Hierarchy of Storage

- Cache
- RAM
- Disk
- Optical
- Tape

Higher Cost → Lower Cost
Smaller Capacity → Larger Capacity
Faster Speed → Slower Speed
Capacity Growth

IBM HDD Evolution

Ed Grochowski at Almaden
Storage Costs
Disk Architecture

- Platters
- Spindle
- Track
- Cylinder
- Sector
- Arms with read/write heads

Rotation
RAID

• Redundant Arrays of Inexpensive Disks

• Two orthogonal concepts:
  – data striping for performance
  – redundancy for reliability

• Striped arrays can increase performance, but at the cost of reliability (next page)
  – redundancy can give arrays better reliability than an individual disk
Reliability of Striped Array

![Diagram showing the reliability of a striped array with the number of disks on the x-axis and system reliability on the y-axis. The reliability decreases as the number of disks increases.]
## RAID Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Additional Disks</th>
<th>Failures Tolerated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Non-redundant striping</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>Mirrored</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Memory-style ECC</td>
<td>$1 + \lg n$</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Bit-Interleaved Parity</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Block-Interleaved Parity</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Block-Interleaved, Distributed Parity</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>P+Q Redundancy</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
RAID: 4x Small Write Penalty
Log-Structured File Systems

• Based on assumption that disk traffic will become dominated by writes
• Always writes disk data sequentially, into next available location on disk
  – no seeks on write
• Eliminates problem of 4x write penalty
  – all writes are “new”, no need to read old data or parity
• However, almost no examples in industry file systems
Networked Storage

- Storage attached by general-purpose or dedicated network

- Motivations:
  - homogenous and heterogeneous file sharing
  - centralized administration
  - better resource utilization (shared storage resources, pooling)

- Dedicated Networks:
  - Fibre-Channel: FCP (SCSI over FC)
  - iSCSI: SCSI over IP
  - InfiniBand
Networked Storage

- Can mean many things:
  - NAS (Network-Attached Storage): file server appliances serving NFS and/or CIFS (for example, Network Appliance)
  - NASD (Network-Attached Secure Disk): intelligent, network-attached drives w/ security features (also, Network-Attached Storage Device)
  - SAN (Storage Area Network): network for attaching disks and computers, usually dedicated only to storage operations
    - OBSD (Object-Based Storage Device): similar to NASD
A SAN File System

Control Network (IP)

Win2K IFS w/cache
AIX IFS w/cache
Solaris IFS w/cache
Linux IFS w/cache

SAN

Data

HSM & Backup

NFS
CIFS
FTP
HTTP

Security assists

Win2K
AIX
Solaris
Linux

Meta-data Server
Meta-data Server
Meta-data Server
Storage Management Server

Meta-data
Additional Reading

• Hennessy & Patterson: Chapter 6
• Rosenblum & Ousterhout: The design and implementation of a log-structured file system. ACM Transactions on Computer Systems, Feb. 1992, 26-52
• http://www.almaden.ibm.com/cs/storagesystems/stortank/