6.033 Spring 2018

Lecture #14

- Reliability via Replication
  - General approach to building fault-tolerance systems
  - Single-disk failures: RAID
How to Design Fault-tolerant Systems in Three Easy Steps

1. identify all possible faults
A fatal exception 0E has occurred at 0028:C011E36 in VXD VMM(01) + 00010E36. The current application will be terminated.

* Press any key to terminate the current application.
* Press CTRL+ALT+DEL again to restart your computer. You will lose any unsaved information in all applications.

Press any key to continue
Your PC ran into a problem and needs to restart. We're just collecting some error info, and then we'll restart for you.

25% complete

For more information about this issue and possible fixes, visit http://windows.com/stopcode

If you call a support person, give them this info:
Stop code: CRITICAL_PROCESS_DIED
You need to restart your computer. Hold down the Power button for several seconds or press the Restart button.

Veuillez redémarrer votre ordinateur. Maintenez la touche de démarrage enfoncée pendant plusieurs secondes ou bien appuyez sur le bouton de réinitialisation.

Sie müssen Ihren Computer neu starten. Halten Sie dazu die Einschalttaste einige Sekunden gedrückt oder drücken Sie die Neustart-Taste.

コンピュータを再起動する必要があります。パワーボタンを数秒間押し続けるか、リセットボタンを押してください。
How to Design Fault-tolerant Systems in Three Easy Steps

1. identify all possible faults
2. detect and contain the faults
3. handle the fault
quantifying reliability
dealing with disk failures
# Barracuda 7200.10

Experience the industry’s proven flagship perpendicular 3.5-inch hard drive

<table>
<thead>
<tr>
<th>Specifications</th>
<th>750 GB¹</th>
<th>500 GB¹</th>
<th>400 GB¹</th>
<th>320 GB¹</th>
<th>250 GB¹</th>
<th>160 GB¹</th>
<th>80 GB¹</th>
</tr>
</thead>
</table>
| **Model Number**       | ST3750640A  
ST3750640AS | ST3500630A  
ST3500630AS | ST3400620A  
ST3400620AS | ST3320620A  
ST3320620AS | ST3250620A  
ST3250620AS  
ST3250820A  
ST3250820AS | ST3160815A  
ST3160815AS  
ST3162015A  
ST3162015AS  
ST310615AS  
ST310615AS | ST380815AS  
ST380215A  
ST380215AS |
| **Interface Options**  | Ultra ATA/100  
SATA 3Gb/s NCQ  
SATA 1.5Gb/s NCQ | Ultra ATA/100  
SATA 3Gb/s NCQ  
SATA 1.5Gb/s NCQ | Ultra ATA/100  
SATA 3Gb/s NCQ  
SATA 1.5Gb/s NCQ | Ultra ATA/100  
SATA 3Gb/s NCQ  
SATA 1.5Gb/s NCQ | Ultra ATA/100  
SATA 3Gb/s NCQ  
SATA 1.5Gb/s NCQ | Ultra ATA/100  
SATA 3Gb/s NCQ  
SATA 1.5Gb/s NCQ |
| **Performance**        |         |         |         |         |         |         |       |
| **Transfer Rate, Max Ext (MB/s)** | 100/300 | 100/300 | 100/300 | 100/300 | 100/300 | 100/300 | 100/300 |
| **Cache (MB)**         | 16      | 16      | 16      | 16, 8   | 16, 8   | 8, 2    | 8, 2   |
| **Average Latency (msec)** | 4.16    | 4.16    | 4.16    | 4.16    | 4.16    | 4.16    | 4.16   |
| **Spindle Speed (RPM)** | 7200    | 7200    | 7200    | 7200    | 7200    | 7200    | 7200   |
| **Configuration/Organization** |         |         |         |         |         |         |       |
| **Heads/Disks²**       | 8/4     | 6/3     | 5/3     | 4/2     | 3/2     | 2/1     | 2/1    |
| **Bytes per Sector**   | 512     | 512     | 512     | 512     | 512     | 512     | 512    |
| **Reliability/Data integrity** |         |         |         |         |         |         |       |
| **Contact Start-Stops**| 50,000  | 50,000  | 50,000  | 50,000  | 50,000  | 50,000  | 50,000 |
| **Nonrecoverable Read Errors per Bits Read** | 1 per 10¹⁴  
1 per 10¹⁴ | 1 per 10¹⁴  
1 per 10¹⁴ | 1 per 10¹⁴  
1 per 10¹⁴ | 1 per 10¹⁴  
1 per 10¹⁴ | 1 per 10¹⁴  
1 per 10¹⁴ | 1 per 10¹⁴ |
| **Mean Time Between Failures (MTBF, hours)** | 700,000 | 700,000 | 700,000 | 700,000 | 700,000 | 700,000 | 700,000 |
| **Annualized Failure Rate (AFR)** | 0.34%   | 0.34%   | 0.34%   | 0.34%   | 0.34%   | 0.34%   | 0.34%  |
| **Limited Warranty (years)** | 5       | 5       | 5       | 5       | 5       | 5       | 5      |

700,000 hours ≈ 80 years

© Seagate Technology LLC. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [https://ocw.mit.edu/help/faq-fair-use.](https://ocw.mit.edu/help/faq-fair-use.)
dealing with disk failures
RAID 1 (mirroring)

😊 can recover from single-disk failure
😢 requires 2N disks
RAID 4 (dedicated parity disk)

sector i of the parity disk is the xor of sector i from all data disks

- can recover from single-disk failure
- requires N+1 disks (not 2N)
- performance benefits if you stripe a single file across multiple data disks
- all writes hit the parity disk
RAID 5 (spread out the parity)

- Can recover from single-disk failure
- Requires N+1 disks (not 2N)
- Performance benefits if you stripe a single file across multiple data disks
- Writes are spread across disks
• Systems have faults. We have to take them into account and build reliable, **fault-tolerant systems**. Reliability always comes at a cost — there are tradeoffs between reliability and monetary cost, reliability and simplicity, etc.

• Our main tool for improving reliability is **redundancy**. One form of redundancy is **replication**, which can be used to combat many things including disk failures (important, because disk failures mean lost data).

• **RAID** replicates data across disks in a smart way: RAID 5 protects against single-disk failures while maintaining good performance.