



Education

A Crash Course In Wide Area Data Replication

Jacob Farmer, CTO, Cambridge Computer

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
 - ◆ Any slide or slides used must be reproduced without modification
 - ◆ The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.

➤ A Crash Course in Wide Area Data Replication

- ◆ Replicating data over a WAN sounds pretty straight-forward, but it turns out that there are literally dozens of different approaches, each with its own pros and cons. Which approach is the best? Well, that depends on a wide variety of factors! This class is a fast-paced crash course in the various ways in which data can be replicated with some commentary on each major approach. We trace the data path from applications to disk drives and examine all of the points along the way wherein replication logic can be inserted. We look at host based replication (application, database, file system, volume level, and hybrids), SAN replication (disk arrays, virtualization appliances, caching appliances, and storage switches), and backup system replication (block level incremental backup, CDP, and de-duplication). This class is not only the fastest way to understand replication technology it also serves as a foundation for understanding the latest storage virtualization techniques.

Course Outline

- Basic Replication Concepts and Terminology
- Some Thoughts on Setting Business Objectives
- The Storage I/O Path and Insertion Points for Replication Logic
 - ◆ Host-Based Replication
 - ◆ SAN Replication
 - ◆ LAN-Based Replication
- Backup System Replication
 - ◆ CDP
 - ◆ De-Duplication

Not Covered In This Session

- The following topics are not discussed in this session, but could be relevant to your wide area storage solution:
 - ◆ Metropolitan Area Networks and WDM
 - ◆ IP Channel Extenders / FCIP / FCoE
 - ◆ Wide Area File Services / Distributed Locking
 - ◆ WAN Accelerators



Check out “Recent Advances in WAN Acceleration Technologies” and “FCoE: Fibre Channel Over Ethernet”

Basic Concepts and Terminology

- **Source**
 - ◆ The storage resource that is to be replicated
- **Target**
 - ◆ The storage resource that receives the replication stream.
- **Bandwidth**
 - ◆ The maximum throughput of your WAN connection.
- **Latency**
 - ◆ Networking delays introduced by distance, protocols, and/or networking equipment.
- **Quiesce (No one is really sure how to spell this!)**
 - ◆ Pause application and flush buffers to ensure application-consistent data on disk.
- **Storage Virtualization**
 - ◆ A generic term that describes inserting abstraction layers in the I/O path between applications and hard drives.

Asynchronous v. Synchronous

- Synchronous – Data successfully arrives at target before the writes are acknowledged on the source.
 - ◆ Zero data loss
 - ◆ Synchronous replication can adversely affect performance
 - › Bound by WAN latency. 40ms or so.
- Semi-Synchronous – You specify how far behind the replication can get before performance is affected.
- Asynchronous
 - ◆ The application does not wait for replication. Performance is not hindered.
 - ◆ Replication catches up during lighter activity
 - ◆ It is possible to lose some data in the face of a catastrophic failure

Setting Business Objectives

- RTO – Recovery Time Objective
 - ◆ How long can you wait to access your data?
- RPO – Recovery Point Objective
 - ◆ How much data are you willing to lose?
- What level of readiness do you need at the other side of the wire?
 - ◆ Do you just need data to be in a safe place.
 - › Maybe you can worry about how / where to restore it later?
 - ◆ Do you need for applications to fail over?
 - ◆ Do you have servers available to you on the other side of the wire?
 - ◆ If you have servers at the remote location, do they match the servers you have at your primary location?
- What can you tolerate for performance impact?
 - ◆ At the primary site
 - ◆ At the fail-over site

Some General Hurdles to Overcome

- How do you initially get your data over to the remote site?
 - ◆ What happens if you add more data to the replication environment?
- How will you access your data in the event that you have to roll over to the target site?
 - ◆ What will it take to bring up and access your applications at the remote site?
 - ◆ How will you boot up your servers?
 - ◆ Will performance be suitable for running the business?
- How will you fail-back to the primary site?
- How do you test the target without disrupting on-going replication?
- Documentation
 - ◆ How will you document the environment and ensure that people know what to do to bring the applications up at the remote site.

Storage Virtualization 101

The Storage I/O Path

Common Places to Insert Replication Logic

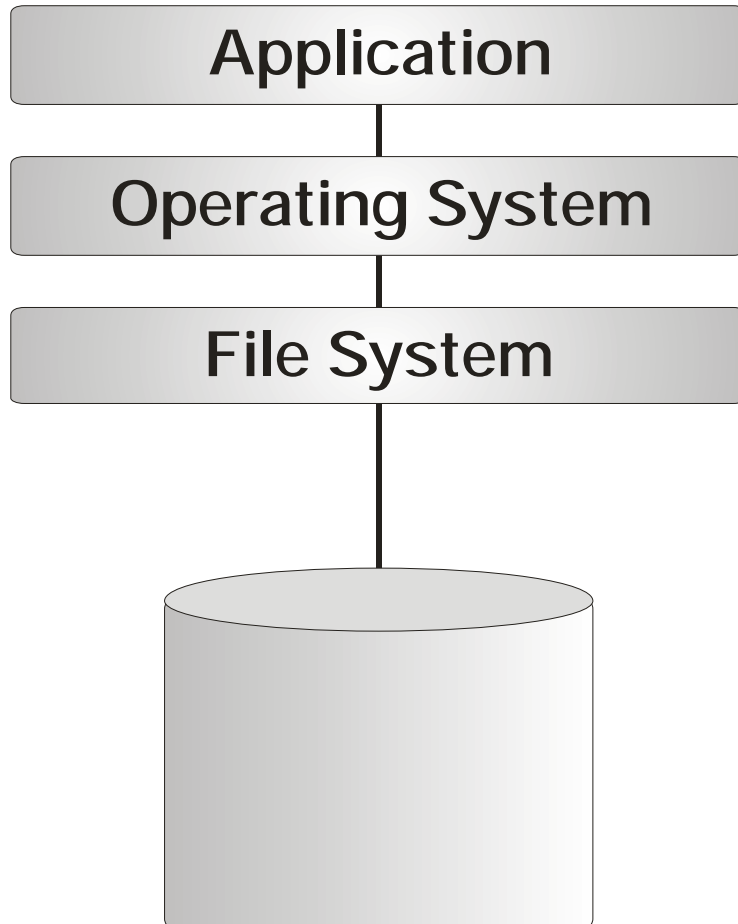
Files v. Blocks

➤ **Blocks**

- ◆ Least common denominator in conventional storage technologies.
- ◆ A block is a unit of data storage.
- ◆ Hard drives and RAID arrays serve requests for blocks.

➤ **Files**

- ◆ Objects consisting of multiple blocks.
- ◆ Blocks are organized into files by file systems, which are like databases of all of the files, their attributes and records of the blocks that make up the files.



Storage has a layered architecture, very much like a network stack.

Disk drives store data in blocks. Each block has a unique numerical address.

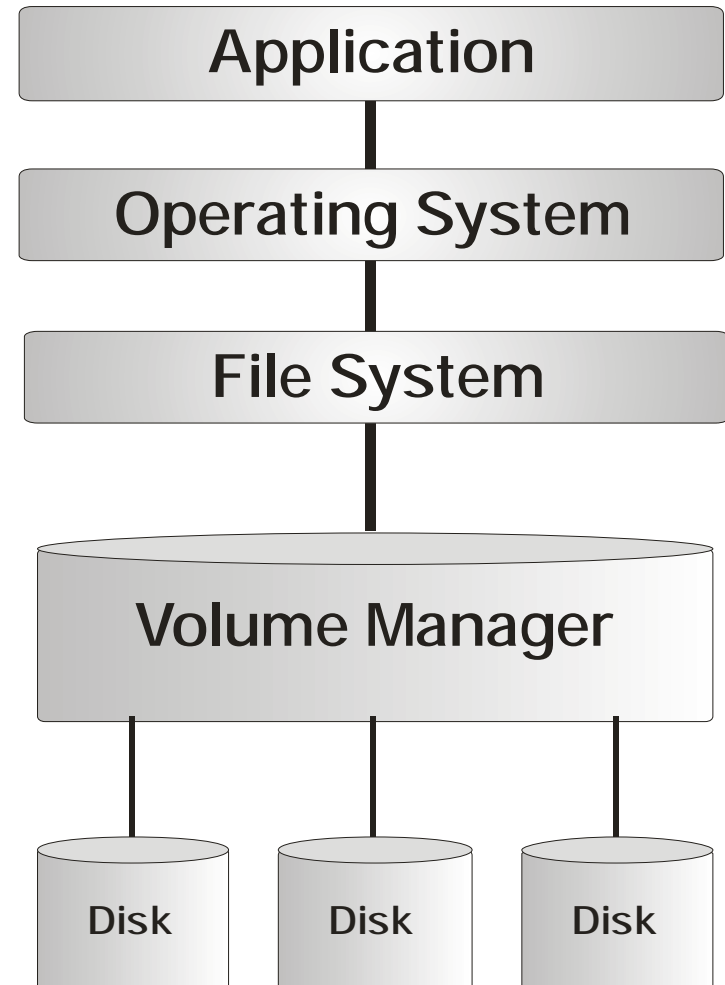
Disk devices (hard drives, RAID systems, etc.) are like “block servers”, meaning you ask them to perform operations on specific blocks.

Volume Manager: Block-Level Abstraction

Abstraction of the physical disk.

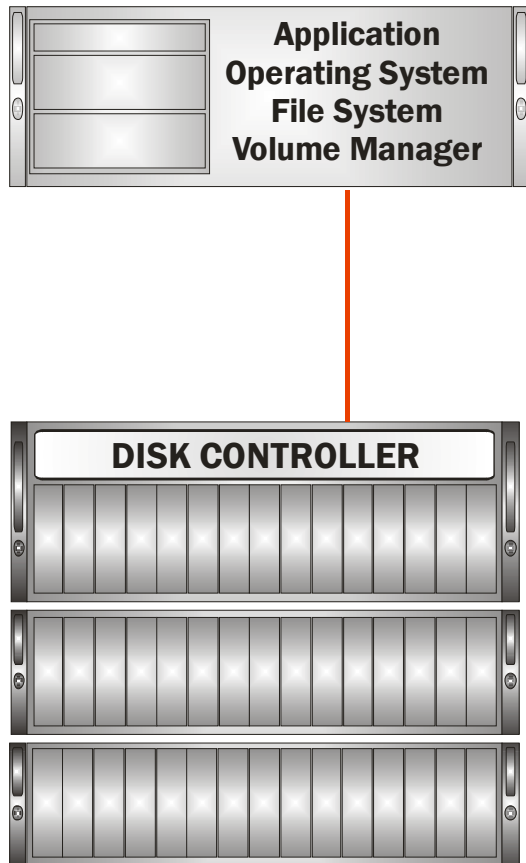
The file system asks for specific block addresses and the volume manager fulfills request from the actual disks.

- Software RAID
 - Hard Drive Fault Tolerance
 - Spindle Aggregation

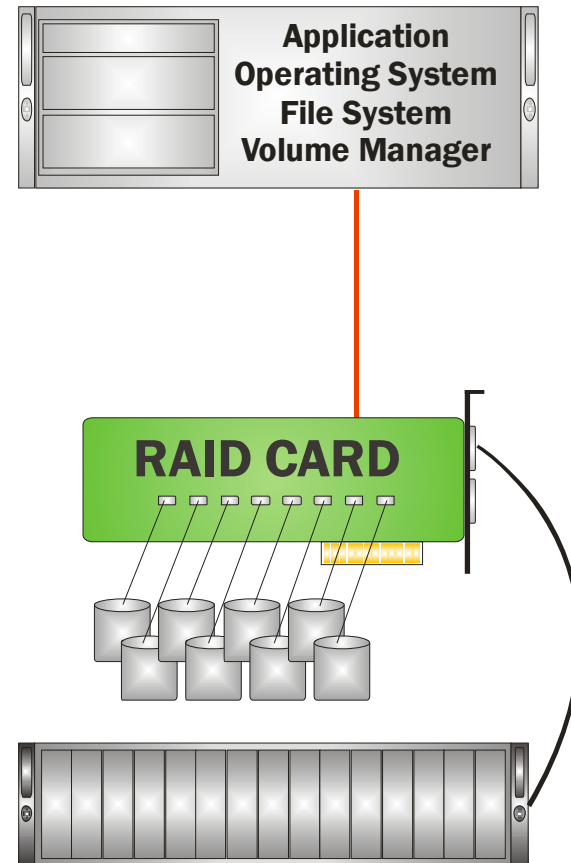


RAID Arrays: Hardware-Enabled Block Abstraction

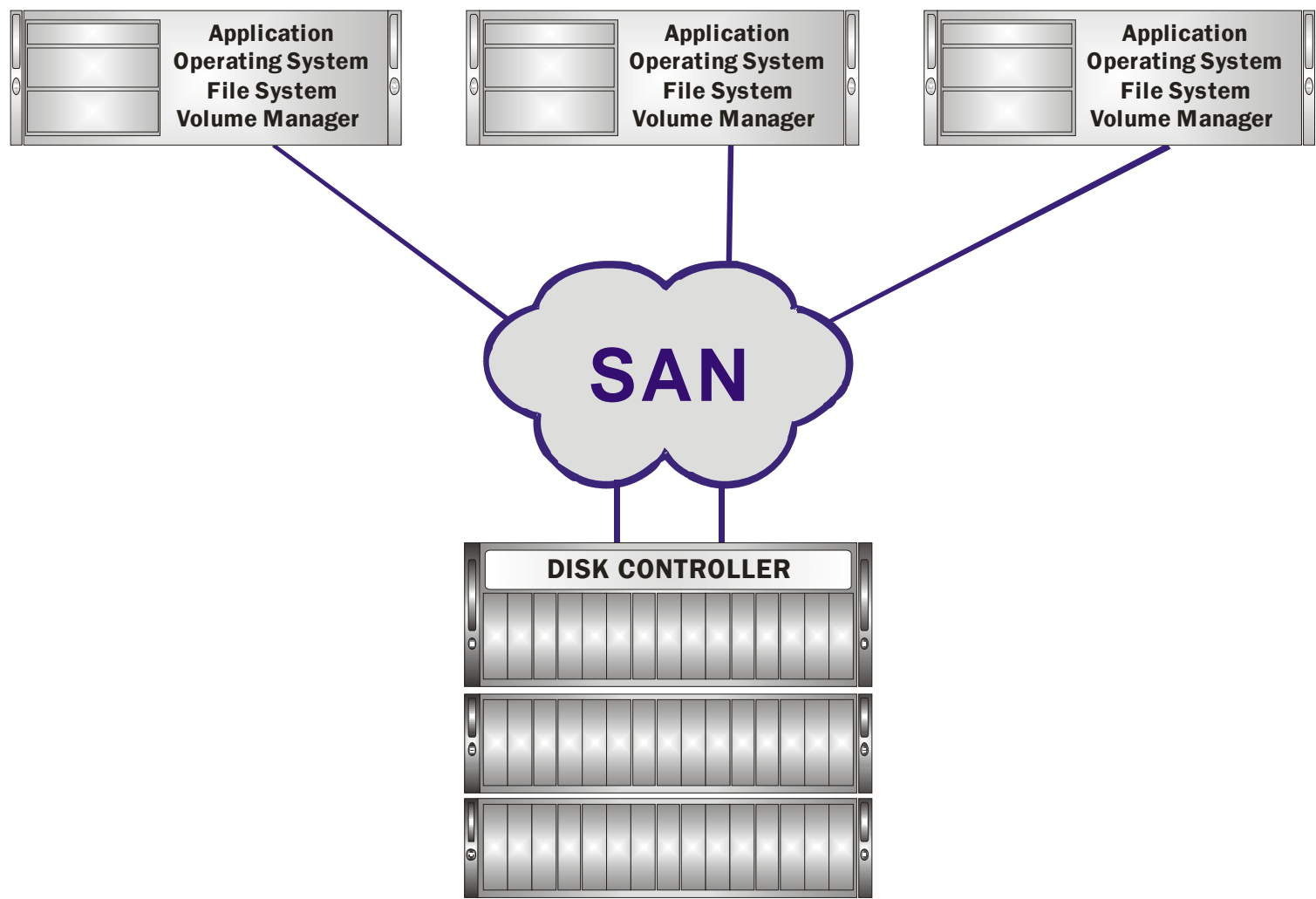
External Array



Internal Array



SAN Array: Centralized, External Abstraction

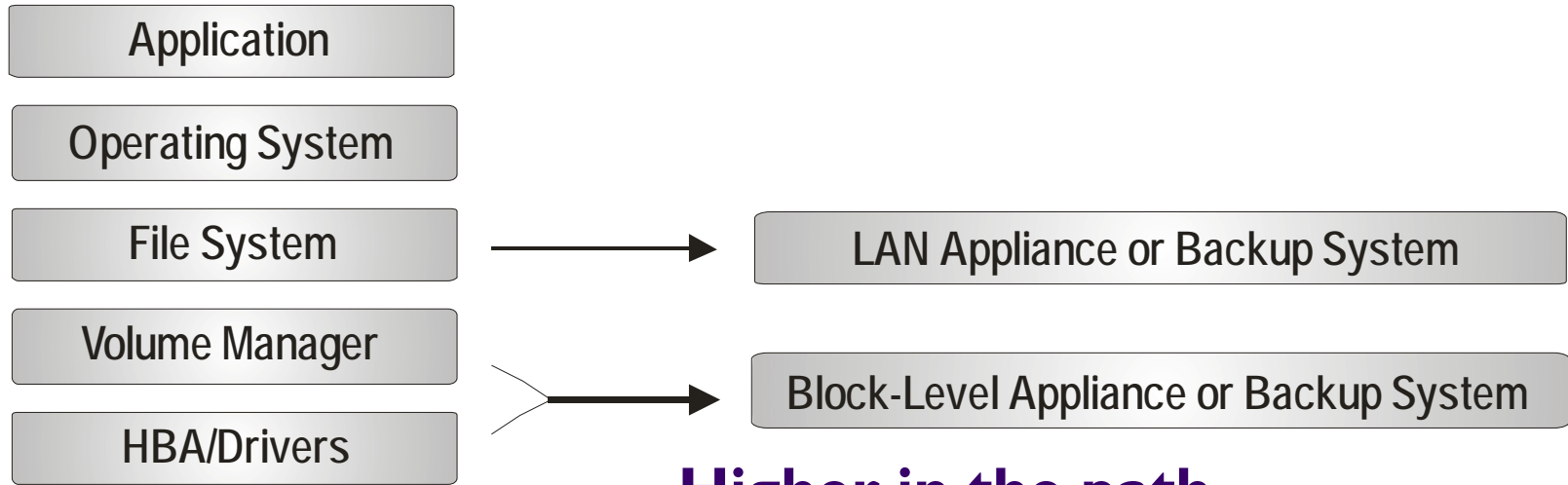


Where Else Can You Insert Storage Logic?

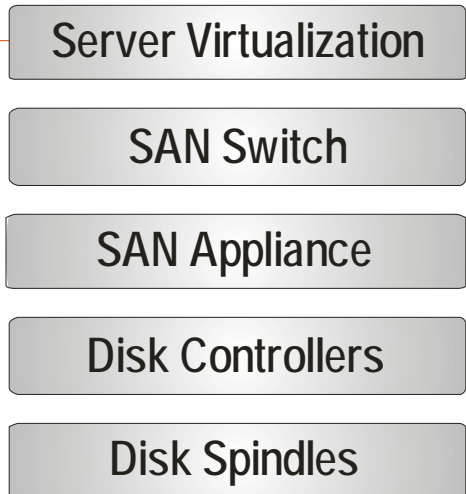
- **Appliance on the SAN**
 - ◆ Spliced in between your hosts and your storage device
- **Out-of-Band Appliance**
 - ◆ Device that sits on LAN or SAN, but is not in the I/O path.
 - ◆ Requires a driver that will sit in I/O path on the host.
- **Inside a storage switch**
 - ◆ Modern switches offer APIs that allow 3rd party applications to access the storage I/O path from within the switch.
- **Appliance on the LAN**
 - ◆ A device that sits between a file server or mail server, captures I/O and replicates the data stream.
- **Stay tuned: clever new products come to market all the time!**

Replication and the Storage I/O Path

Host-Based



SAN-Based



Higher in the path

- ◆ More specialized, “intelligent”

Lower in the path

- ◆ Brute force
- ◆ Device / Application Independent

Host Based Replication – Many Incarnations

- **Application**
 - ◆ Replication integrated into the application
- **Database replication**
 - ◆ Log shipping and other log-based technologies
- **Middleware Replication**
 - ◆ Example DICOM replication in PACS applications
- **File System Replication**
 - ◆ “Block-level”
 - ◆ Object-level
- **Volume-level Replication**
- **Out-of-band SAN Replication**

Generic Advantages of Host-Based v. SAN

- Storage device independence
 - ◆ You can have different storage on both sides of the wire
- Not all SAN arrays offer built in replication.
- You might be striping multiple SAN volumes together for greater spindle performance.
 - ◆ In this case it is more appropriate to manage replication at the logical volume level as opposed to the physical disk volume level.
- Opportunity to work at a higher level in the storage I/O path.
 - ◆ This might get you a better result for a given application.
 - ◆ This might allow tighter integration with application fail-over.
 - ◆ This might enable you to access the target storage during replication.

File System Replication

- File system operations are captured and replayed on the replication target.
- Operates transparently to application.
- Replication engine can be aware of files and directory structures
 - ◆ Exclude/ include based on file masks
 - ◆ Replicate between volumes of different sizes
 - ◆ Replicate to and from sub-directories
 - ◆ Only replicates data, blank space is not replicated
 - ◆ Optionally do not replicate deletions

Advantages of File System Replication

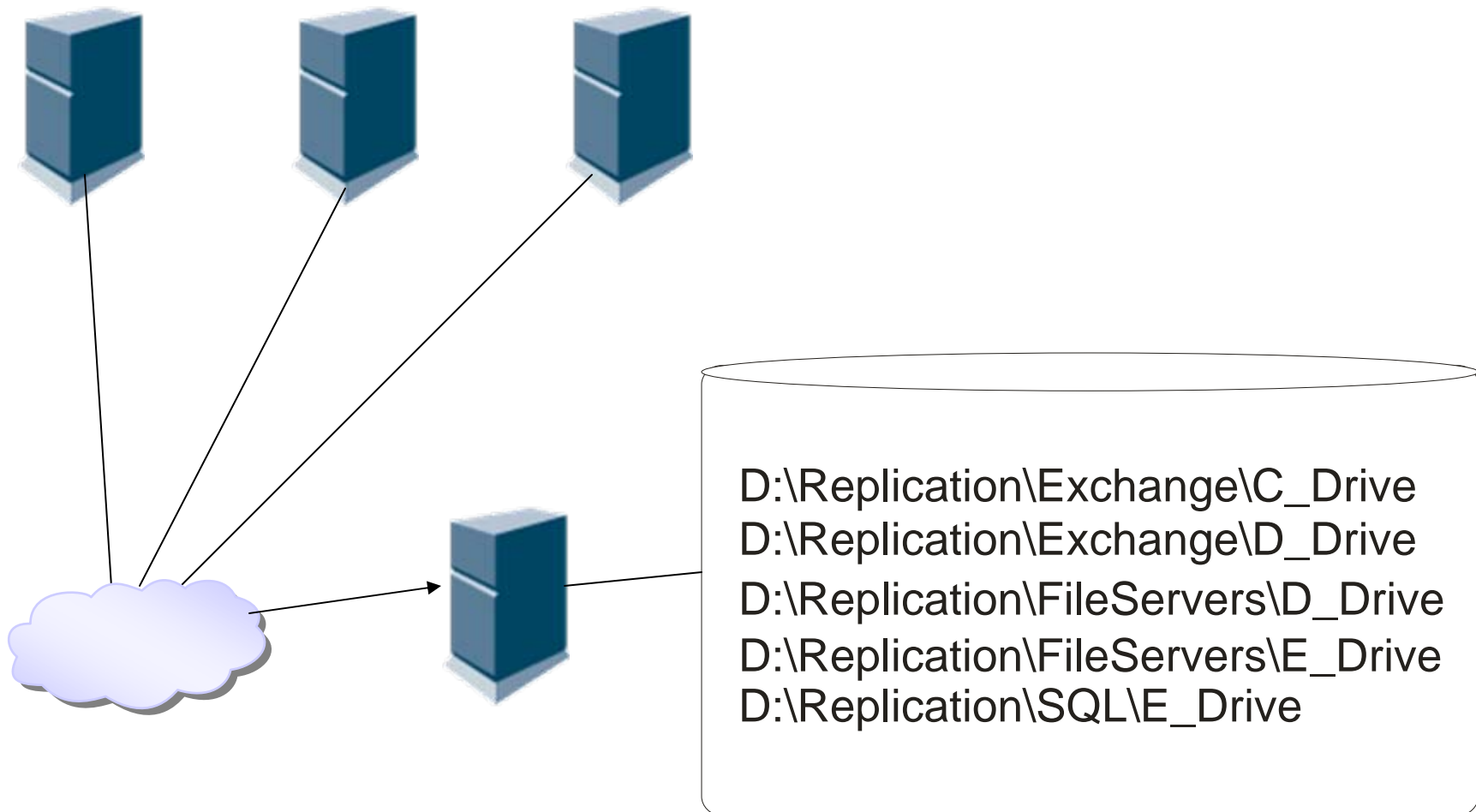
- Target file system is mounted
 - ◆ Facilitates application fail-over
 - ◆ Target system could be used for queries, backups, etc.
- Target capacity only has to match actual capacity in use.
- Easier to ensure a consistent, healthy state of data because such functionality is inherent in the file systems.
 - ◆ Lends well to roll-back in the event of corruption.
- Possibility for application awareness by analyzing file content on the fly.
- Relatively easy to add new file systems after the target has been shipped to a remote location.
 - ◆ Send a tape or disk with the files. Copy them to target server and resynch.
 - ◆ File system metadata facilitates identifying and reconciling differences between source and target file systems.

Many-To-One Writing to Subdirectories

File Server

SQL

Exchange



File Devices with Integrated Replication

- Host-based replication could be packaged within a turnkey storage device or software solution.
- Some commercial file systems have integrated replication.
- Many proprietary file systems are packaged as NAS appliances.
 - ◆ It is very common for these devices to have integrated replication.
- **CAS – Content Aware Storage**
 - ◆ Many CAS solutions offer integrated object-level replication.

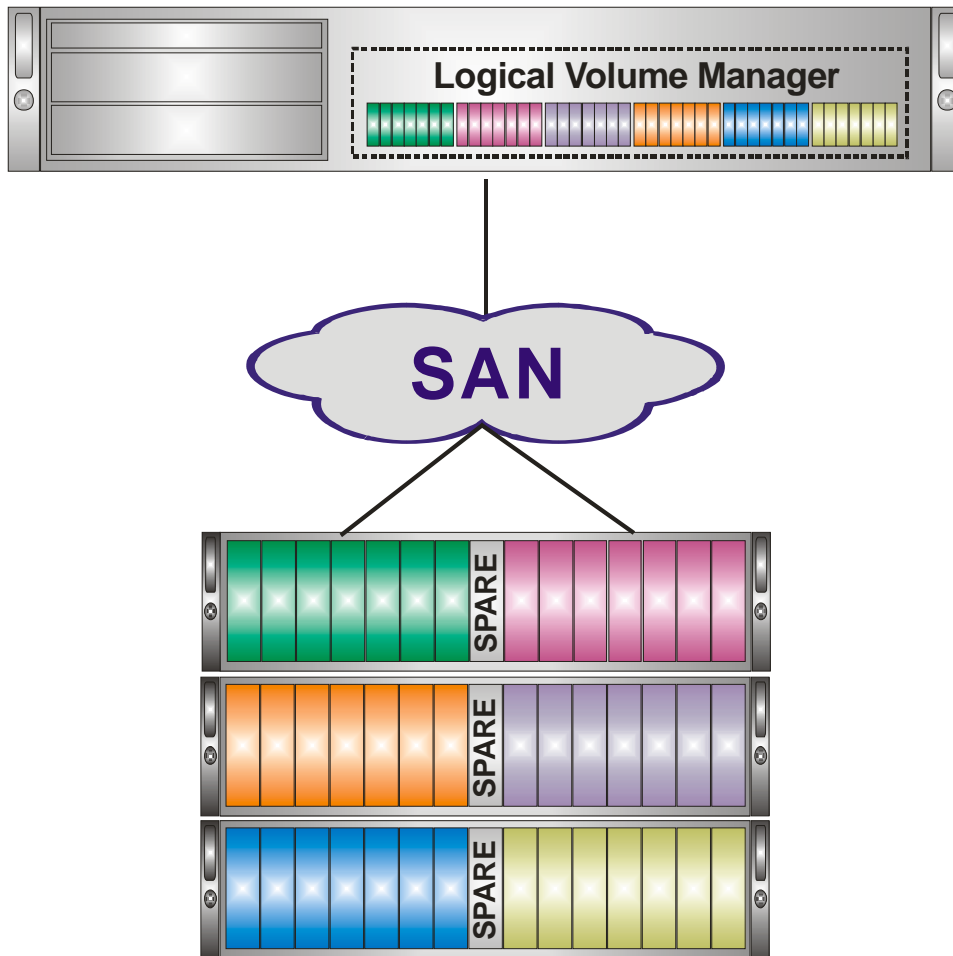


Check out “High Availability and Disaster Recovery for NAS Systems”

Advantages of Volume-Level Replication

- Able to perform synchronous replication from the host.
- Preserves the volume image.
 - ◆ It is feasible to replication boot volumes
- Does not require mounted file system.
 - ◆ Might be able to replicate to generic target that is OS independent
 - ◆ Most suitable for when you do not have servers available at remote site.
 - › For instance, many third party disaster recovery sites offer shared server hardware. That is, you only get access to the server hardware when you are testing or when you have a disaster.
- Very efficient – changes are captured at the sector level.
 - ◆ These replication solutions tend to have minimal performance impact.
- Often able to run over storage channels (SCSI, iSCSI, Fibre Channel) as well as over TCP/IP.

Logical Volume Manager to Aggregate Spindles



When a logical volume is assembled at the host, it is seldom practical to replicate individual (LUNs) from within the SAN array. The best practice is to replicate at the layer of the logical volume.

SAN Replication Choices

- **Replication from within the disk array**
 - ◆ Many arrays offer replication built in or as an add-on feature.
- **Replication inserted between host and array**
 - ◆ Volume-level “virtualization” appliances.
 - ◆ Replication engine running in a switch.
 - ◆ Replication engine in a caching appliance.
 - › Allows remote disk to be mounted as read/write device.
- **Out-of-Band Replication or Virtualization Platform**
 - ◆ Device that sits on your SAN and orchestrates data replication.
 - ◆ Note that these types of solutions are arguably host-based applications with external appliance offloading management and doing the heavy lifting.
 - ◆ The trend is for these solutions to offer in-band alternative by using SAN switch to provide access to the I/O path.

Advantages to SAN Replication

- **Device independence**
 - ◆ One platform can replicate all of your servers, regardless of applications and operating systems
- **Centrally managed**
- **No host-based software to load, monitor, and patch.**
- **Target volumes are not mounted. No need for live servers at the remote site.**
- **Many SAN arrays offer management software that makes it very easy to configure and monitor replication.**

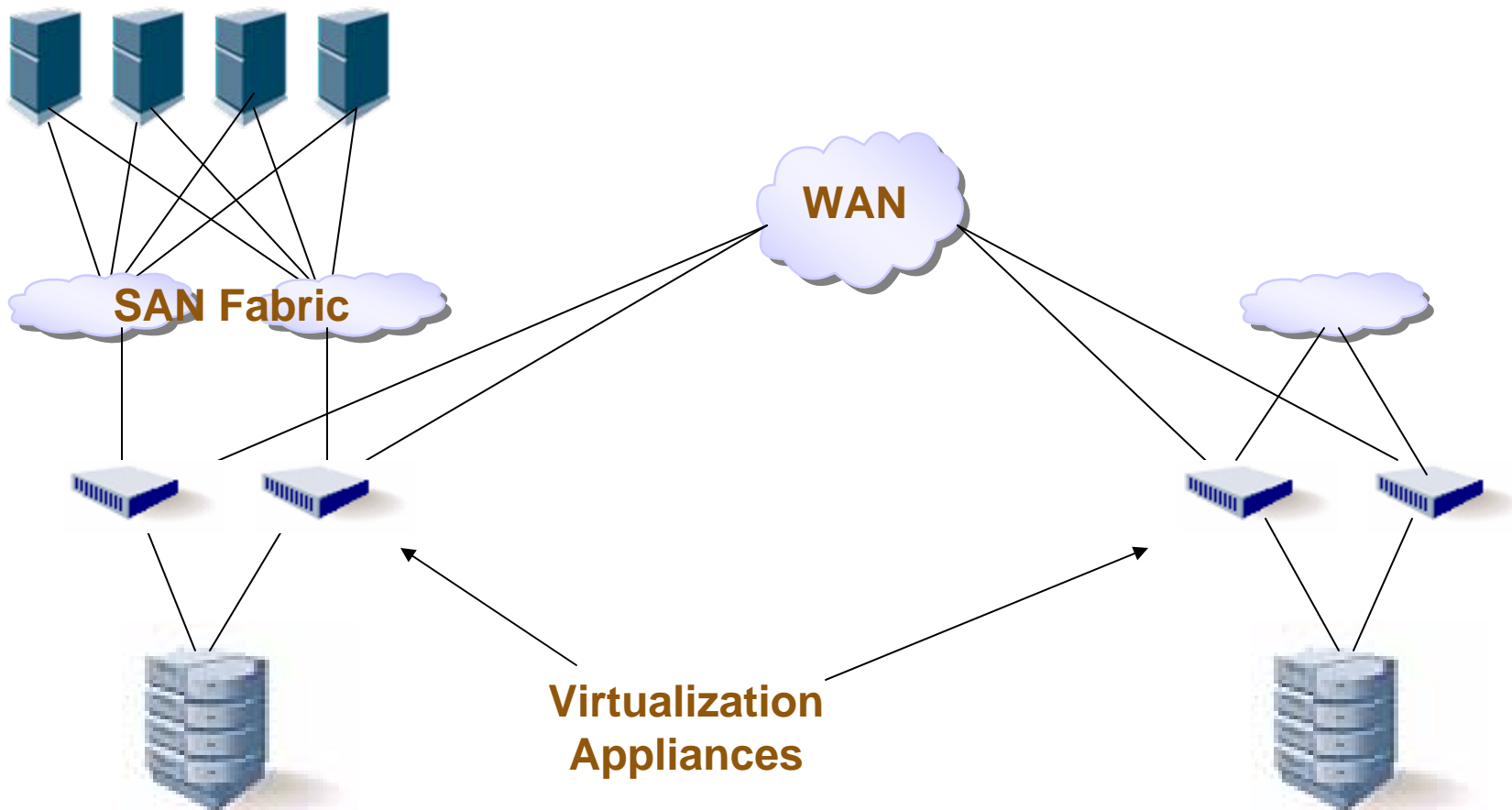
In Band Storage Virtualization

- Devices “splice” into the storage I/O path by way of zoning or other integration within a Fibre Channel fabric or iSCSI network.
- LUNs from the disk arrays are presented to the virtualization layer. The virtualization layer then presents LUNs to the host.
- Most storage virtualization platforms support wide area replication.



Check out “Virtualization I - What, Why, Where and How?” and “Storage Virtualization II - Effective Use of Virtualization”

In-Band Array Virtualization



Boot-from-SAN Considerations

- Many SAN arrays support booting from SAN. Also server virtualization platforms imply booting from SAN.
 - ◆ Booting from SAN can greatly facilitate disaster recovery.
- Remember to Isolate swap files (memory paging files)
 - ◆ Best practice is to put swap files on dedicated LUNS and not to replicate them.
- Reconciling disparate servers at the remote site
 - ◆ Unless your server hardware matches at the remote site or you are using server virtualization, you might have trouble bringing up your applications.
 - ◆ Consider 3rd party applications that can replicate or back up boot volumes and automate device driver substitution.

LAN-Based Replication Appliances

➤ File server virtualization

- ◆ Devices sit on the LAN in between clients and the file servers and NAS appliances.
- ◆ Create virtual network file systems based on the back end file systems.
- ◆ Some of these solutions offer replication within the file server virtualization device.

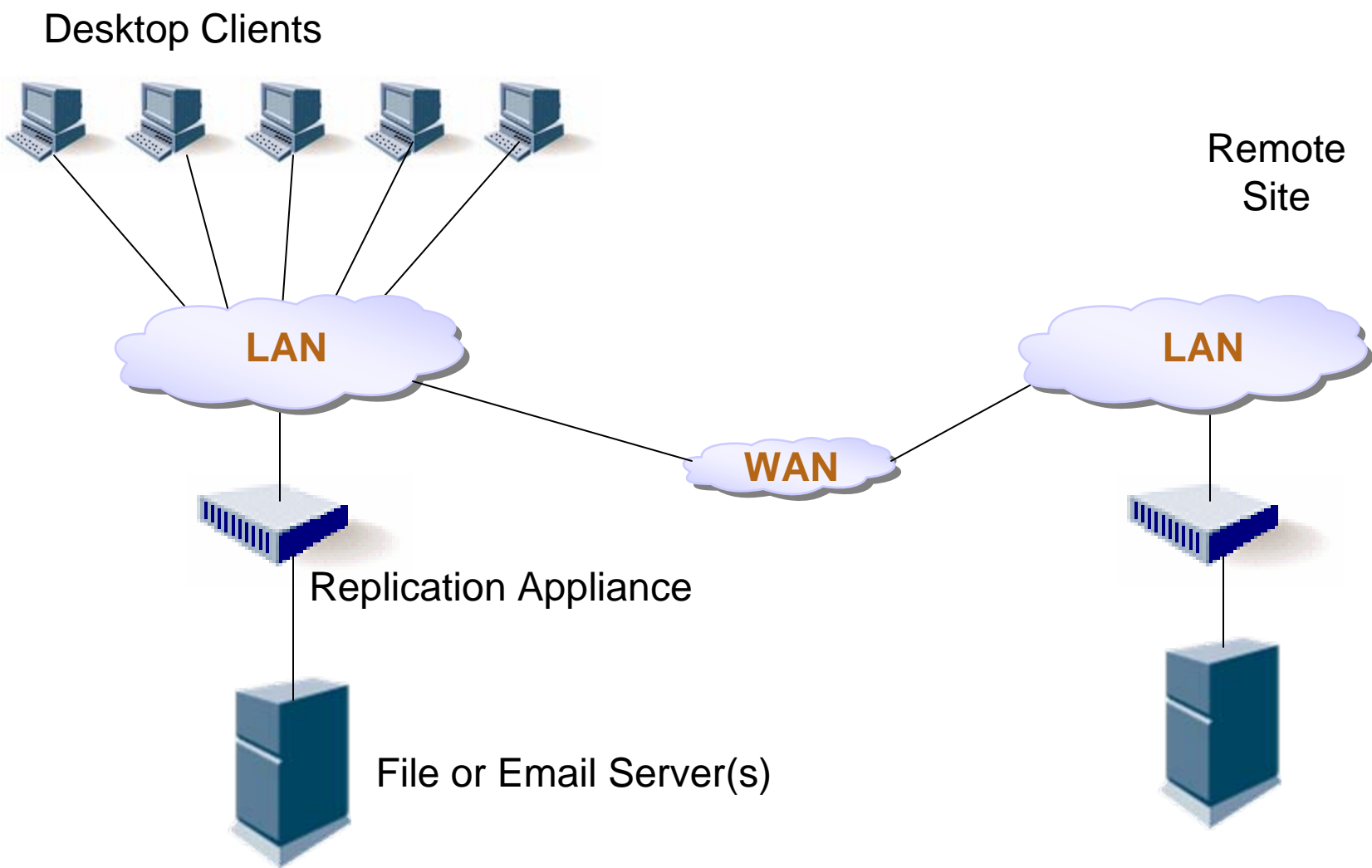
➤ Email High Availability and DR

- ◆ Appliances that sit in front of the email server, capturing mail objects and replicating them offsite.

➤ System State Replication and Conversion

- ◆ Application running on the LAN that captures system states from live running servers, replicates to remote location and updates drivers to compensate for disparate hardware.

In-Band LAN Replication Appliance



Replicating the Backup System

- Replicating conventional backup systems has traditionally been a challenge because backup systems generate a lot of redundant data, all of which would have to traverse the WAN.
- Four possible solutions:
 - ◆ Backup software that minimizes redundancy
 - ◆ De-duplicating storage device
 - ◆ CDP or block-level backup software
 - ◆ SAN array with integrated snapshot and replication

Optimized Backup Software

- **Incremental Forever Backup**
 - ◆ Only new files and changed files are backed up.
- **Synthetic Full Backups**
 - ◆ Full backups are synthesized from previous full and incremental backups. No need to pull a full backup over the WAN.
- **Software based de-duplication**
 - ◆ Redundant information is reduced to a single instance so that it can easily be replicated over the WAN.
 - ◆ Some products work at the file level. Others drill down beneath the individual file.



Check out “Eliminating Backup System Bottlenecks”

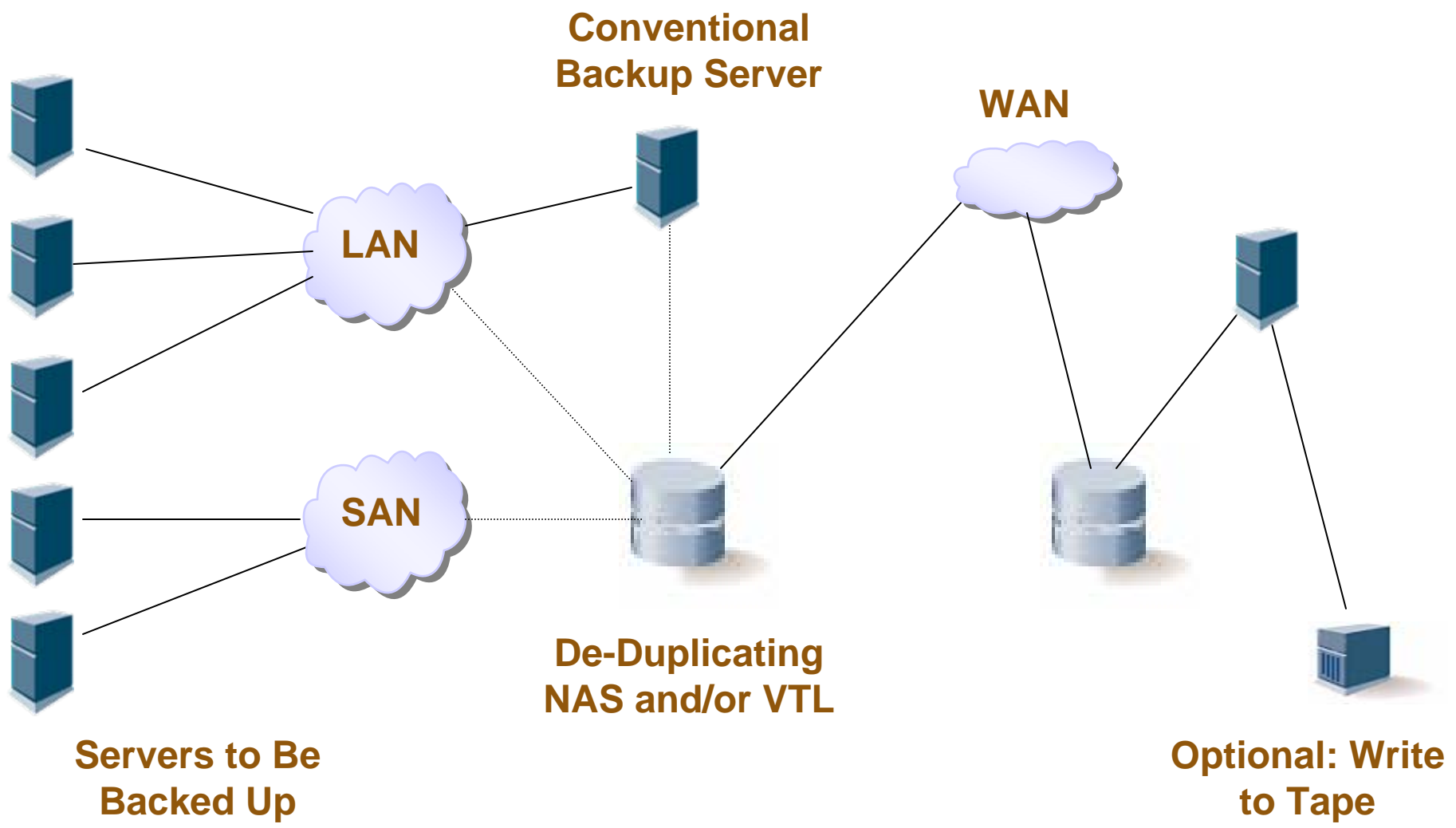
De-duplicated Storage Devices

- Storage devices that eliminate the redundant data generated by conventional backup software.
- Many de-duplicating storage devices include the ability to replicate over a WAN.
 - ◆ Allows your conventional backup system to replicate itself over relatively low-bandwidth connections.



Check out “Data Deduplication Implementation Overview”

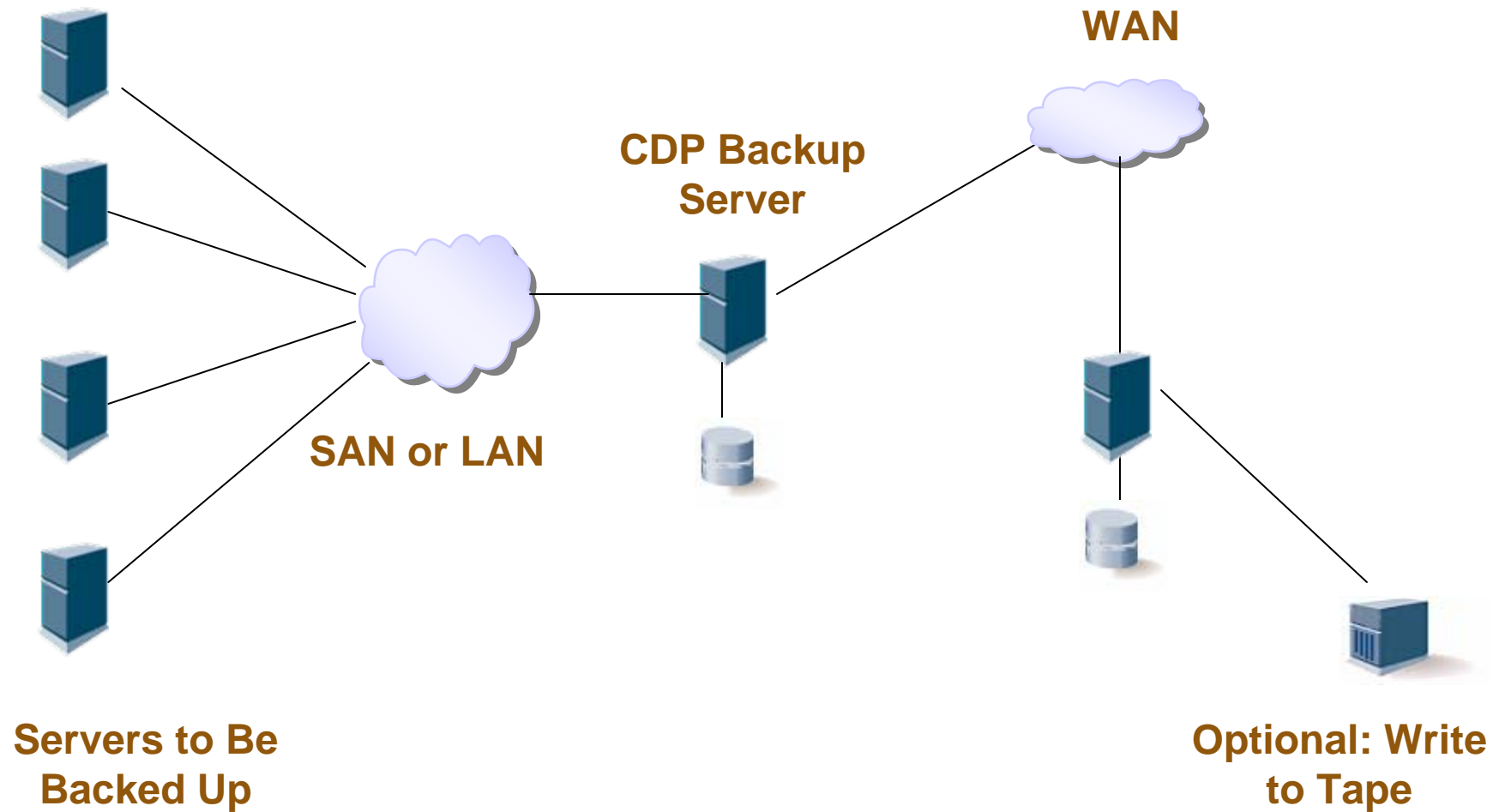
De-Duplication and Replication



CDP / Block Level Backup

- CDP is a generic term that refers to backup that happens on a continuous basis.
 - ◆ What it means to be *continuous* is a source of debate.
 - › Some say every change must be recorded. Others say that continuous simply means more than once per night.
 - › All agree that CDP refers to capturing changes beneath the file level, such that the amount of data being moved to the backup device is minimal.
- Many CDP solutions offer integrated replication.
- Many CDP solutions capture boot images and facilitate bringing systems back on line quickly.

CDP Backup with Replication



- Please send any questions or comments on this presentation to SNIA: trackfilemgmt@snia.org

**Many thanks to the following individuals
for their contributions to this tutorial.**

- SNIA Education Committee

Jacob Farmer, Cambridge Computer