MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

Fault-tolerance

6.033 Lecture 14

Sam Madden

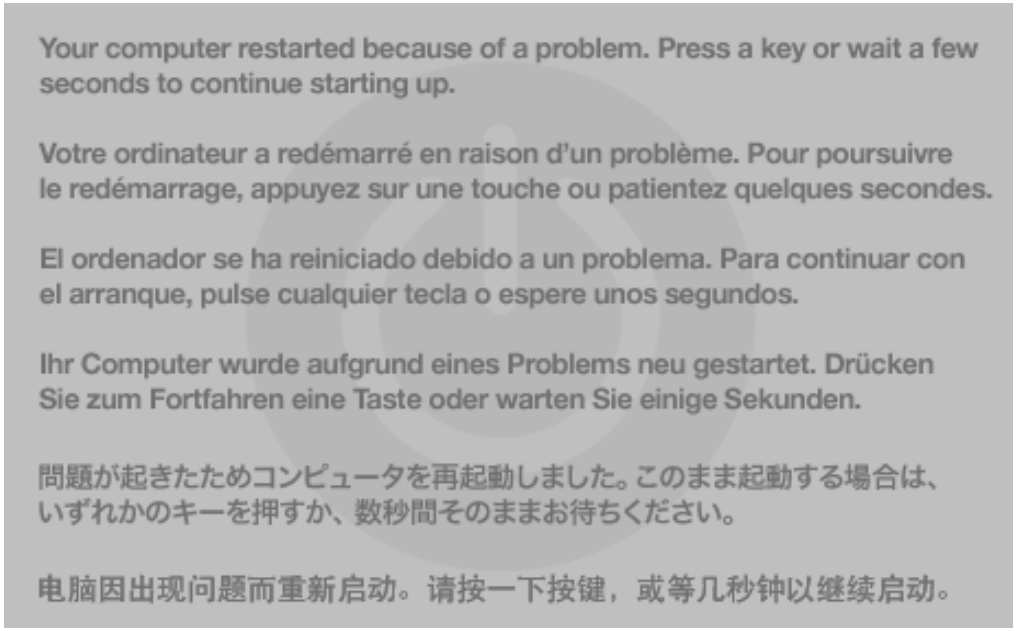
Road Map

- So far: Modularity & Operating Systems; Networked Systems

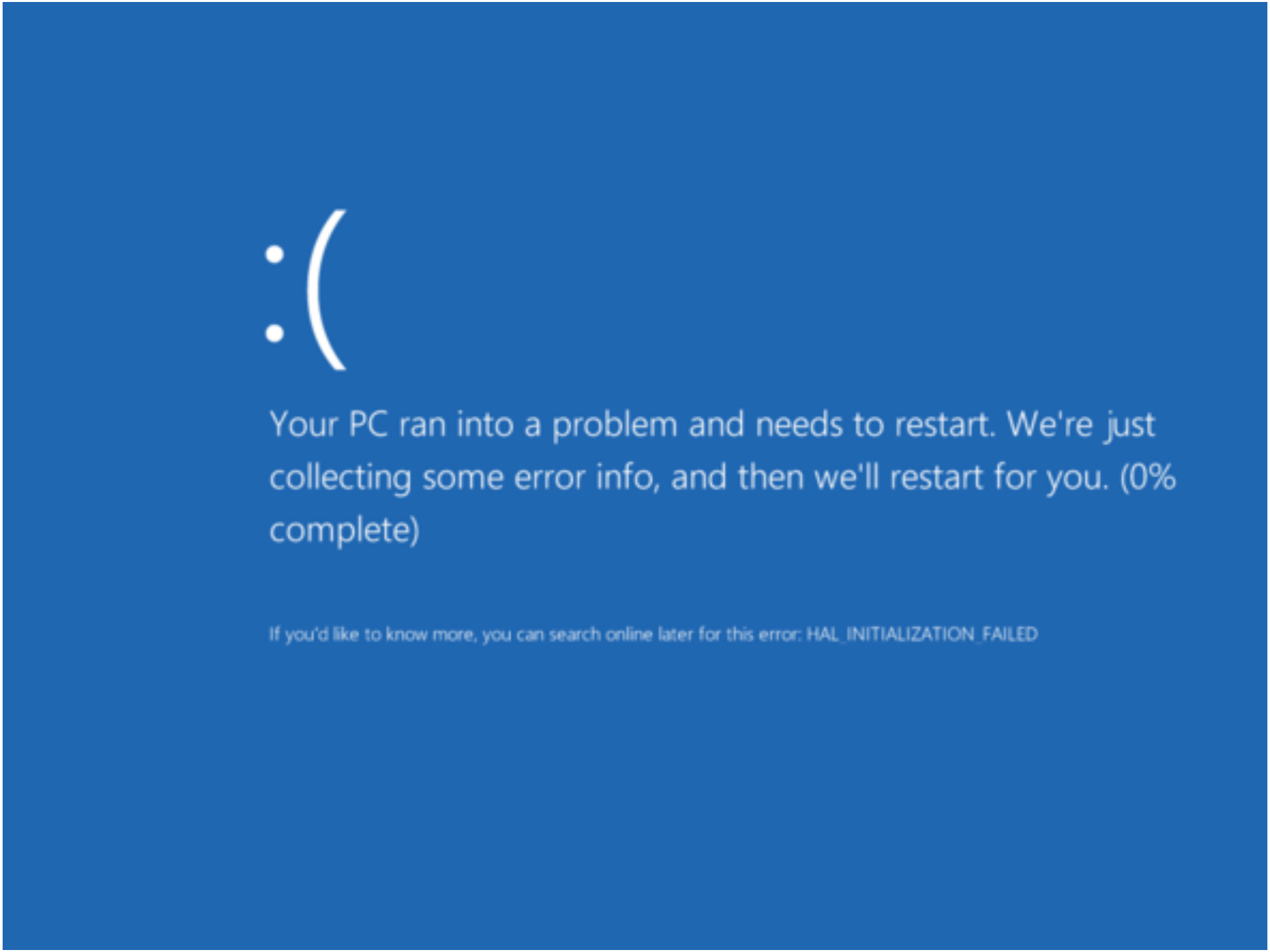
- Rest of this semester:
- - how to keep running despite failures
- - broaden class of failures to include malicious ones

Fault Tolerance Plan

- General introduction: today
  - Recovery/Replication
- Transactions: next 4 lectures
  - updating permanent data in the presence of concurrent actions and failures
- Replicated state machines: 2 more
  - Keep computing despite failures


Your computer restarted because of a problem. Press a key or wait a few seconds to continue starting up.
Votre ordinateur a redémarré en raison d'un problème. Pour poursuivre le redémarrage, appuyez sur une touche ou patientez quelques secondes.
El ordenador se ha reiniciado debido a un problema. Para continuar con el arranque, pulse cualquier tecla o espere unos segundos.
Ihr Computer wurde aufgrund eines Problems neu gestartet. Drücken Sie zum Fortfahren eine Taste oder warten Sie einige Sekunden.
問題が起きたためコンピュータを再起動しました。このまま起動する場合は、いずれかのキーを押すか、数秒間そのままお待ちください。
电脑因出现问题而重新启动。请按一下按键，或等几秒钟以继续启动。


Windows
A fatal exception 0E has occurred at 0028:C00068F8 in PPT.EXE<01> +
000059F8. The current application will be terminated.
* Press any key to terminate the application.
* Press CTRL+ALT+DEL 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. (0% complete)
If you'd like to know more, you can search online later for this error: HAL_INITIALIZATION_FAILED

San Francisco plane crash caused by pilot's inexperience with onboard computers

Boeing 777's auto-throttle did not maintain speed as expected

By Aaron Souppouris *on December 12, 2013 03:08 am* *Email* *@AaronIsSocial*

April 25, 1997 7:00 PM PDT

Router glitch cuts Net access

By CNET News.com Staff
Staff Writer, CNET News

Related Stories

Net blackout hits some regions

What started out as a router glitch at a small Internet service provider in Virginia today triggered a major outage in Internet access across the country, lasting more than two hours in some places.

Relative frequency of hardware replacement

COM1	
Component	%
Power supply	34.8
Memory	20.1
Hard drive	**18.1**
Case	11.4
Fan	8.0
CPU	2.0
SCSI Board	0.6
NIC Card	1.2
LV Power Board	0.6
CPU heatsink	0.6

10,000 machines

Pr(failure in 1 year) ~.3

Schroeder and Gibson, FAST 2008

Availability in practice

- Carrier airlines (2002 FAA fact book)
  - 41 accidents, 6.7M departures
  - ✓ 99.9993% availability
- 911 Phone service (1993 NRIC report)
  - 29 minutes per line per year
  - ✓ 99.994%
- Standard phone service (various sources)
  - 53+ minutes per line per year
  - ✓ 99.99+%
- End-to-end Internet Availability
  - ✓ 95% - 99.6%


A19

Data Sheet

Barracuda® 7200.10

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

80 GB to 750 GB • SATA 1.5Gb/s or 3Gb/s and PATA 100

Key Advantages

- First 3.5-inch drive to utilize capacity- and reliability-boosting perpendicular recording technology
- First drive to reach 750 GB—a full year ahead of competition—enabling new solutions for data-intensive applications.
- Industry's most proven and established desktop hard drive available today—more than 16 million shipped to date*
- "One-stop shopping" with a broad range of capacity, cache and interface options for all your computing needs
- Best-in-class environmental specifications and reliability features
- Adaptive Fly Height offers consistent read/write performance from the beginning to the end of your computing workload.
- Clean Sweep automatically calibrates your drive.
- Directed Offline Scan runs diagnostics when storage access is not needed.
- RoHS-compliant design assures an environmentally conscious product.
- Enhanced G-Force Protection™ defends against handling damage.
- Seagate® SoftSonic™ motor enables whisper-quiet operation.

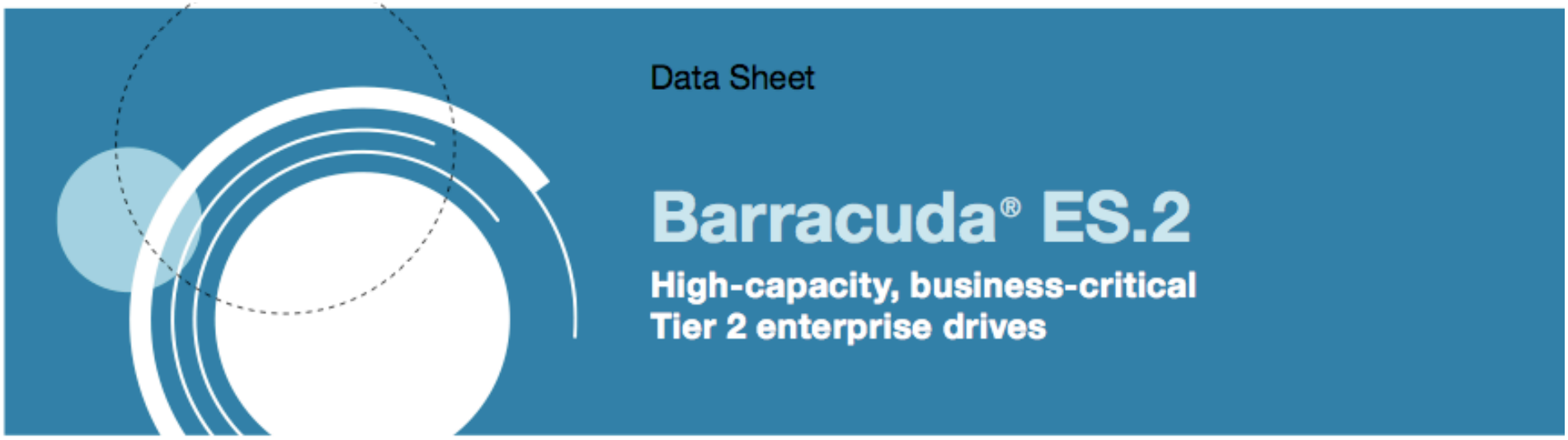
Best-Fit Applications

Desktop and High-Performance PCs

- Gamer PCs
- Workstations
- High-end PCs
- Desktop RAID
- Mainstream PCs
- Point-of-sale devices/ATMs
- USB/FireWire/eSATA personal external storage

*16 million Barracuda 7200.10 drives shipped as of 4/16/07

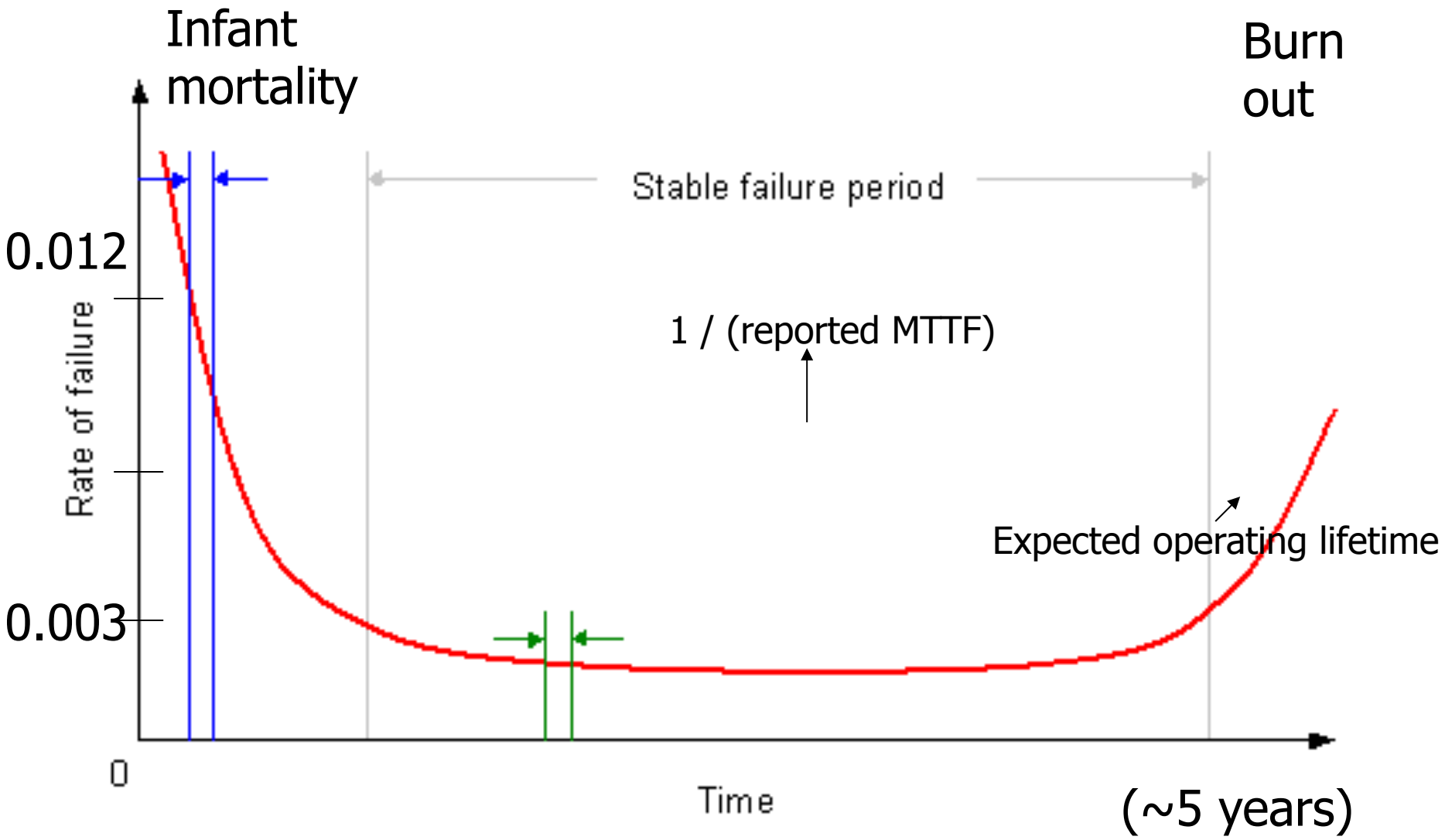
Contact Start-Stops	50,000
Nonrecoverable Read Errors per Bits Read	1 per 10^{14}
Mean Time Between Failures (MTBF, hours)	700,000
Annualized Failure Rate (AFR)	0.34%



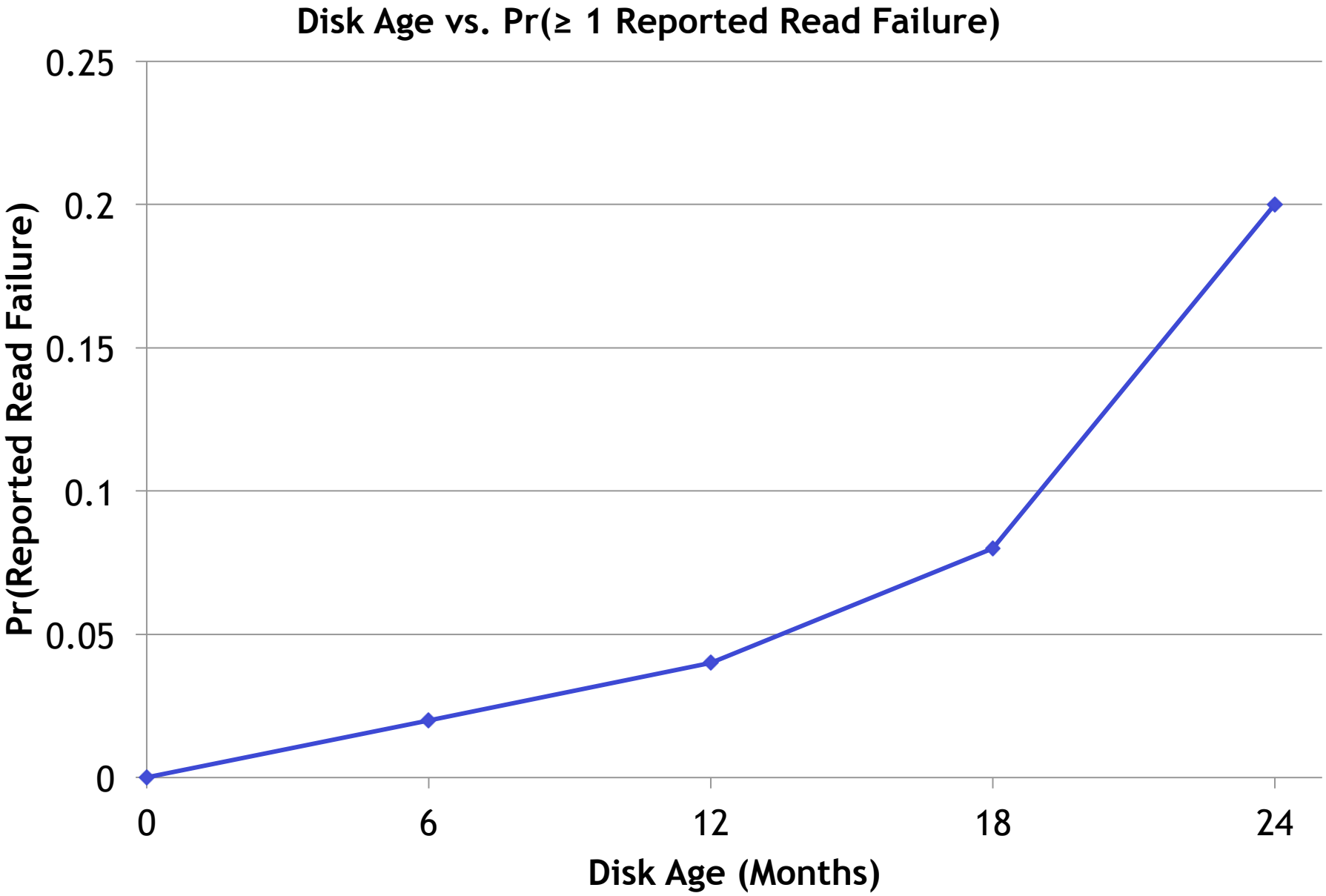
1 TB, 750 GB, 500 GB and 250 GB • 7200 RPM • SATA 3Gb/s, SATA 1.5Gb/s and SAS 3Gb/s

Reliability/Data Integrity	
Mean Time Between Failures (MTBF, hours)	1.2 million
Reliability Rating at Full 24x7 Operation (AFR)	0.73%
Nonrecoverable Read Errors per Bits Read	1 sector per 10E15
Error Control/Correction (ECC)	10 bit
Interface Ports	
SATA	Single
SAS	Dual

Disk failure conditional probability distribution



Bathtub curve



Bairavasundaram et al., SIGMETRICS 2007

Fail-fast disk

```
failfast_get (data, sector) {
        get (s, sector);
        if (checksum(s.data) = s.cksum) {
                data ← s.data;
                return OK;
        } else {
                return BAD;
        }
}
```

Careful disk

```
careful_get (data, sector) {
        r ← 0;
        while (r < 10) {
                r ← failfast_get (data, sector);
                if (r = OK) return OK;
                r++;
        }
        return BAD;
}
```

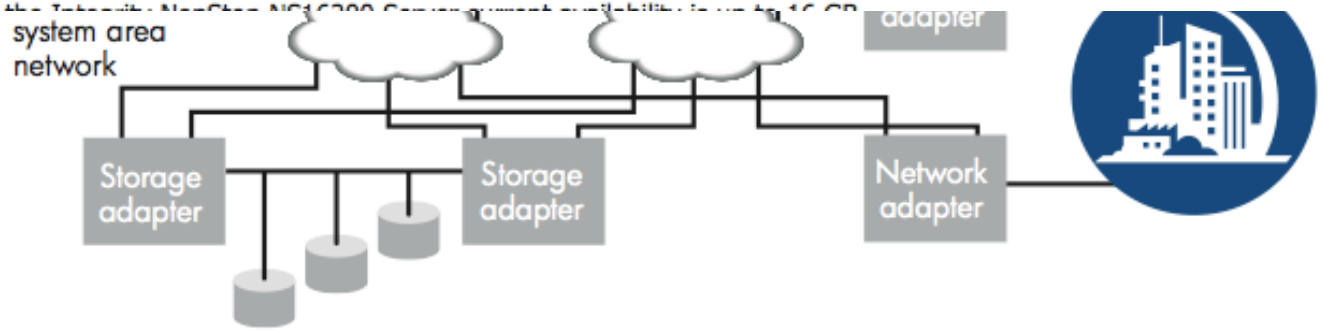
Replicated Disks

```
write (sector, data):
    write(disk1, sector, data)
    write(disk2, sector, data)

read (data, sector):
    data = careful_get(disk1, sector)
    if error
        data = careful_get(disk2, sector)
        if error
            return error
    return data
```

Technical specifications

Processors	2–16 per node Intel Itanium processor 9100 series processors, 1.6 GHz single core processors
Cache	12 MB L3
RAM standard/maximum	Minimum: 4 GB Maximum: 16 GB (32 GB[2])
RAM type/speed	PC2100 ECC registered DDR266A/B
ServerNet I/O	Minimum: 10 Maximum: 60
I/O adapters supported	Fibre Channel, Gigabit Ethernet
Fibre Channel disk modules	14 disks per module
Disk drives supported	146 GB and 300 GB15K RPM Fibre Channel internal hard disk drives HP Disk Array family (e.g., XP24000, XP20000, XP12000, and XP10000 disk arrays)
Standard features	N + 1 power supplies N + 1 fans



How about an error in software?

- Big problem!
- Software for fault tolerant systems must be written with great care
  - Stringent development practices
  - Well-defined stable specification
  - Modeling, simulation, verification, etc.
  - N-version programming is tricky
- Will also be a problem for secure software
- Good design: small fraction is critical