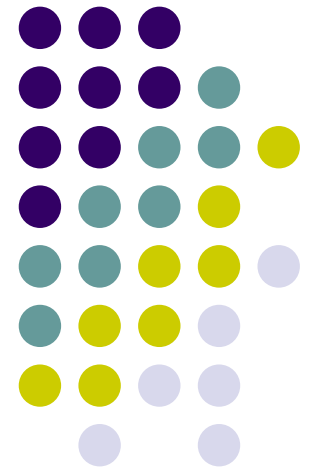


Storage Area Network

File System Tutorial





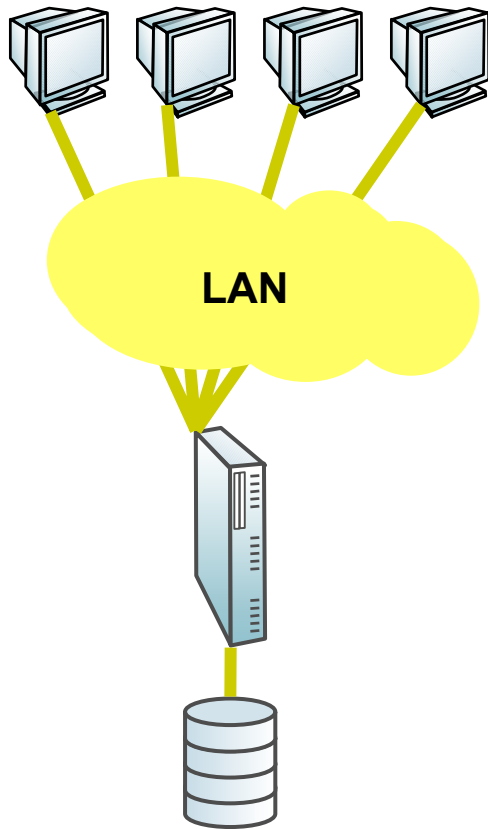
Agenda

- Inside a Traditional File System
- Requirements
- SAN Review
- Inside a SAN File System
- Implementing a SAN File System
- The Future

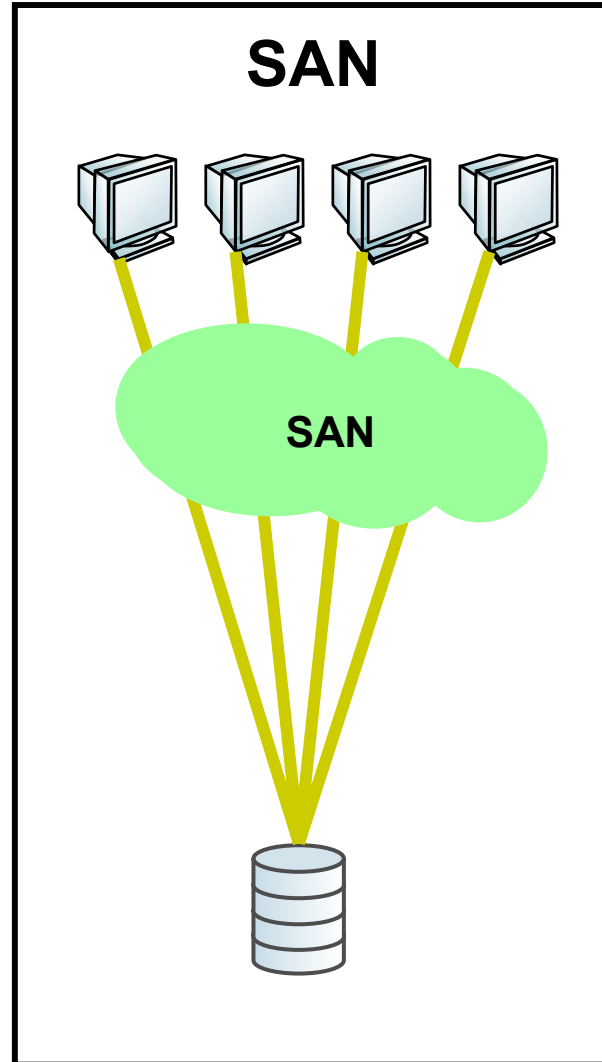
Choice of Architecture



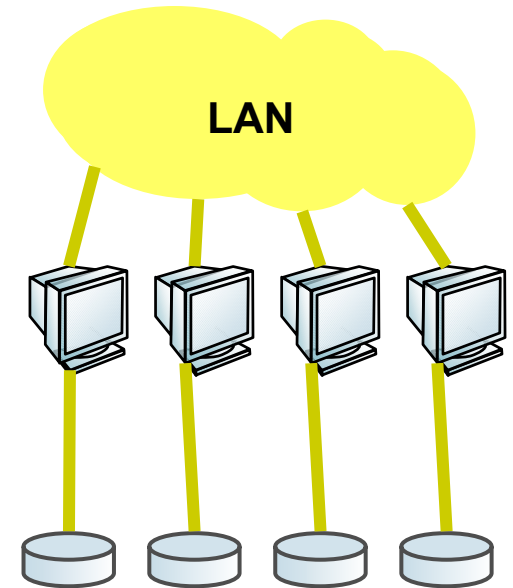
NAS



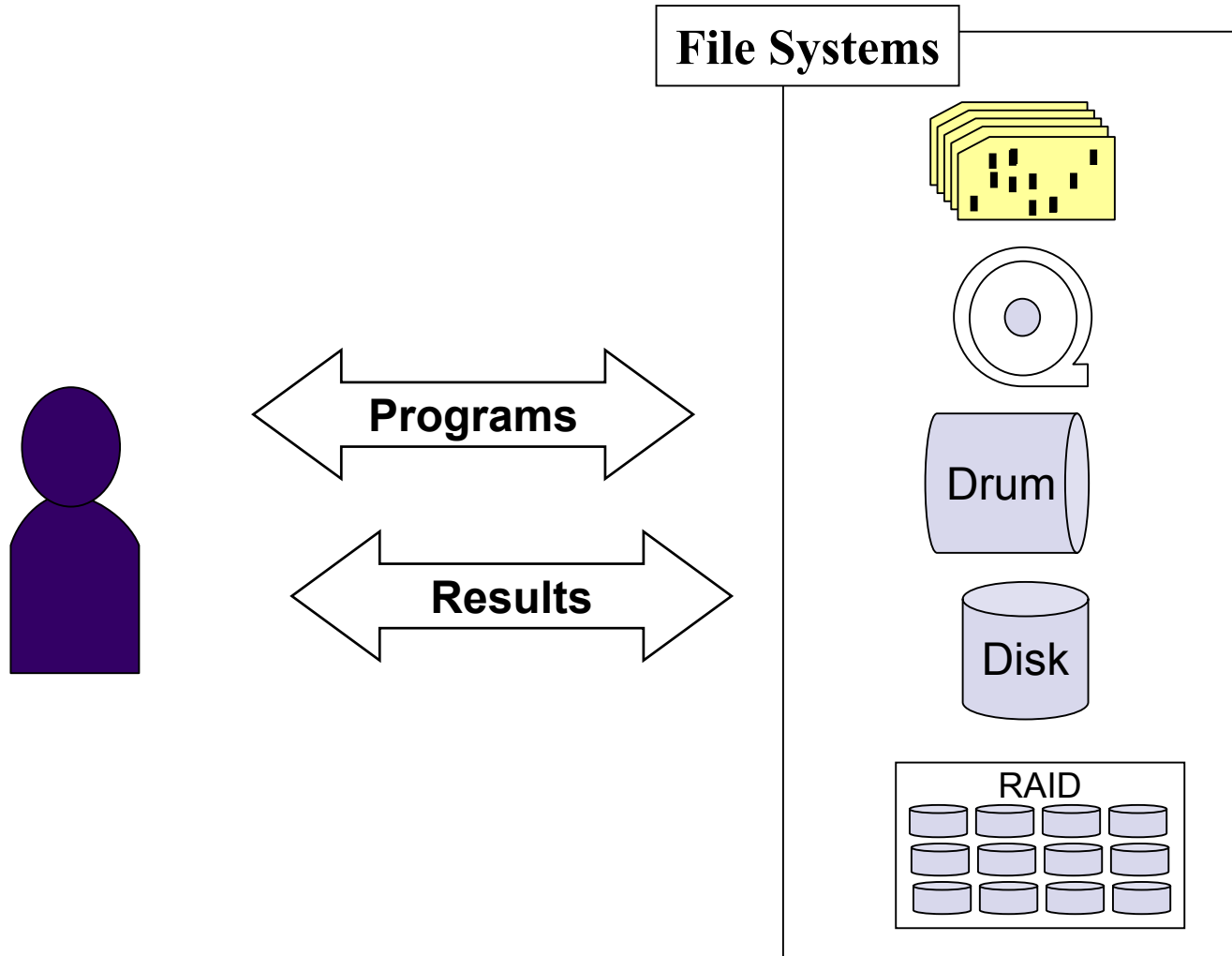
SAN



DAS

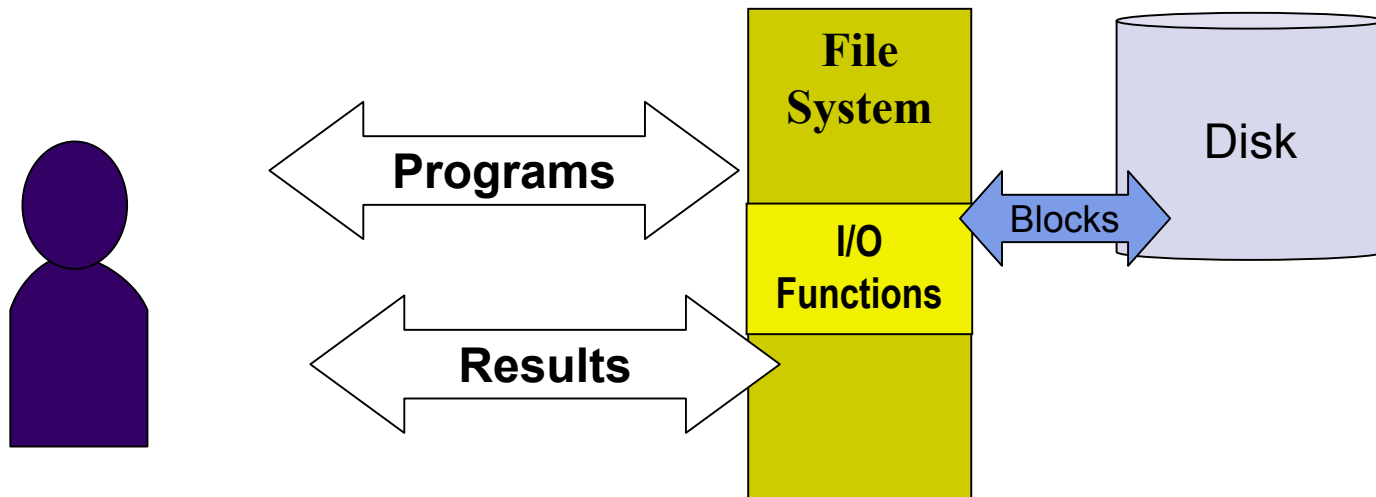


File System Basics



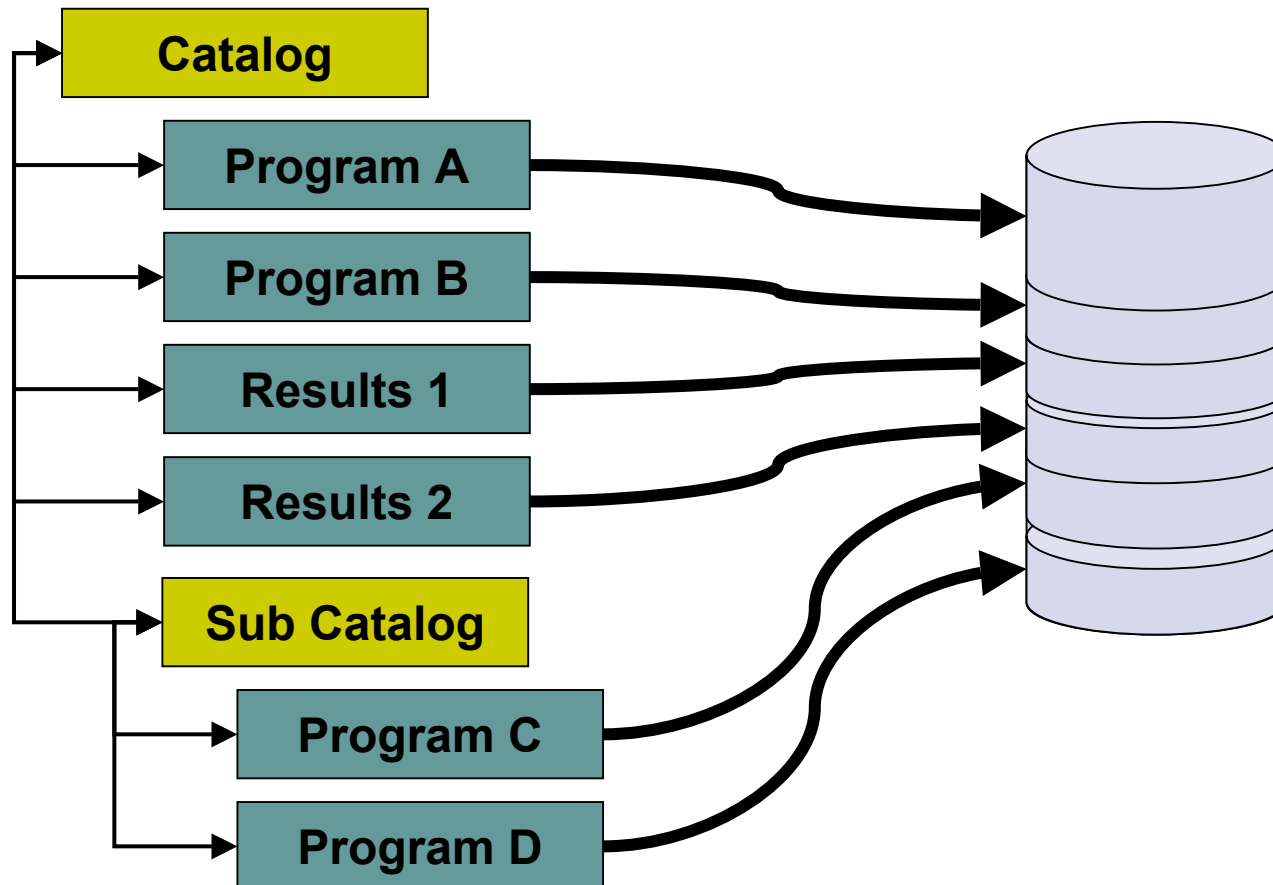
File System Basics

Core Functionality



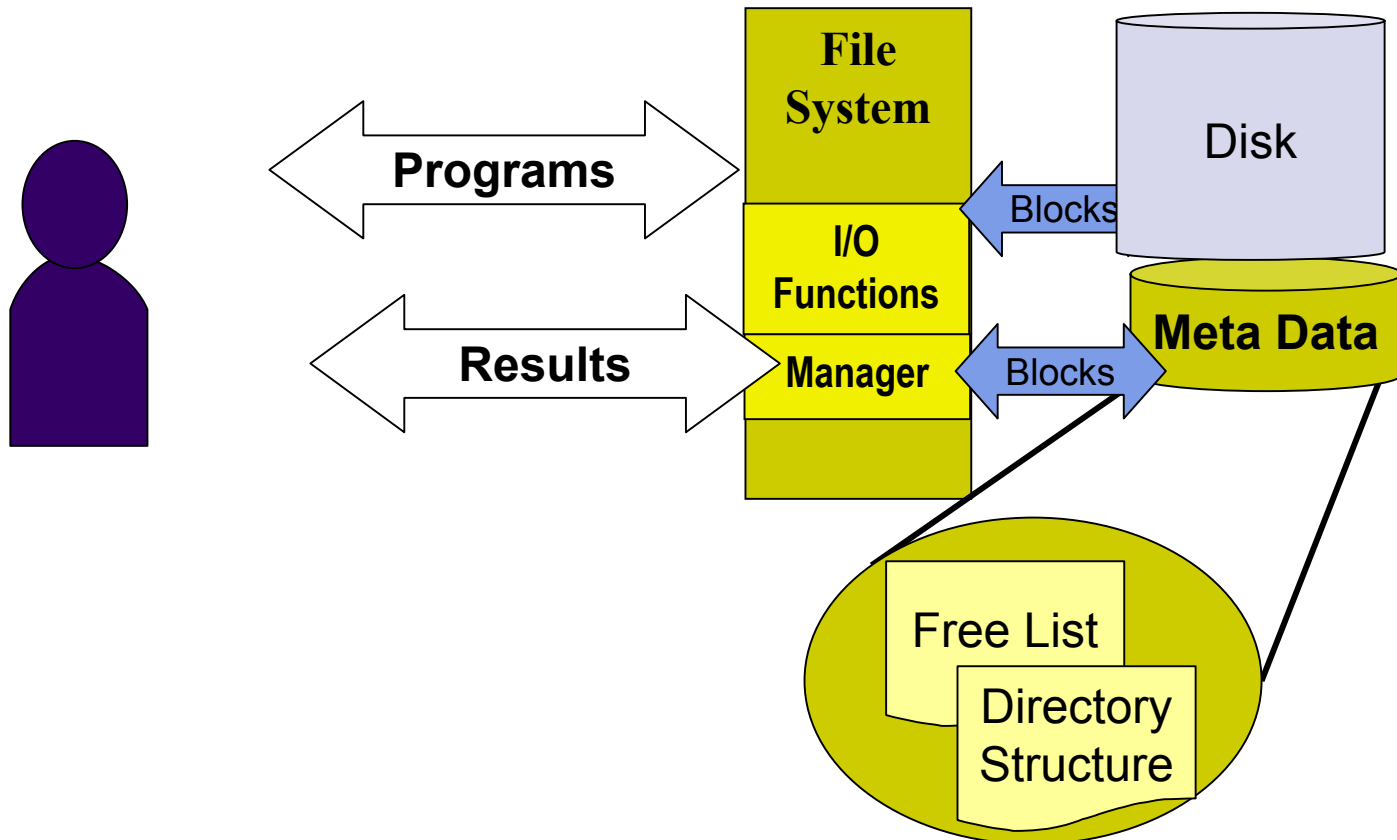
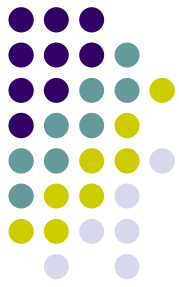
Storing Many Things

Directory Structure



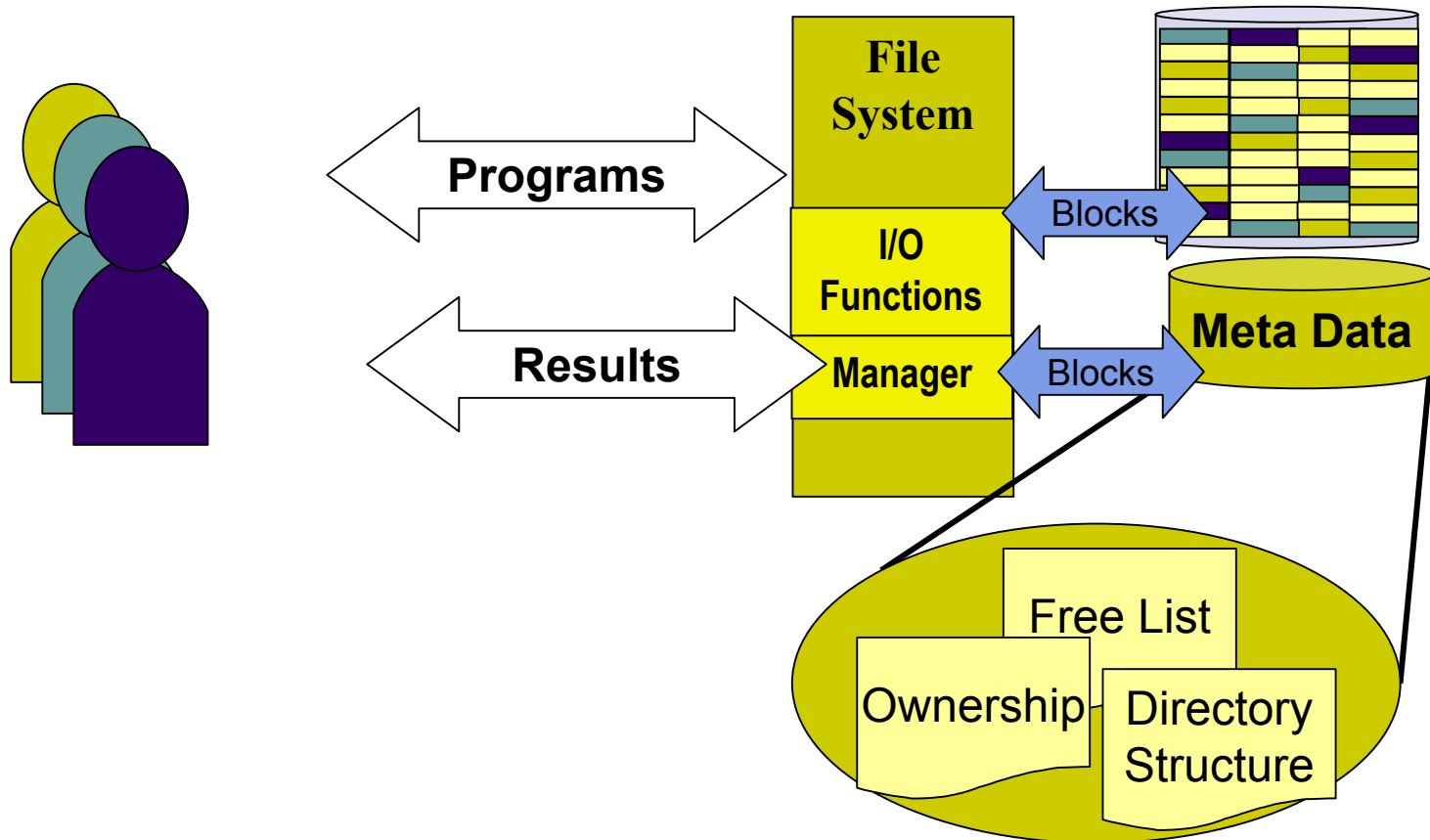
File System Basics

Meta Data



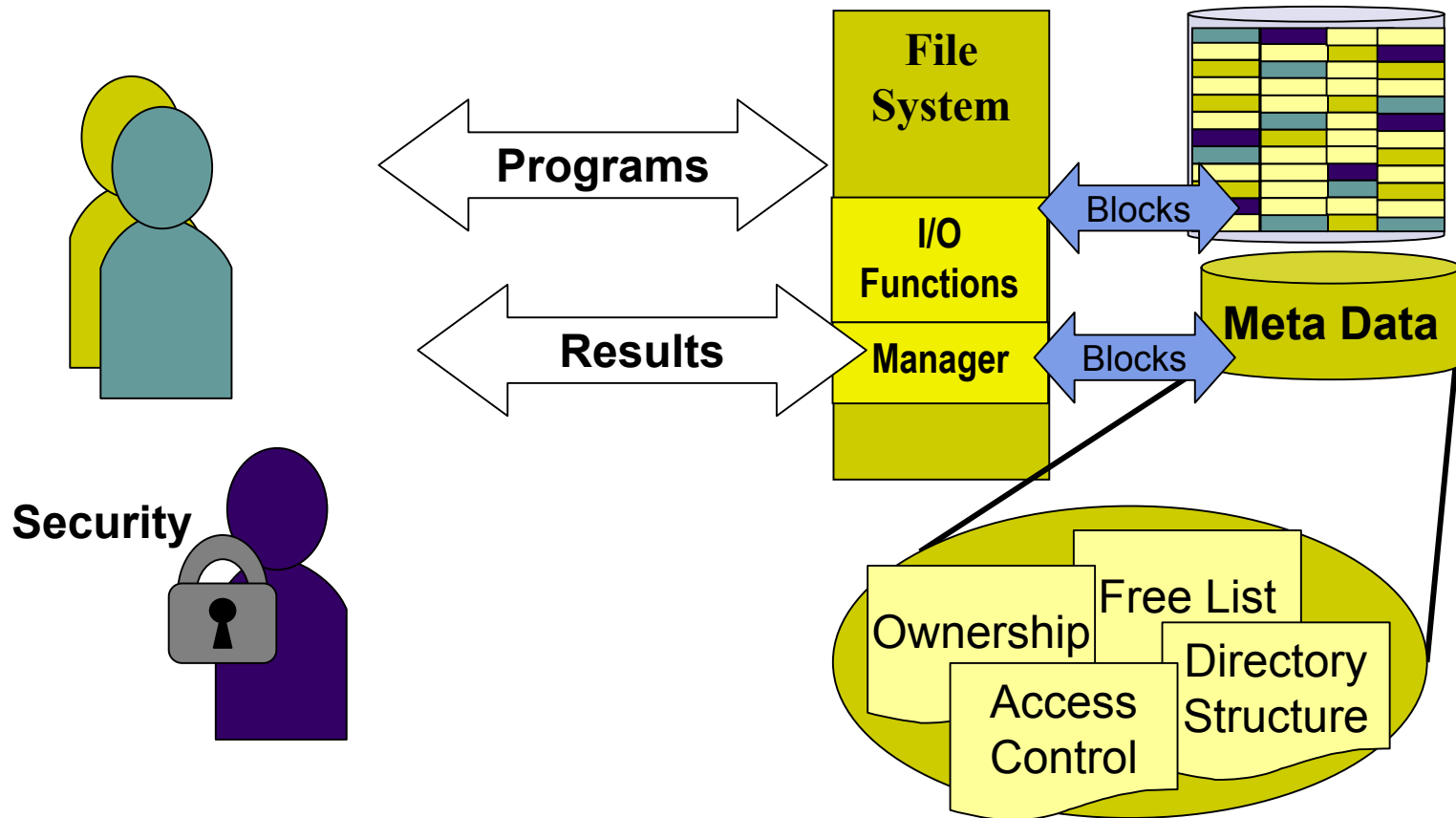
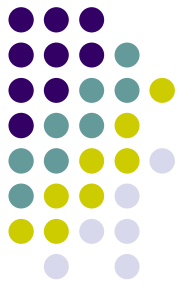
File System Basics

Multiple Simultaneous Users



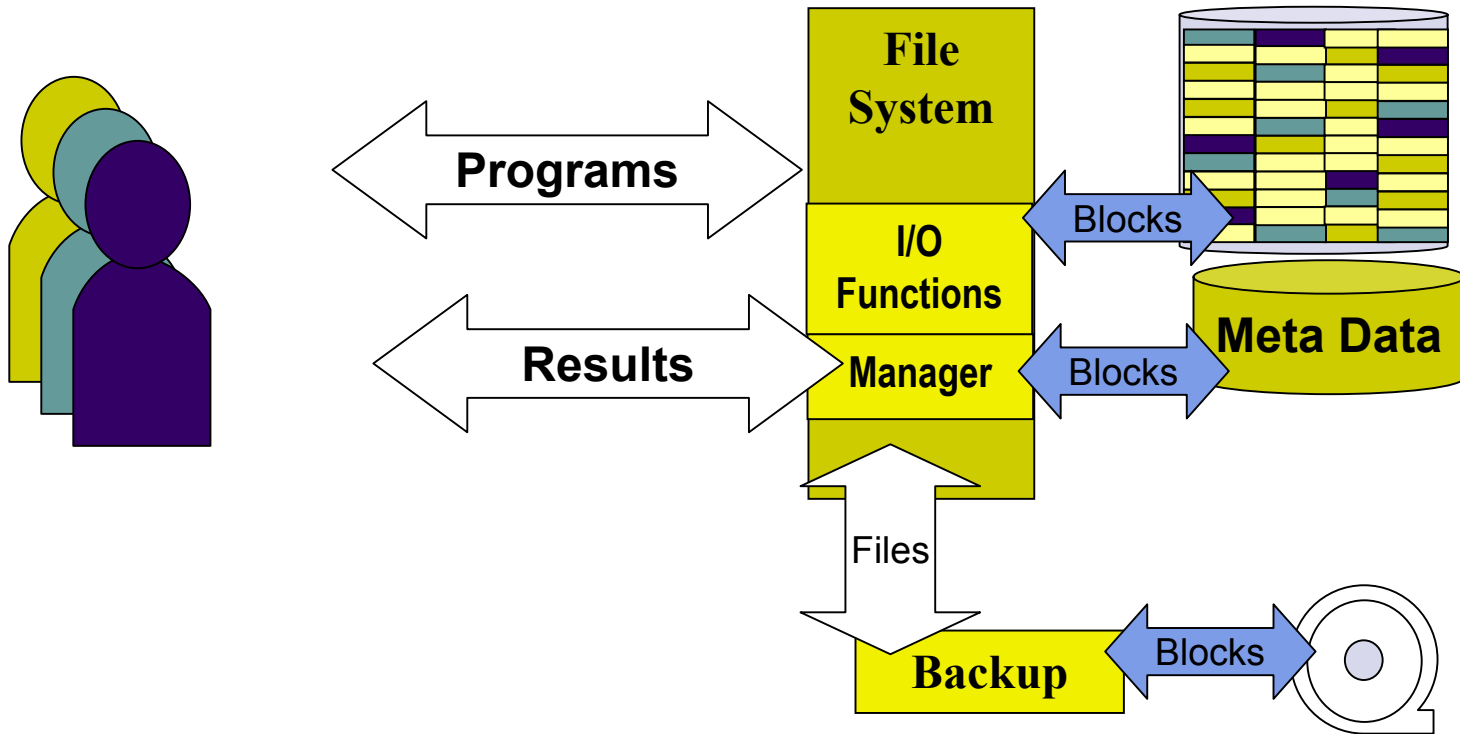
File System Basics

Access Control



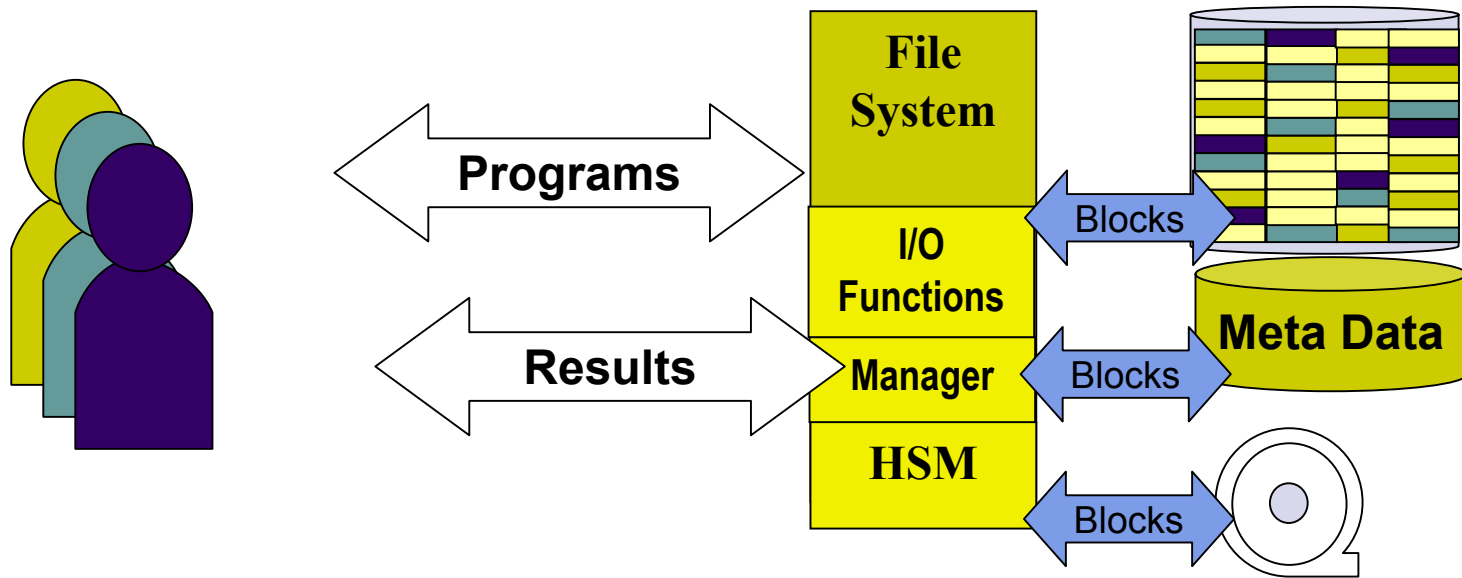
File System Basics

Data Protection



File System Basics

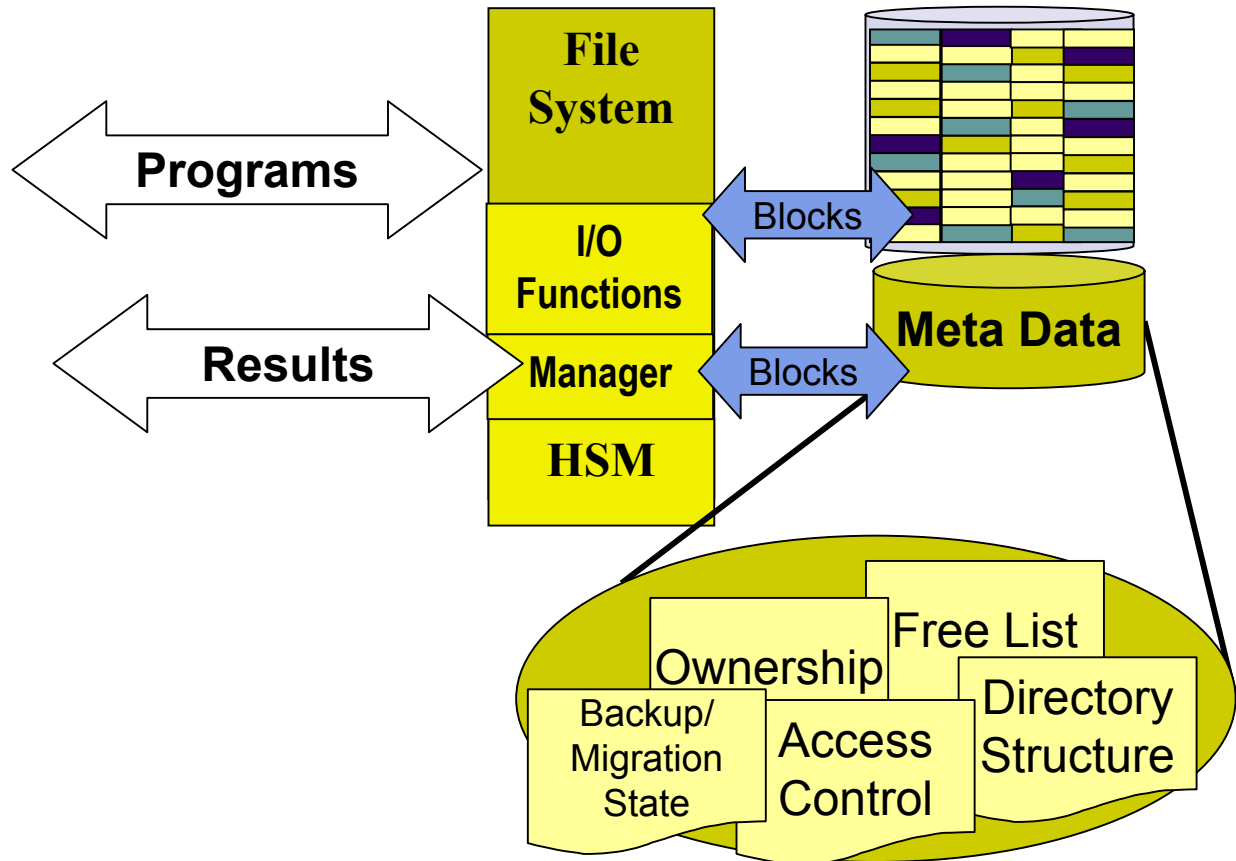
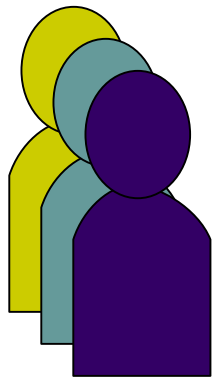
Storage Management

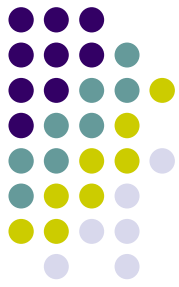




File System Basics

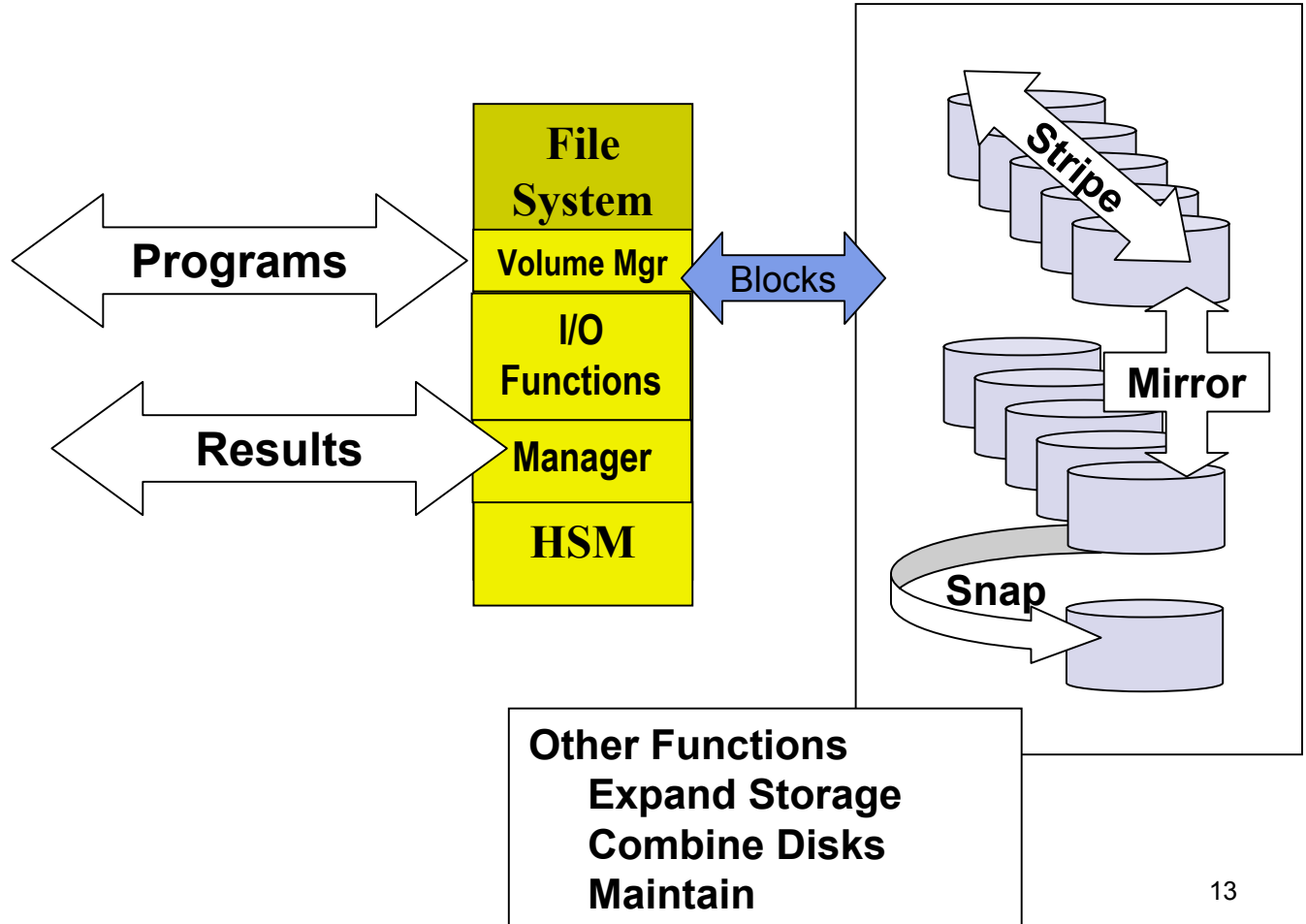
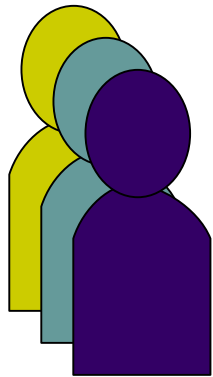
Space Management

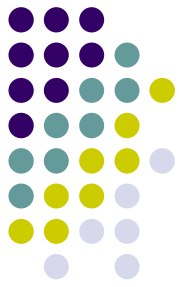




File System Basics

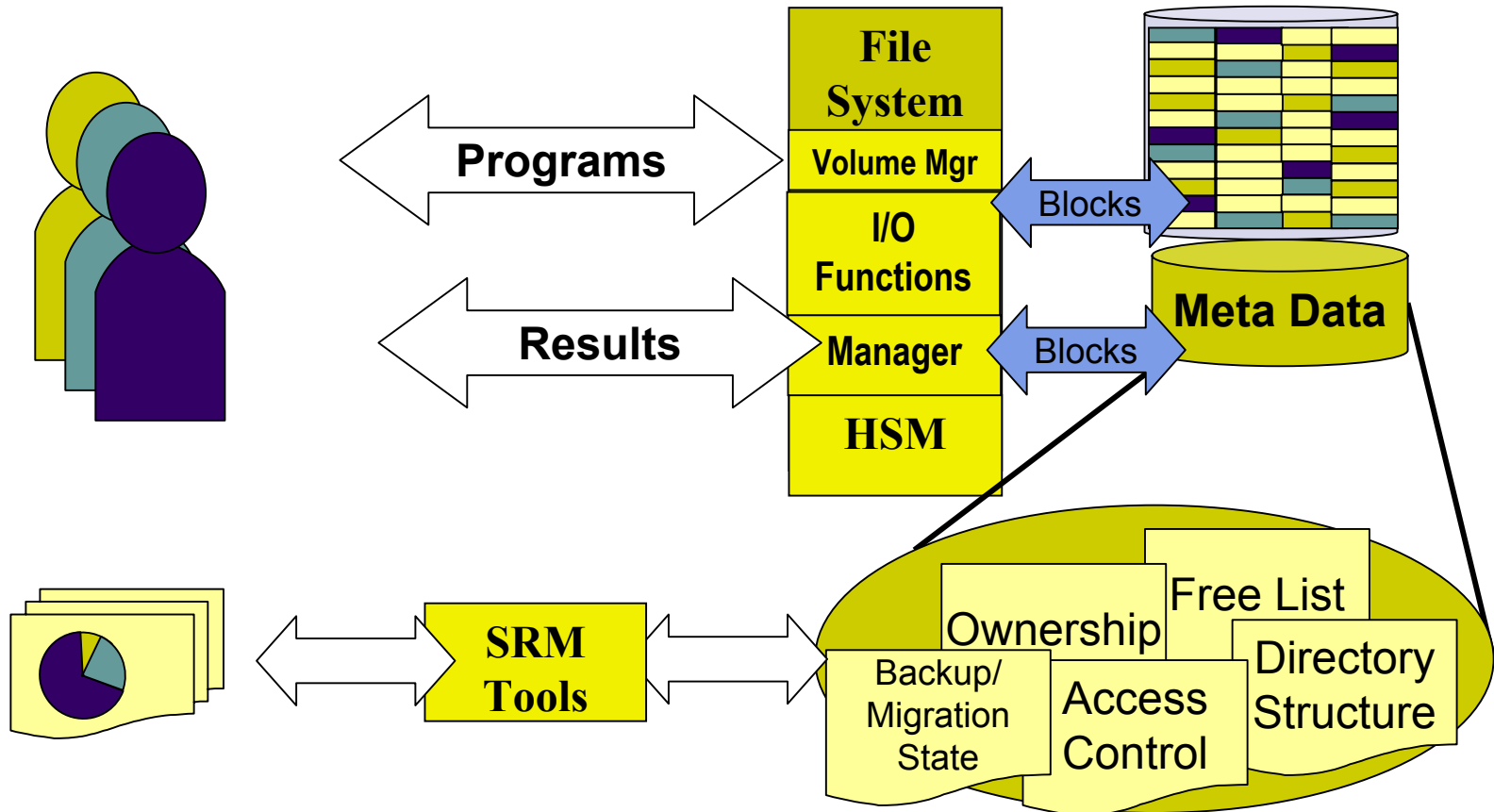
Disk Volume Management





File System Basics

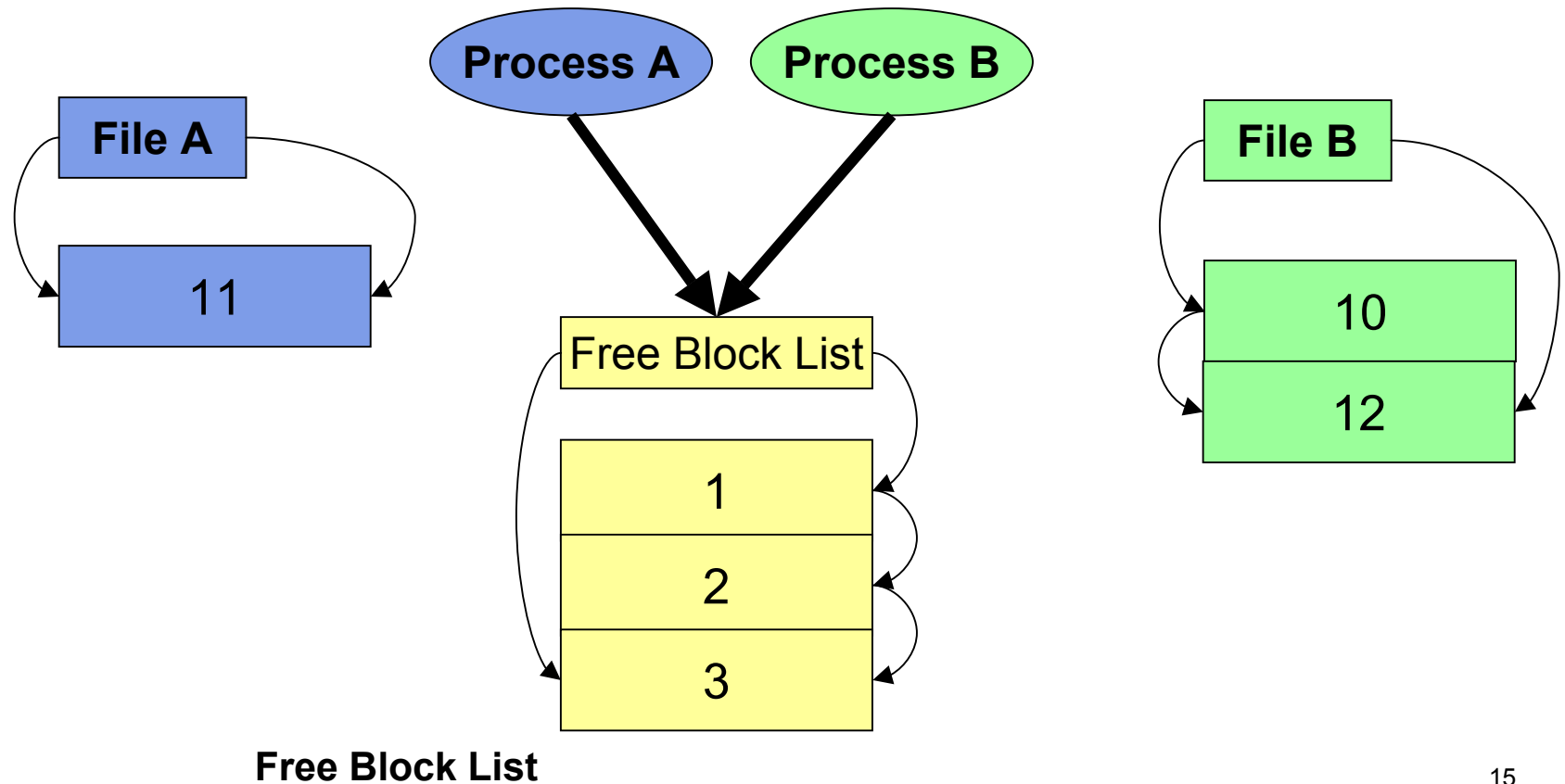
Storage Resource Management





Inside a SAN File System

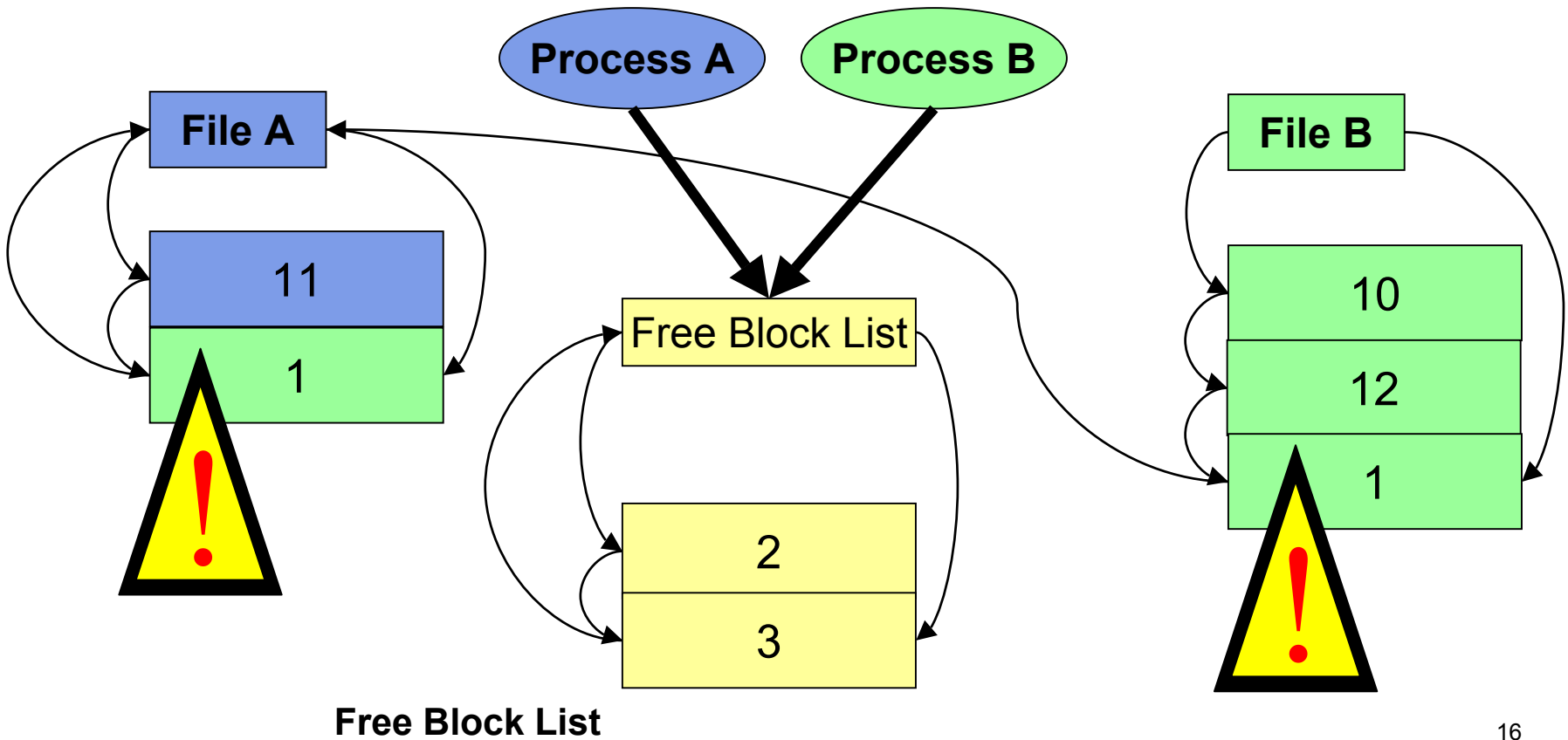
Meta Data Manager Locking

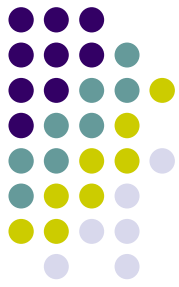




Inside a SAN File System

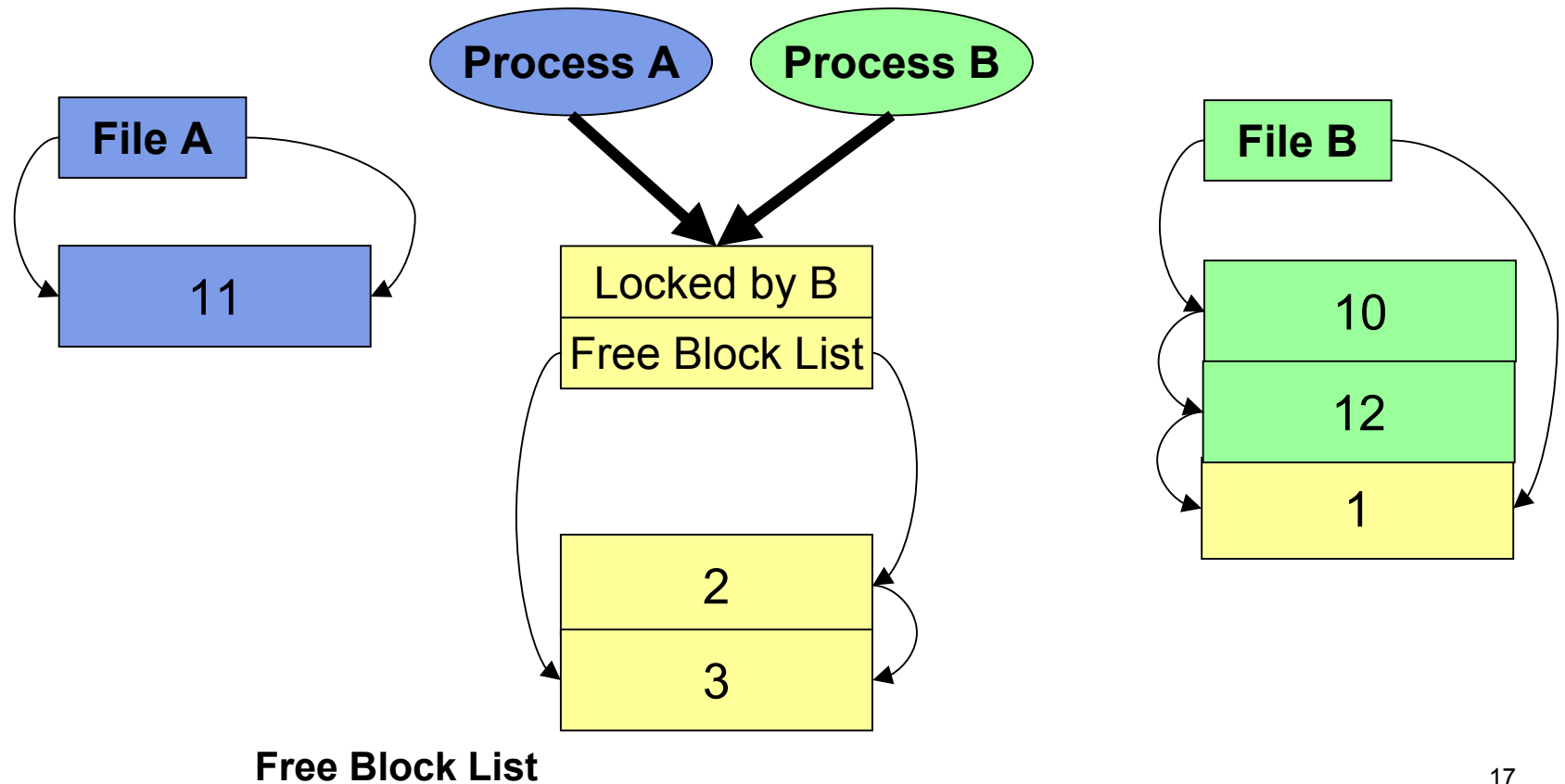
Meta Data Manager Locking





Inside a SAN File System

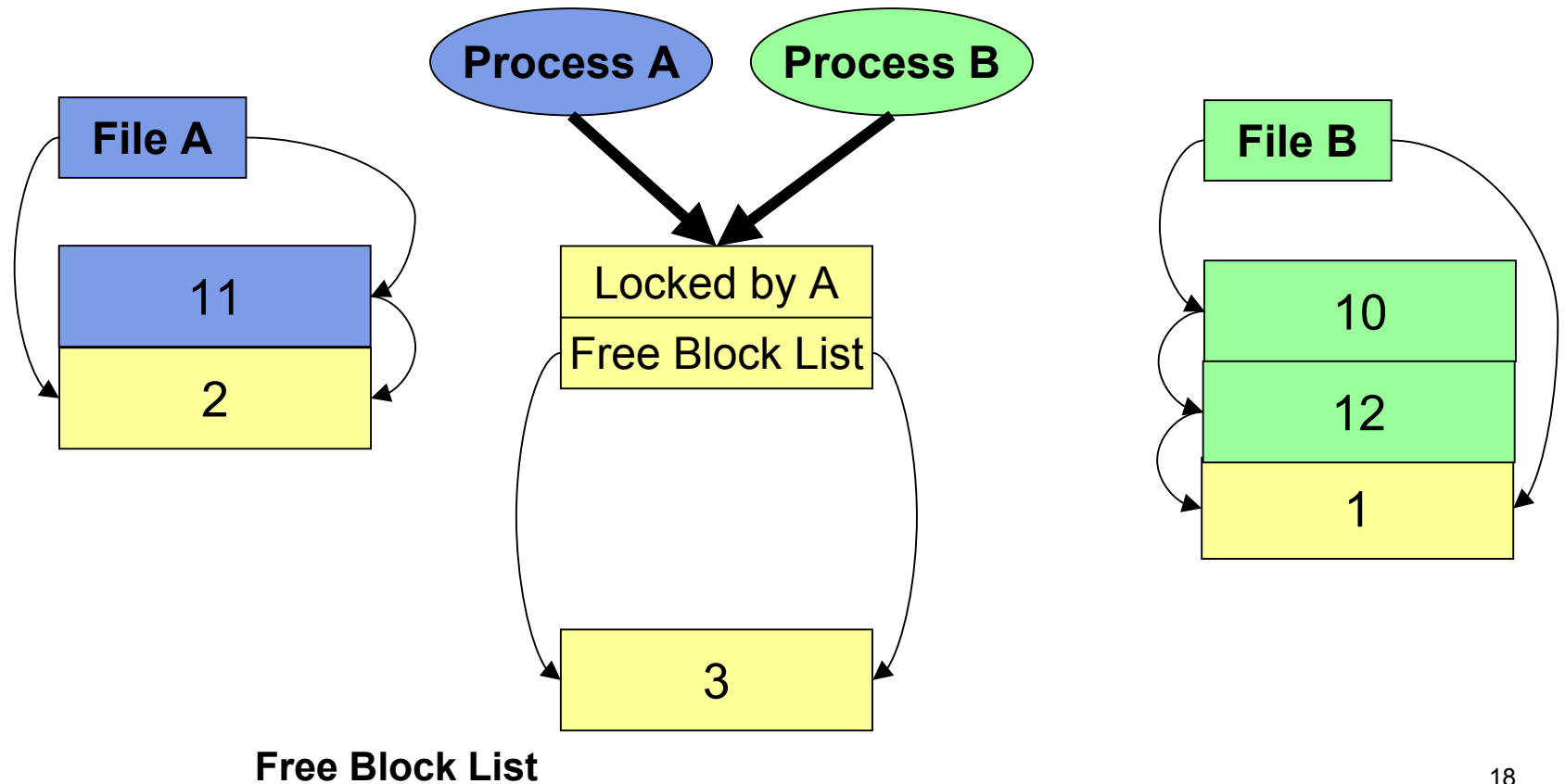
Meta Data Manager Locking



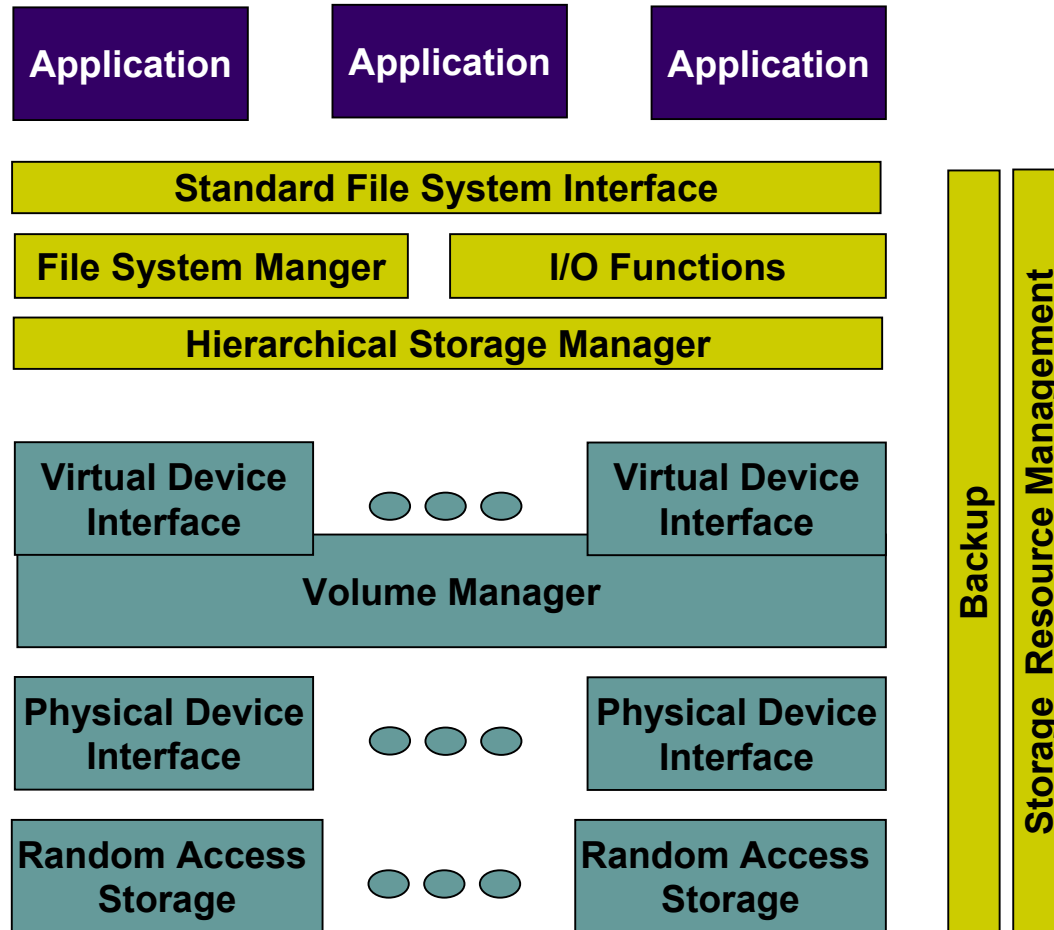


Inside a SAN File System

Meta Data Manager Locking

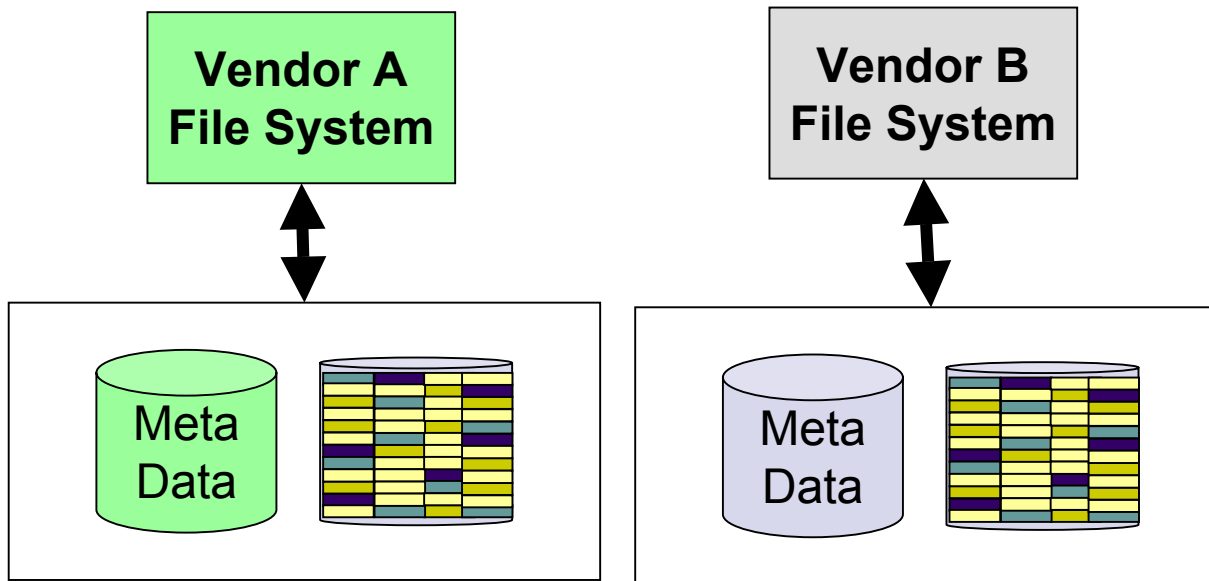


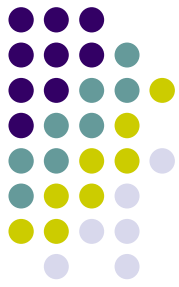
Inside a File System



File System Basics

No Standard On Disk Format

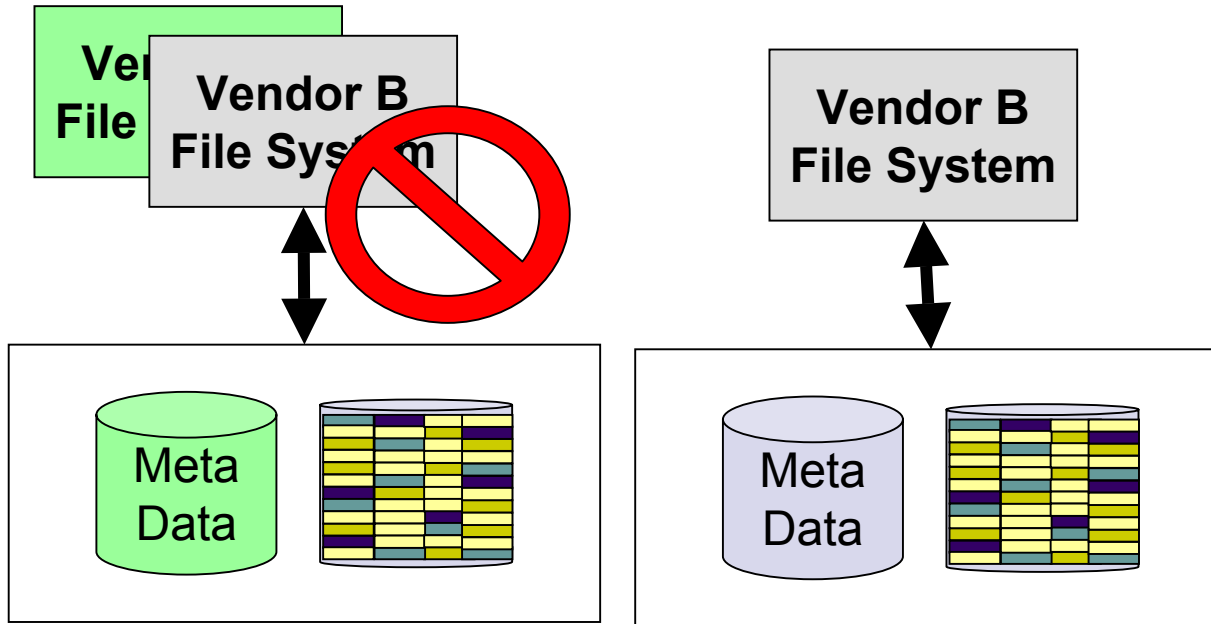


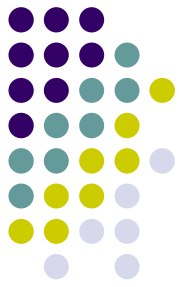


File System Basics

No Standard On Disk Format

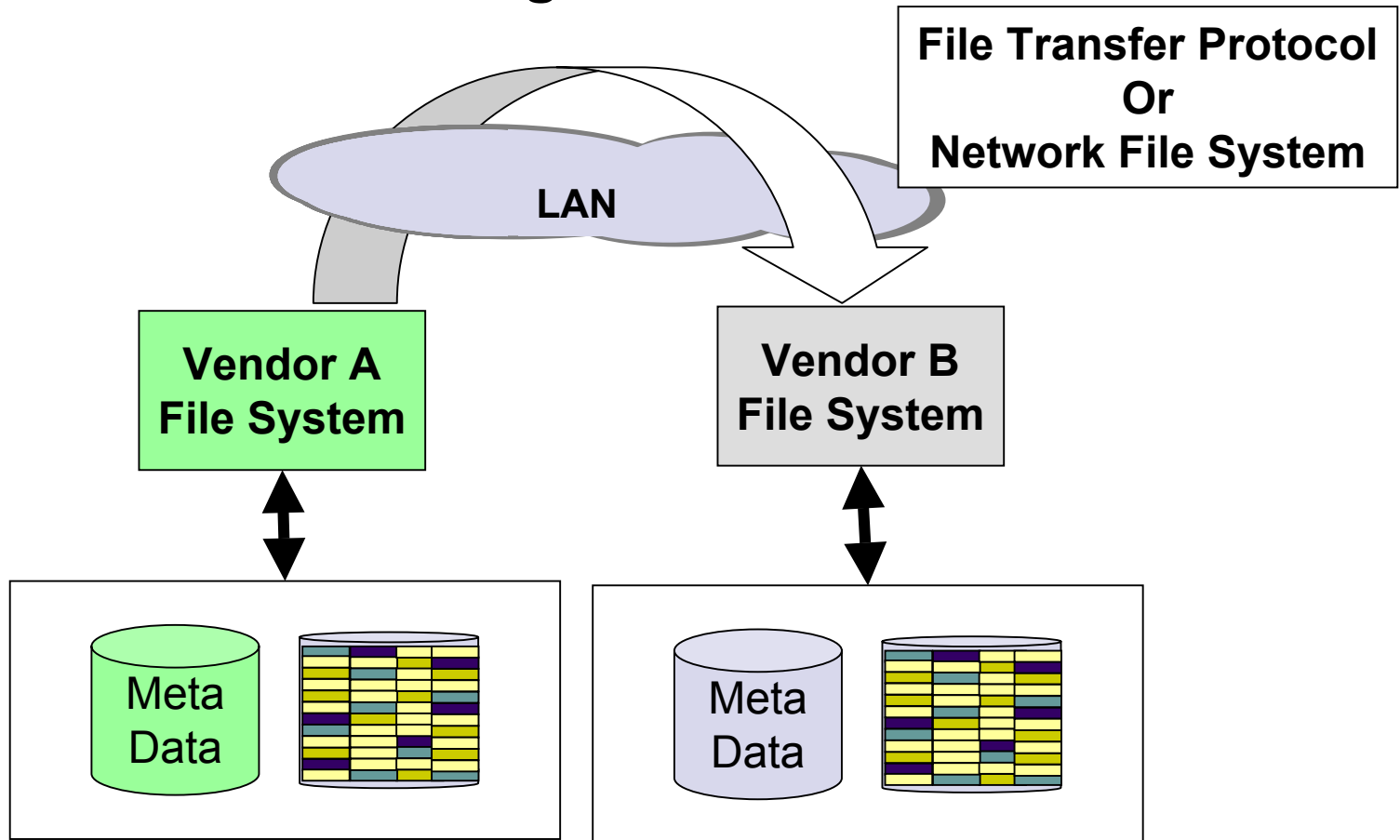
Cannot Simply Replace Vendor A With Vendor B



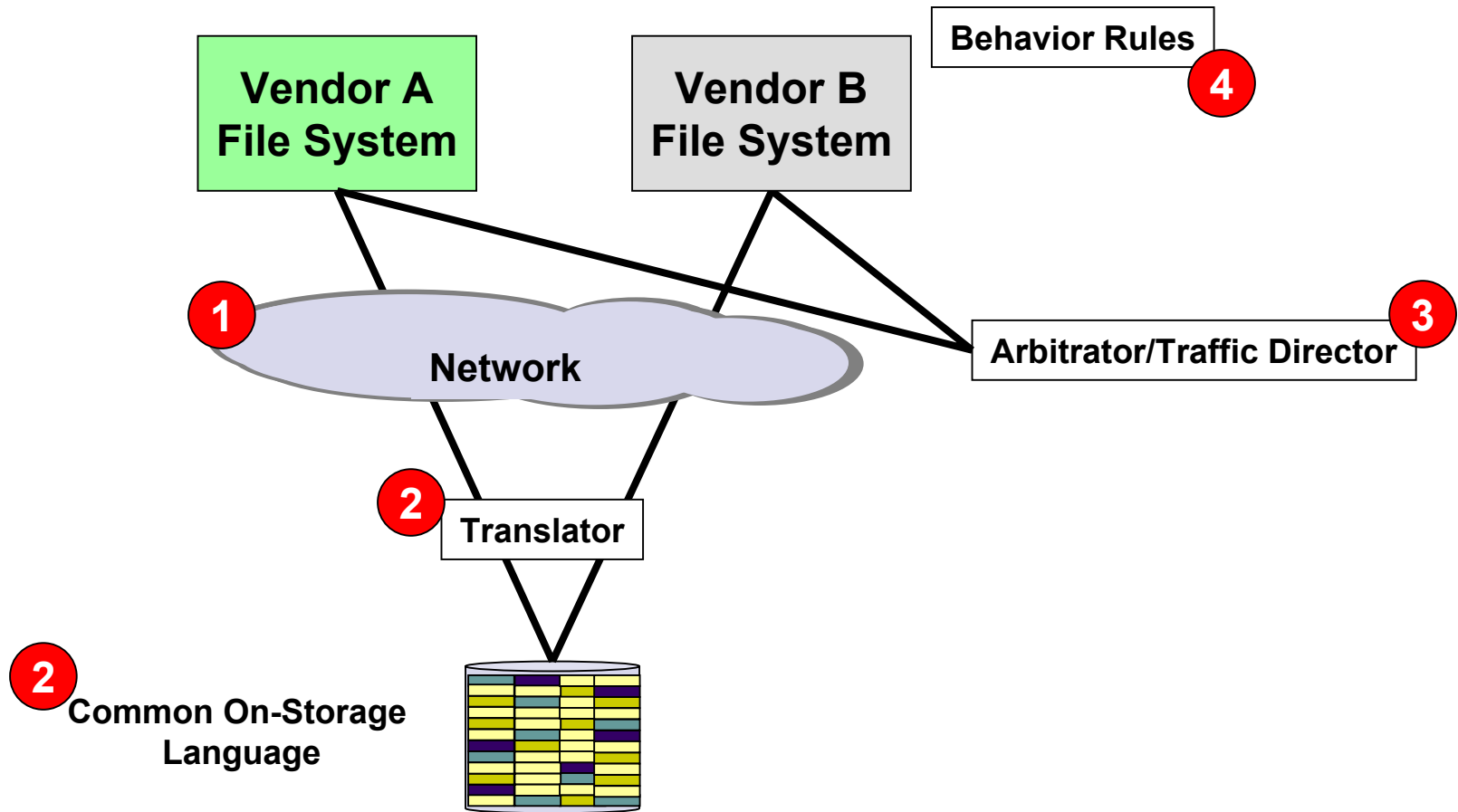


File System Basics

Network Based Sharing



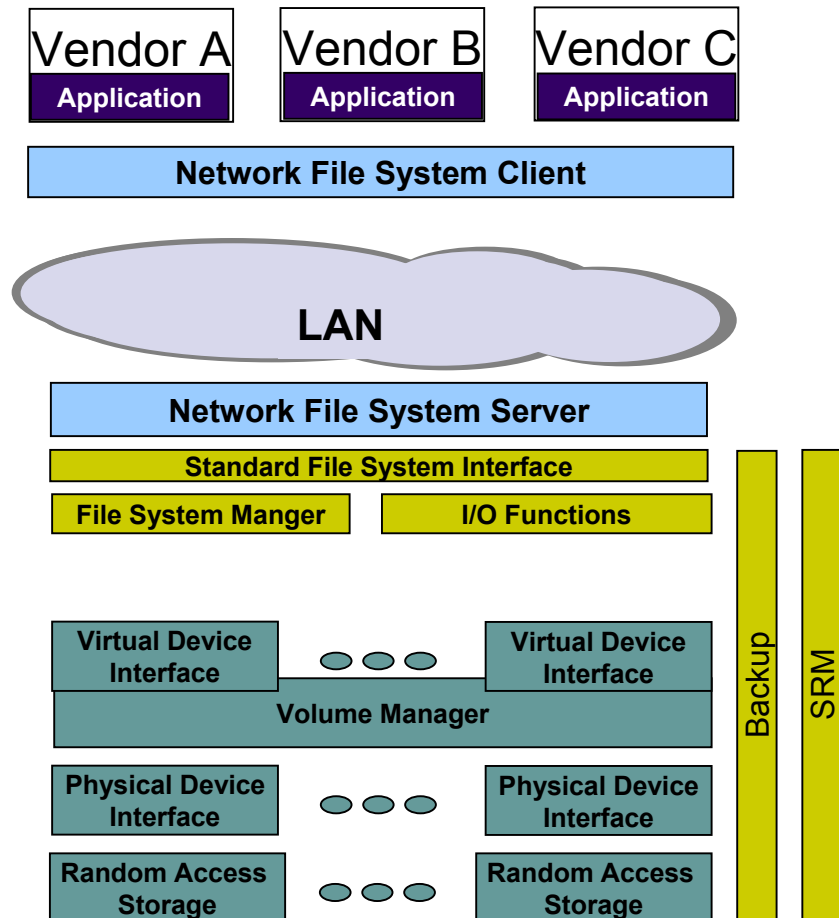
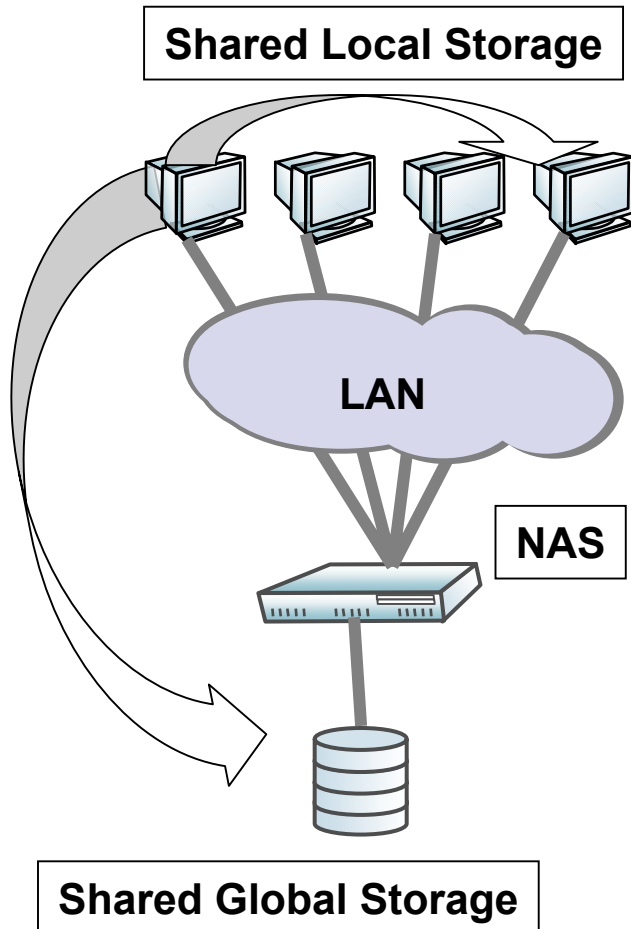
File Sharing

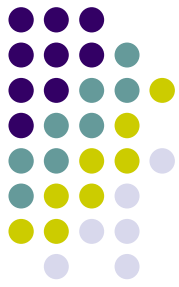




File System Basics

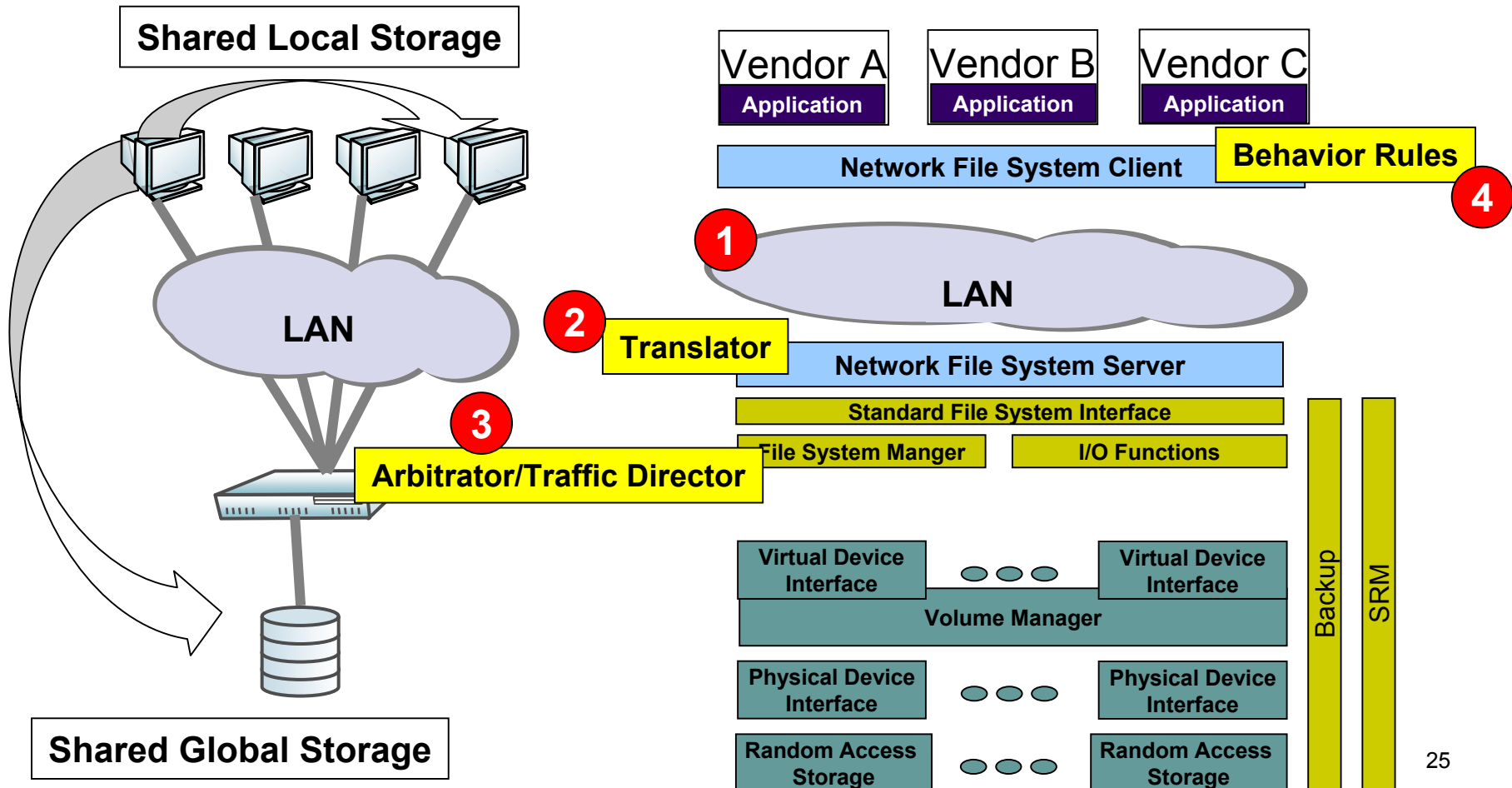
Network File Systems





File System Basics

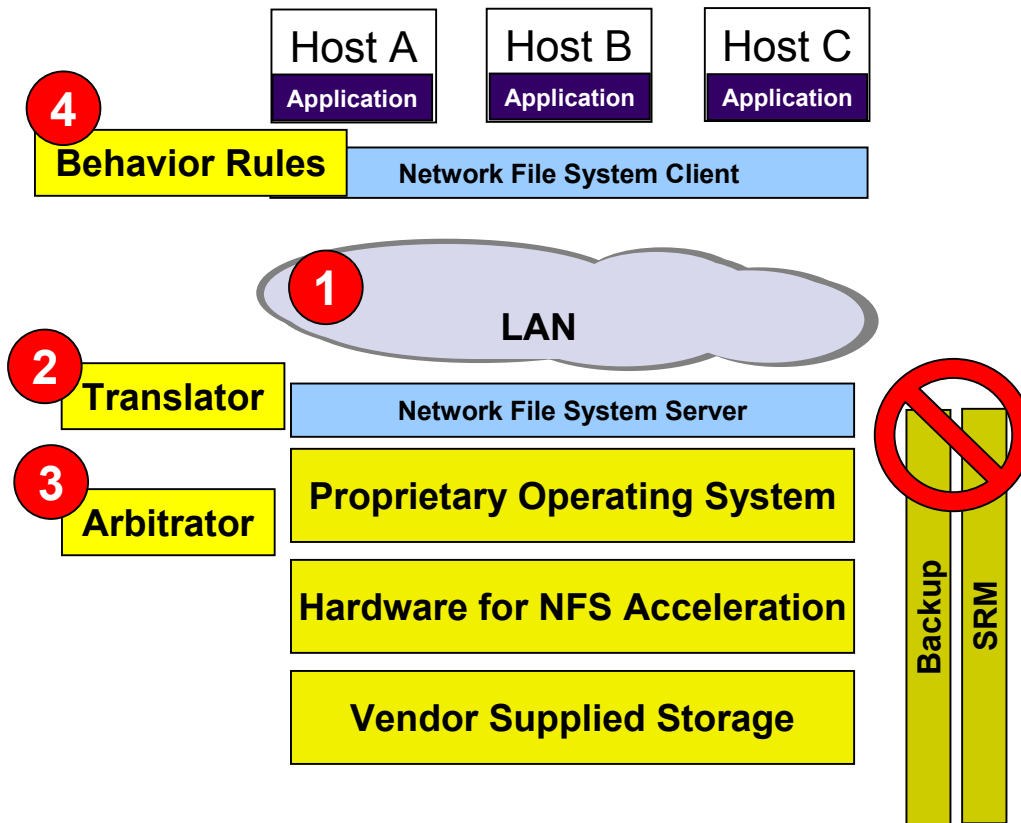
Network File Systems





File System Basics

Network Attached Storage (NAS)



Proprietary System

- Shared Data
- Consolidated Storage
- Great NFS Performance
- Moderate Bandwidth
- No Support for Standard Tools
- Vendor Provided Storage



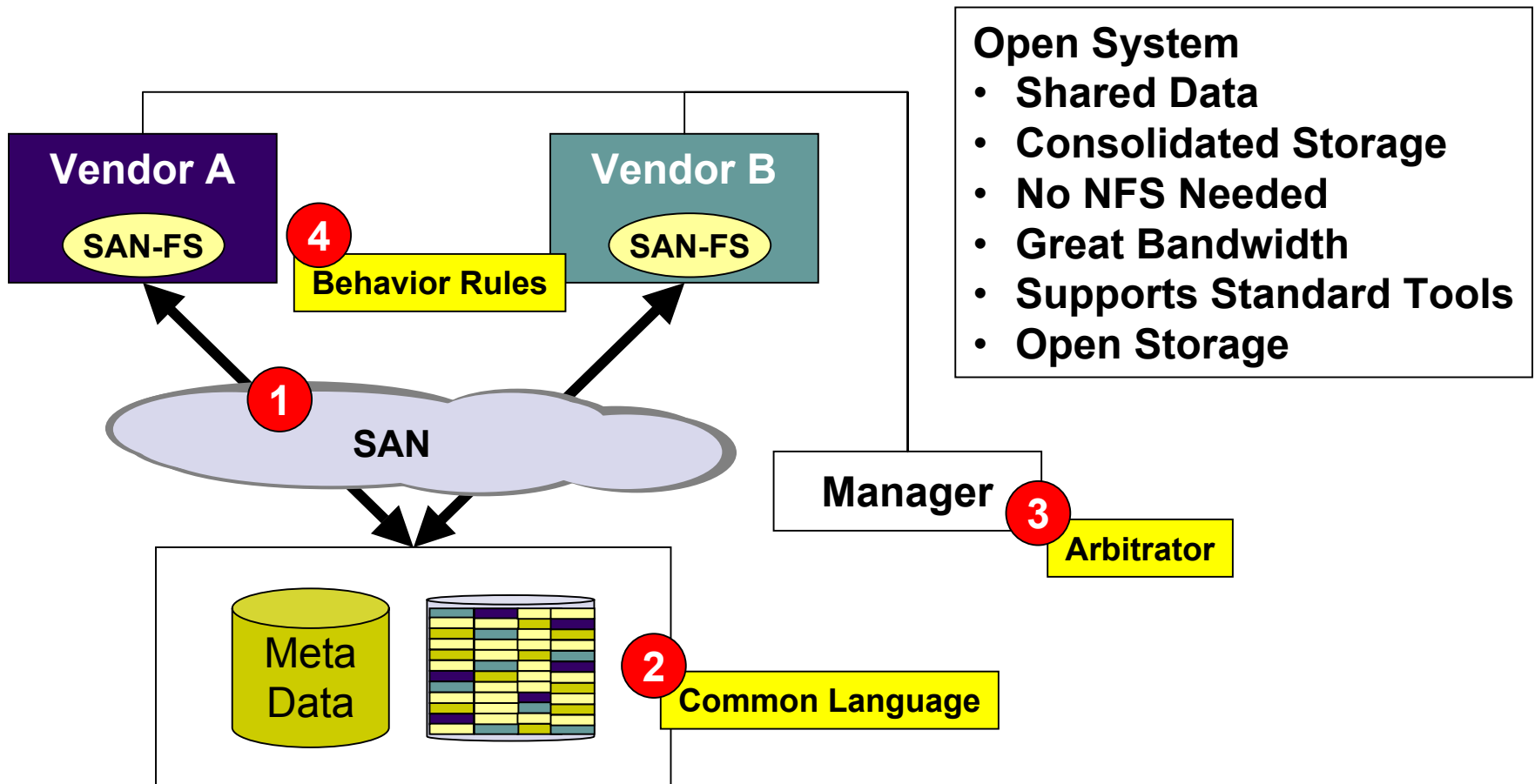
SAN Review

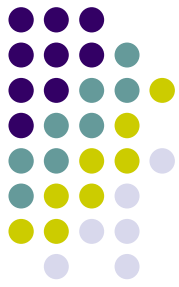
- Driving Need
 - Improved Availability
 - Continuous operation drives need for redundancy
 - Higher Performance
 - Faster Processors, Larger Data
 - Improved Scalability
 - As business grows so must the system
 - Easier Management
 - Manage more with less



SAN File Systems

A New Architecture

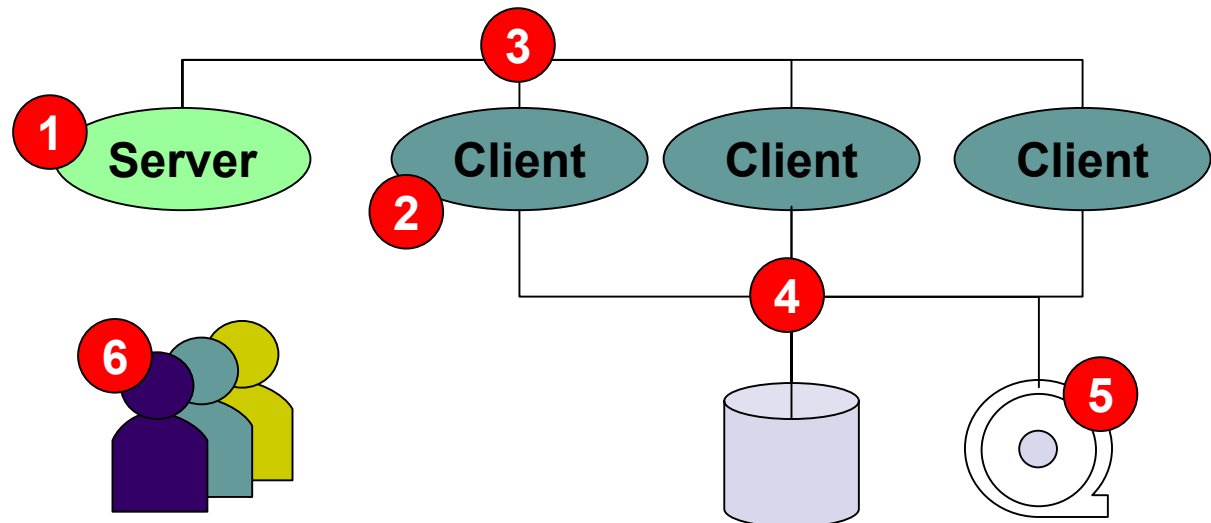




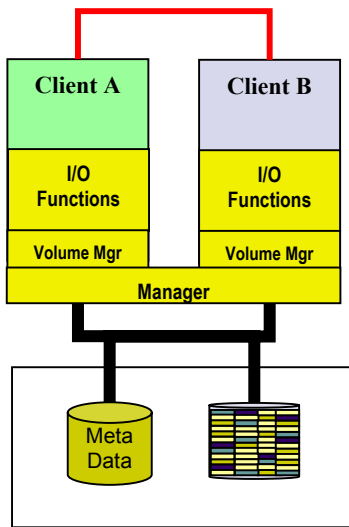
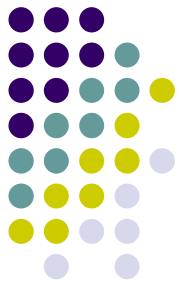
Inside a SAN File System

Key Technologies

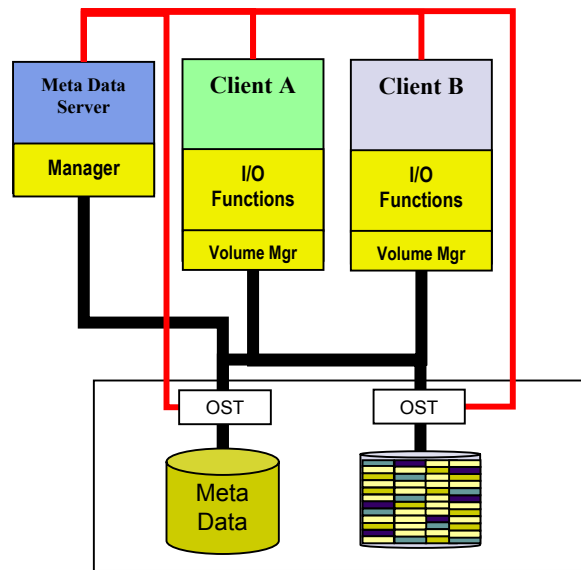
1. Server Architecture
2. Client Architecture
3. Control Path
4. Data Path & Storage
5. Data Management/Protection
6. User/Security Management



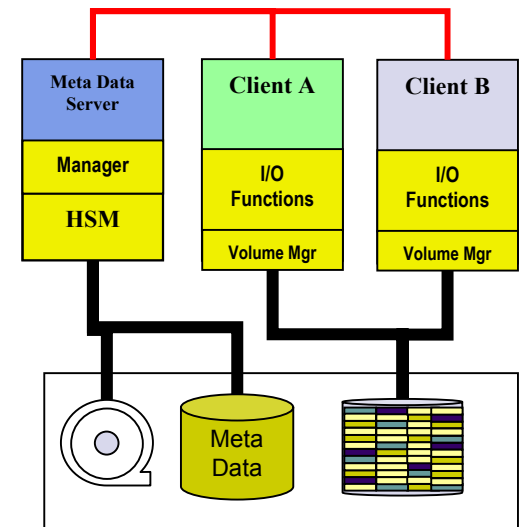
SANFS Software Architectures



Distributed Manager



Object Storage

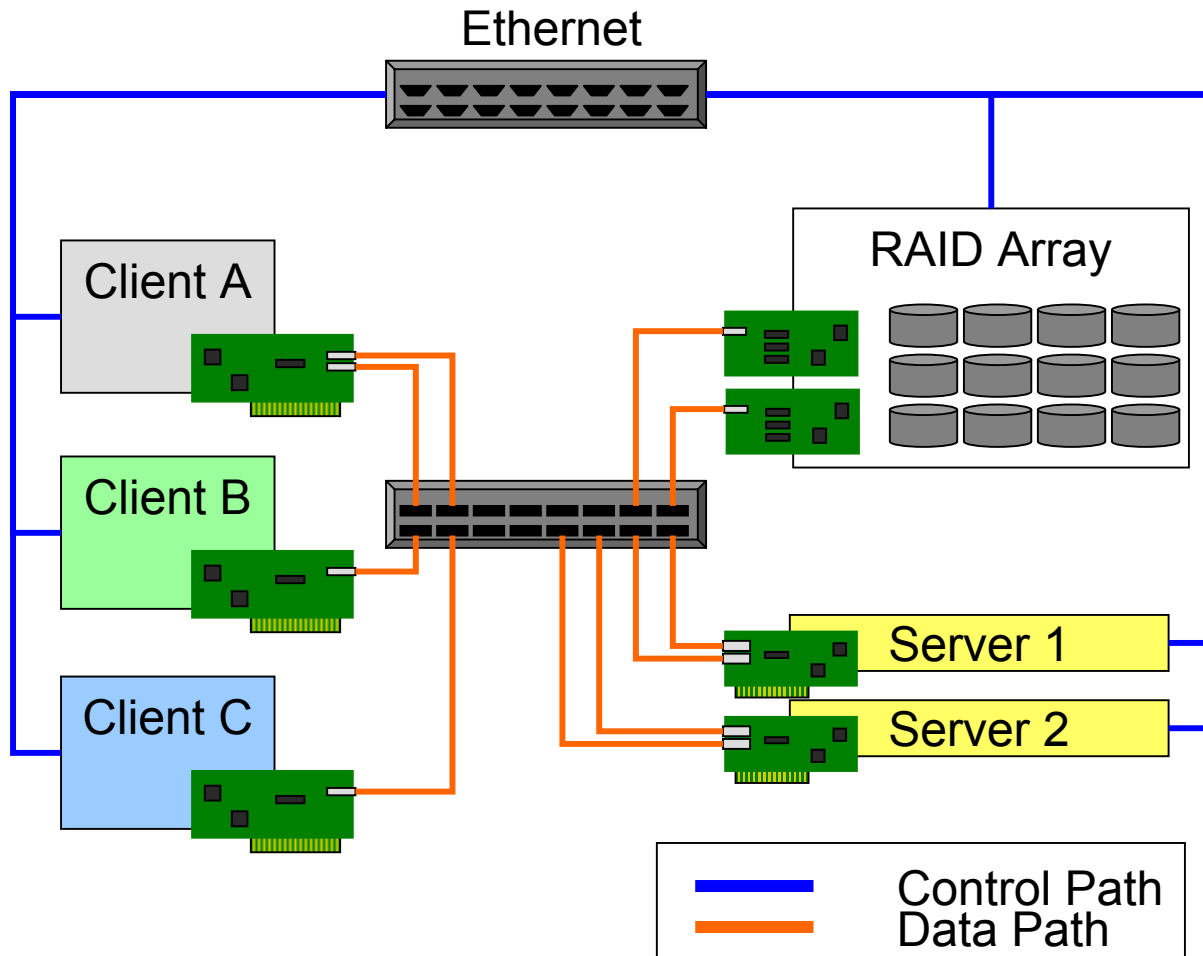


Central Manager



SAN File Systems

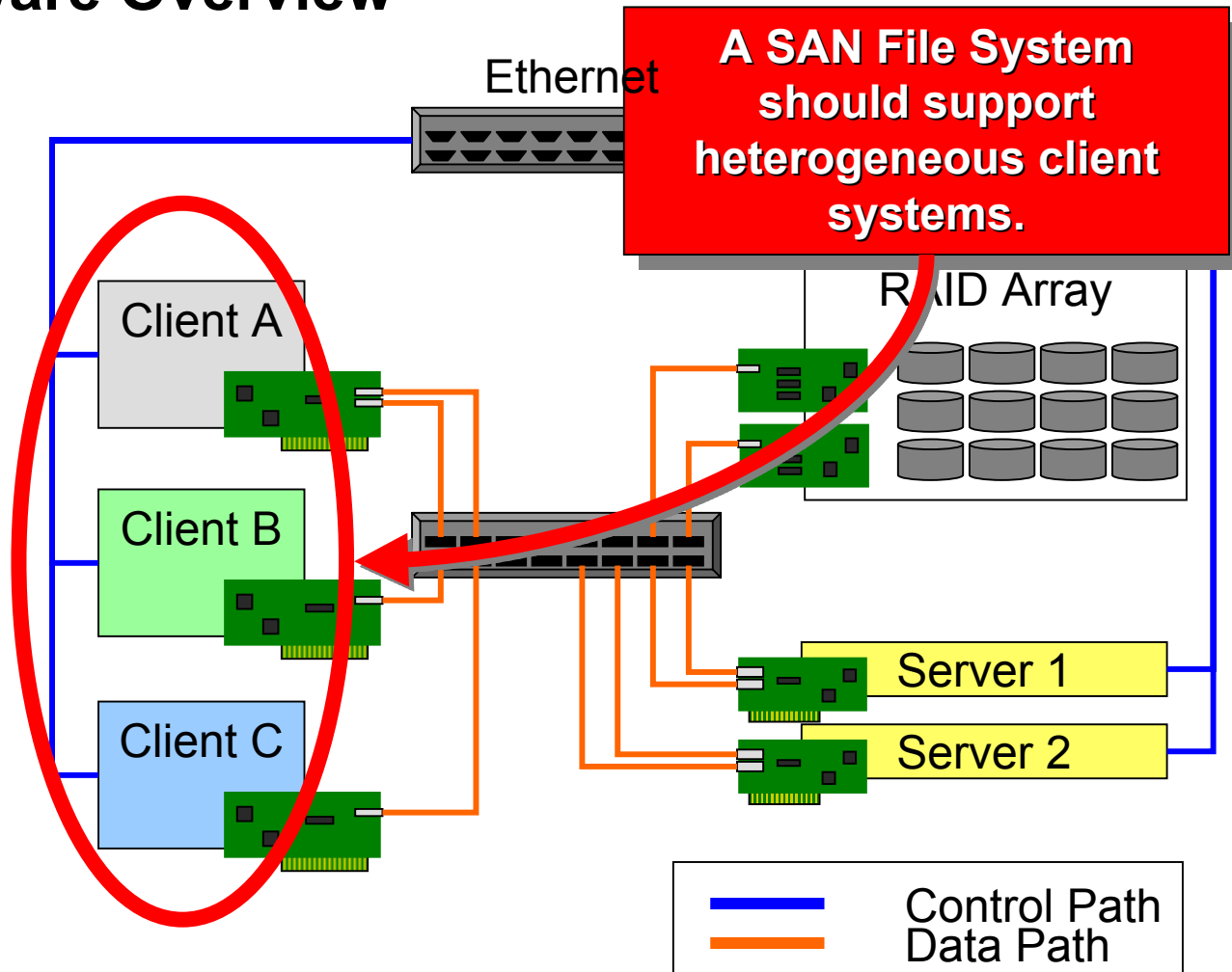
Hardware Overview

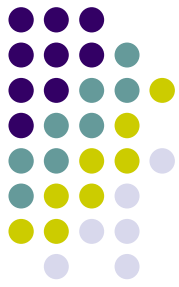




SAN File Systems

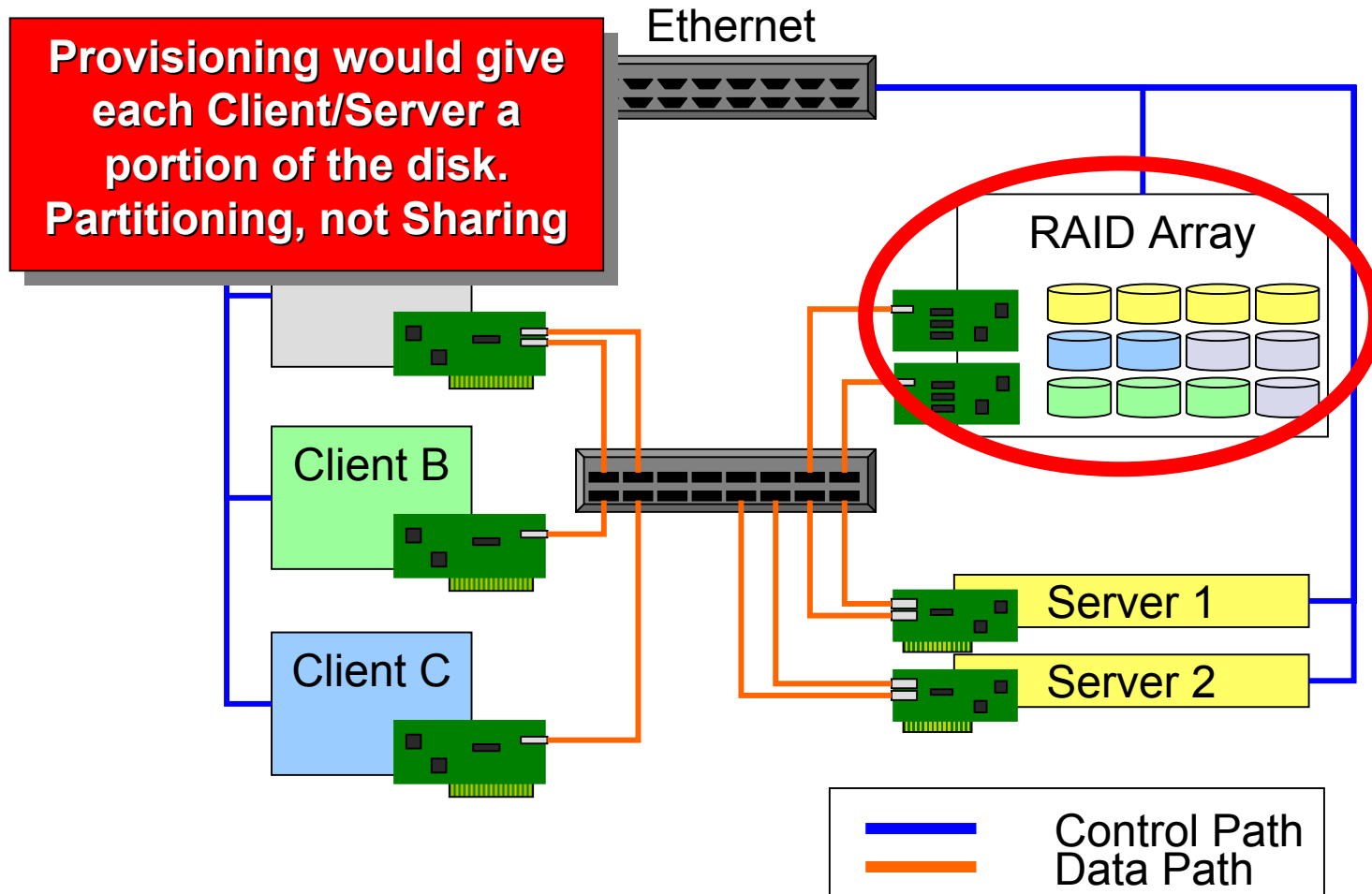
Hardware Overview





SAN File Systems

Provisioning

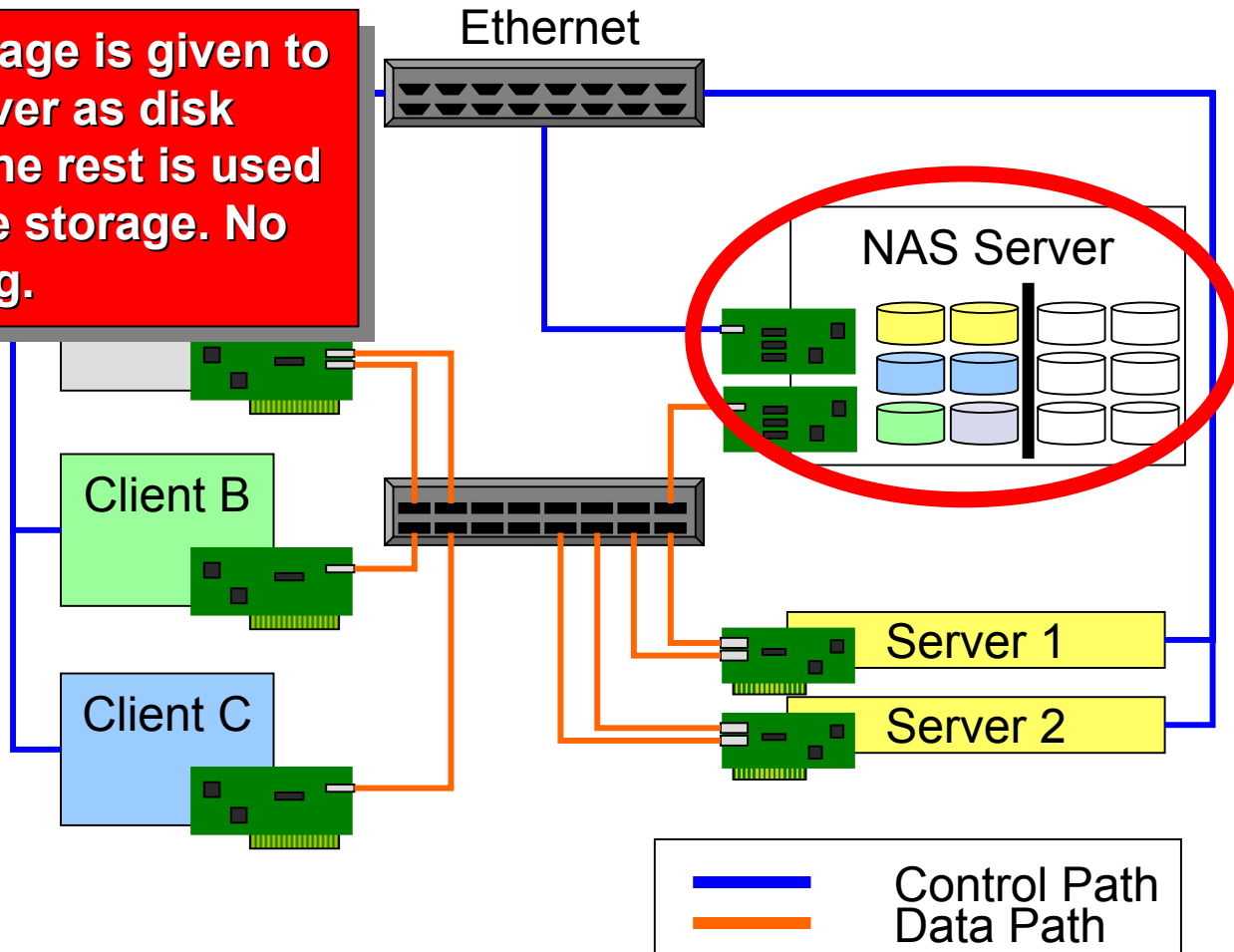




SAN File Systems

Combined NAS and Provisioning

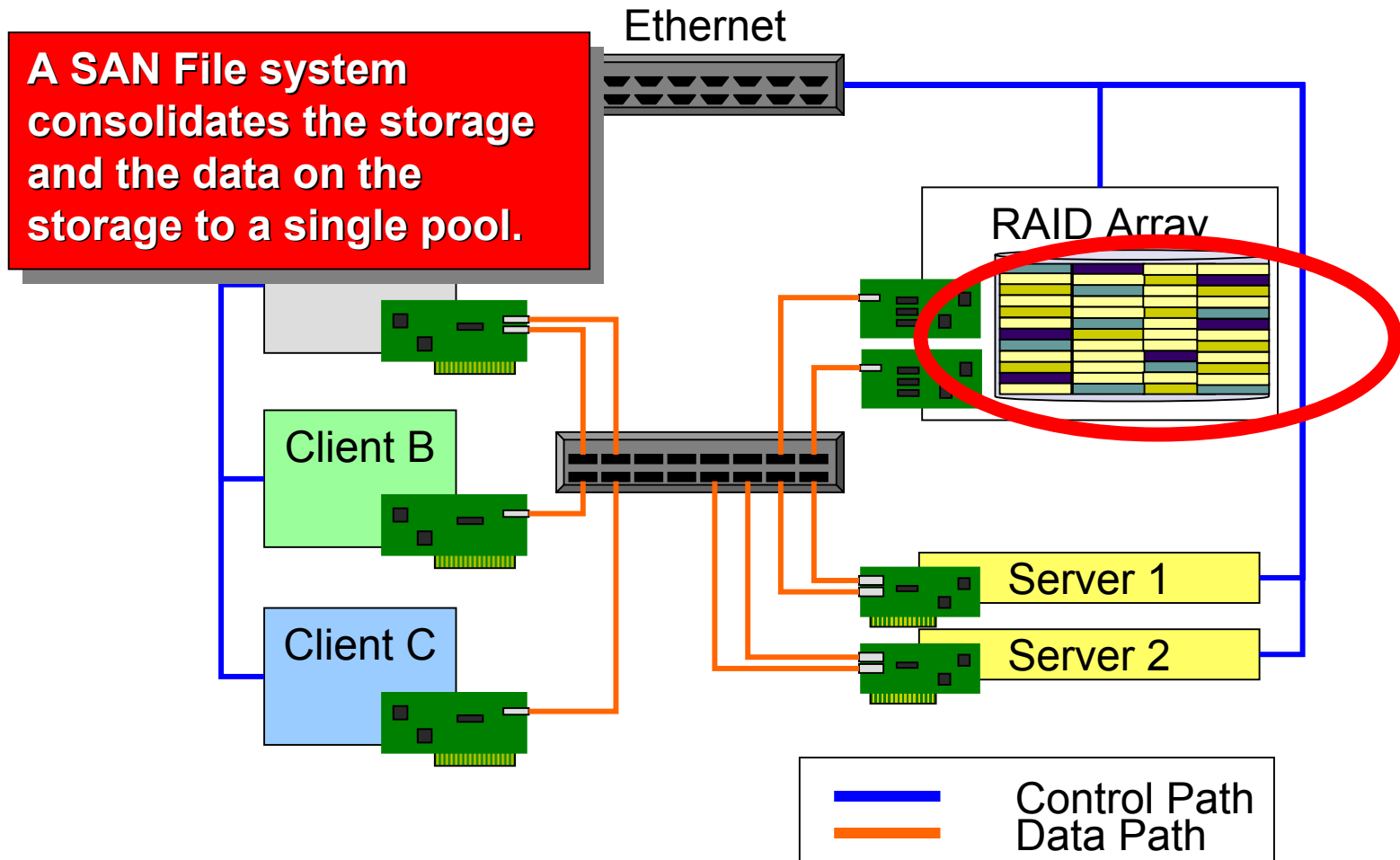
Part of Storage is given to Clients/Server as disk volumes. The rest is used for NAS File storage. No data sharing.





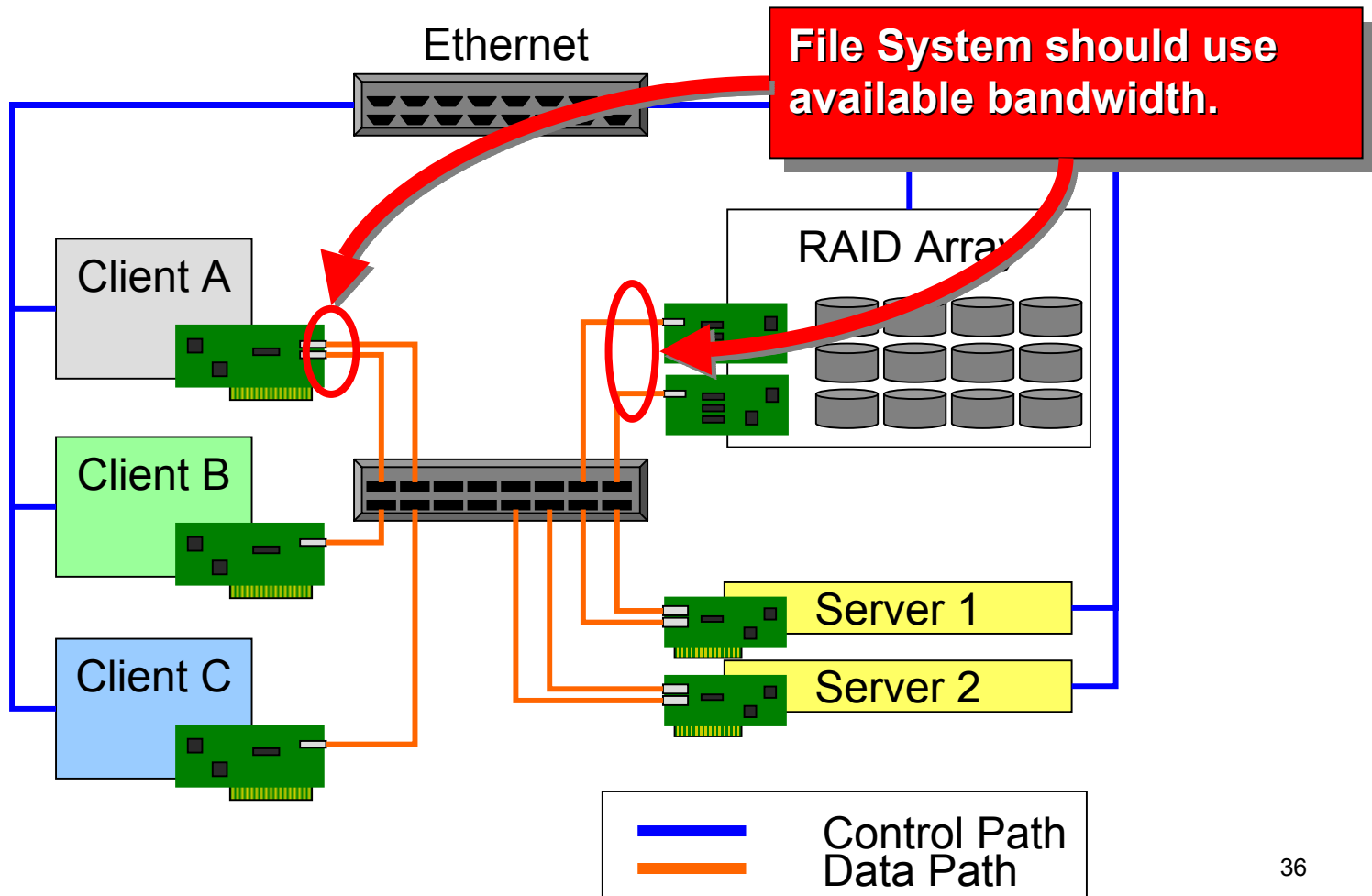
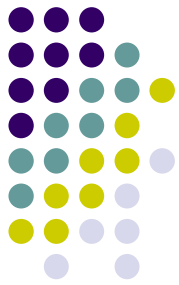
SAN File Systems

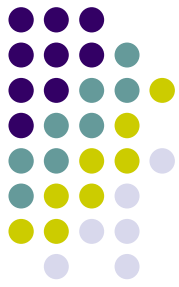
Hardware Overview



SAN File Systems

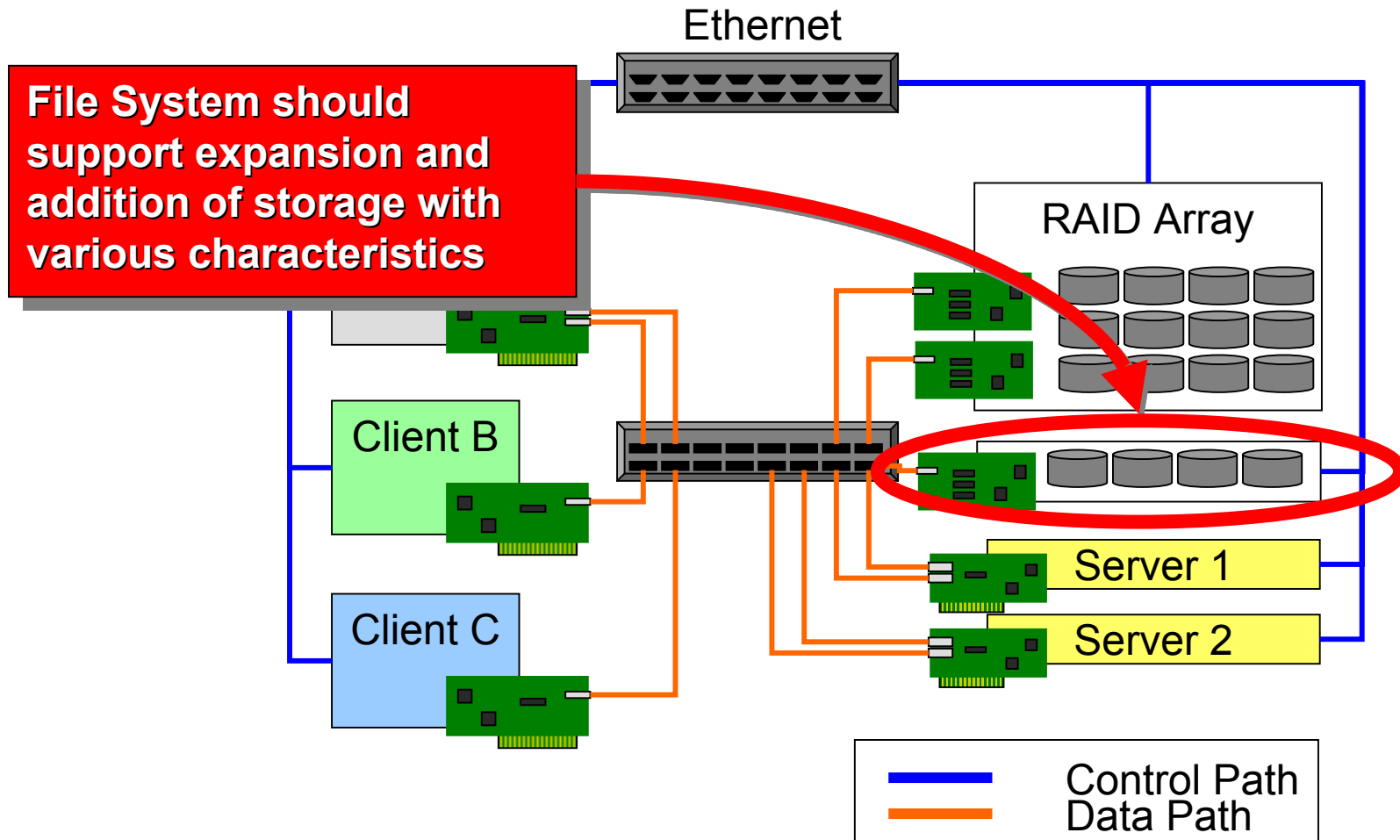
Bandwidth Aggregation / Port Fail over

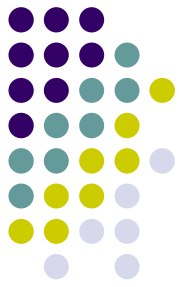




SAN File Systems

Expansion and Heterogeneous Storage





Inside a SAN File System

- Meta Data Server Locking
 - Allocating Disk Blocks
 - Updating Meta Data
 - Last Access, Size, Owner, Access Control
 - Token Management
 - Creating/Deleting Files/Directories

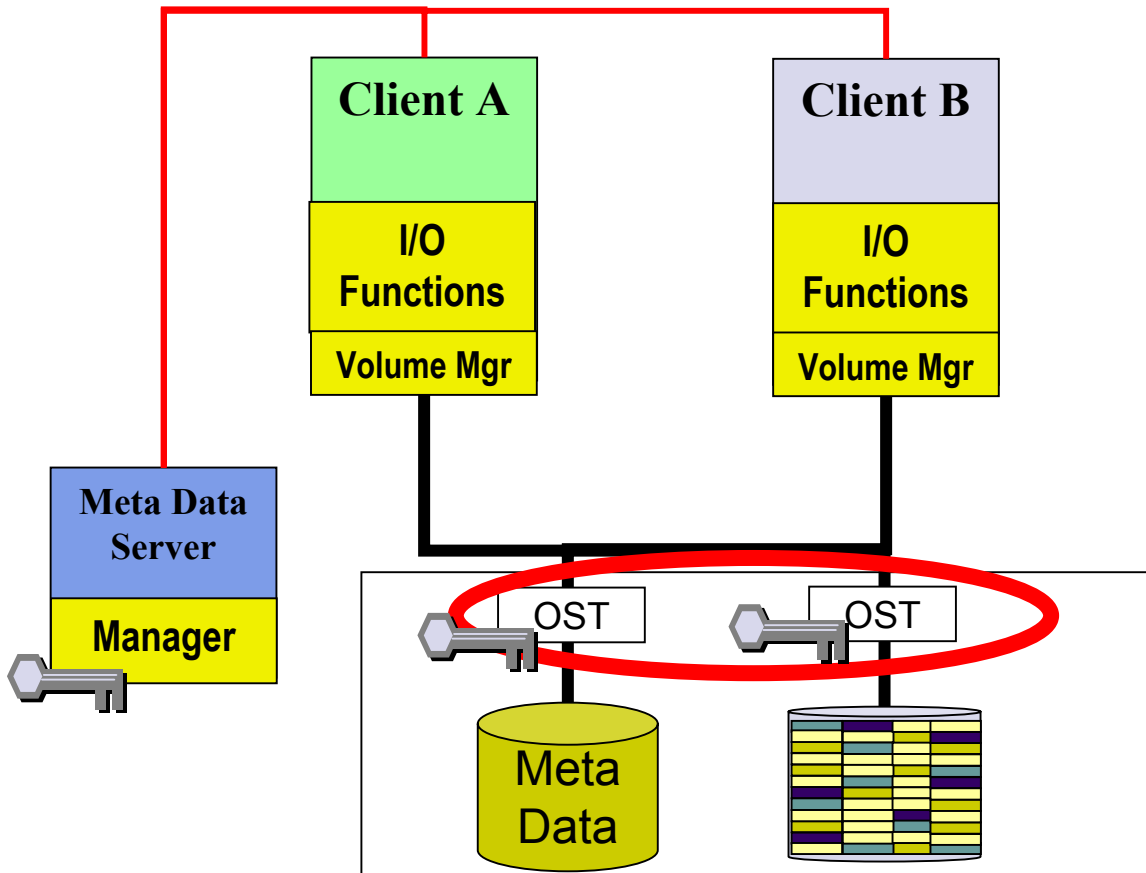


Inside a SAN File System

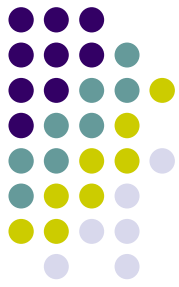
- Locking Techniques
 - In a Single Server
 - Hardware Test/Set Semaphores
 - In a Cluster
 - Central Lock Manager
 - Multiple Lock Managers
 - Distributed Lock Manager

SAN File System Architecture

Object Storage



- Distribute Meta Data Workload
- In-line Security
- New Abstraction for Disk



Inside a SAN File System

- Object Storage Devices
 - Like a file system in a disk
 - Simple Flat Name Space
 - Object Relative Addressing
 - Extended SCSI Command Set
 - Object_ID is key
 - Meta Data Manager Creates Object_ID
 - OSD allocates space on write

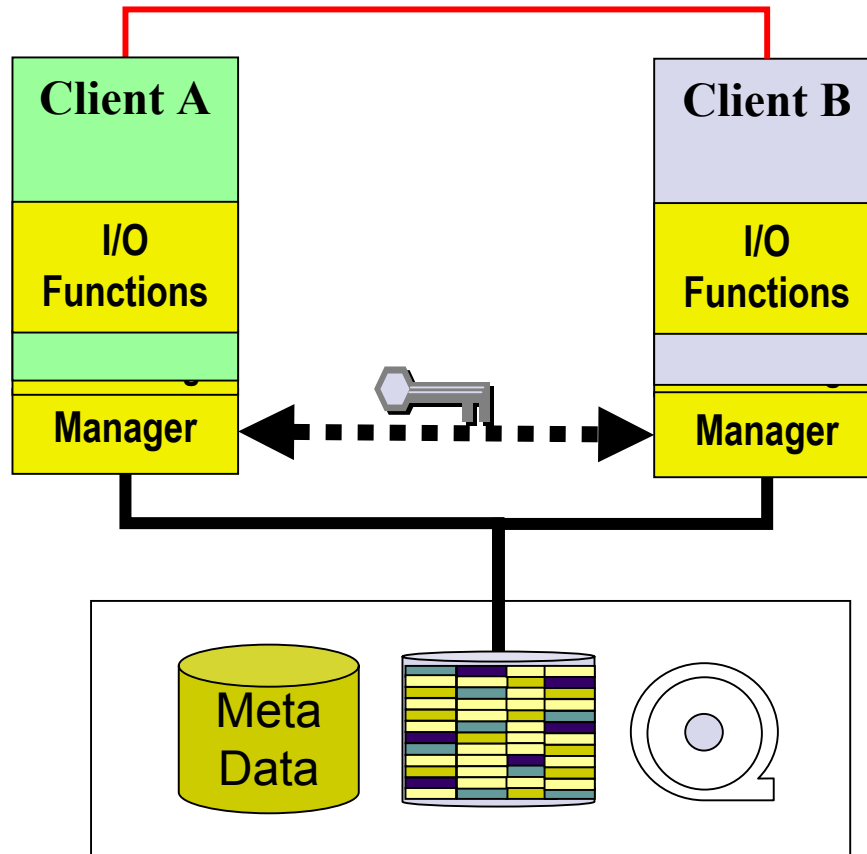


Inside a SAN File Server

- Object Servers versus Block Access
 - One Possible Advantage is Locking
 - Allocation of disk blocks is handled by OST
 - Reduces Communication to Manager
 - Multiple OSTs allows for Less Waiting for Locks
- Ultimately, this is a Technology Issue
 - Need a Low Latency “Lock” method
 - In the Disk or a special appliance?

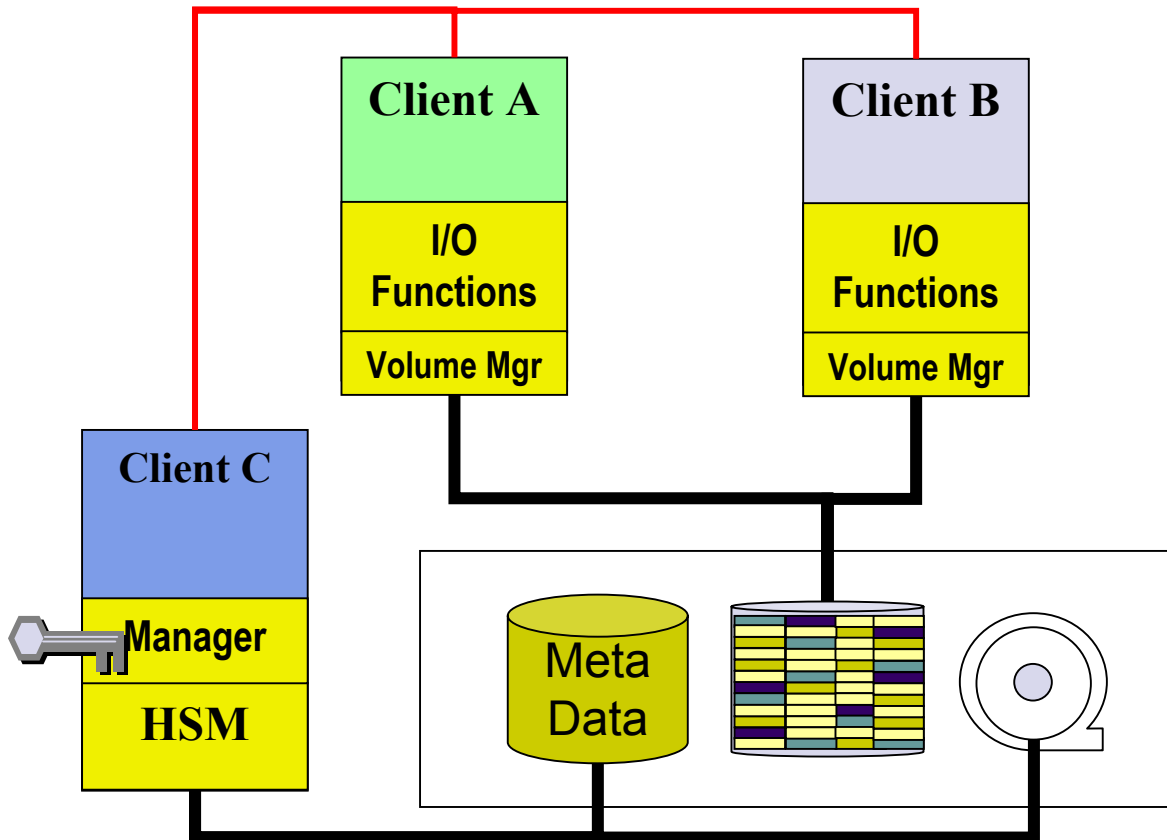
SAN File System Architecture

Distributed Meta Data Server



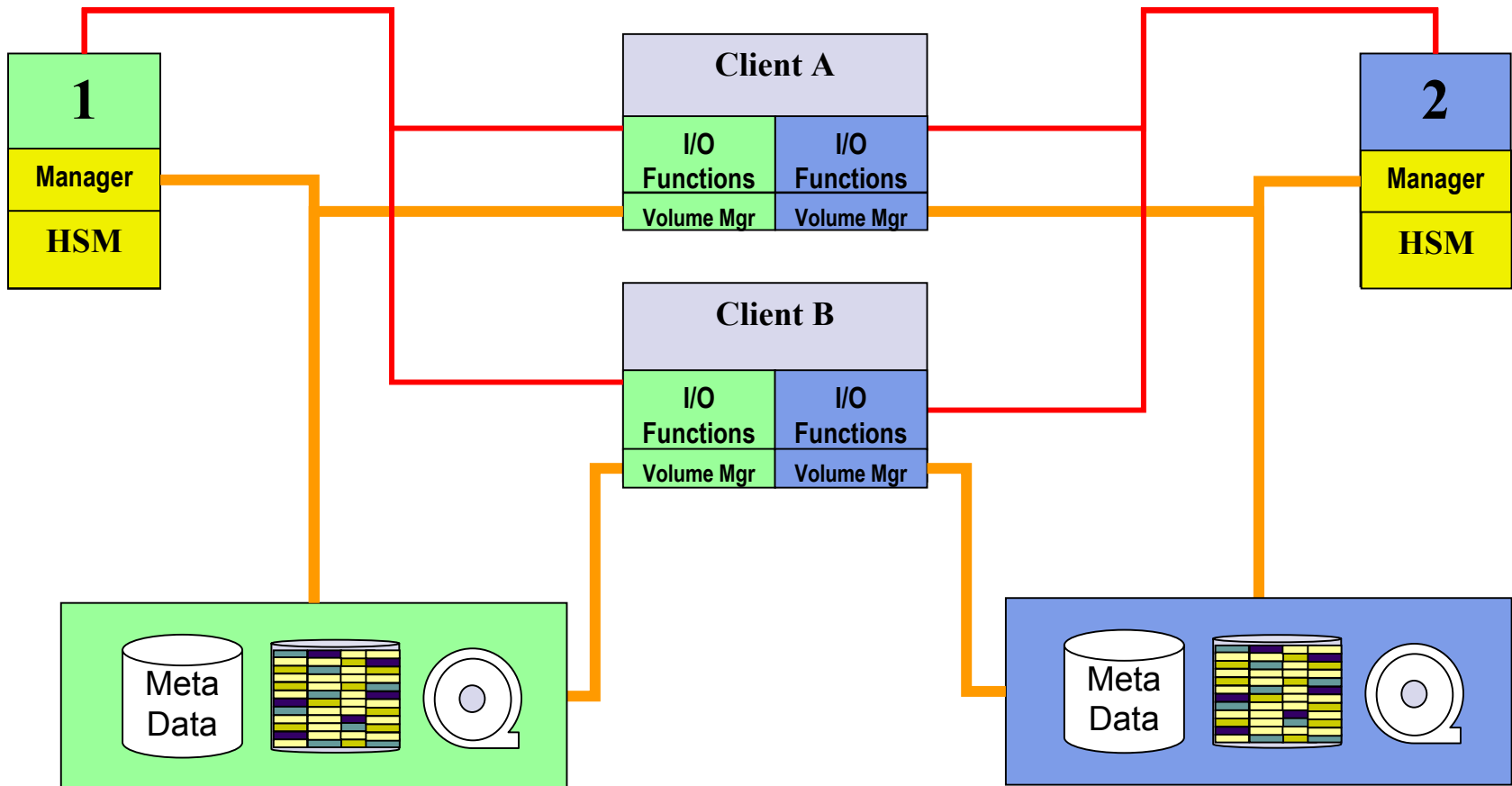
SAN File System Architecture

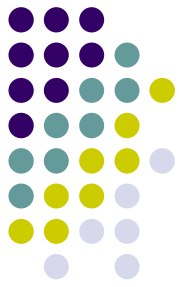
Central Meta Data Server



SAN File System Architecture

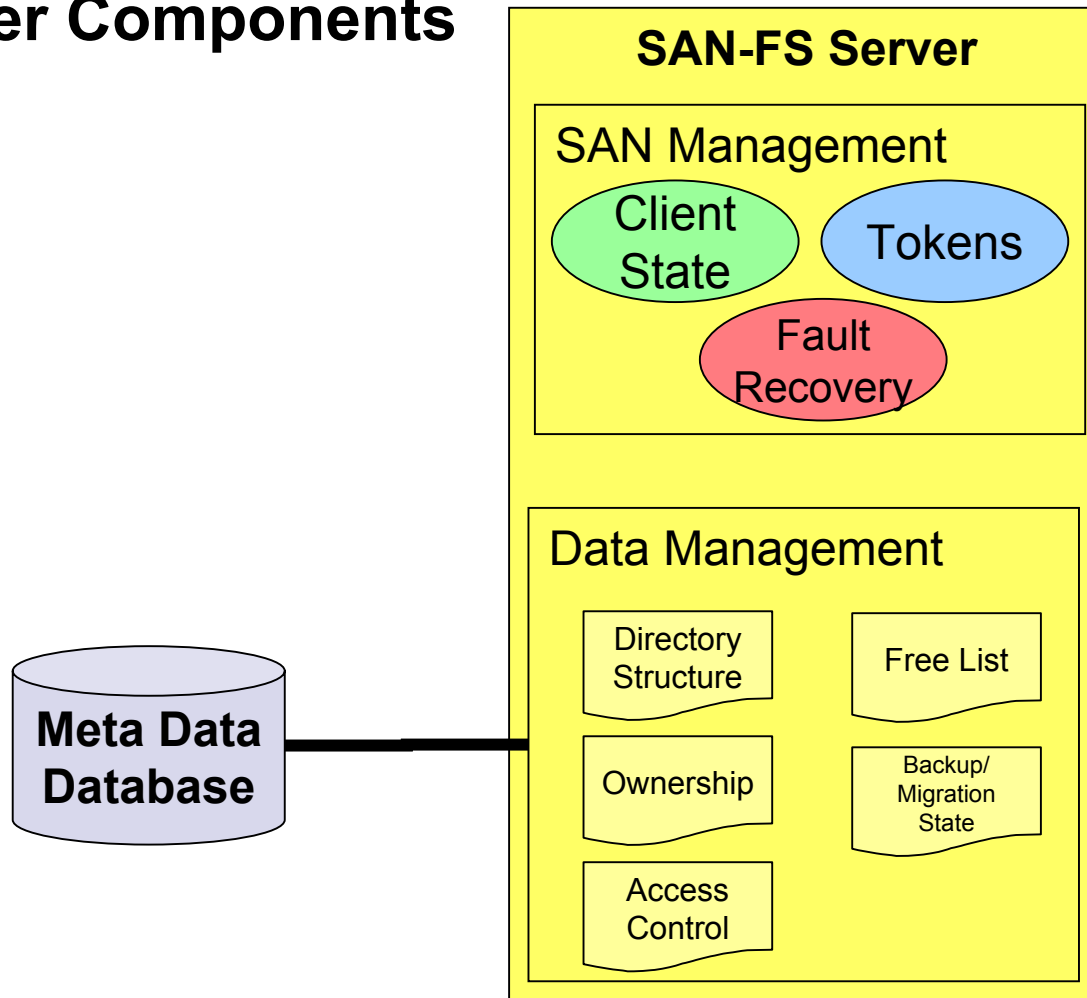
Multiple File Systems





Inside a SAN File System

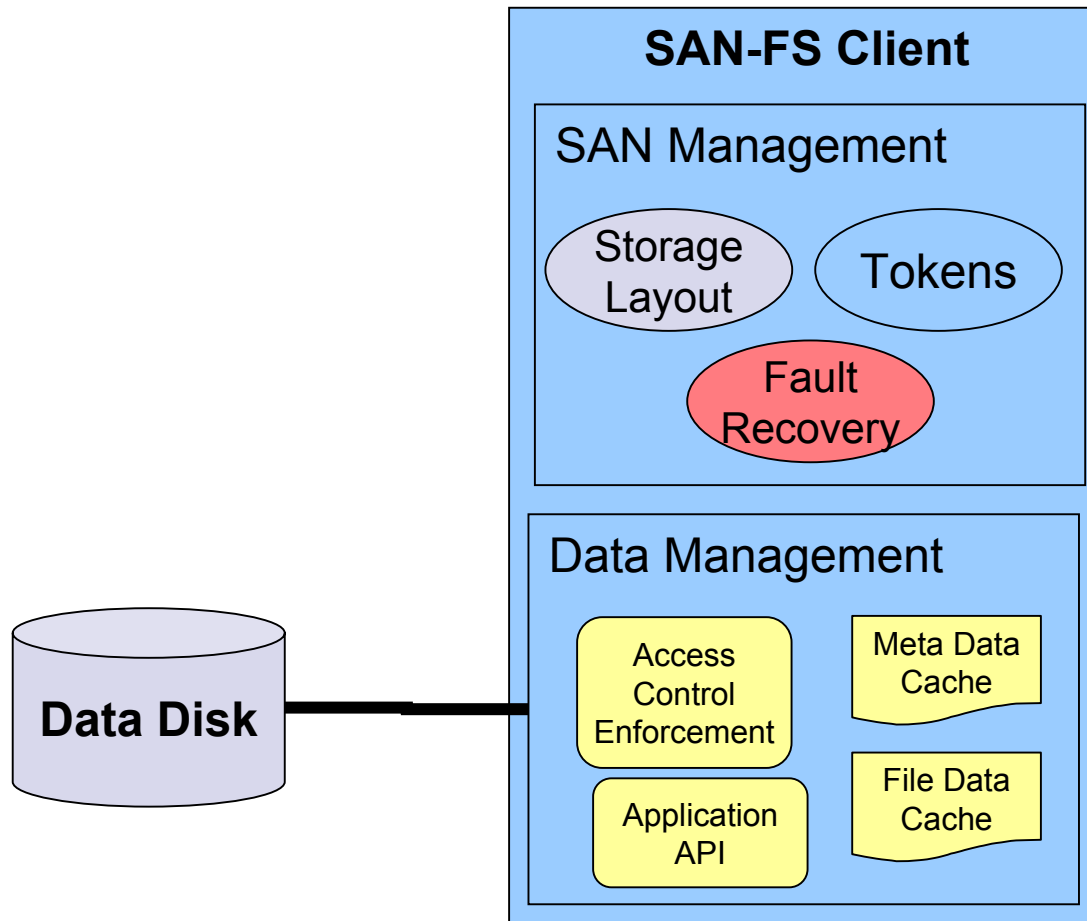
Server Components





Inside a SAN File System

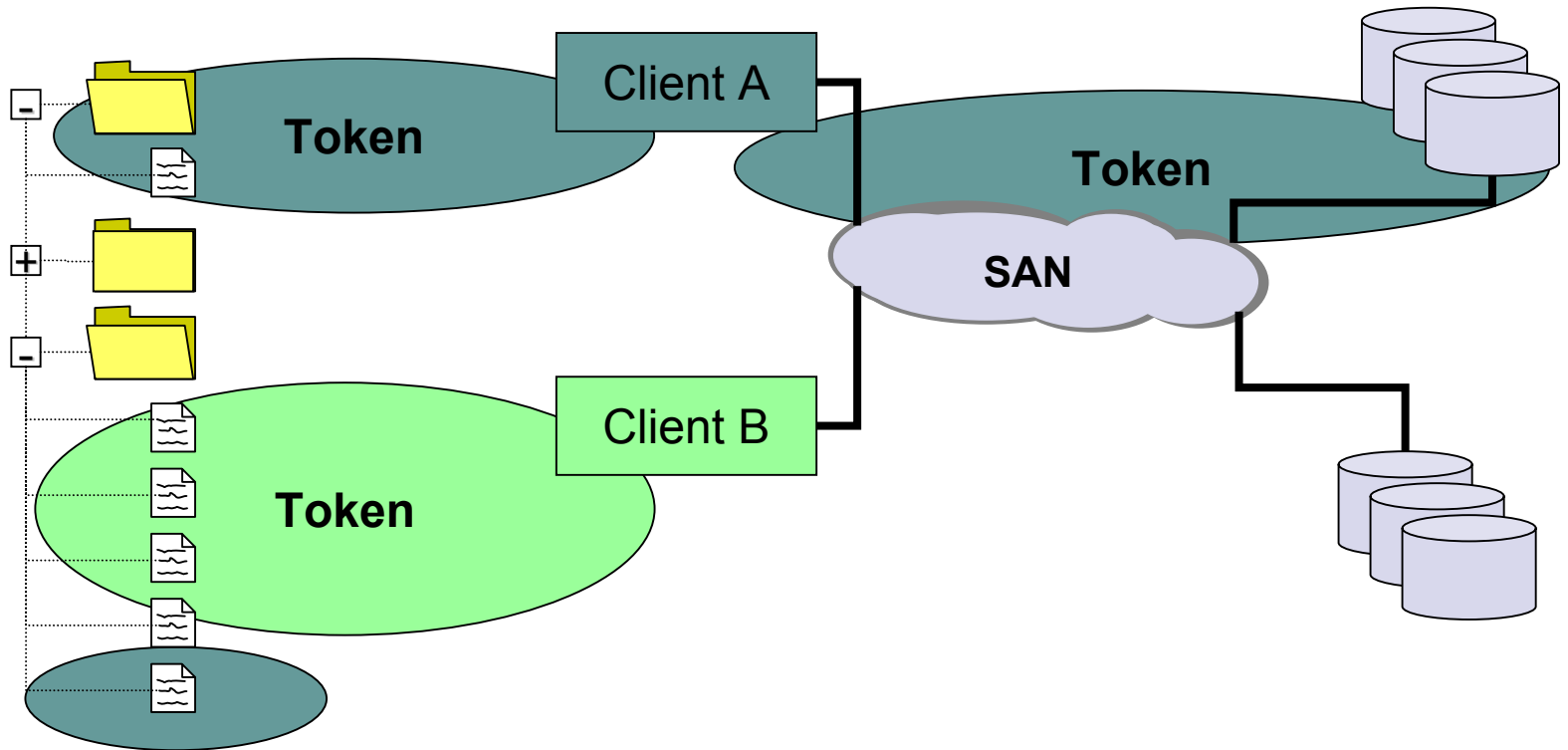
Client Components

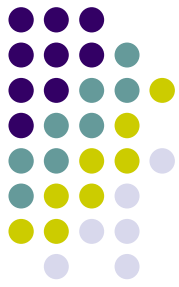


Inside a SAN File System



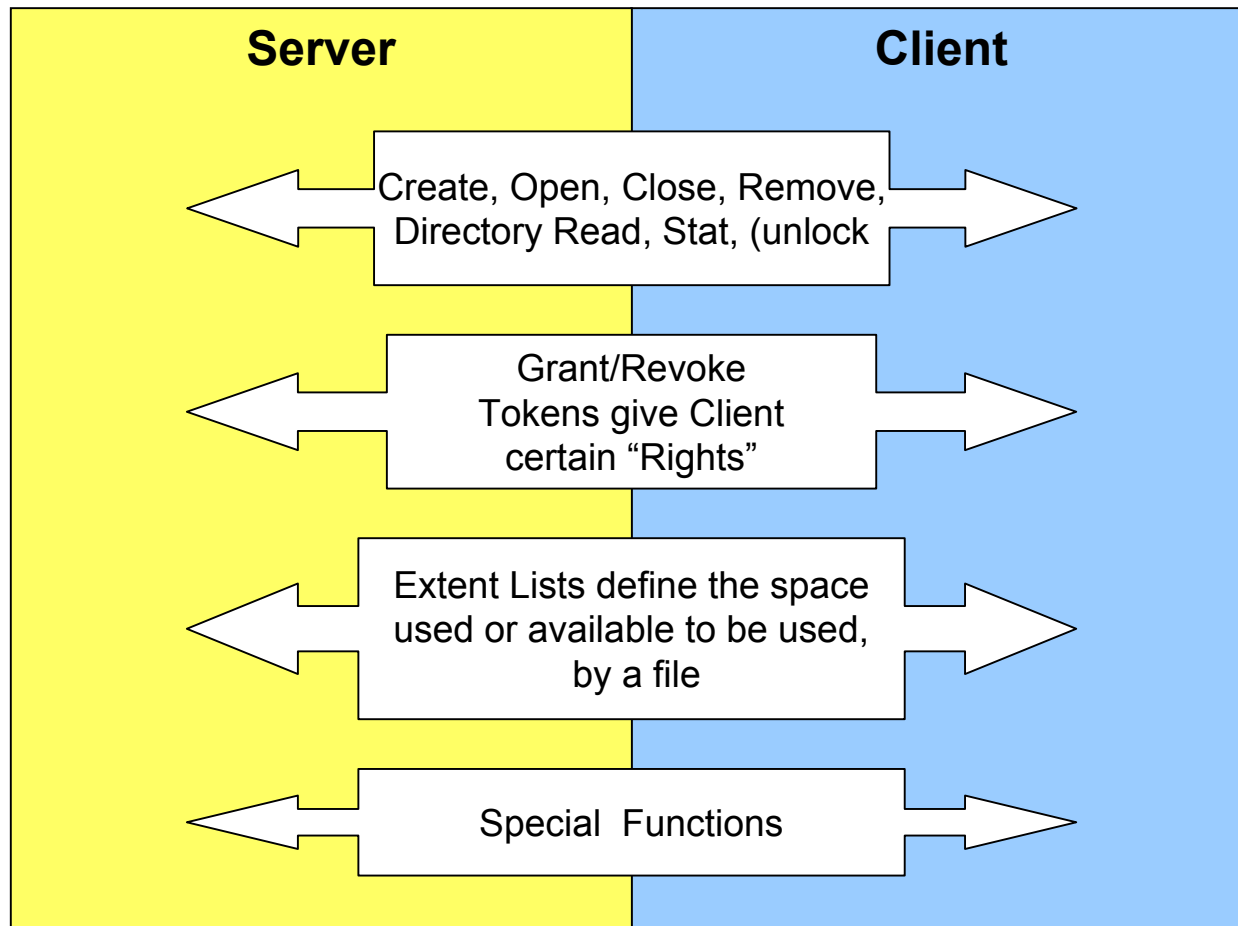
Tokens





Inside a SAN File System

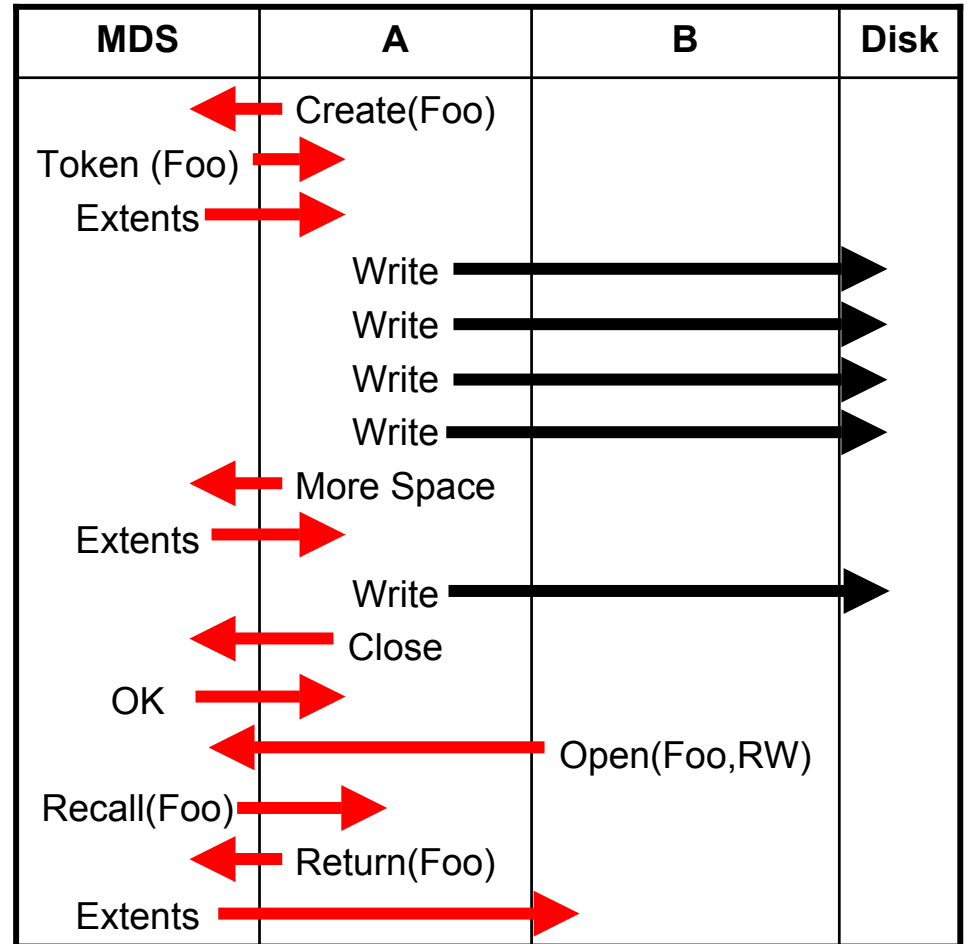
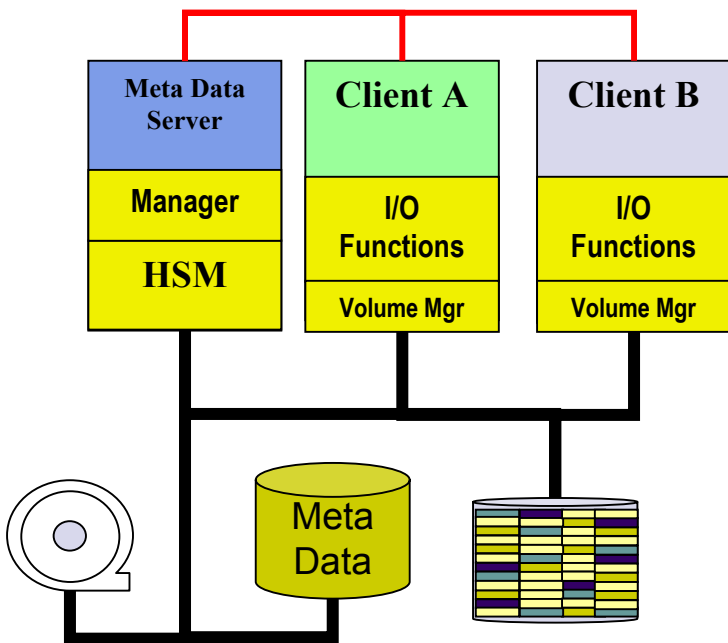
Client Server Interaction





Inside a SAN File System

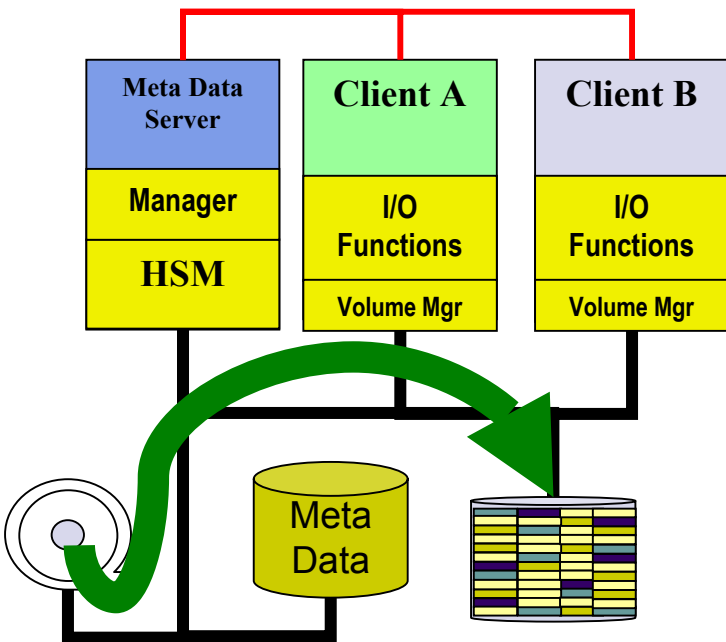
Sample Operation





Inside a SAN File System

HSM Operation

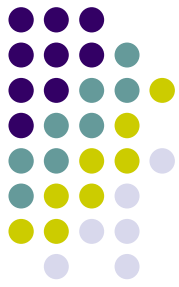


MDS	A	B	Disk
	Open(Foo)		
Offline(Foo)			
HSM			
Online(Foo)			
Token(Foo)			
Extents			
	Read		

Inside a SAN File System



- Scalability Concerns
 - Large Number of Clients in a Single File System
 - Central Locking Versus Distributed
- Similar Problems with NAS
- Does the SAN Allow New Alternatives?

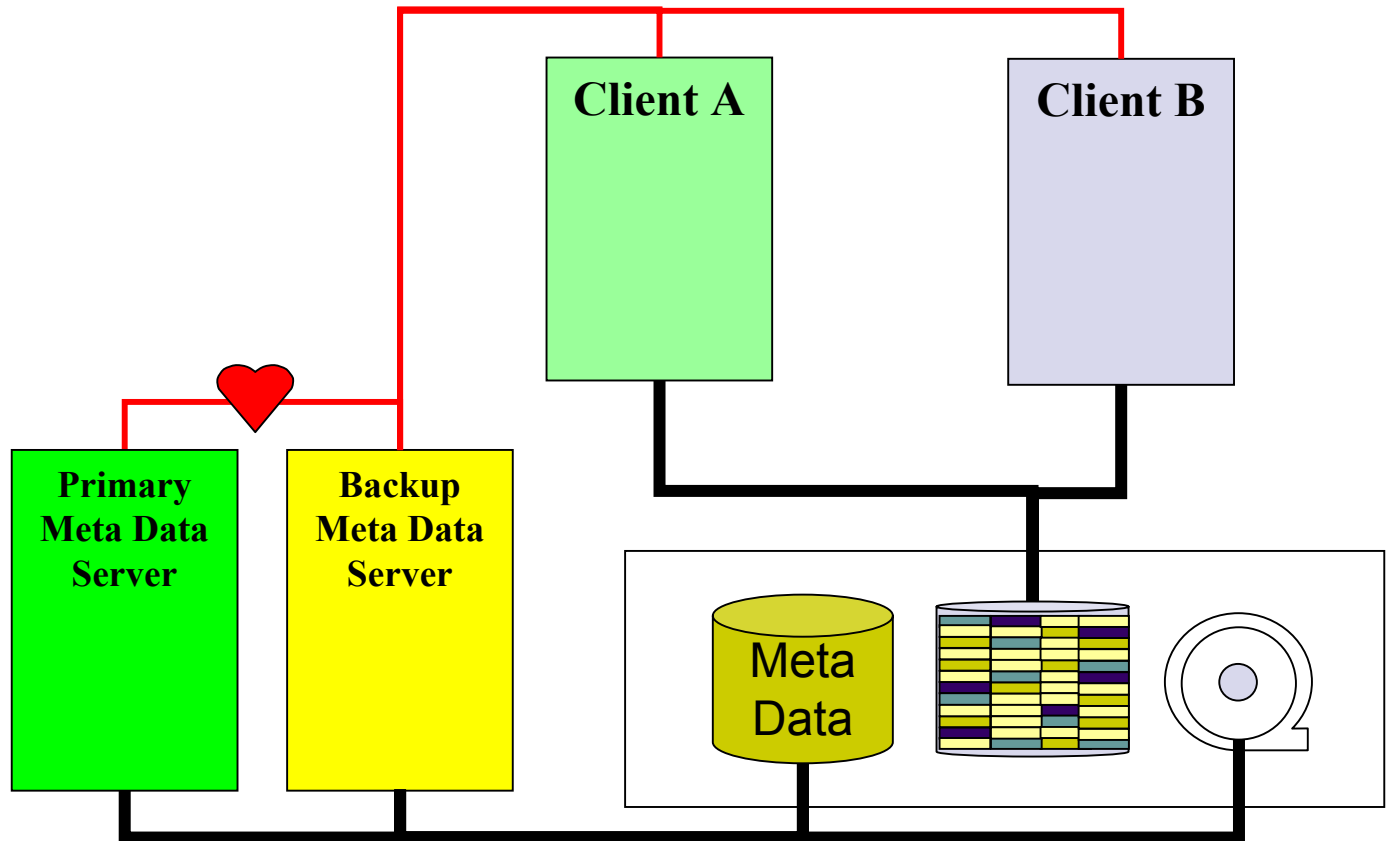


Base Requirements

- Reliability
- Transparent to Applications
- Secure Files
- Manage Space
- Share Files
- Performance
- Protection of Data
- Storage Management
- Administration

SAN File System Architecture

Non-Stop Operation





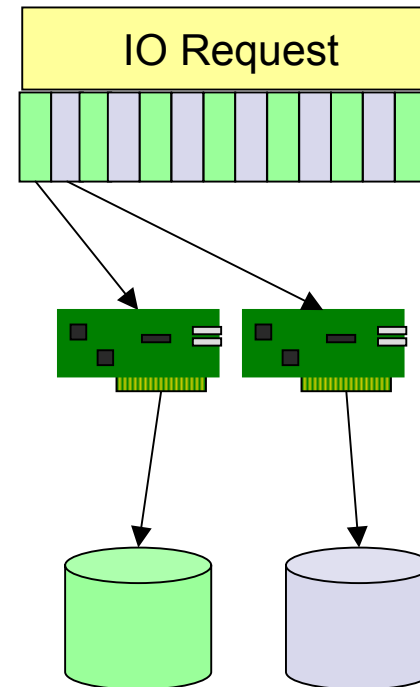
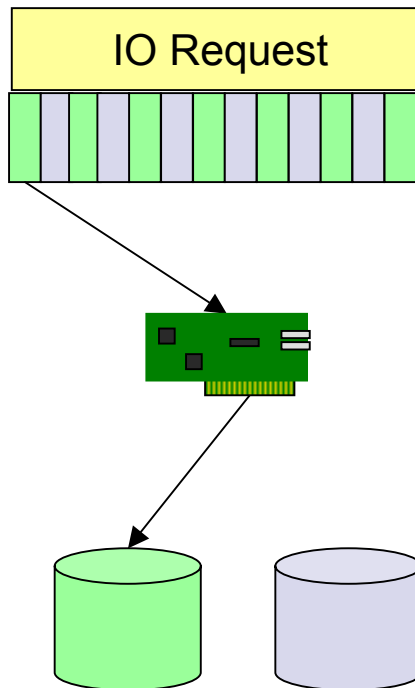
Performance

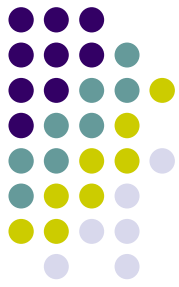
- Deliver Majority of Bandwidth
- Low Latency Access
 - Meta Data
 - Storage
- Multi-threaded for Multi-Processor
- Millions of Files



Inside a SAN File System

Bandwidth Aggregation

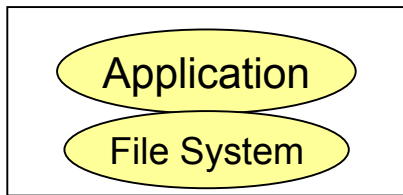




Inside a SAN File System

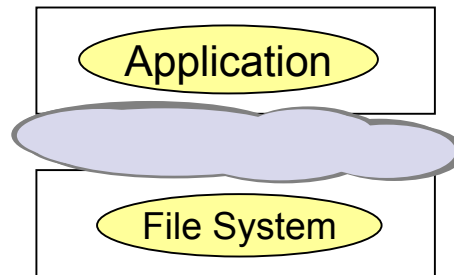
Meta Data Latency

Server



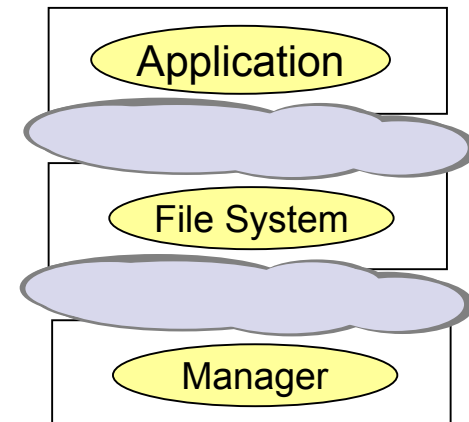
Application and File System on Same machine

NAS



Network between application and File system

NAS client to SAN FS



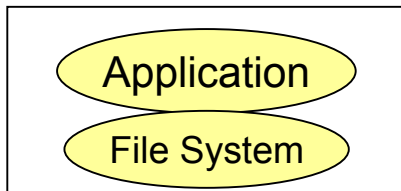
SAN Clients must contact SANFS Manager for meta data



Inside a SAN File System

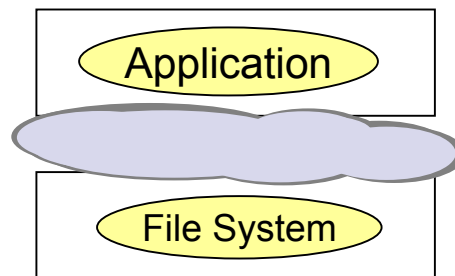
I/O Latency

Server



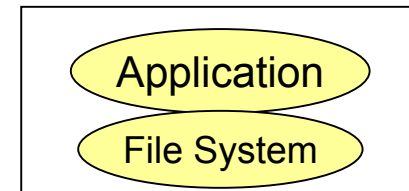
Application and
File System on
Same machine

NAS



Network between
application and
File system

SAN File System

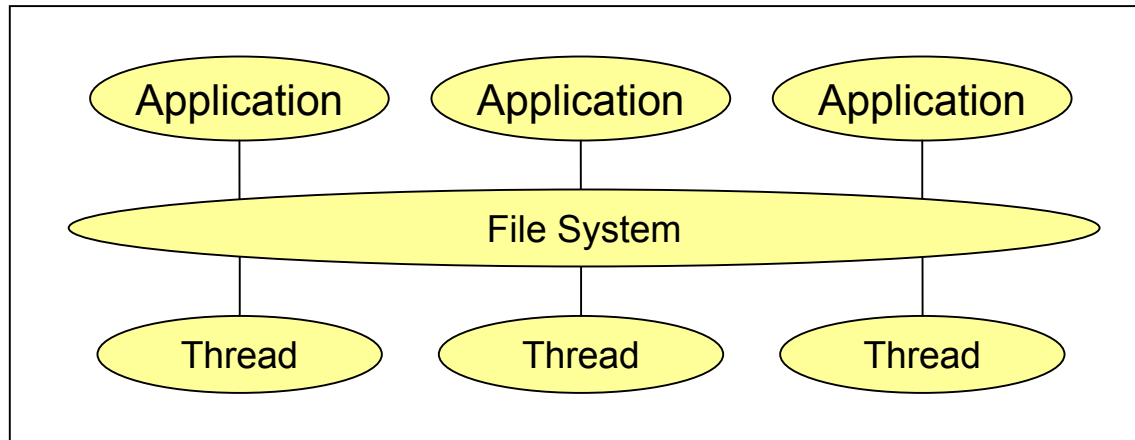


I/O path is
Identical to local
File system



Inside a SAN File System

Multi-threaded

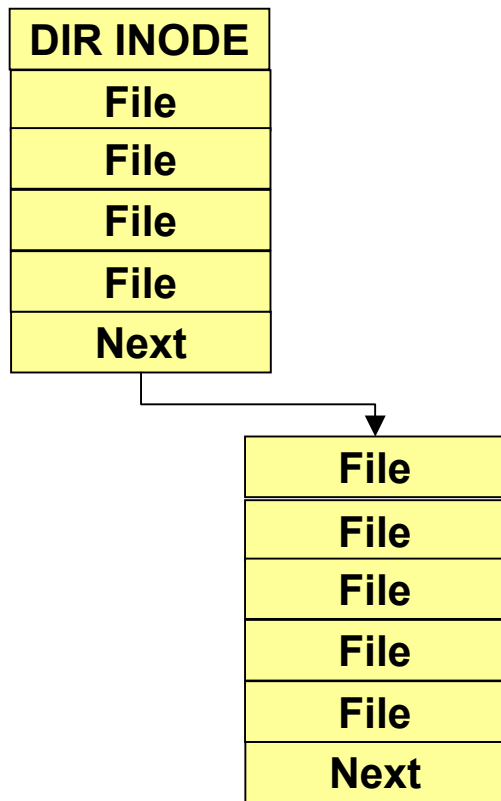


On a 3 CPU system all three I/O request
Will run at that same time

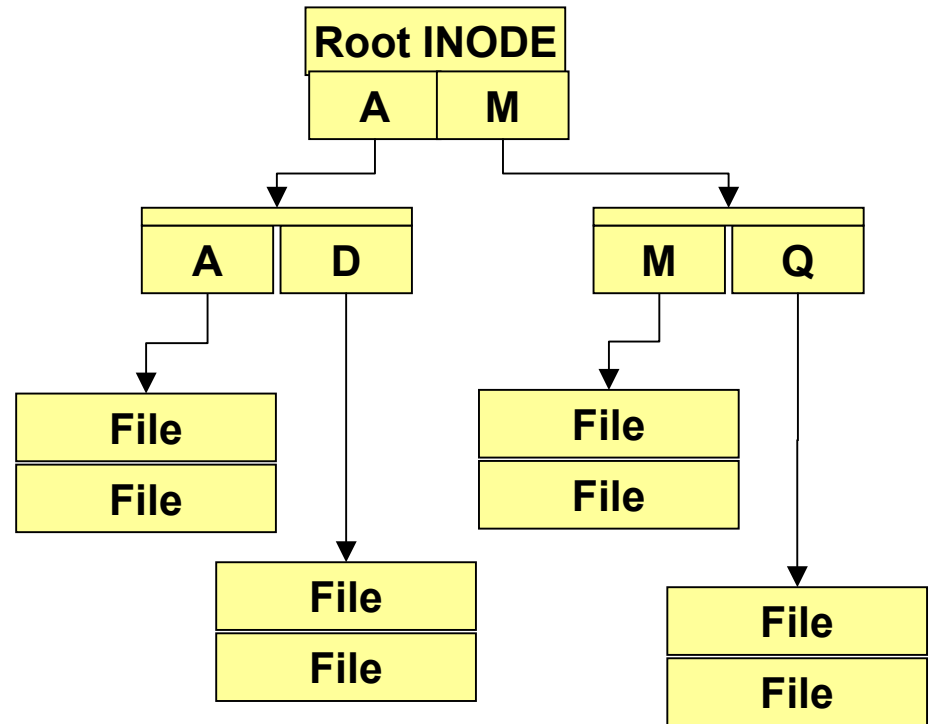


Inside a SAN File System

Millions of Files



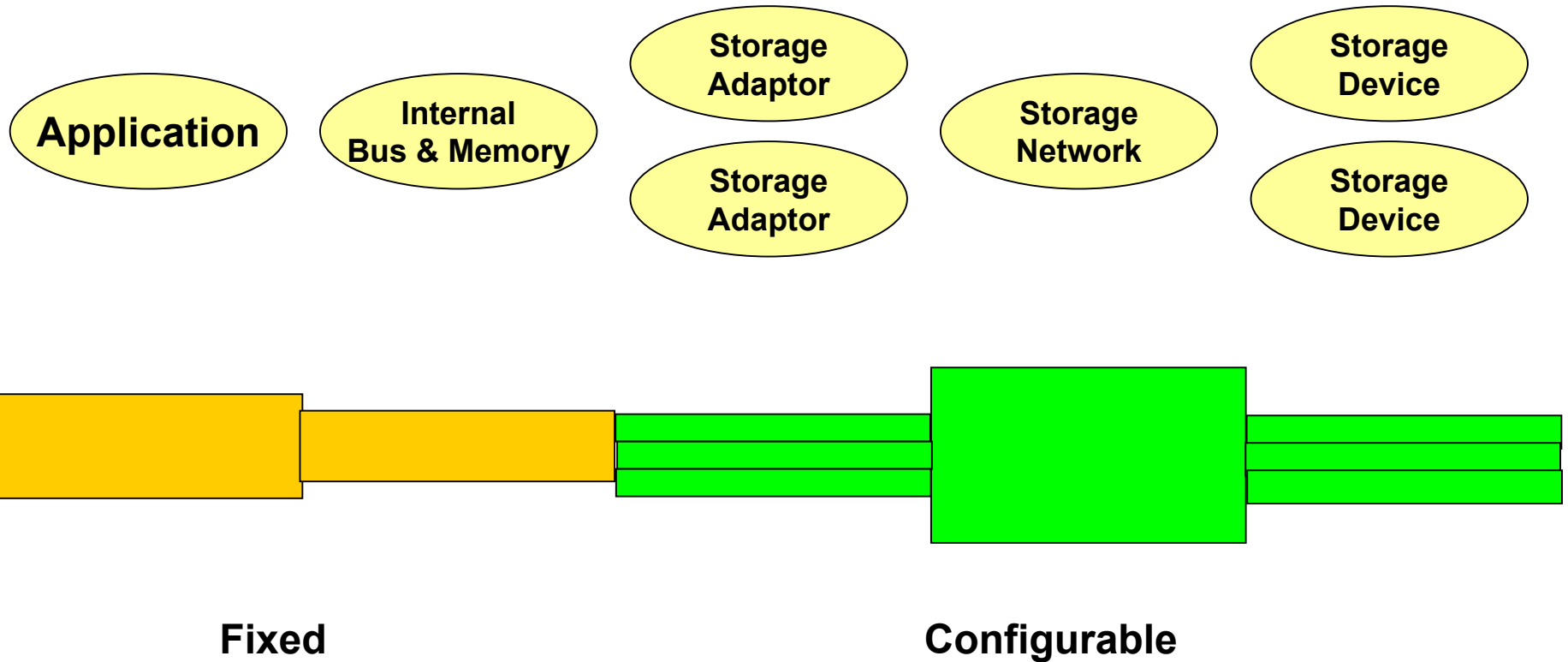
Old Structure

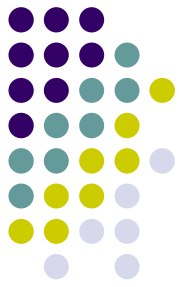


B+ Tree Structure

Bandwidth

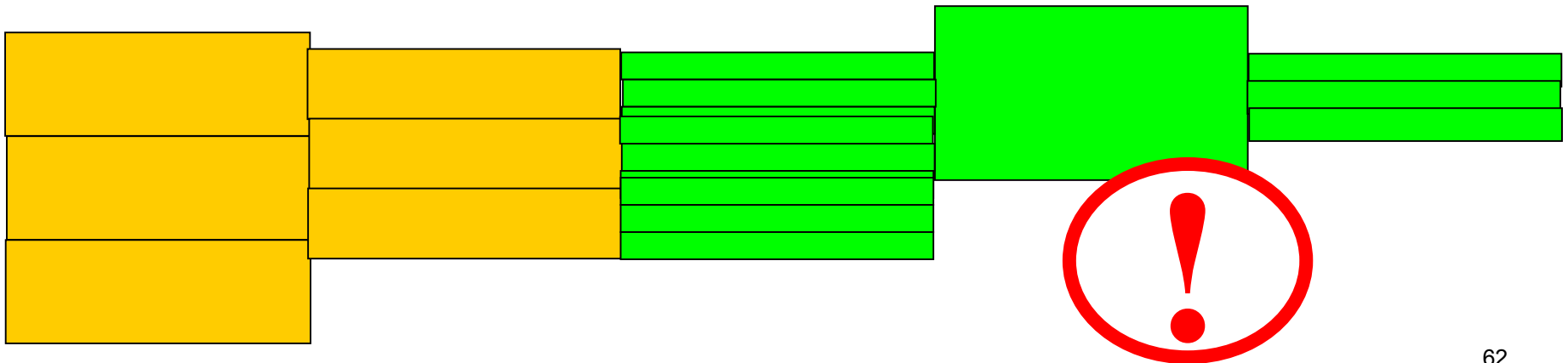
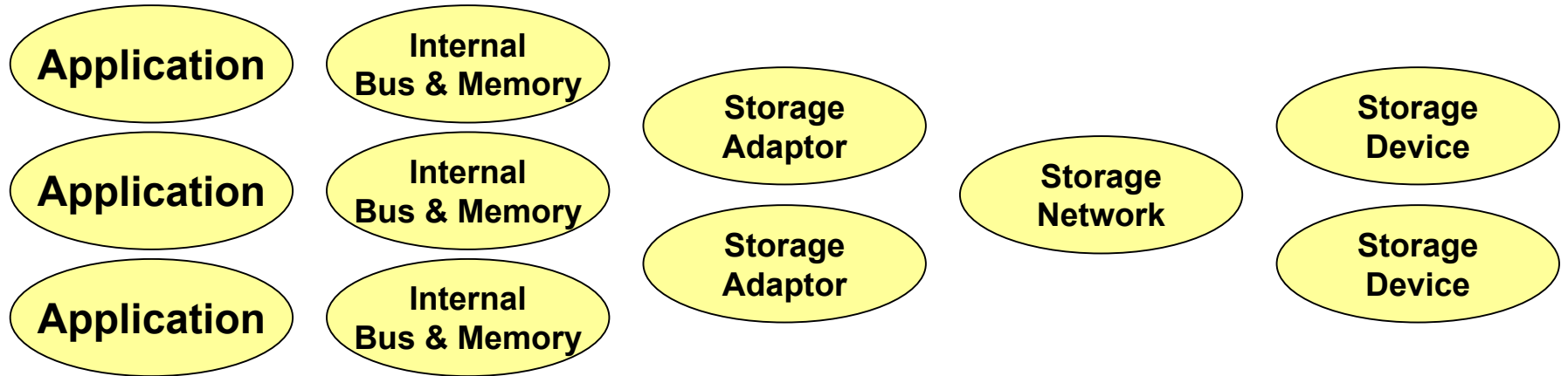
End-to-End Balance





Bandwidth

End-to-End Balance, Multiple Clients





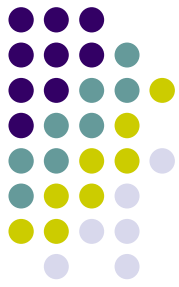
Operations per Second

- Achieved through caching
 - Channels limited to xxK OPS
 - Avoid Disk Access during transaction
- Read Ops are easier
 - Last Access Time
- Write Ops require more
 - Must guarantee that write will complete
 - Battery backup
 - Non-Volatile RAM



Operations per Second

- Multi-headed NAS Servers
 - More heads for Network and I/O Bandwidth
 - Introduces cache coherency issue (SAN-FS??)
 - More heads = more op/s
- NAS-FS Gateways
 - NAS head for a SAN File System
 - Challenges on the OP/s Side
 - Need SAN bandwidth and NAS OP/s
 - True SAN/NAS Convergence



Steered Performance

- Applications request a specific QoS
 - Application Needs 80 MB/s
 - System Throttles other I/O
- Applications request specific storage type
 - This file should be placed on storage of this type
- Applications give hints
 - What they need next for storage
 - How they will read a file
 - How files are associated



Storage Management

- Logical Volume Management
 - LUN Management
 - Heterogeneous Storage
 - Snapshots
 - Mirrors
- Hierarchical Storage Manager
 - Space Management
- Load Balancing
- Fault Tolerance

Data Management

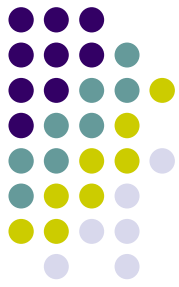
- Relocation
- Migration
- Backup
- Replication





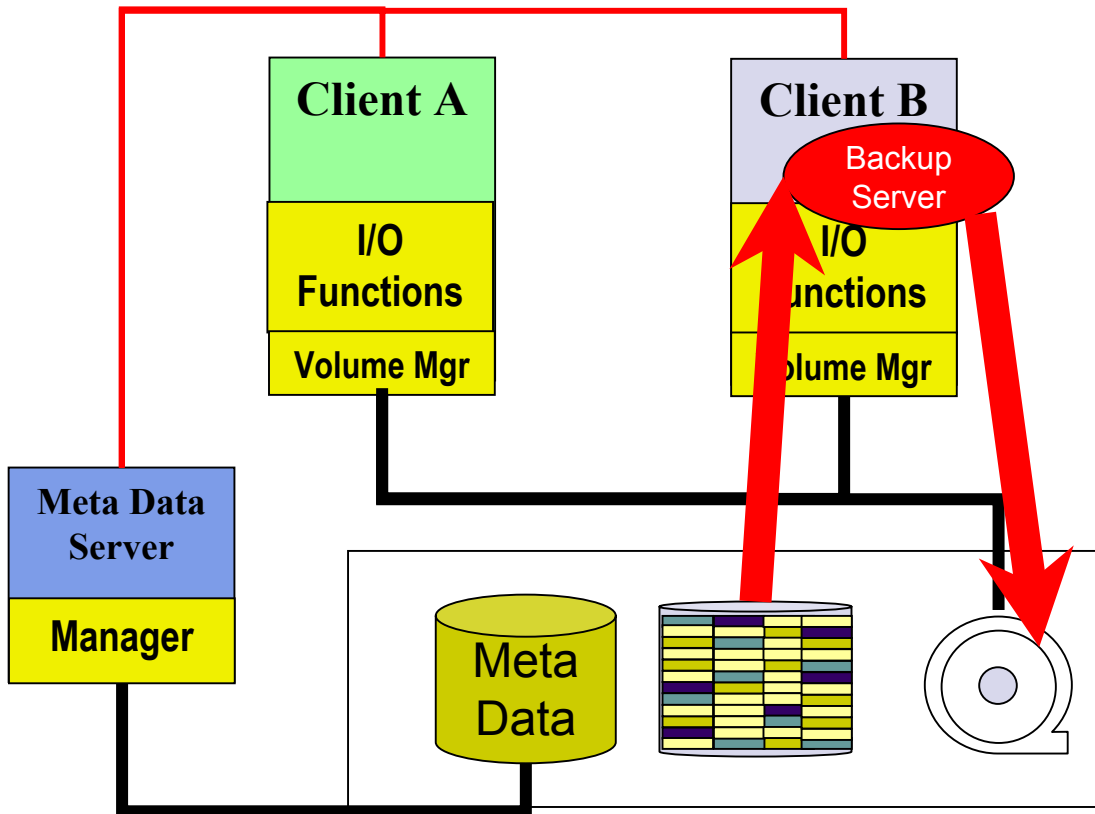
Protect Data

- Backup
 - Standard Backup Tools Should Work
- Business Continuance Support
 - Snap Shot Support
 - Synch and Pause
- Replication
 - Directory and File Level
 - Local and Remote
 - Synchronous and Asynchronous



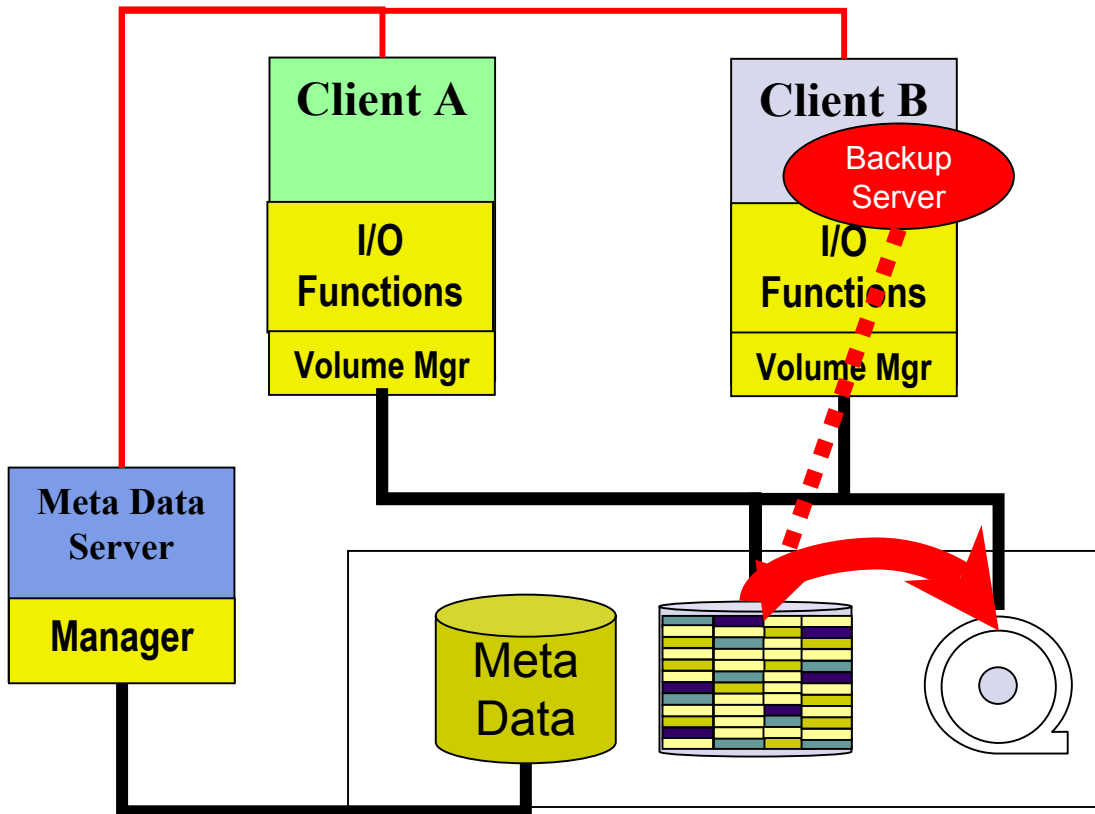
SAN File System Architecture

Backup Runs on SAN-FS Client



SAN File System Architecture

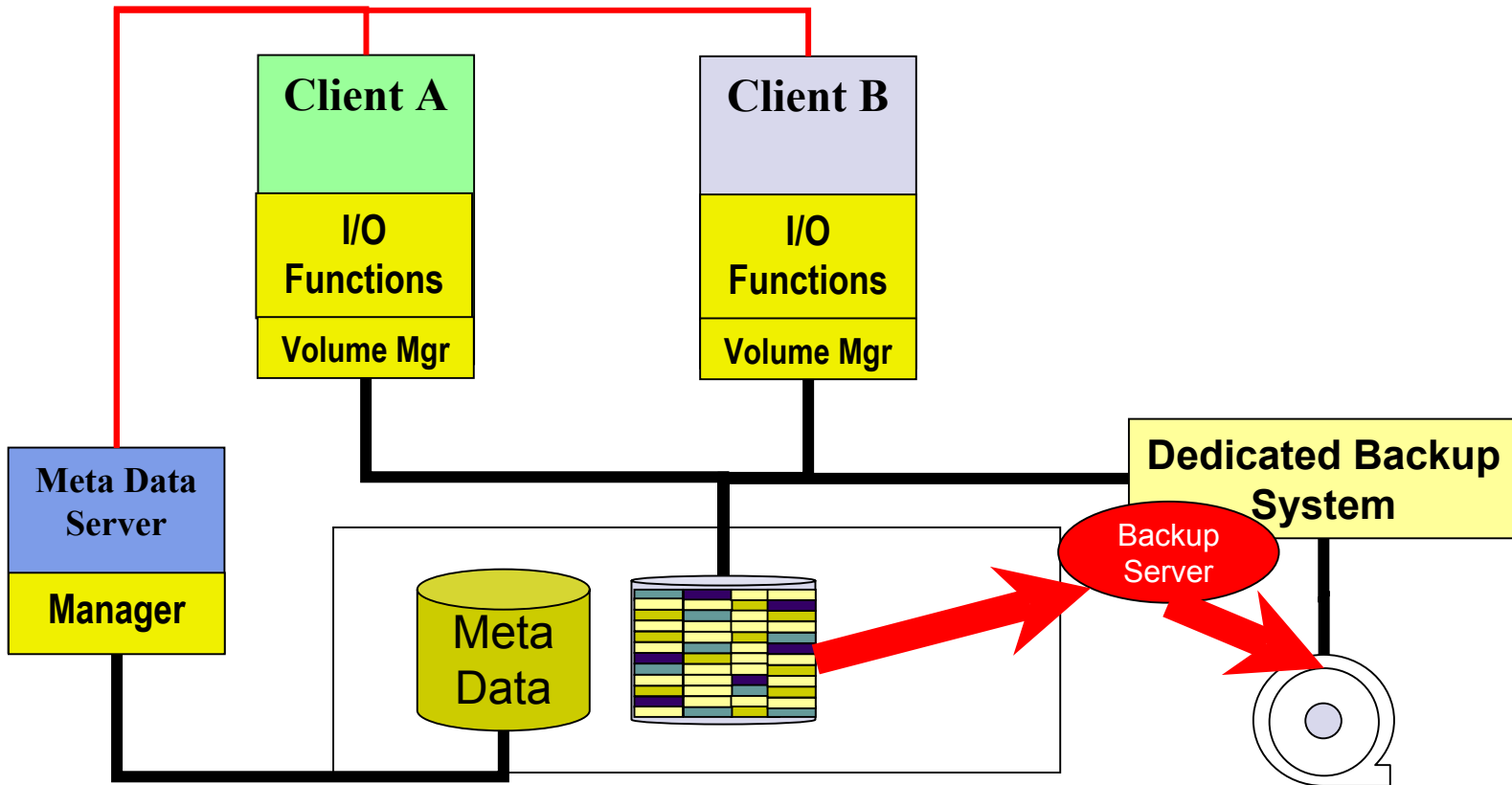
Backup Using 3RD Party Copy





SAN File System Architecture

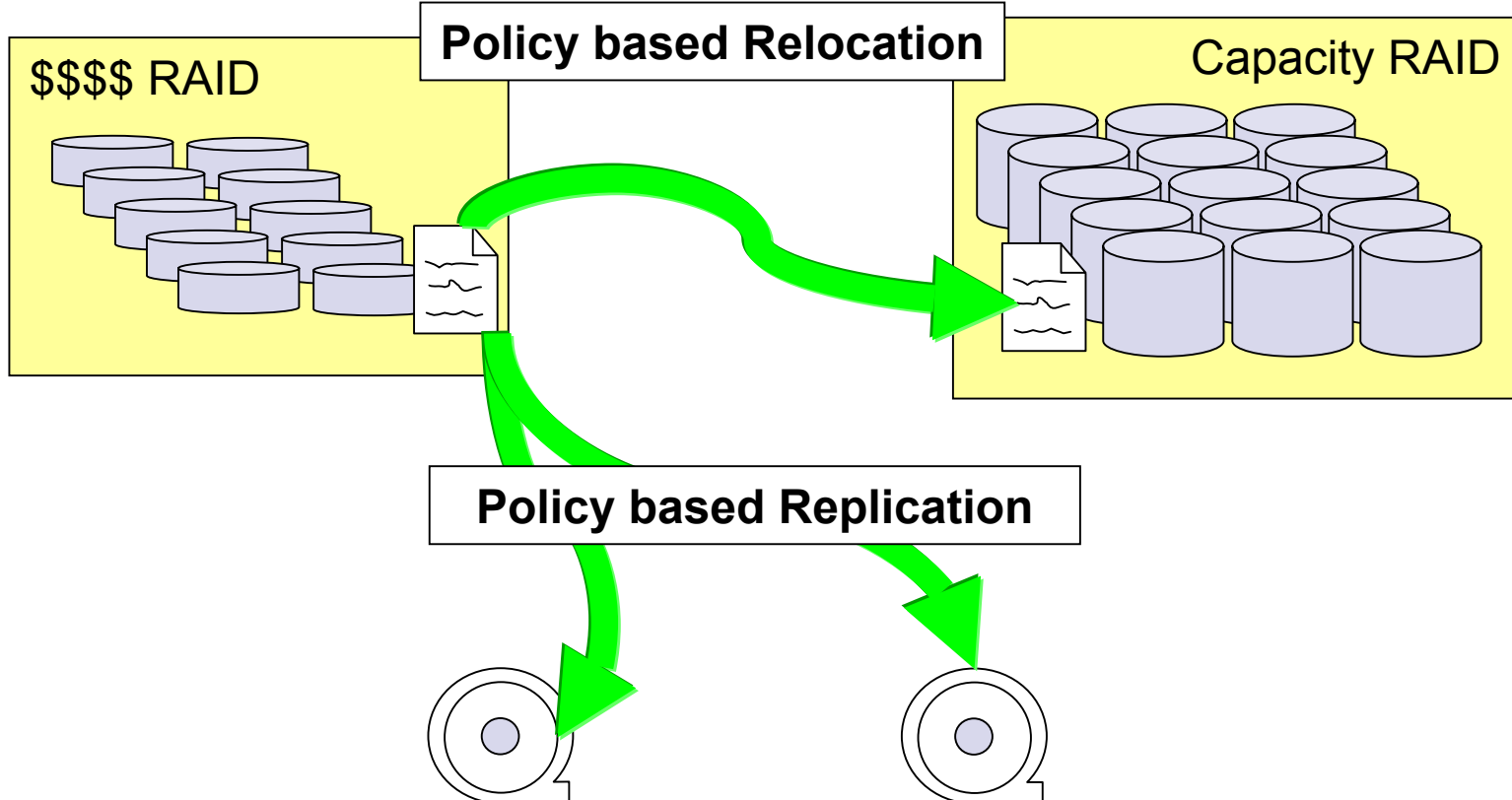
Backup Runs on Dedicated System

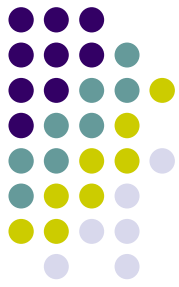




Relocation/Replication

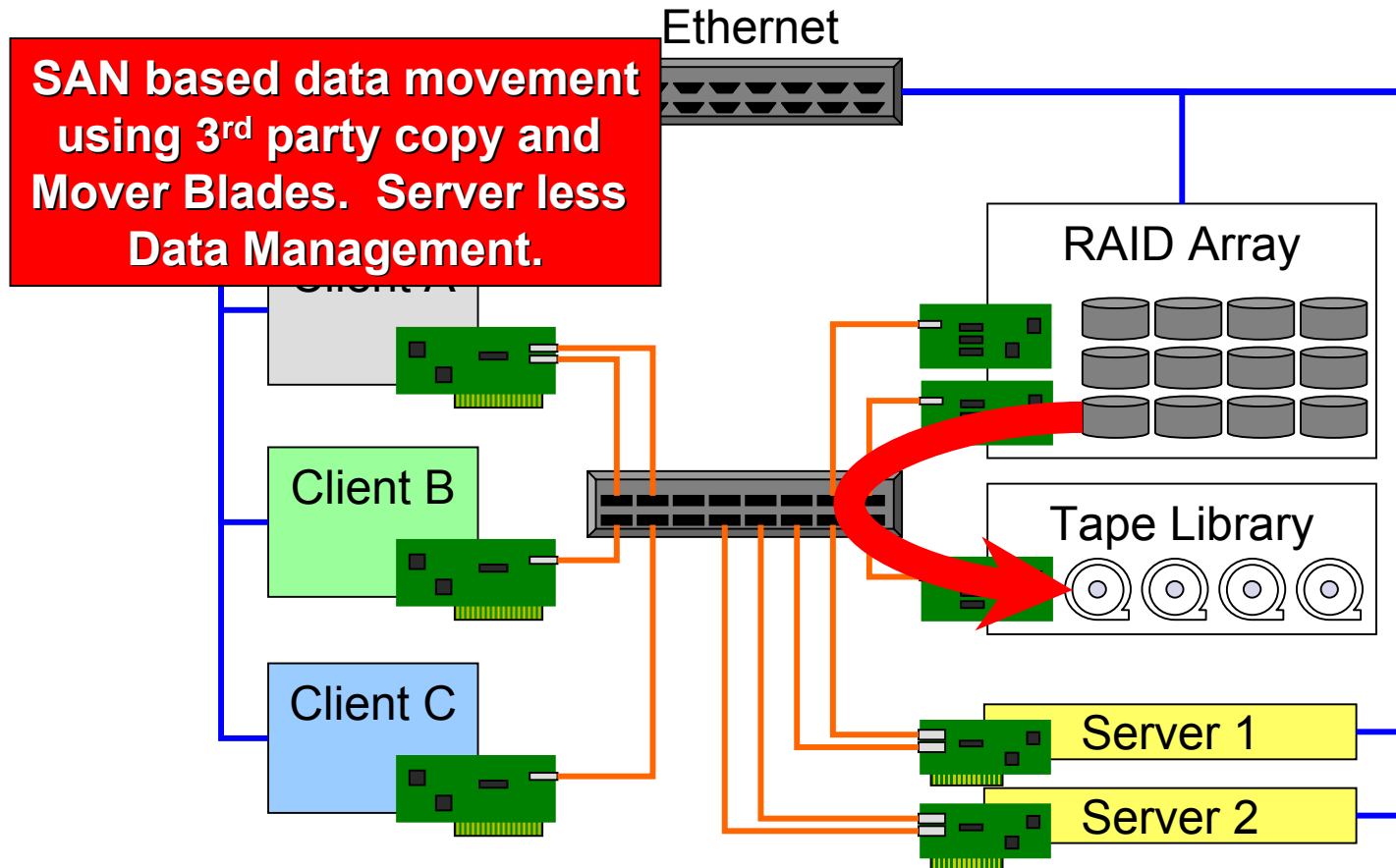
Single Name Space





SAN File Systems

Replication/Migration



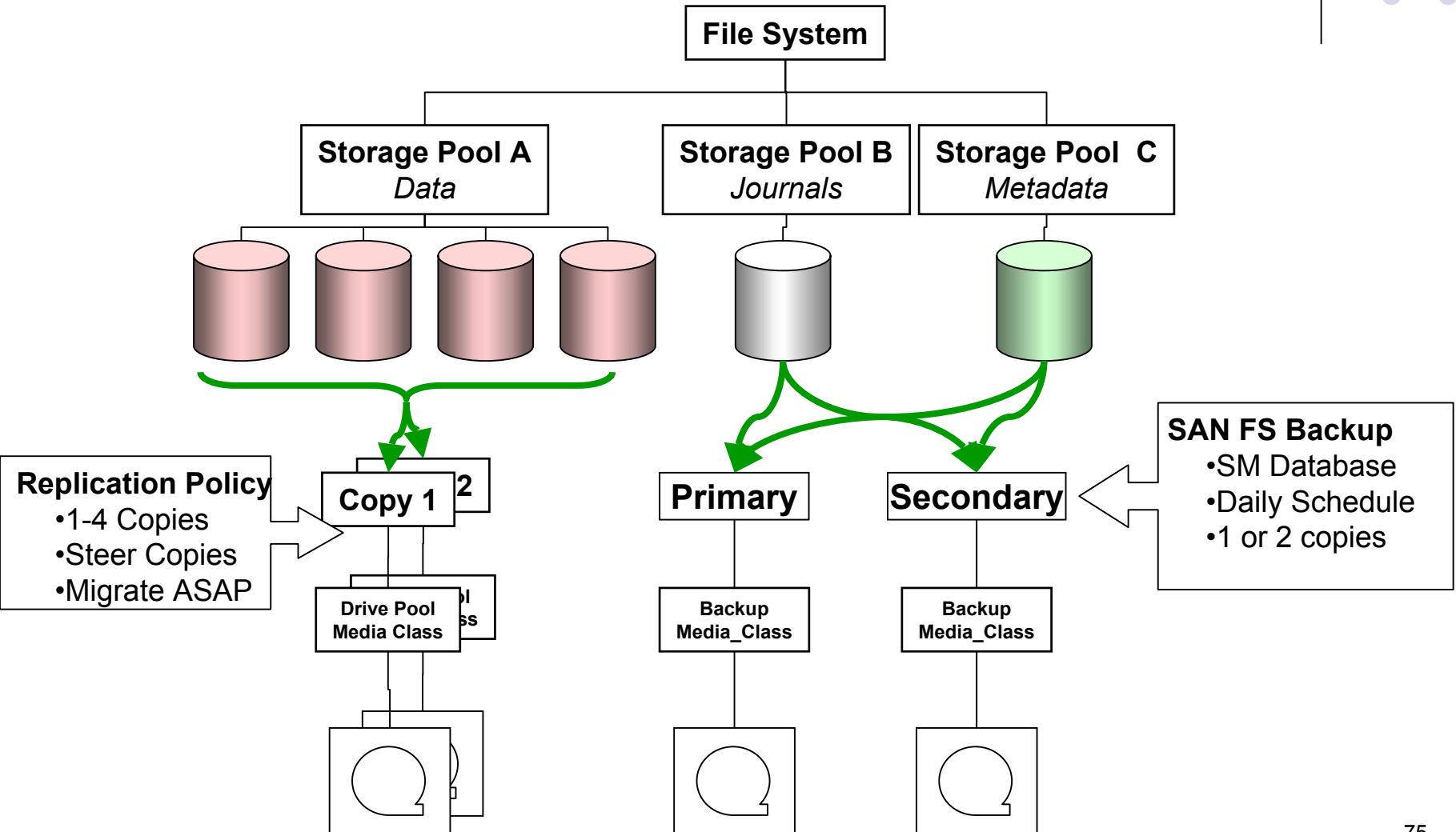


Backup Revisited

- Change Migration to Replication
- Data is replicated continuously
- Multiple copies are made
- Versions are kept
- Recovery is different



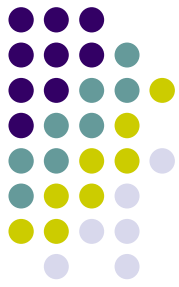
SAN FS Data Protection



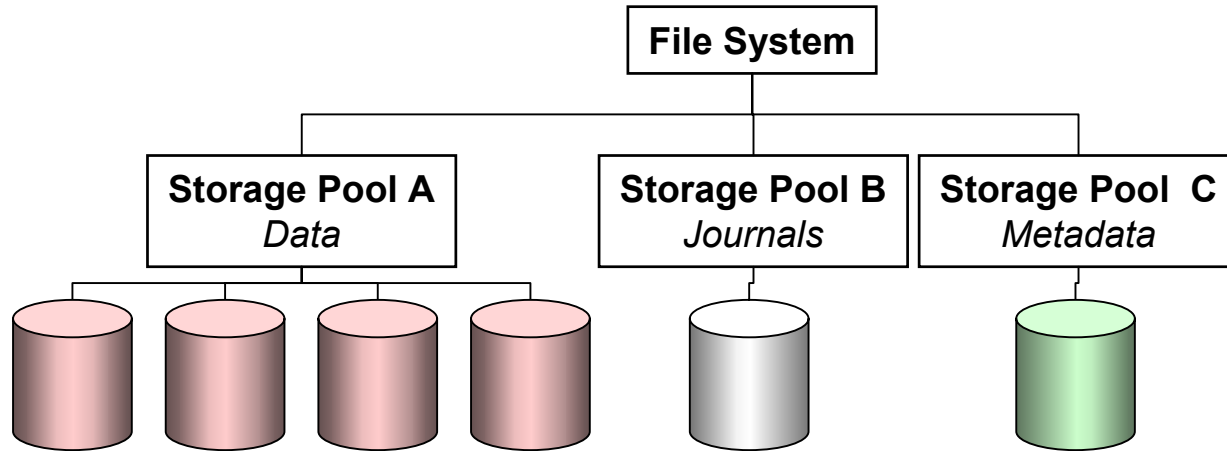


Recovery

- Backup requires full restore from tape
- Snap Shots do not do DR
- Integrated DP
 - Restore of Meta Data
 - Restore only Recently accessed files
 - Tape mounts \leq Backup
 - Data Read \lll Backup

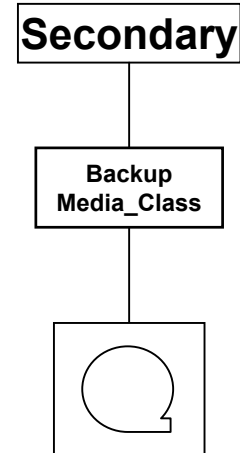
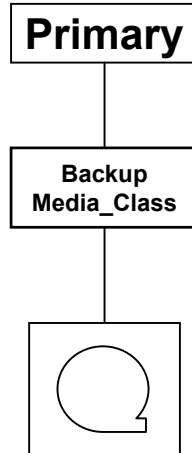
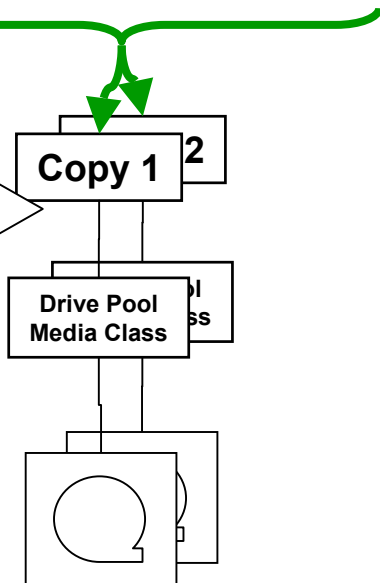


SAN FS Data Protection



Replication Policy

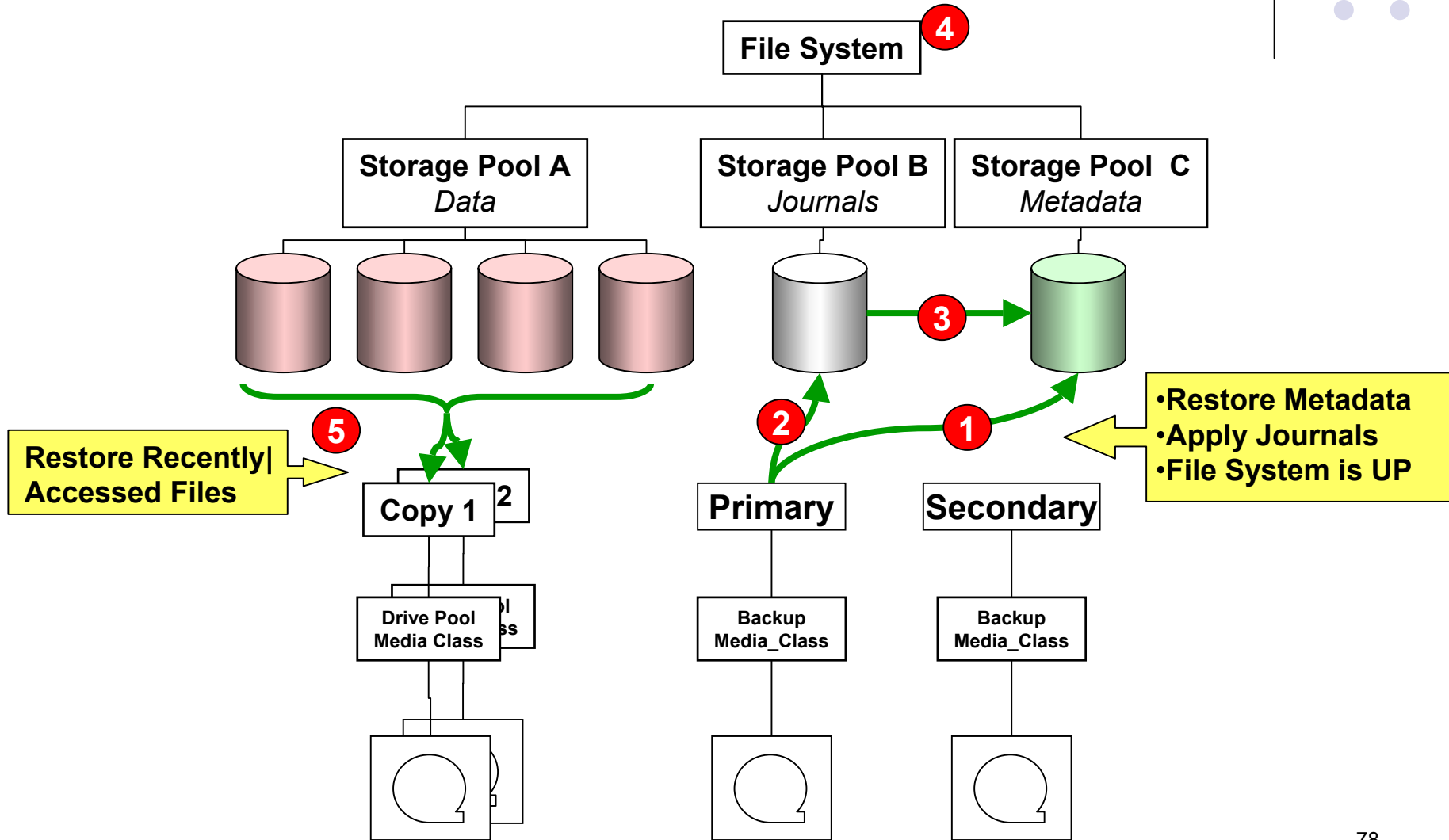
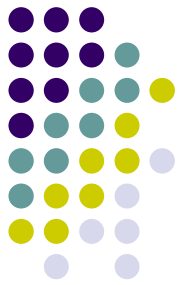
- 1-4 Copies
- Steer Copies
- Migrate ASAP



SAN FS Backup

- SM Database
- Daily Schedule
- 1 or 2 copies

SAN FS Data Protection





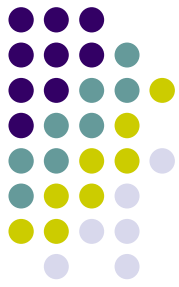
RAS

- Reliability
 - Checks and Balances
 - Recoverability
- Availability
 - Redundancy
 - Reduced Performance
- Serviceability
 - Expansion
 - Code Updates



Reliability

- Multi-Path IO
- I/O Fencing
 - Restrict Access
- State Cleanup
- Storage Failure
 - Alternate Allocation Storage Pools



Availability

- Fail over of Manager
 - Active-Passive
 - Active-Active
- Client Fail over
 - Network
 - Application
- Client Independence
 - Run without Server for Delta Time
- Degraded Mode
 - Storage Pool Can Go Down, System Stays UP

Serviceability



- Software Upgrades on the Fly
- Test New Releases in Parallel
- Fix Hardware Without Stopping System
- Logging of all Activity

Security

- Trust
- Authentication
- Group & User Management
- Authorization
- Cross Domain Management
- File Encryption
- Auditing
- SAN Security

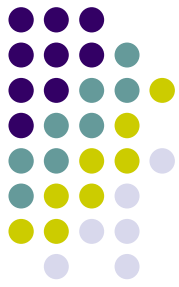




SAN-FS Security

User Management & Authentication

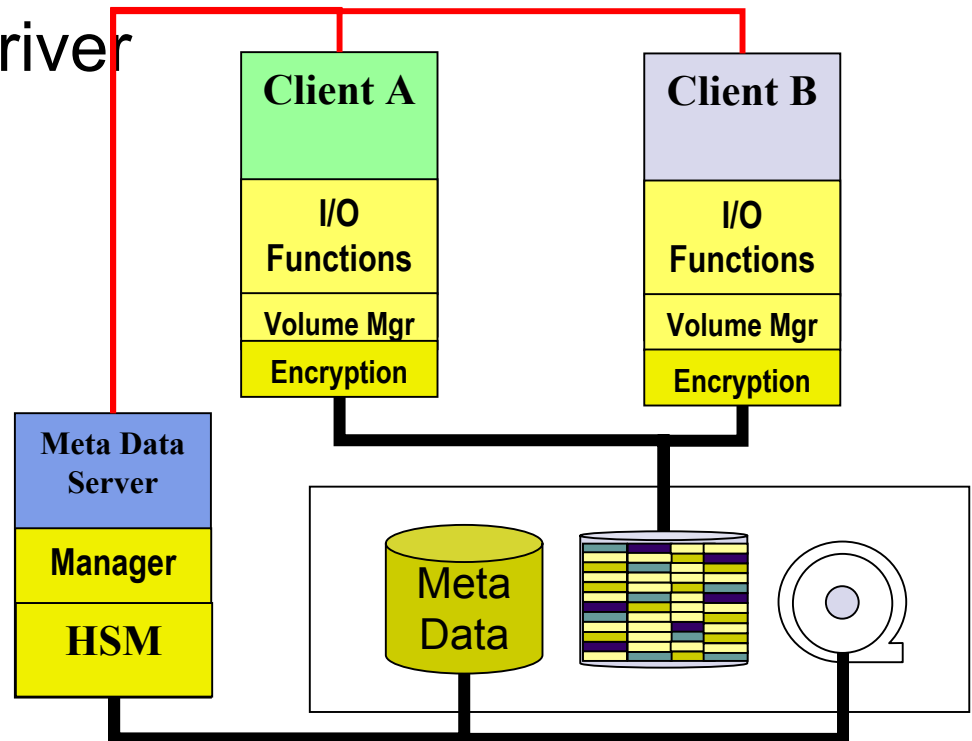
- Must Encode Owner/Access Control
 - Use Unix or Windows or ???
 - UID vs. SID, Posix ACLs vs. Windows ACLs
 - Customers Sites Vary
 - So the answer is ALL OF THE ABOVE and the Future
- Components of the System Must Authenticate
 - Local File System with Distributed Components
- Standardization Needed
 - LDAP
 - XOpen GSS-API



SAN-FS Security

Encryption

- Multiple Methods
 - Layered Device Driver
 - Hardware Bridge
 - File-by-file



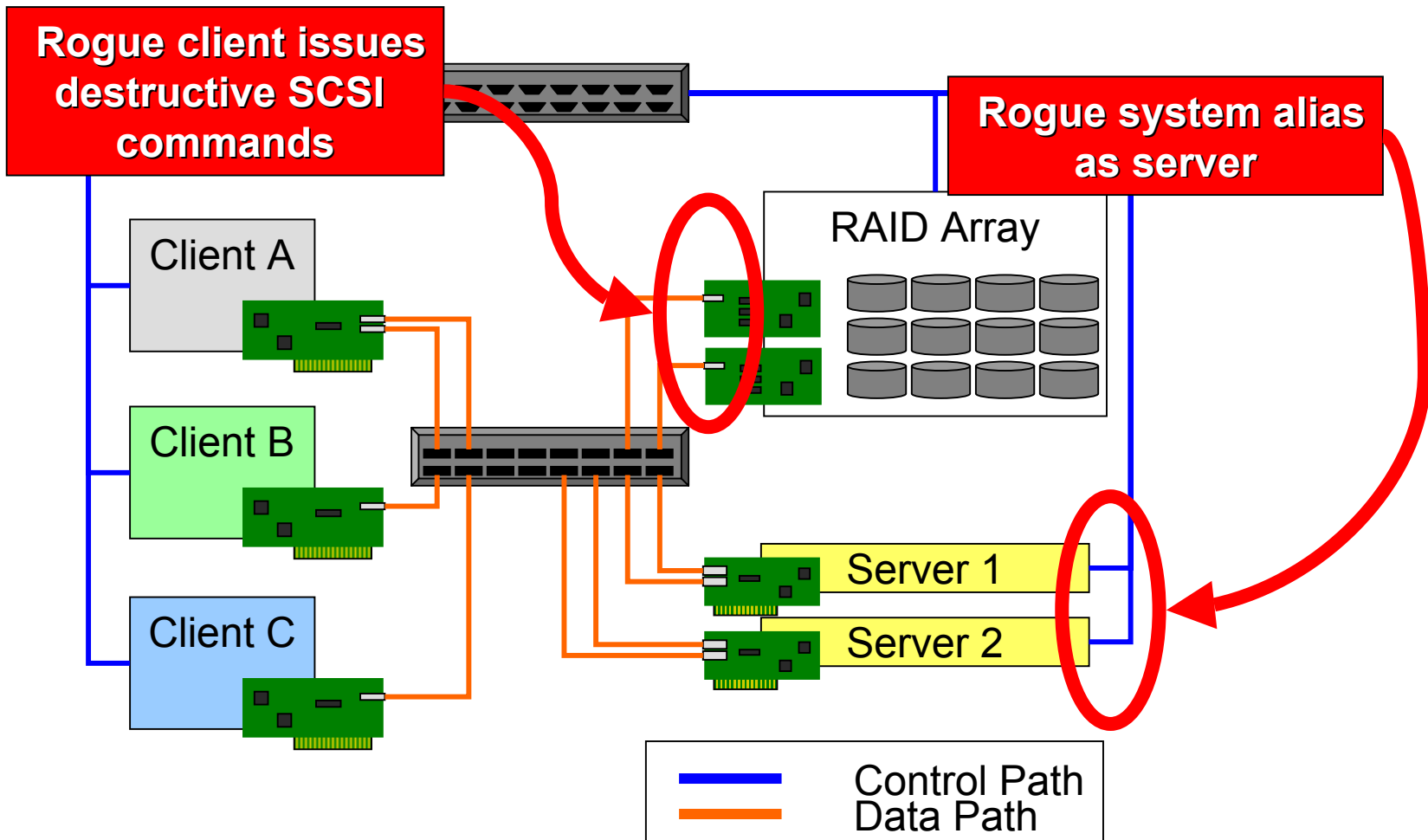
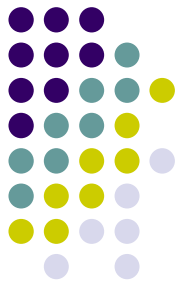
SAN-FS Security

- Auditing
 - Accountability
 - Attempts
 - Unauthorized access
 - Misuse of privileges



SAN-FS Security

SAN Security



True SAN/NAS Convergence

