Scalable Storage for Life Sciences

Presented By:

Jacob Farmer, CTO
Cambridge Computer

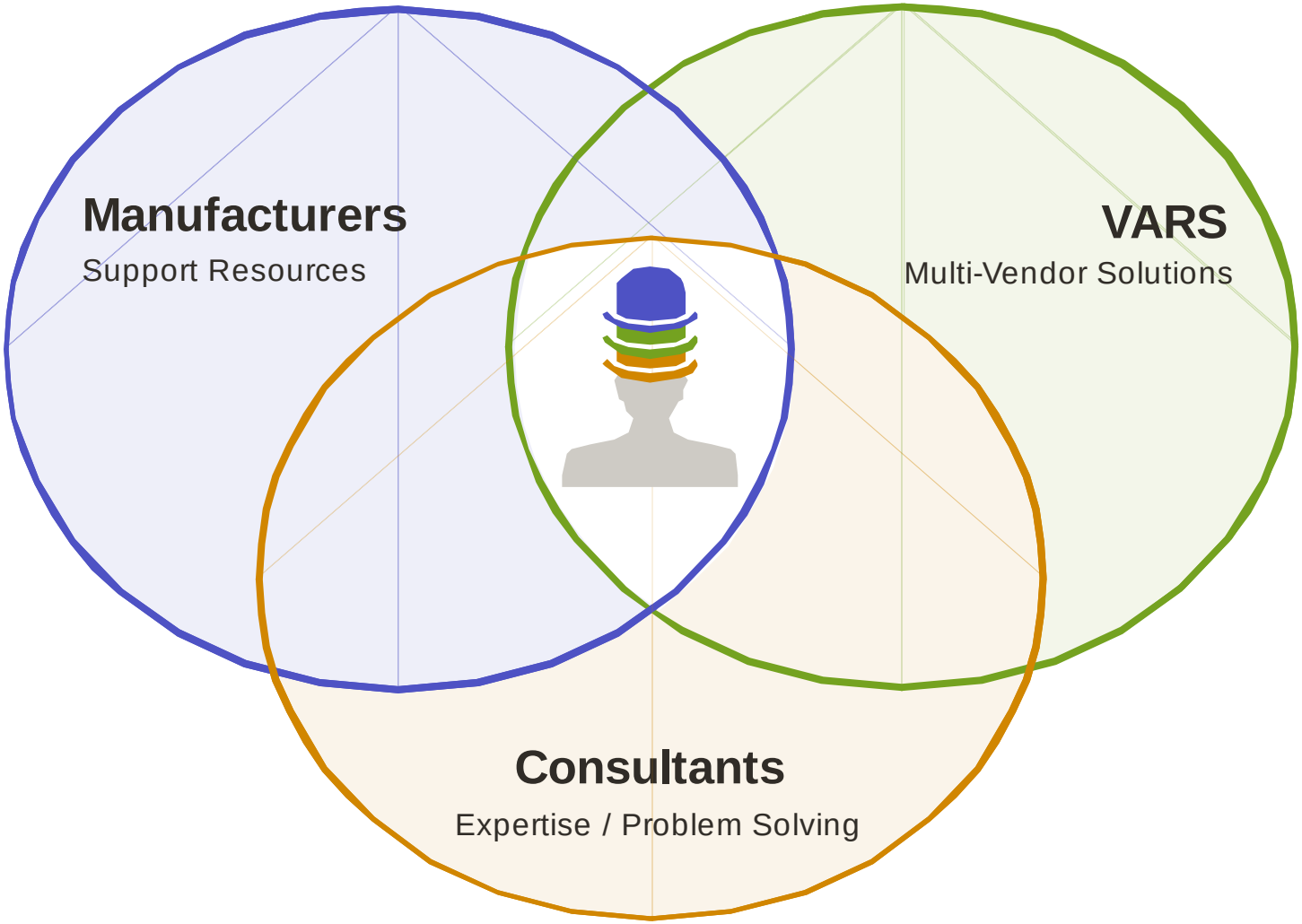
About Your Lecturer

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Jacob Farmer, CTO, Cambridge Computer
- 20+ years of experience with data protection, archiving, and storage management
- Hybrid of industry analyst and consultant to end-users.
  - Spend 25% of my time working in the industry, going to conferences, meeting with vendors.
  - 75% of my time customer-facing, helping the sales and services departments design solutions for end users.
- Lecturer at major trade shows and conferences
  - Usenix, Interop, Storage Networking World, BioITWorld, others
  - Travelling lecturer for Usenix (Usenix On-The-Road Lecture Series)
- Email: jfarmer@CambridgeComputer.com

About My Company: A Different Way to Shop for Storage Solutions

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



Agenda

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- A Crash Course in Storage Networking
  - Storage 101
- Trends and New Developments in Storage
  - Solid State Storage Devices
  - File Systems
- Storage Tiers
  - To Tier or Not to Tier
  - Where to Insert Tiering Logic
- Backup and Archiving
  - Various Topics on Backup
  - Tape/Archival File Systems

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

Storage 101

A Crash Course in Storage Networking

Files v. Blocks

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

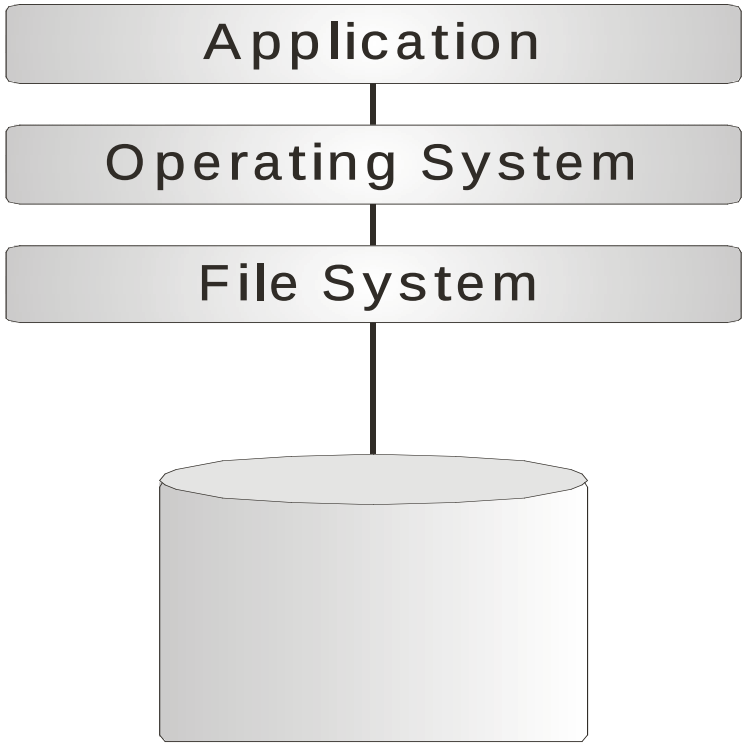
- Blocks
  - Least common denominator in conventional storage technologies.
  - A block is a unit of data storage.
  - Hard drives and RAID arrays serve requests for blocks.
- Files
  - Objects consisting of multiple blocks.
  - Blocks are organized into files by file systems, which are like databases of all of the files, their attributes and records of the blocks that make up the files.

Metadata

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Data that describes other data
  - File system metadata consists of the names of the files and directories, the file attributes, permissions, etc.
  - Backup system metadata are indexes (logs) of all of the files that were backed up, their location on tape, and perhaps copies of relevant file system metadata.

The **Storage** I/O Path

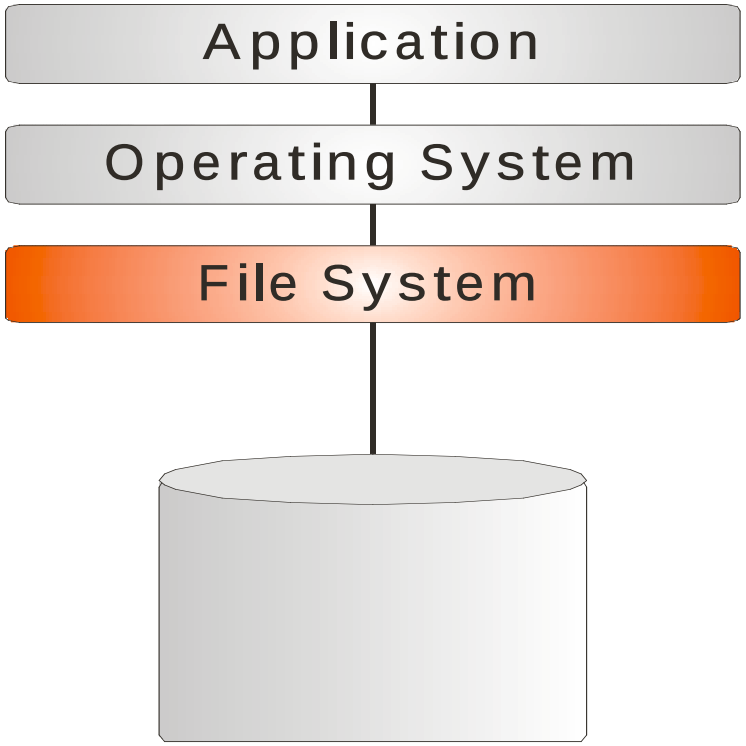


Storage has a layered architecture, very much like a network stack.

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.

File System Abstraction



- Abstraction layer or redirector is inserted above or next to the file system.
- Read and write commands are handled or filtered by the file system redirector.
- Applications do not know or care where the file resides.
  - As long as they get the data they were asking for!

Why Mess with Your File System?

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Network file services.
- Hierarchical and near-line file systems
  - Allow files to be managed on removable media and jukeboxes.
- SAN file systems and Parallel file systems
  - Allow file sharing at high speed and low latency over a SAN.
  - Parallelize network file system I/O processing.
- Capacity Optimization
  - Deduplication, Compression
- Security / Regulatory Compliance
  - Encryption, Immutability (write-once)
- More robust or journaling file systems
  - FAT → FAT32 → NTFS.
  - Reiser, Ext3, VxFS, etc.

NAS - Network Attached Storage

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- NAS = Fancy word for File Server Appliance
- Appliance = Fancy word for “proprietary”
  - Proprietary = two different meanings
    - Fancy word for “you buy it all from me at whatever price I dictate”.
    - Fancy word for “something better than you can build by yourself”.
  - Hardware and Software are bundled together
    - Software is NOT a perpetual license
    - If you replace the hardware, you might forfeit the software.
    - If you want to change the software, you forfeit the hardware.
- NAS is a marketing term that allows storage vendors to play in the file server marketplace by packaging file servers as storage appliances.
  - Storage vendor can compete for revenue with the server vendor
  - Storage vendor can compete for revenue with the operating system vendor

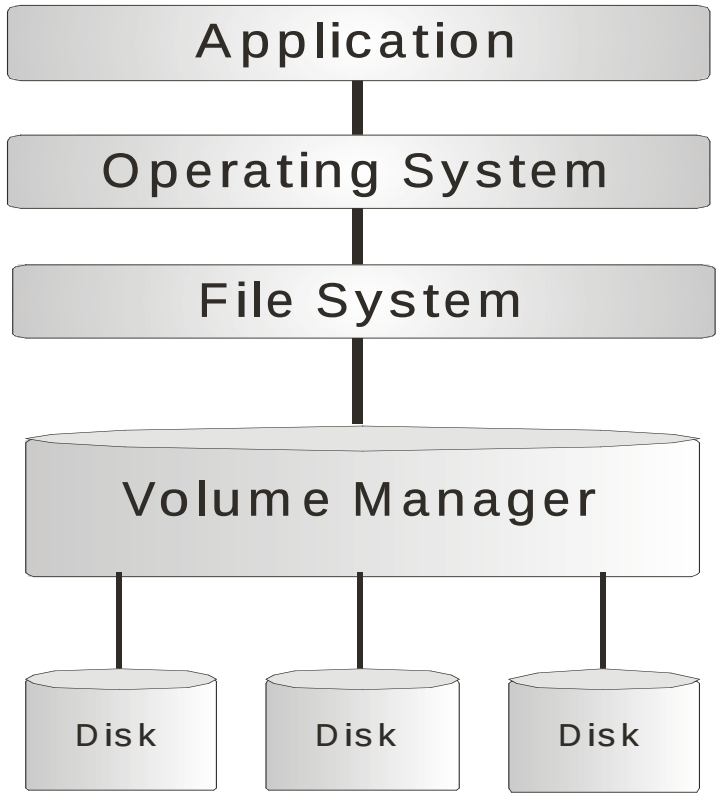
Block-Level Abstraction

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE

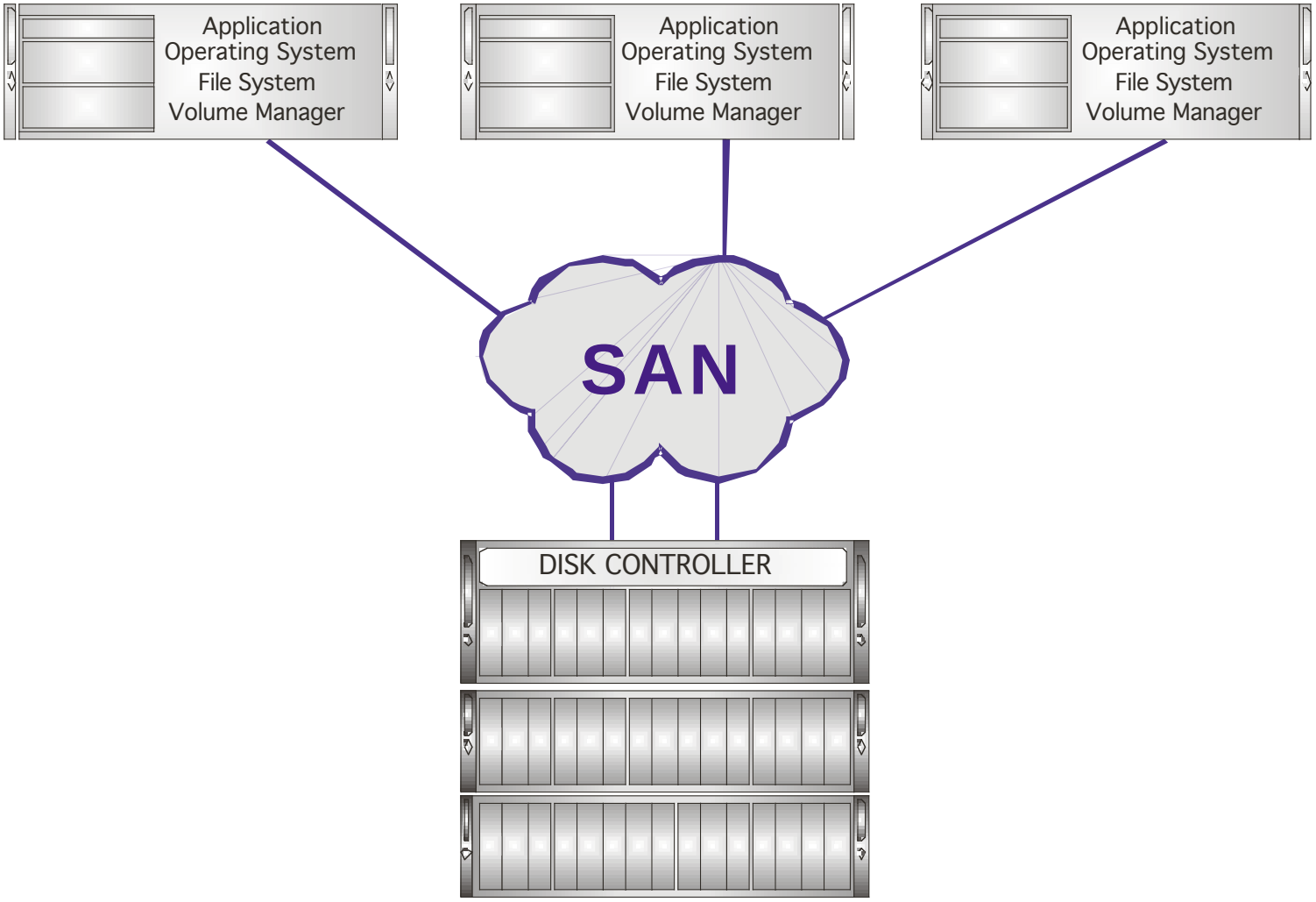
Abstraction of the physical disk.

File system asks for specific block addresses and volume manager fulfills request from the disks.

- Software RAID
  - Hard Drive Fault Tolerance
  - Spindle Aggregation
- Host-based mirroring
- Volume-level snapshots
- Volume-level replication

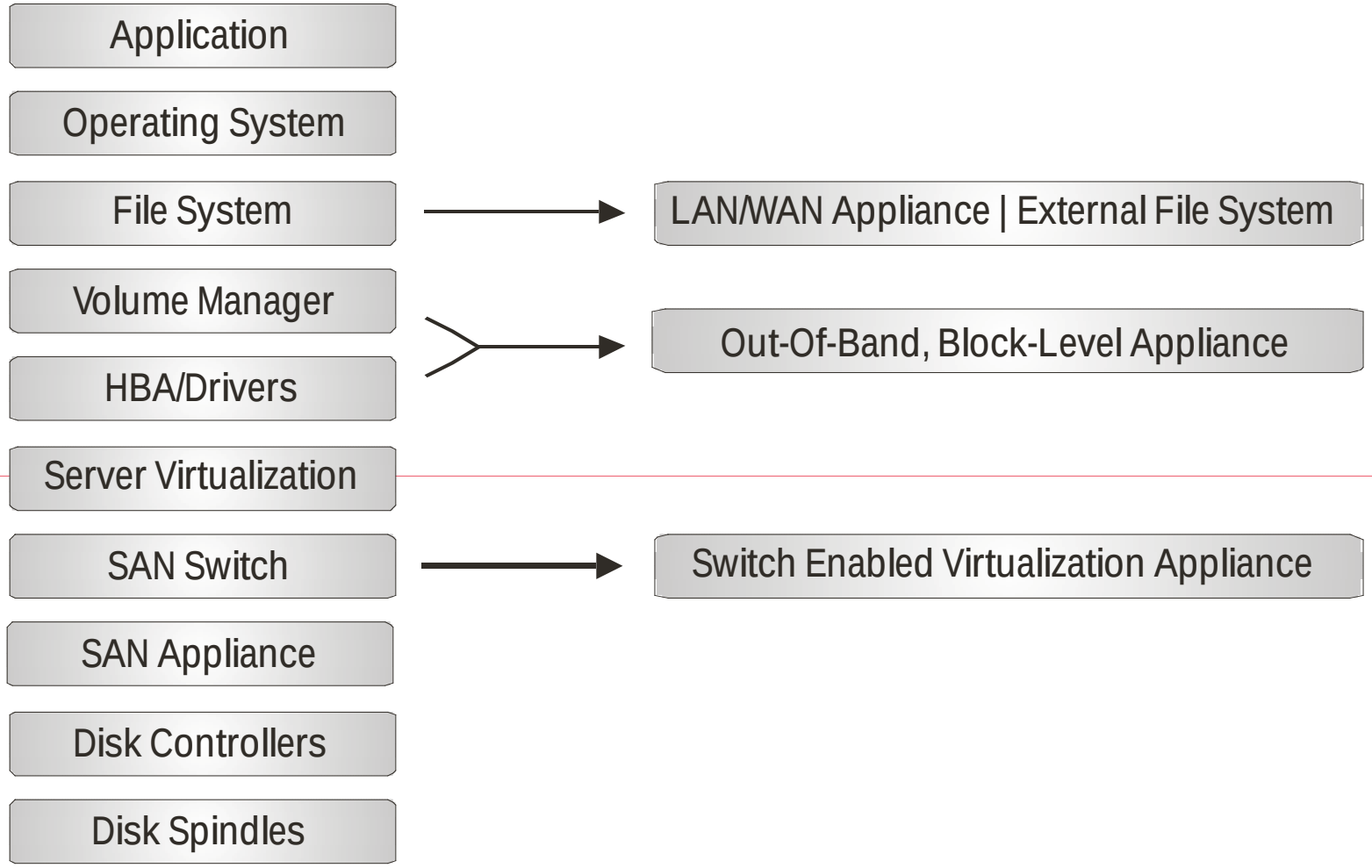


SAN Array = Centralized, External Abstraction



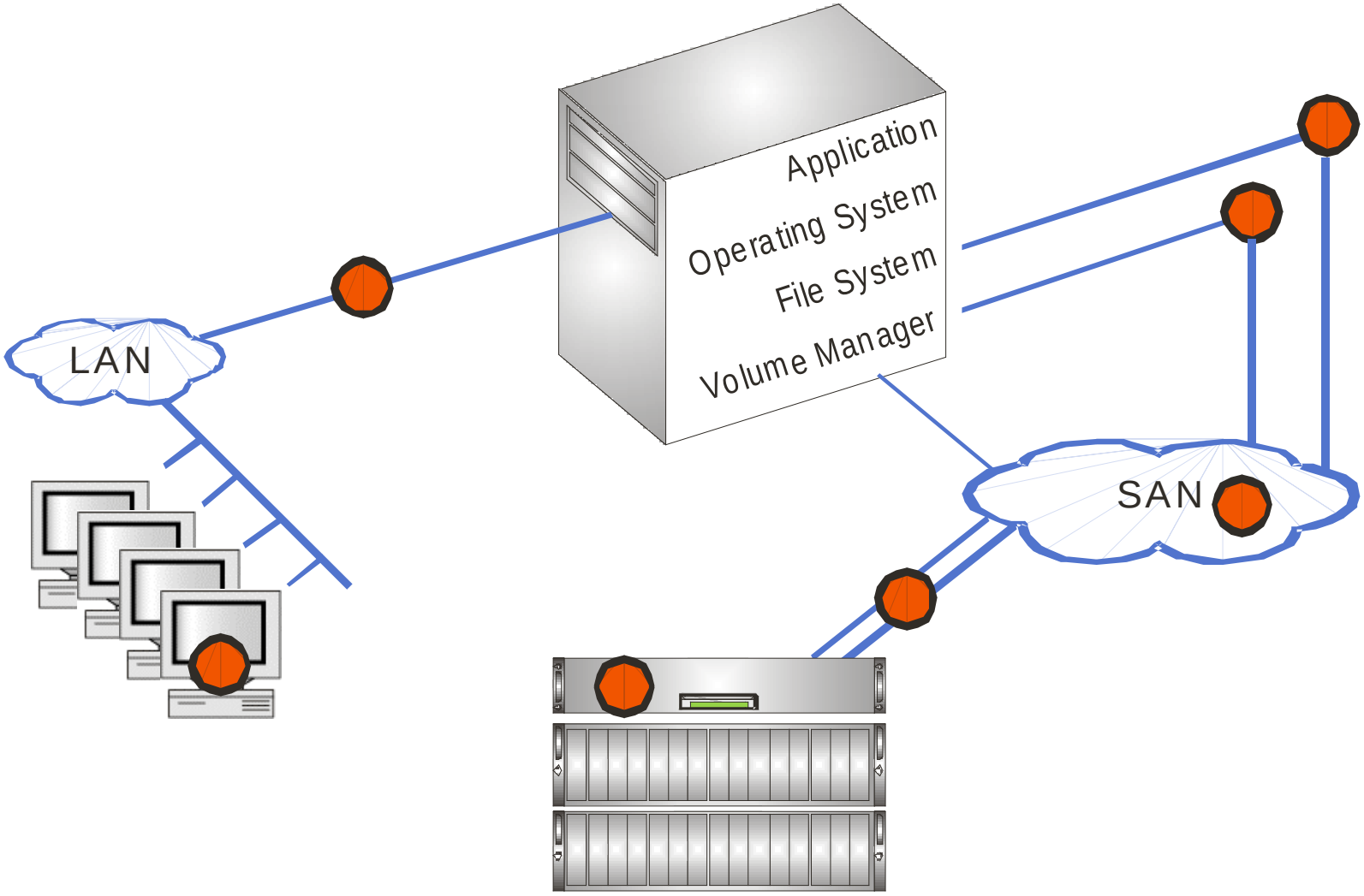
Summary of the Storage I/O Path

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE



Inserting “Virtualization” Logic into the Storage I/O Path

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

Trends and New Development

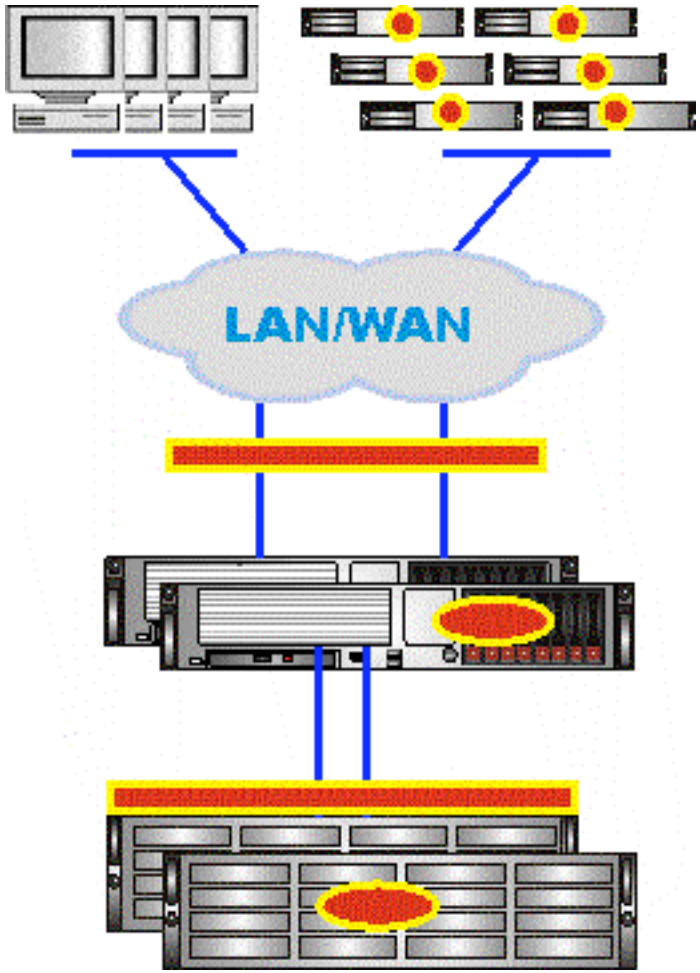
Storage is a Vibrant Field

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Solid State Disks – New Affordable Choices
  - Performance characteristics that are 3x to 500x that of conventional hard drives
  - The industry is looking for novel ways to insert SSD into storage system architectures
- File Systems
  - Scalability without becoming “brittle”
  - Self-healing
  - Data integrity assurance
    - Ability to detect data corruption
    - Ability to correct data corruption

Inserting SSD into the Mix

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

Tiered Storage, Workflow, and Life Cycle Management

Definitions

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Tiered Storage
  - Managing multiple types of storage devices to better batch costs, capabilities, and properties of storage devices to the data being stored on them
- Life cycle management
  - Movement of data between tiers based on frequency of access or recentness of access
- Workflow management
  - Movement of data automatically between stages of data processing

To Tier or not to Tier?

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Can all of your needs be met by a single tier solution?
  - If you can meet all of your needs in a single tier, than do it.
    - Example, SATA array with a heavily cached file system.
  - Sometimes the premium you pay for tiering technology outweighs the cost of growing the top tier.
- Tiers are typically thought of as defined by drive types: 15K, 10K, SATA, maybe SSD.
  - But a tier can be any set of properties
    - Data integrity assurance
      - Addressing silent data corruption comes at a performance price.
    - Fault tolerance
      - Fault tolerance comes at a monetary price.
    - Data Protection Policy
      - Backups, snapshots, replication ,etc.

Common Tiers

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE

- Scratch space
  - Funds are prioritized around performance instead of fault tolerance and data protection.
    - Set expectations with users that if it breaks, they lose their data.
  - As scratch systems get bigger, data loss might start being a serious problem.
    - How much does it cost to re-run 250TB worth of data?
- User work space
  - Home directories and collaboration space.
  - Needs to be backed up. Snapshots are nice.
- Archive
  - A place to move data that is not in active use.
- Deep archive
  - A place to bank your raw data in case you need to rerun experiments in the future.

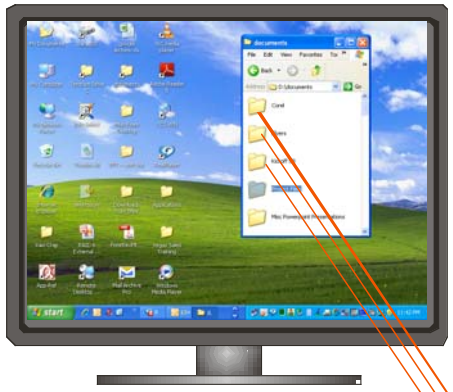
Where Can You Insert Tiering Logic?

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE

- In-band in the data network between clients and file servers
- Out-of-band in the data network
- In a file system or NAS
  - But be careful to weigh the costs and consider the long-term commitment to the specific product
- In a disk array
  - Some disk arrays can manage tiered storage at the block-level

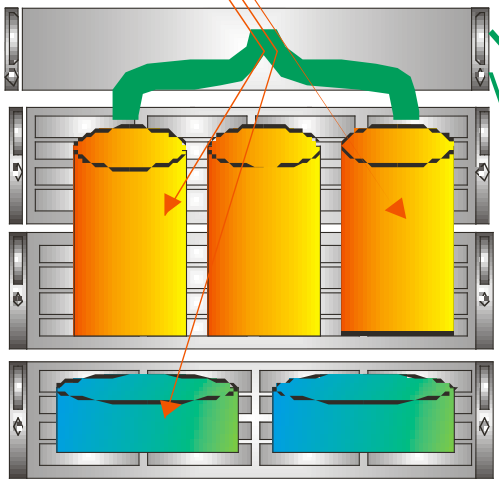
NAS Appliance with Integrated Virtual Namespace

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



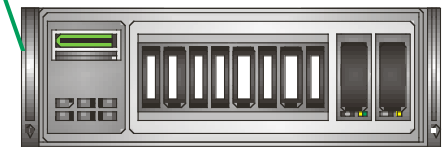
Files are automatically moved between tiers based on policies

NAS Appliance with integrated virtual namespace and tiered storage



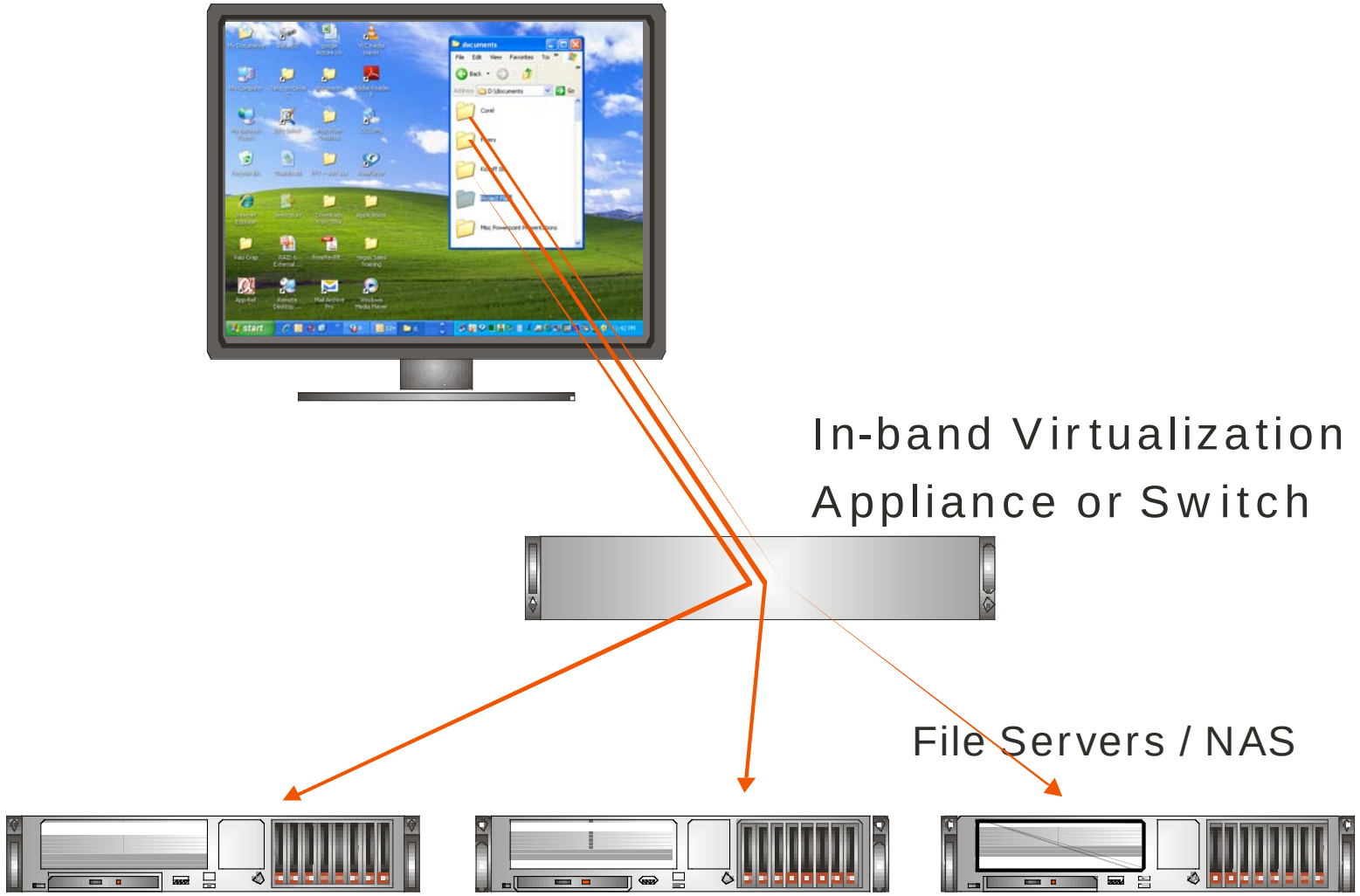
Bulk Disk or CAS

Tape Library



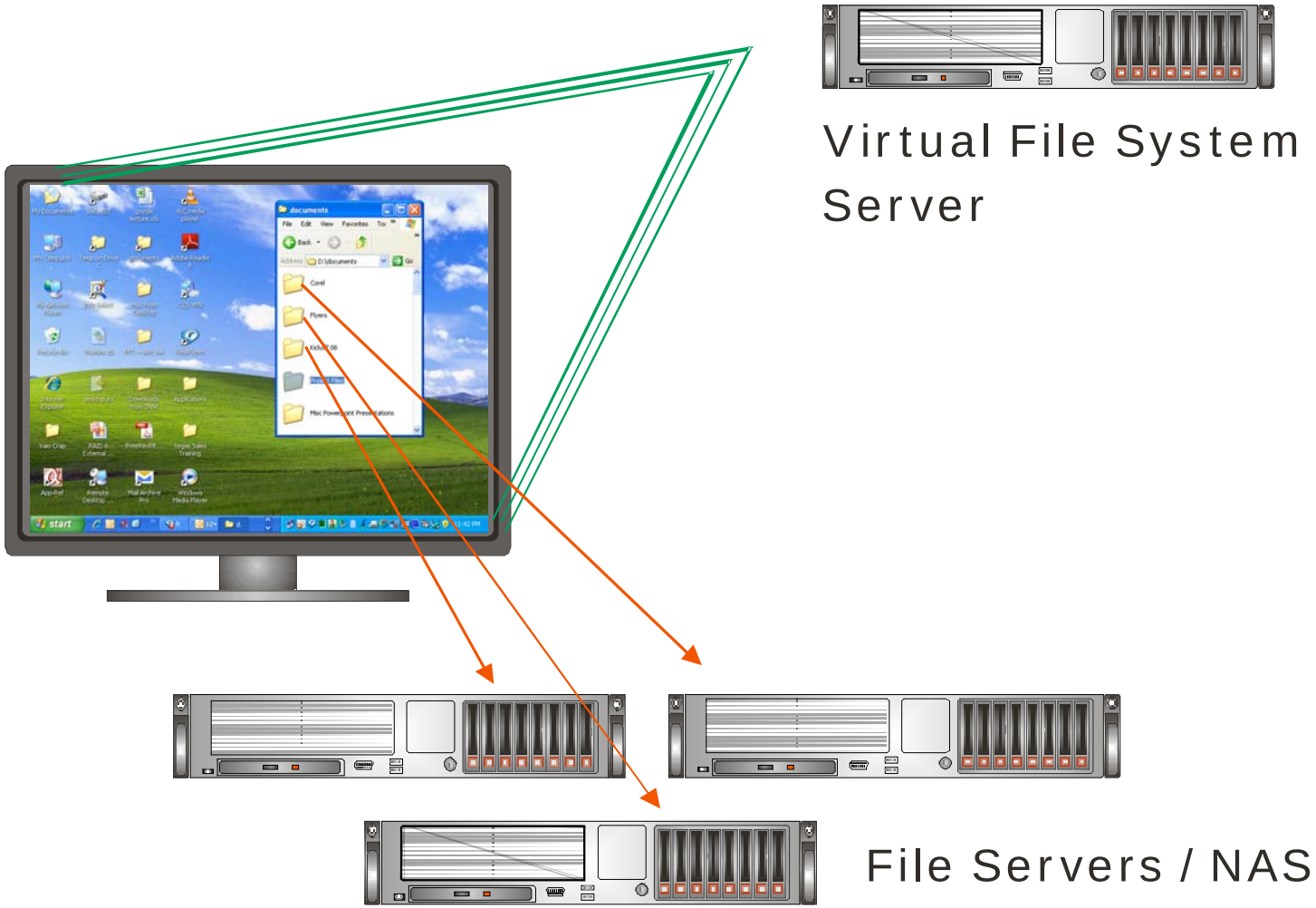
In-Band File System Virtualization

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



Out of Band File System Virtualization

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



General Recommendations on Tiered Storage

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE

- Minimize the number of tiers you have
  - The fewer tiers the easier to manage
  - You can always add more tiers and management later
- Have at least two tiers:
  - General purpose storage
  - Secondary copy
    - Maybe a scaled down version of your general purpose storage
    - Maybe an archival/backup file system at a much lower cost
- Keep an eye on SATA-SSD combinations
  - NOTE: "SSD" might not be called out by name. Many caching file systems get the same benefit as SSD drives.

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

Backup and Data Protection

The Importance of Backup

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Fault tolerance storage devices break all the time!
  - Drive failures, controller failures, data corruptions
  - File systems get “brittle” as they grow large
    - The bigger they are the harder they fall
- Ideally, your backup system enables you to move a copy of the data off-site
- Your backup copy could eliminate doubt about the integrity of the data
- Backing up scratch space
  - Many people don’t, but there could be real lost productivity if scratch space were compromised.
  - How much is sequencer time worth?

Backup Requirements / Policies

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE

- What are your backup requirements?
  - RPO – Recovery Point Objective
  - RTO – Recovery Time Objective
  - Retention
- Do you need to retain versions of files?
  - How do you plan to access those files?
  - Can the users help themselves?
  - Can you do it through a file system snapshot?
- Are you worried about file system corruption and having to roll back the whole file system?
- Do all of your file systems or data types have the same backup requirements?
  - Can you “tier” your storage around data protection requirements?

Backup / Offsite Storage Options

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Rsynch and its commercial equivalents
  - Replicate/copy your primary file system to a secondary file system
  - Even better if target file system understands versions or can perform snapshots
- NAS with integrated snapshot and replication
- Conventional enterprise backup
- Mirrored or multi-site file system
  - Sometimes these are described as replicated “opbject-based” storage systems

Limitation of Enterprise Backup Systems for Research Data

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE

- Enterprise backup systems (Legato, NetBackup, etc.) perform repetitive full backups, even when data has not changed
  - Enterprise users struggle with file systems as small as a few hundred Gigabytes
    - Research data sets are too large for repetitive full backups
    - The vast majority of research data does not typically change, so there is no benefit to repeated full backups.
  - Note: TSM from IBM is based on incremental backups. It is thus more suitable for research data, but still problematic.
- Data has to be restored before it can be accessed
  - If you lose data, you have to restore it back to your disk before you can access it.
    - This takes too long
    - If you disk is broken, you have to first fix the disk before you can access data

Conventional Tape Backup/Restore is Too Slow

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE

- An LTO-4 tape drive at full tilt, can theoretically back up 10TB per day
  - One tape drive could theoretically back up a Petabyte in 100 days
  - Yes, you can run multiple tape drives, but tape performance does not scale linearly and ultimately it becomes hard to manage
- If it takes 100 days to backup a Petabyte to a single LTO-4 drive, how long is going to take to restore?
  - We typically estimate restore time at 1.5x to 2x the backup time
- It is difficult to prioritize restore processes
  - Generally, you restore everything before people go back to work
  - Wouldn't it be nice if you could prioritize restore around the most recently accessed files?
    - You can theoretically do this, but it would be a highly manual process.

Tape is Not Your Enemy

CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

- Tape is not the enemy
  - Conventional enterprise backup software is the enemy
- Tape is really inexpensive
  - Incremental cost per TB with LTO-4 is as low as $50.00
    - It is practical to keep multiple copies of your data
  - Yes, of course, there is a cost to tape libraries, management software, and labor, but the costs are very low compared with disk systems
- Power and cooling
  - Tape sits idle most of the time
  - Idle tape libraries consume power akin to light bulbs
    - Disk systems are akin to hair dryers

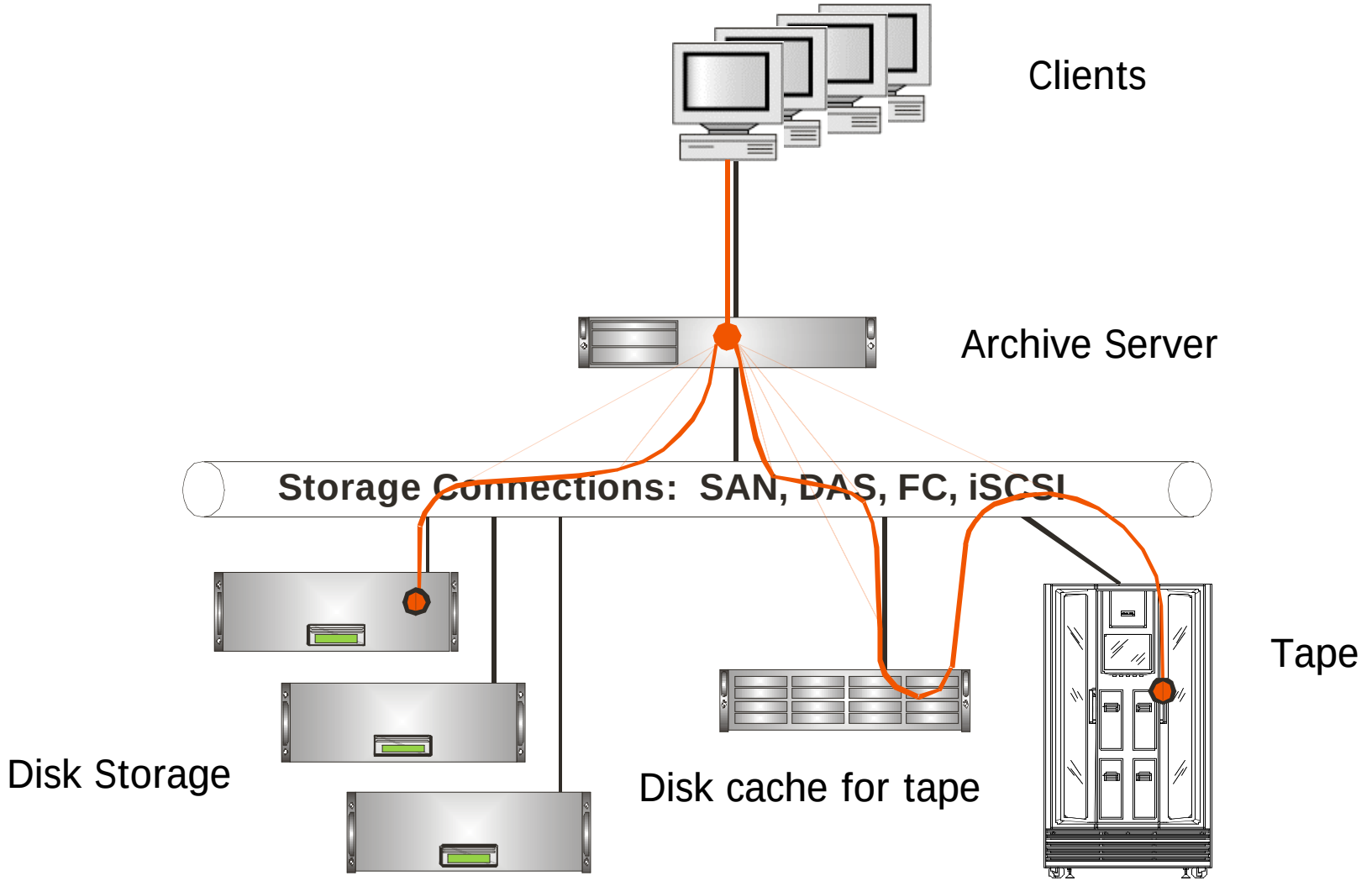
Tape File Systems

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE

- Tape storage systems that emulate conventional file systems
- No need to restore files
  - Access them directly from tape
  - Perhaps with the aid of a disk buffer
    - Disk buffer can be large enough to accommodate likely working set of data
- Tape file systems can be used for backup as well as long-term redundant storage
  - Tapes can be made redundant and a set can be stored off-site
- Guaranteed data integrity
  - Data is broken up into shards and checksum is calculated
  - If a file failed checksum test it is pulled from alternative media
- Ingest rates of 3-5TB per day with just two tape drives

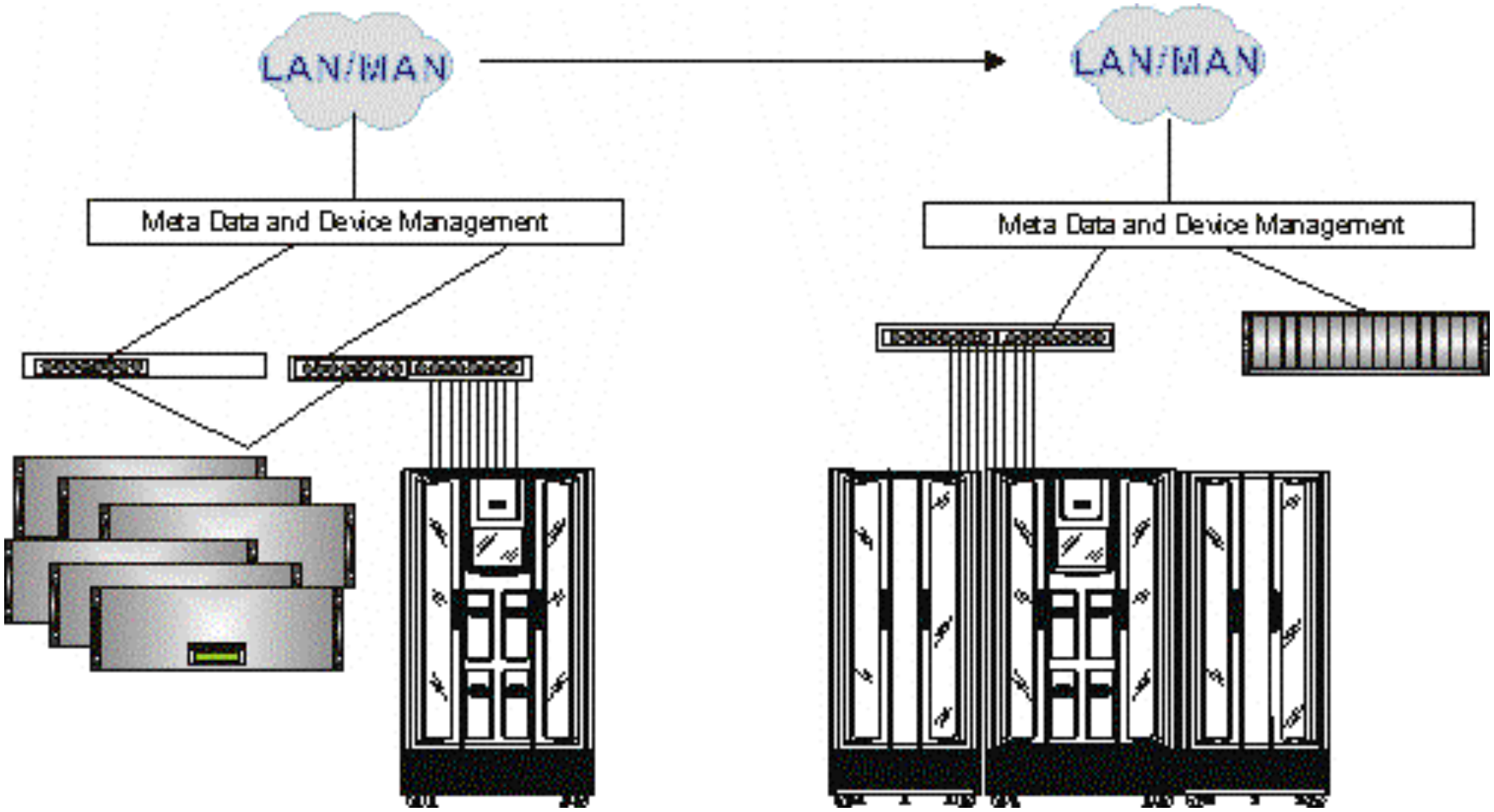
Archival File System with Both Disk and Tape

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



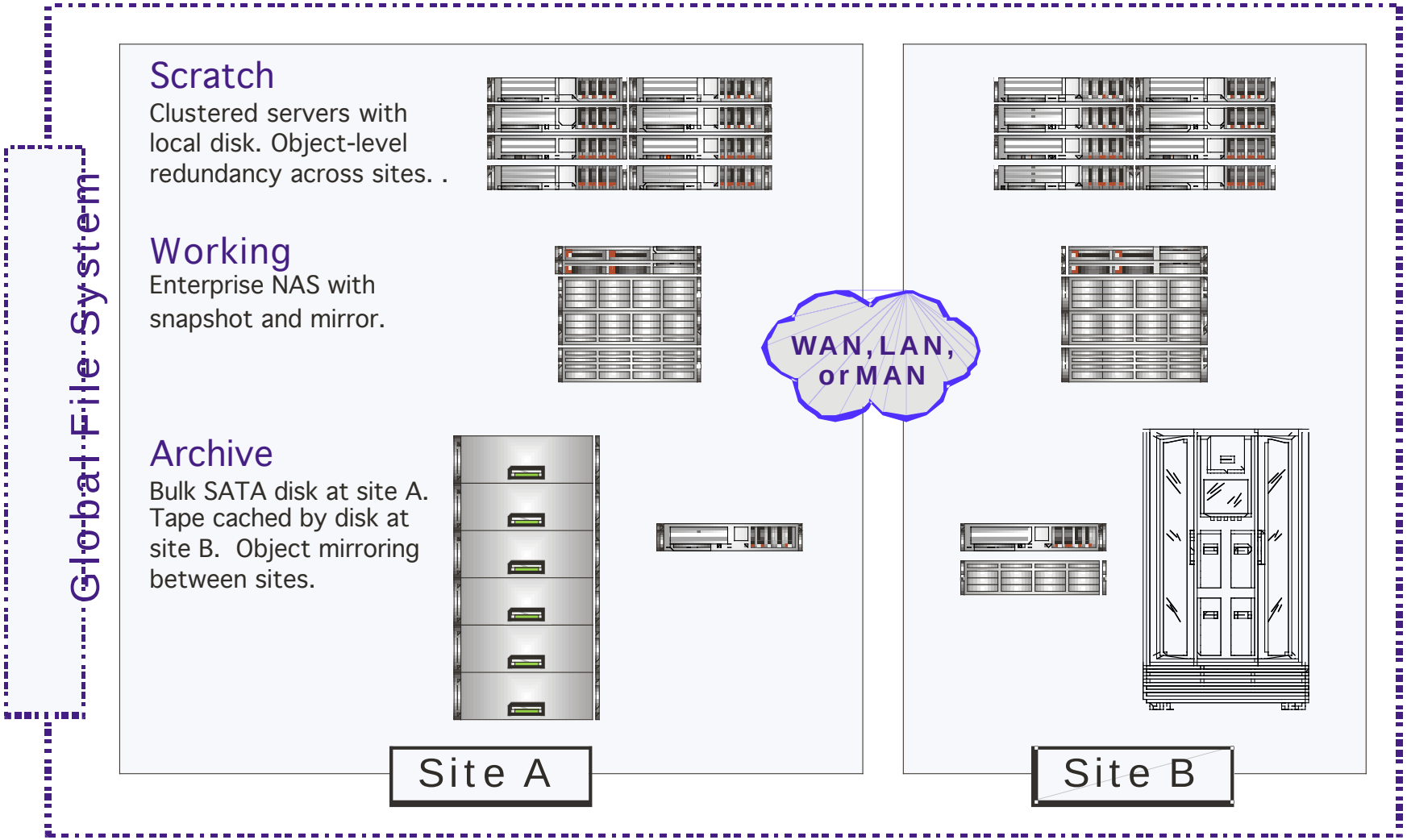
Archival Tape File System Extended Between Two Sites

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



Multi-Tier Global File System – Redundant Across Two Sites

CAMBRIDGE Computer
ARTISTS IN DATA STORAGE



CAMBRIDGE
Computer
ARTISTS IN DATA STORAGE

Questions?