you today.

This presentation is a result of many requests for a "tutorial like" overview of material presented at previous THIC conferences.

The next thirty minutes will provide a solid explanation about the architectural issues related to both Full-Fabric Fibre-Channel and Storage Area Networks (SAN) technology.

These technologies are an exact fit to the theme of this conference, "Tailoring Technology to Requirements A Look at User Needs and Emerging Products".



When discussing SAN architectures, it is inevitable that the topic of Heterogeneous and Homogeneous capability arises.

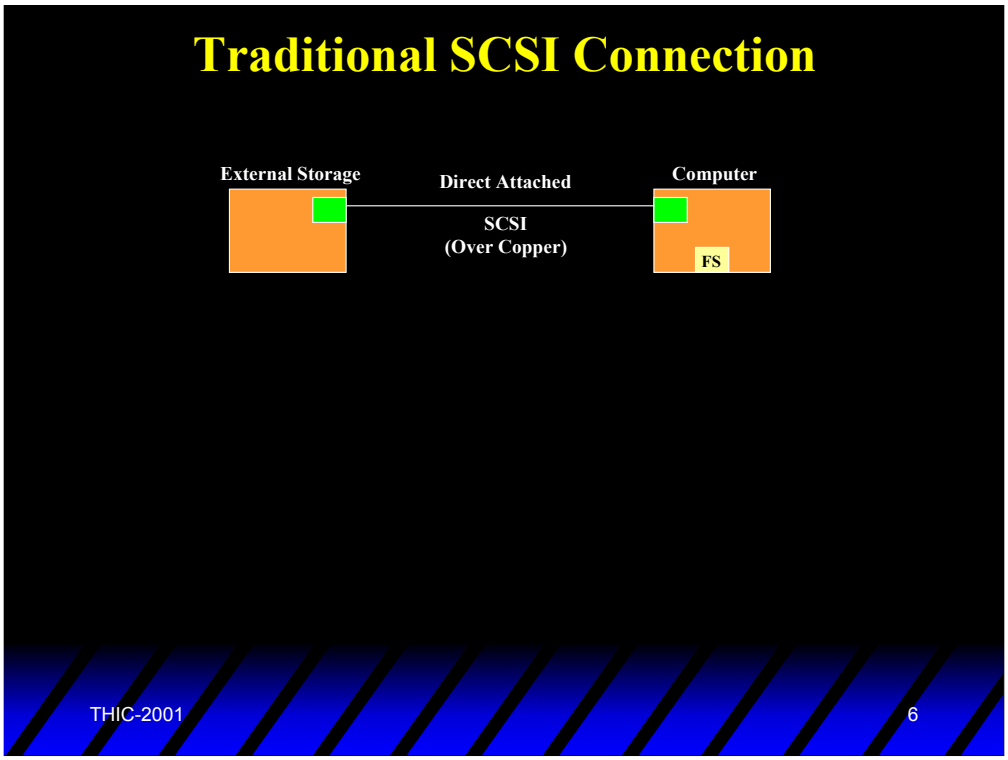
If one has a requirement to incorporate computing resources from a variety of manufacturers, it is a "heterogeneous" environment.

If on the other hand, the architecture utilizes only a single brand of computers, it is a "homogeneous" environment.

As we will see, it is the File System (FS) management software, not the interconnecting "fabric" hardware, that must deal with this issue.

Jumping right in.

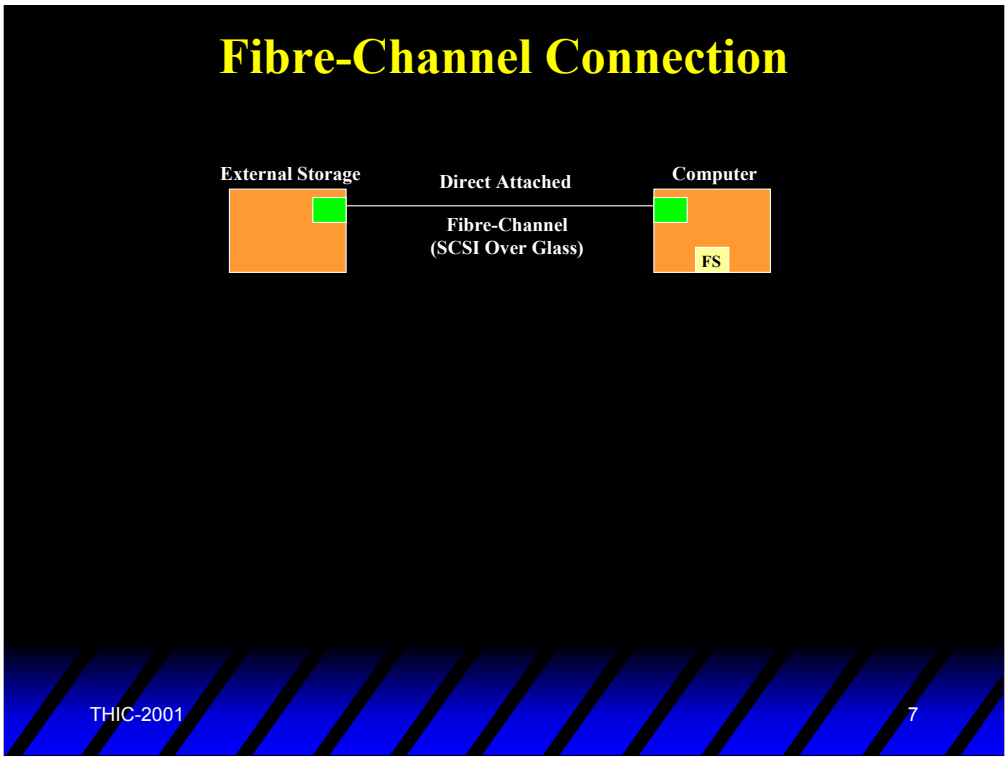
First, we will develop an understanding of how traditional networks “move data”.



Here we see a computer which is "direct attached" to an external storage device. Please make note of the term "direct attached".

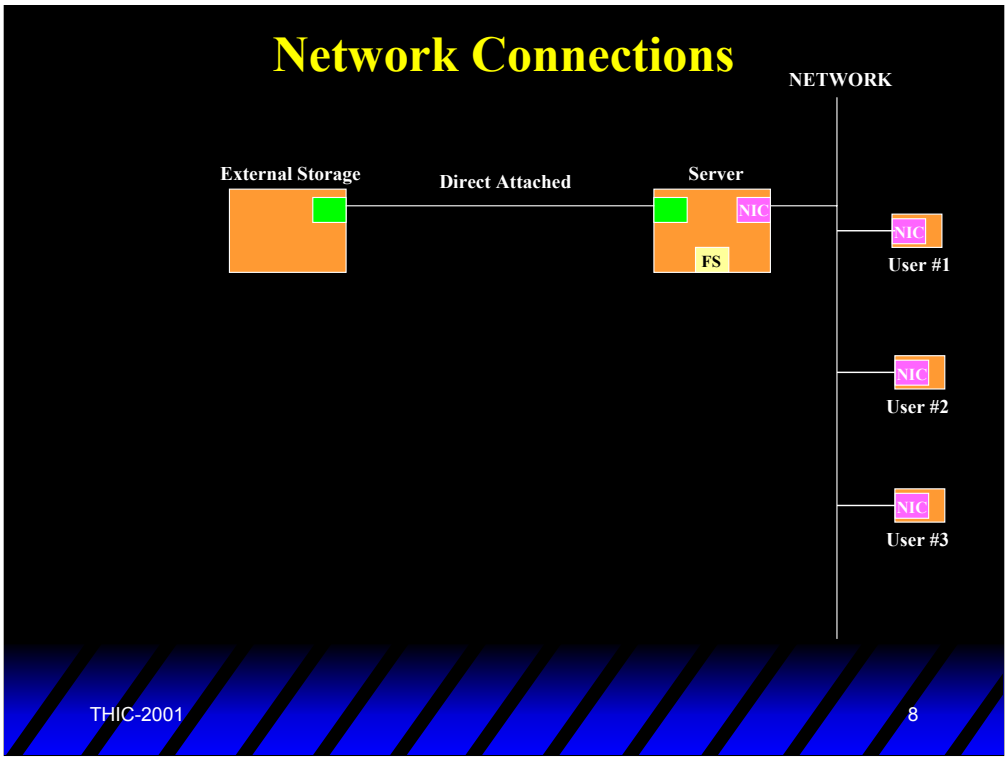
We have all heard the term SCSI and most of us have likely seen/used SCSI with a "copper" implementation. That is, all protocol and hand-shaking is accomplished over copper wires.

Also please note, the computer incorporates a "file-system" (FS) to manage data being stored on the storage device. Such a FS must keep track of the exact location of the data. Sometimes this is thought of as the "path" and the "name" of the data file.



More modern technology implements a variation on the copper SCSI connection by utilizing fiber-optic cables (sometimes called "glass"). This implementation is still "direct attached", but uses "commands" to perform all protocol and hand-shaking duties. One can think of this as "SCSI over glass".

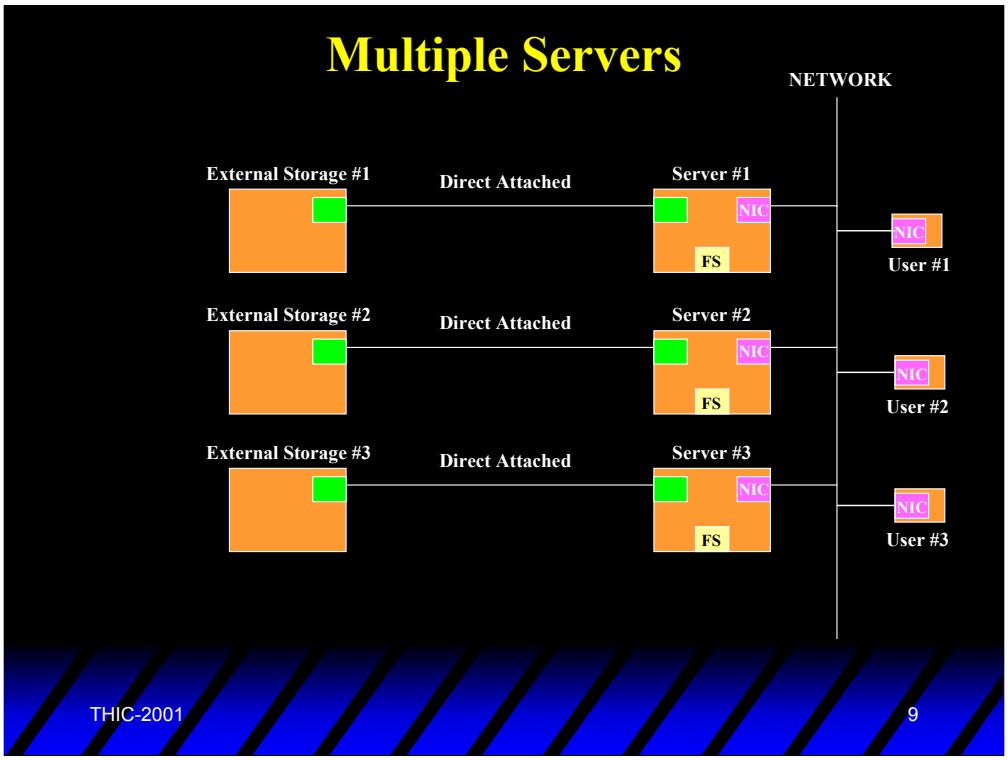
Using glass (rather than copper) is transparent to the FS, but naturally requires more sophisticated "peripheral drivers" in the computer and storage device.



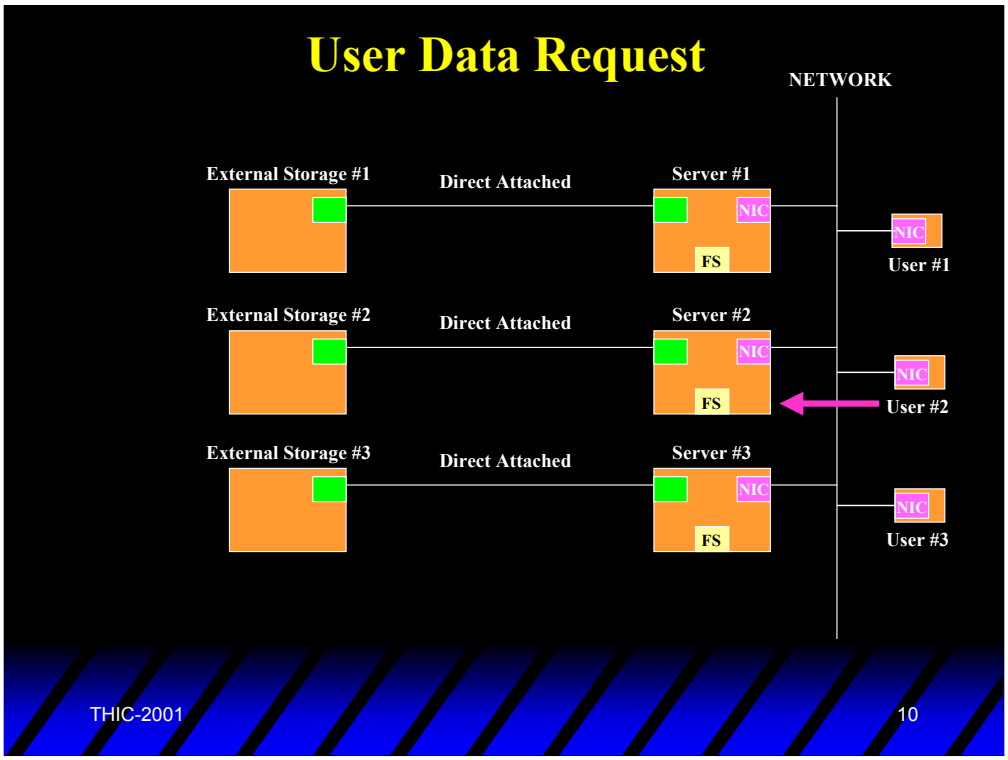
Now, let's introduce a “traditional network".

Here we see three (3) users (our computer being used as a "server”) with the external storage being controlled and managed by the server.

This is a very classic network implementation and has utility for any office environment.



We now depict a more complicated traditional network environment, where multiple servers (each with direct attached storage) are implemented.

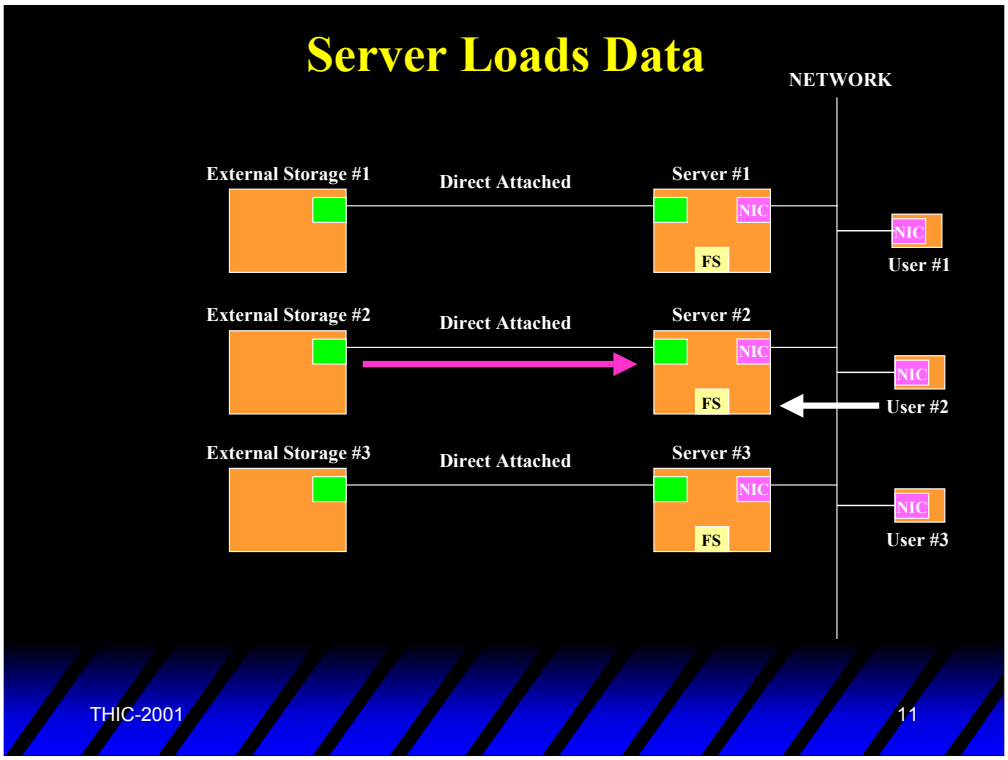


Let's follow simple transaction.

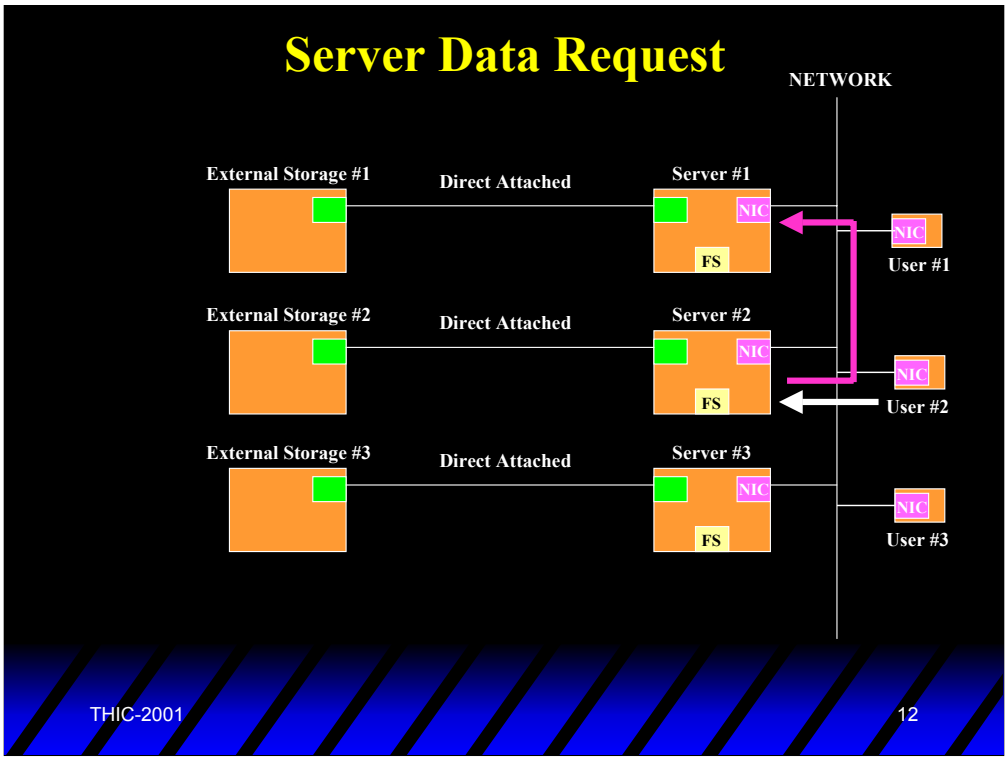
A user wants to edit an existing document.

[Modify a PowerPoint briefing via the MS server.]

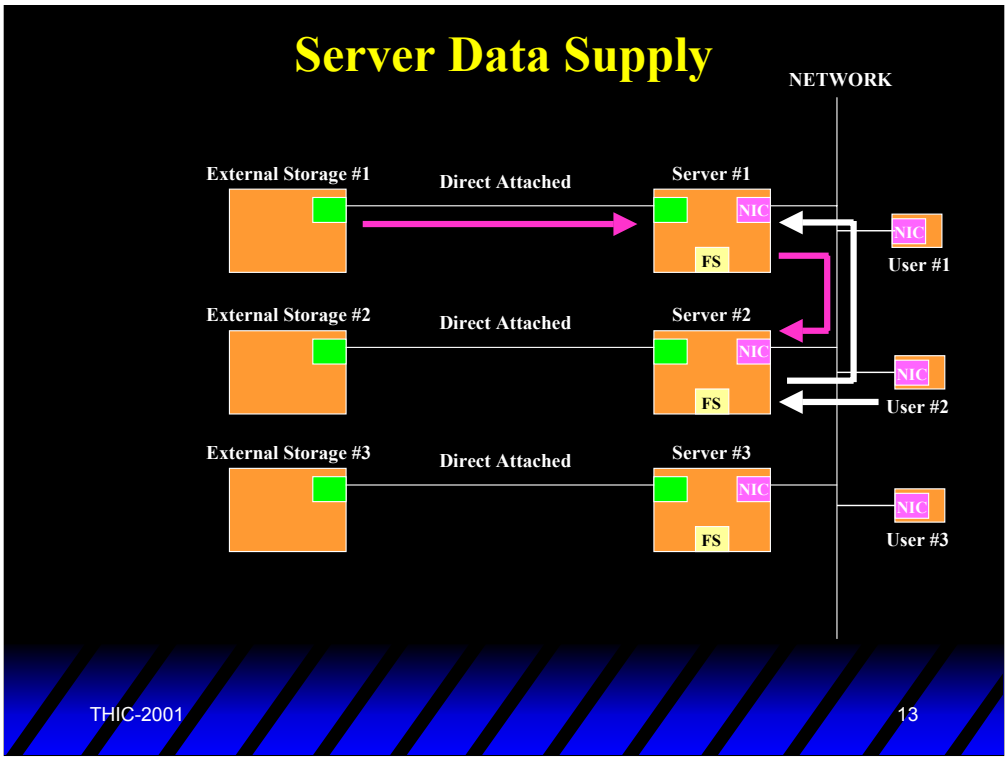
User #2 requests a transaction from Server #2. The Server #2 FS locates the data file.



Server #2 loads the data from the local storage if the file resides on External Storage #2.

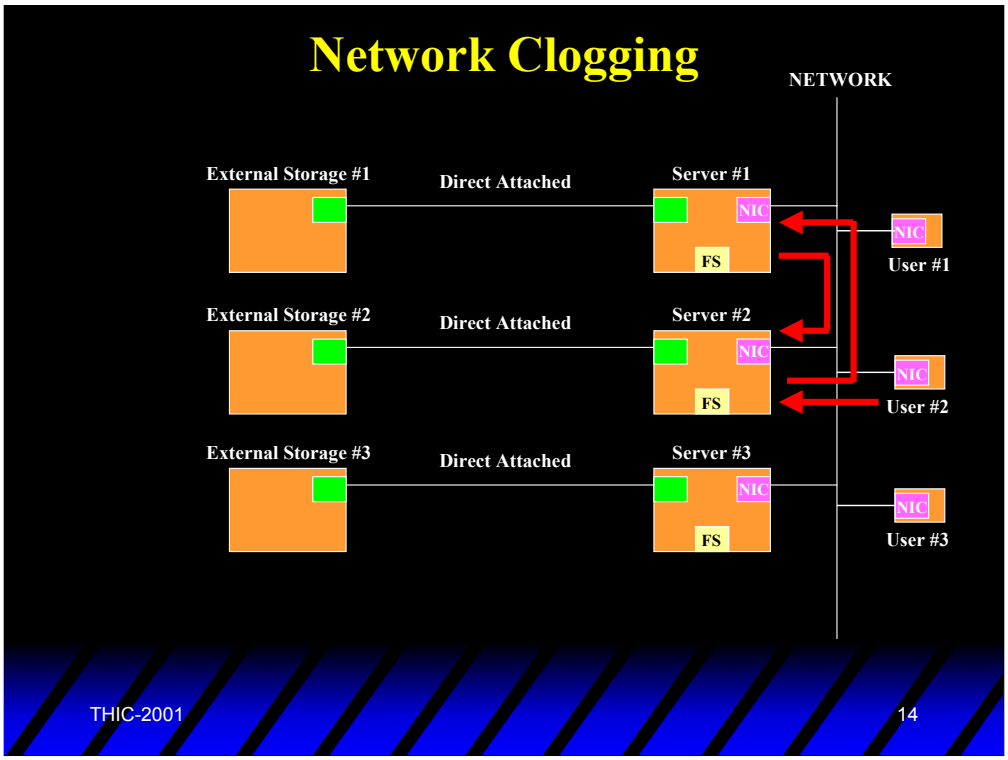


If for some reason the data file is not local, Server #2 requests the data from an alternate location. File "path" information assistance in this process.



The alternate server (Server #1) provides the requested data file from its direct attached storage (External Storage #1).

Please note, the data file being transferred consumes some of the the Server #1 backplane bandwidth, some of the network bandwidth and some of the Server #2 backplane bandwidth.



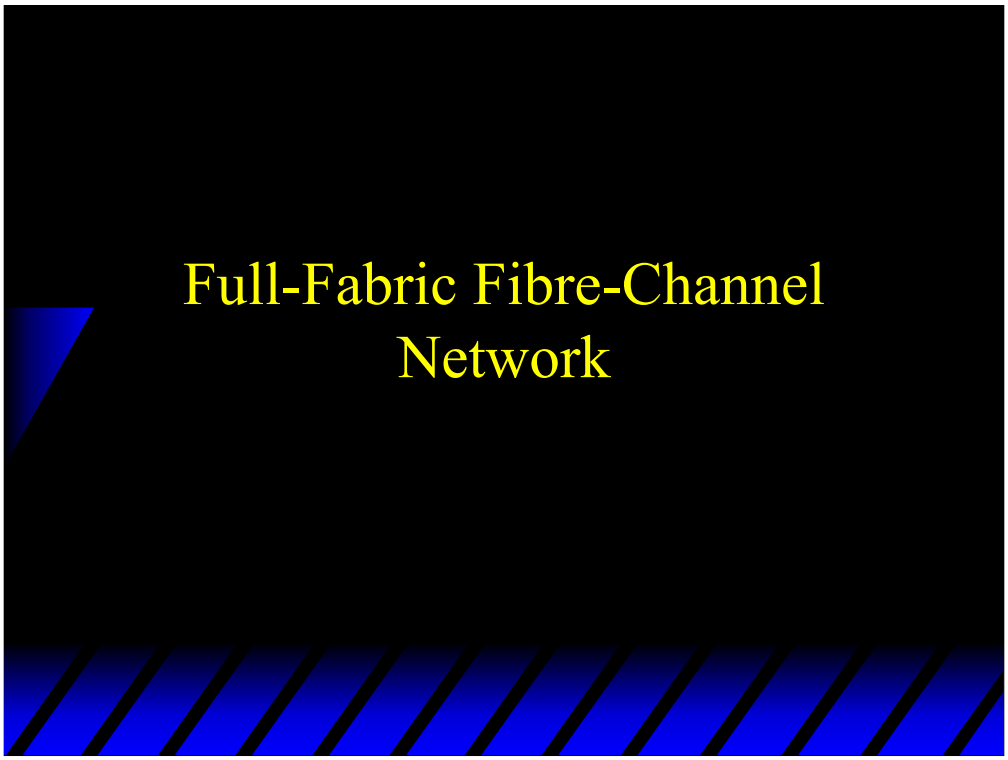
Please take note of the network traffic workload.

Thus far, the user's desire to modify the file has not taken place. Merely the retrieval of the data.

As the number of users increase and the size of data files increase, it is easy to see how transferring data over traditional networks can become "congested".

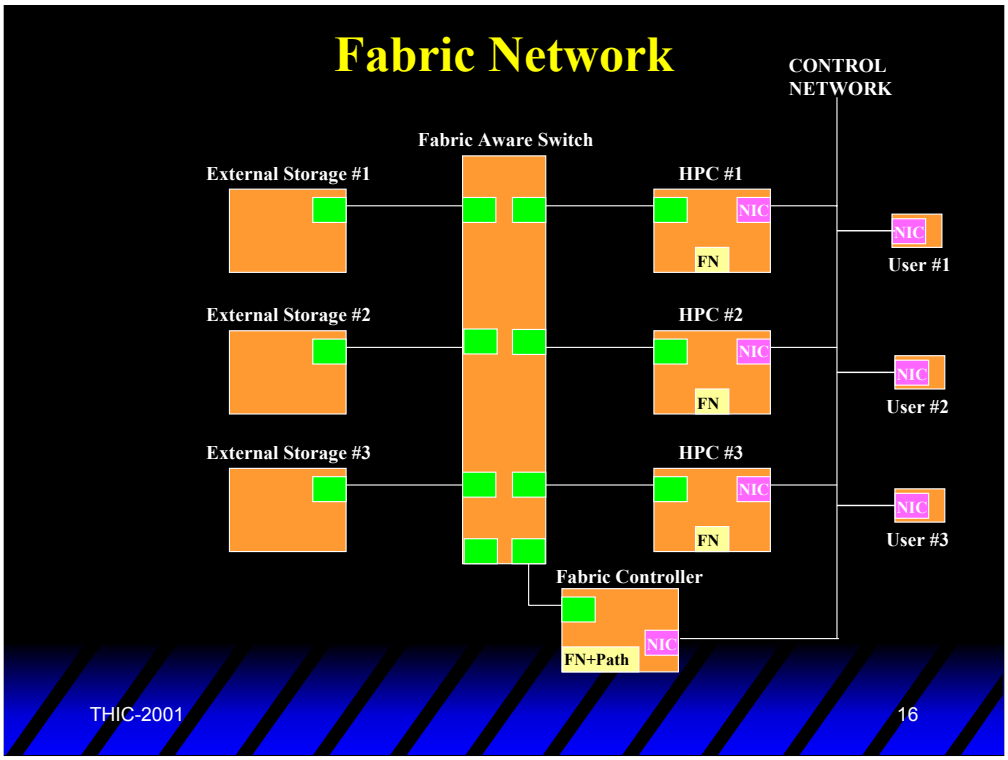
Unfortunately, the common architectural remedy is usually, "increase the bandwidth" of the network. In ATM terms, go from OC-3 to OC-12 to OC-48. In Ethernet terms, go from 10base to 100base to gigabit.

Each of these solutions does provide more bandwidth, but at the penalty of more complexity, more administration burden and more cost.



Now let's take a look at an alternate architecture. One which implements a "data network" in addition to the traditional network. This implementation utilizes Full-Fabric Fibre-Channel technology.

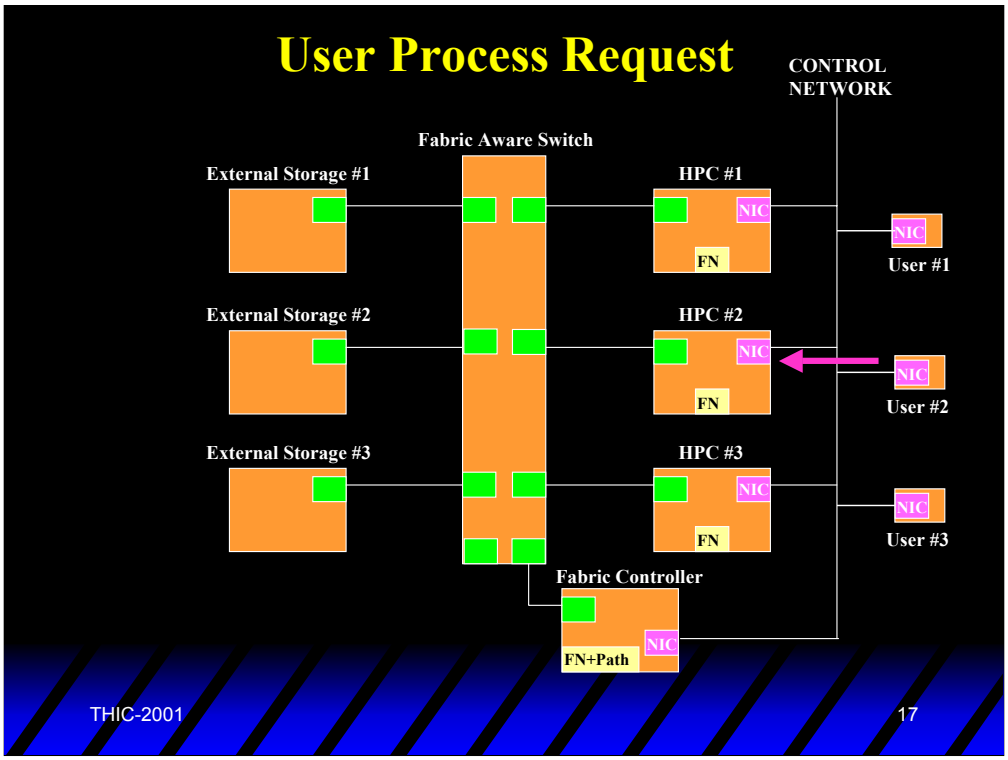
Full-Fabric is in contrast to the more traditional Fibre-Channel implementation using "Arbitrated Loop". Sometimes called FC-AL (pronounced, F-cal).



Please note; we have the same users, the same traditional network, the same computers and the same storage devices. We have simply added a non-blocking switch and a switch controller and a new file management approach.

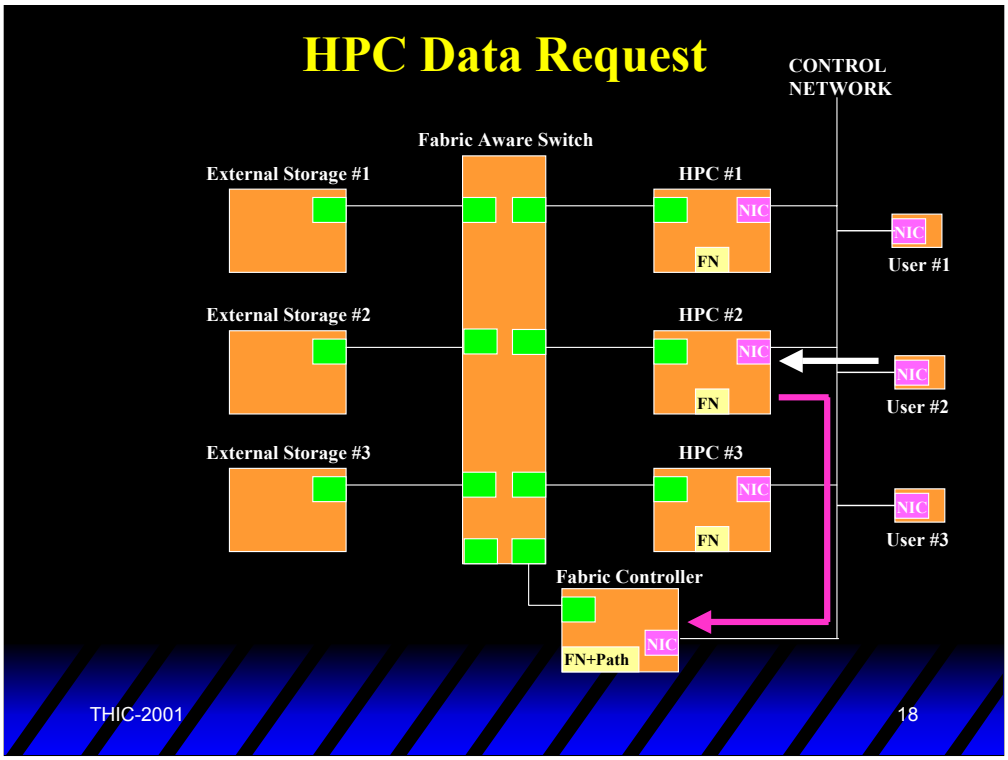
Also, we will call the computers HPC rather than servers. What is the subtlety in this name change. Often in the scientific community, large computers are implemented to perform a "computationally intense" process. Thus, the HPC does not function as a server, rather as a computation engine for a family of specific algorithms.

In this architecture, only the "Fabric Controller" knows the path to the data. All computers know the file names (FN) of all available files.



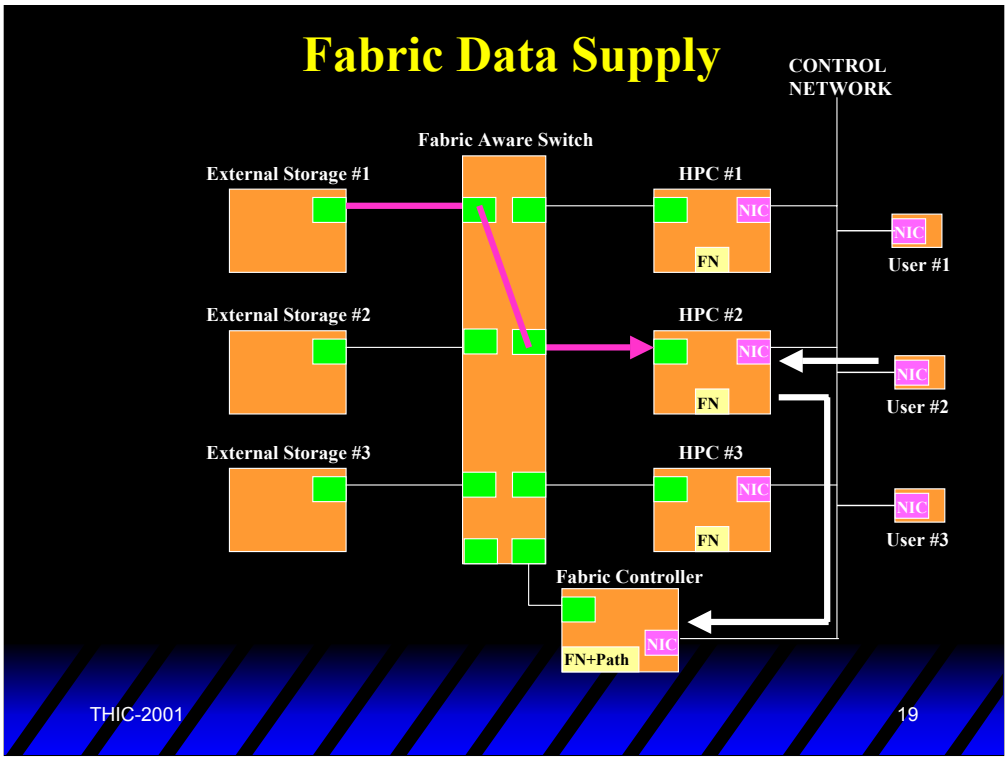
How does the new architecture move data?

Again, a request is made by User #2.



HPC #2 requests the data from the fabric controller.

Please understand, both of these “transaction request” messages are very small in size and consume little of the control network bandwidth.

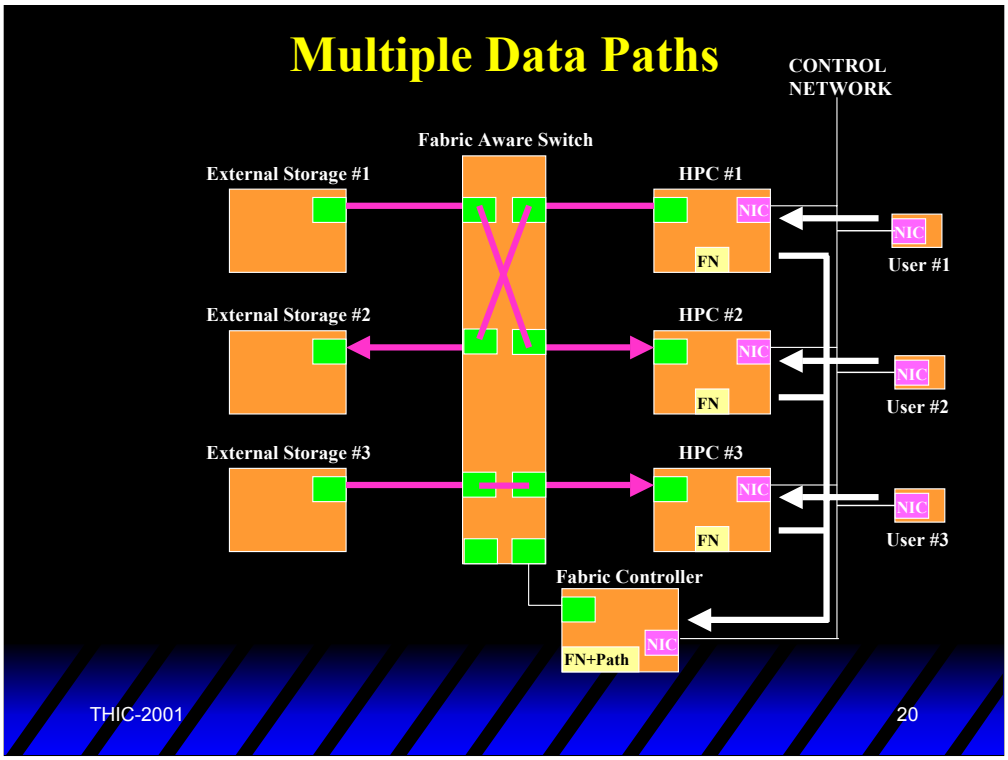


The fabric controller identifies the data location and provides a "connection" that allow External Storage #1 to supply the data to HPC #2.

This " file transfer" is accomplished entirely on the "data network" and not on the "control network".

Imagine the bandwidth savings to the control network, when many multi-gigabyte files require transfer.

It is also important to note; control network functions are not "blocked" while data is being transferred on the "data network".

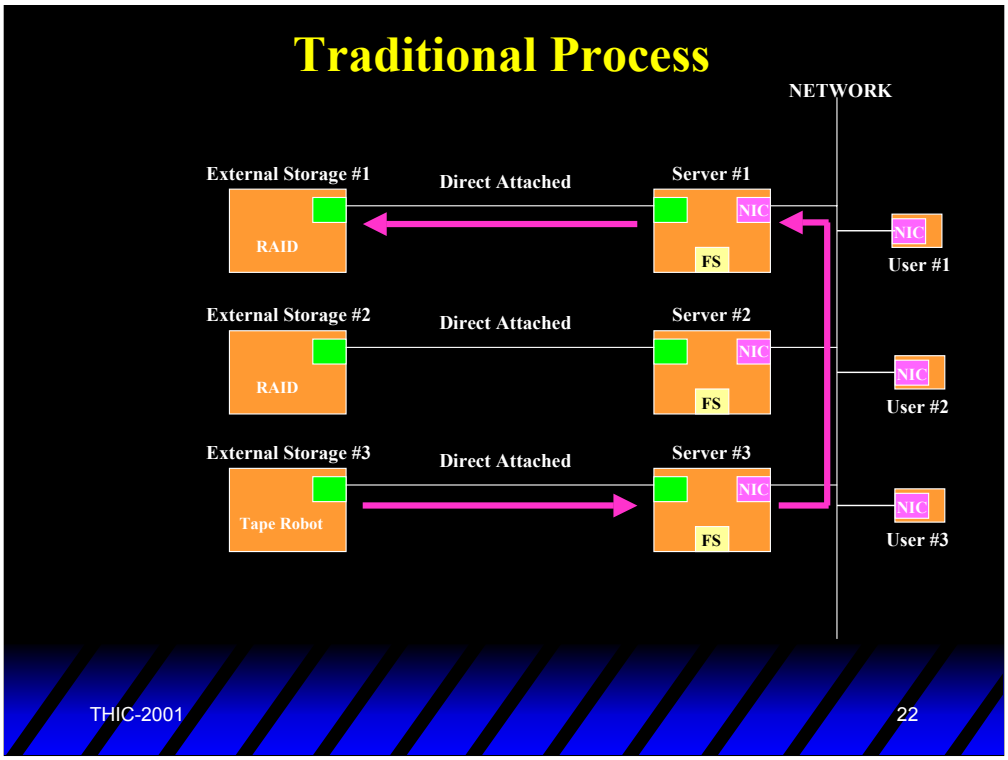


The beauty of the "non-blocking switch" is, multiple SIMULTANEOUS paths can be utilized. Remember, this is NOT a loop or ring topology, it is truly a "fabric".

I trust it is now evident; how the term “fabric” relates to this architecture.

This is an excellent time to discuss” file management” a little.

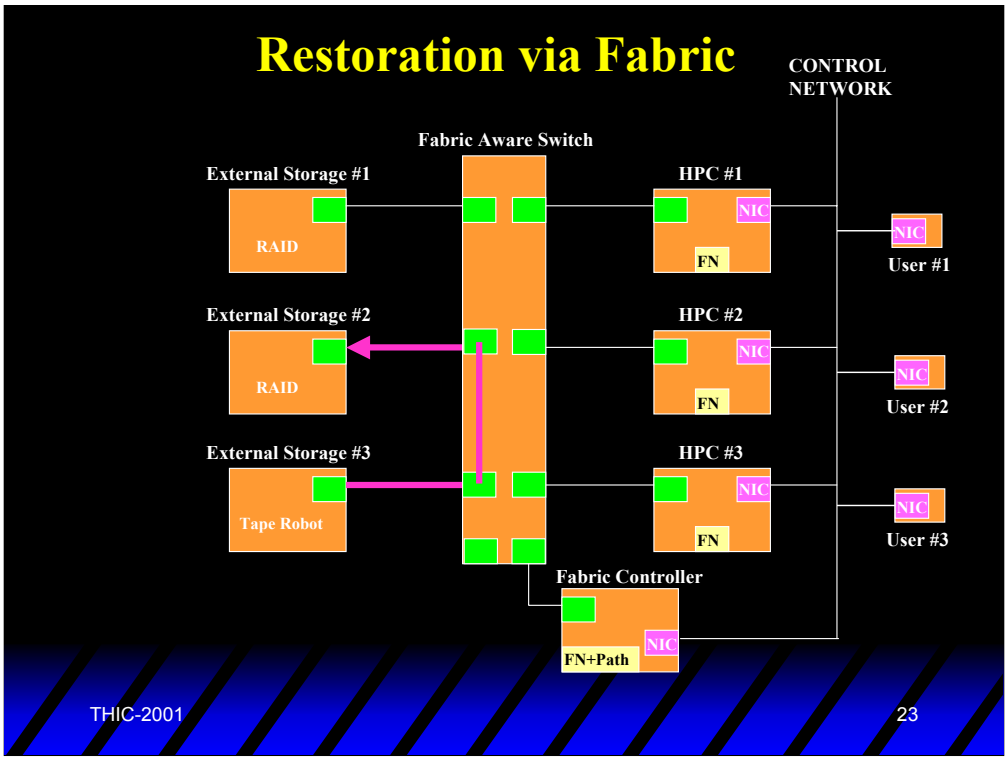
How can "fabric" aid with data archive and recovery?



Here we depict a "recovery" process. Note, External Storage #3 is a Tape Robot and External Storage #1 is a RAID.

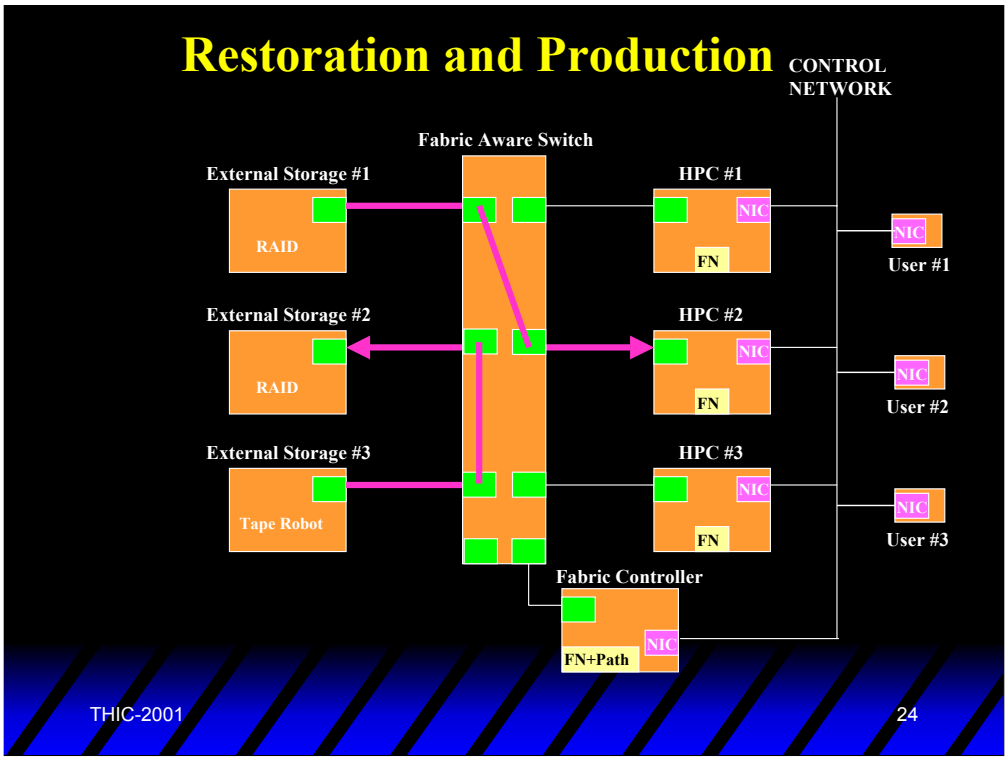
In a traditional network architecture, ALL data moves across the user network. Thus, this network must be "sized" to accommodate archive/recovery processes as well as production activities.

What usually happens is, archive is performed in "off peak" hours. This strategy leaves data "unprotected" for long periods of time.



Here we see, the restoration process does not affect any HPC, any User or the "control network". The point being, the control network need not be “oversized” to accommodate archive/restoration functions.

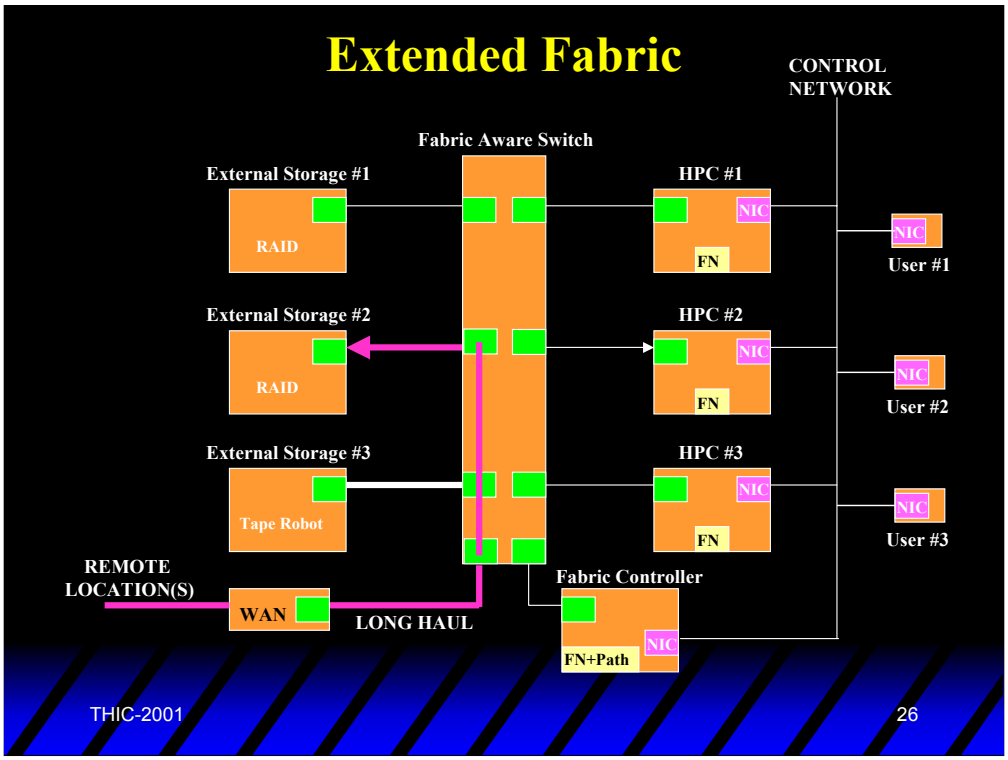
Yes, contention for External Storage #2 assets can exist and restoration could be delayed depending on the “priority” strategy.



Being a "fabric", restoration and production can occur SIMULTANEOUSLY in many circumstances.

Traditional networks require multiple processes to be “interleaved”.

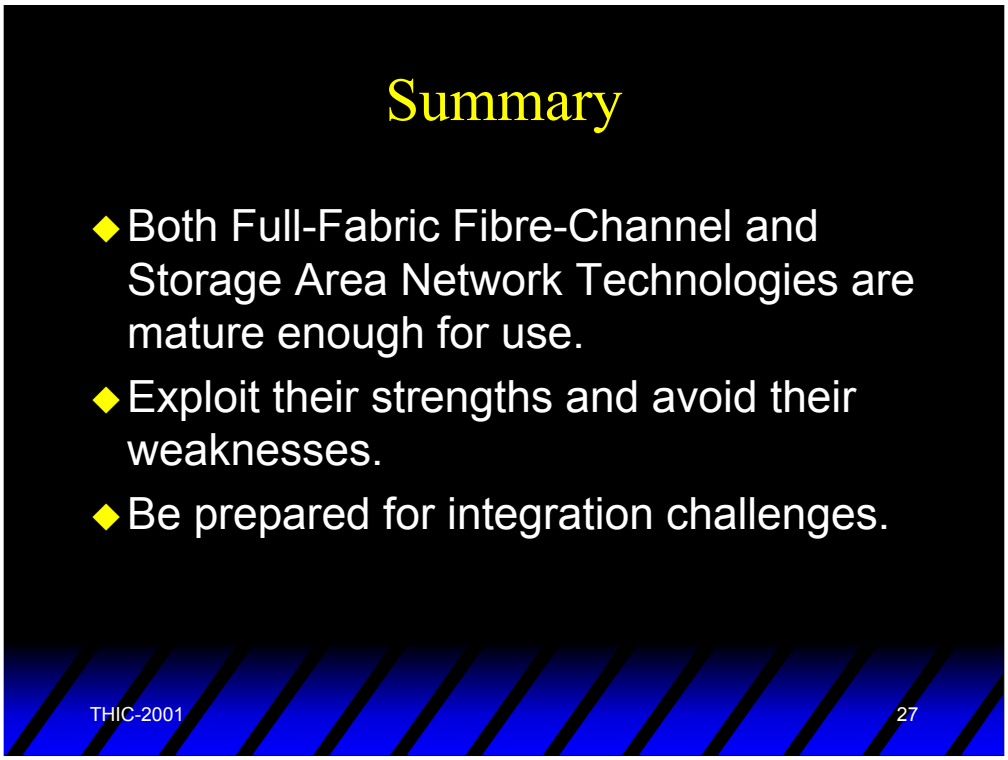
What happens if we have multiple locations?



To extend a fabric, "Long Haul" is possible.

This is very analogous to a Wide Area Network (WAN) connection. Naturally, the bandwidth/priority of the long haul connection will determine responsiveness.

Also, "file management" must consider multiple locations. Unique path/file names are required (inter/intra-site) for optimum performance.



Both fabric and SAN technologies are ready for implementation.

They are not the panacea for all applications. One must choose the features which assist the architecture. Good "Systems Engineering" is required.

Because these products (multiple vendors) are relatively new, each integration activity presents new challenges.