CS/ECE 5381/7381 Computer Architecture Spring 2023

Dr. Manikas

Computer Science

Lecture 22: Apr. 20, 2023

Exam 3

- Exam will be administered using Lockdown Browser (same as for previous exams)
- Exam format also same as previous exams
  - 25 questions, 2 hours
  - The exam will be available from **Thursday, Apr 27 at 12 am**
  - The exam must be completed and submitted by **Saturday, Apr 29 at 11:59 pm**

Exam 3

- **Exam 3 will cover the following materials:**
  - Modules: 9 - 12
  - Quizzes: 8 - 10
  - Text: Ch. 2.3 – 2.4, 4.1 – 4.2, 5.1 – 5.2, 6.1 – 6.5, 7.1 – 7.4
- **MATERIALS ALLOWED FOR EXAM:**
  - Open book and notes
  - Calculator

NOTE: EXTRA LECTURE THIS WEEK

- Due to the cancelled class of early February, we will have an extra “lecture” this week
- This will be a recorded lecture (no extra class scheduled)
- Recording and notes expected to be made available on Canvas by end of Friday (Apr. 21)
- Therefore, Quiz 10 due date will be extended

Assignments

- Quiz 10 – due **Mon, Apr. 24** (11:59 pm)
  - Covers concepts from Module 11
    - Including extra lecture of Friday

Quiz 10 Details

- The quiz is open book and open notes.
- You are allowed 90 minutes to take this quiz.
- You are allowed 2 attempts to take this quiz - your highest score will be kept.
  - Note that some questions (e.g., fill in the blank) will need to be graded manually
- Quiz answers will be made available 24 hours after the quiz due date.

Warehouse-Scale Computers

(Chapter 6, Hennessy and Patterson)

Note: some course slides adopted from publisher-provided material

Outline

- 6.1 Introduction
- 6.2 Programming Models and Workloads
- 6.3 Architecture
- 6.4 Physical Infrastructure
- 6.5 Cloud Computing

Computer Architecture of WSC

- **WSC often use a hierarchy of networks for interconnection**
- **Each 19” rack holds 48 servers connected to a rack switch**
- **Rack switches are uplinked to switch higher in hierarchy**
  - **Uplink has 6-24X times lower bandwidth**
  - **Goal is to maximize locality of communication relative to the rack**

Storage

- **Storage options:**
  - **Use disks inside the servers, or**
  - **Network attached storage**

  - **WSCs generally rely on local disks**
  - **Google File System (GFS) uses local disks and maintains at least three relicas**

WSC Memory Hierarchy

- **Servers can access DRAM and disks on other servers**

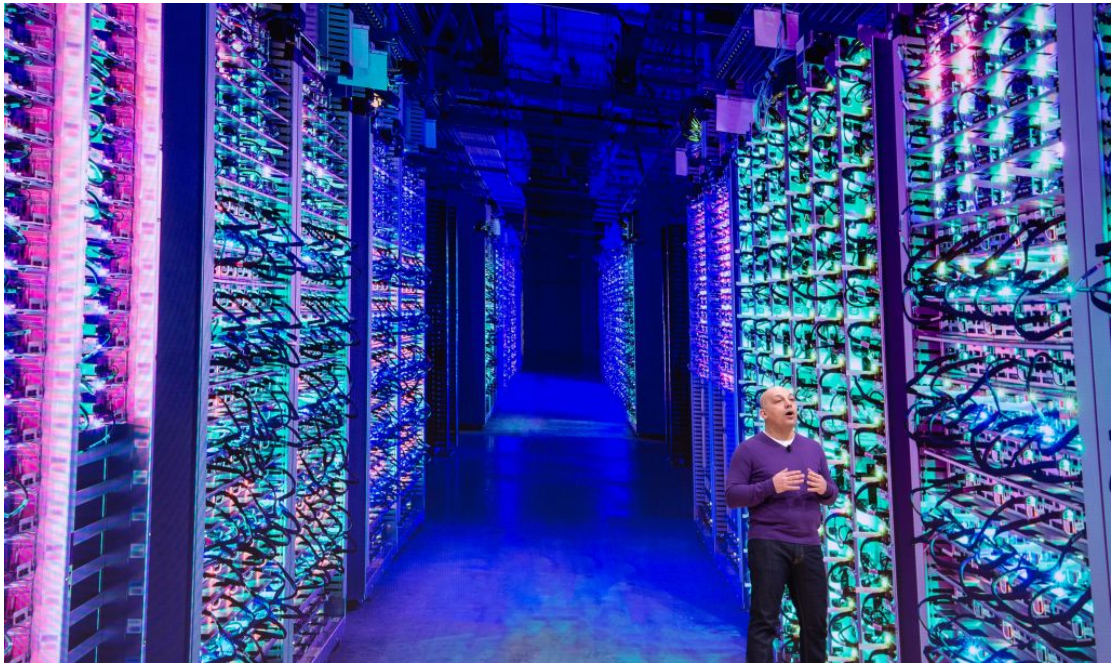
	Local	Rack	Array
DRAM latency (μs)	0.1	300	500
Flash latency (μs)	100	400	600
Disk latency (μs)	10,000	11,000	12,000
DRAM bandwidth (MB/s)	20,000	100	10
Flash bandwidth (MB/s)	1000	100	10
Disk bandwidth (MB/s)	200	100	10
DRAM capacity (GB)	16	1024	31,200
Flash capacity (GB)	128	20,000	600,000
Disk capacity (GB)	2000	160,000	4,800,000

Outline

- 6.1 Introduction
- 6.2 Programming Models and Workloads
- 6.3 Architecture
- 6.4 Physical Infrastructure
- 6.5 Cloud Computing

Warehouse for WSCs

- Building(s) to house computing and storage infrastructure in a variety of networked formats
  - Deliver the utilities needed by housed equipment and personnel, e.g., power, cooling, shelter, and security



Warehouse for WSCs

- Location considerations: proximity to ...
  - Internet backbone optical fibers, low cost electricity, low risk of environment disasters (earthquakes, floods, hurricanes, …), geographical vicinity of large population of users, real estate deals and low property taxes

Warehouse for WSCs

- Main design challenge and costs: delivery of input energy (power distribution) and removal of waste heat (cooling)

Measuring Performance of a WSC

- *Latency* is important metric from user's view
  - Time to find item "on web" (in servers)
- Users' *satisfaction* and *productivity* are tied to response time of a service
  - User productivity = 1 / time of interaction
    t(interaction) = t(human entry) + t(system response) + t(analysis of response)
  - Bing study: users use search ↓ as response time ↑
    - 200 ms longer delay → 500 ms longer time to next click
    - Revenue and user satisfaction drops linearly with increasing delay

Primary Concern: User Satisfaction

- Based on Internet studies...
  - Page load above few tens of milliseconds cause user to switch to another task
  - Page load time must be below 1s or it is deemed broken
    - Users do not come back

Primary Concern: User Satisfaction

- Quantifying influence of response delays
  - Use % requests below a latency threshold instead of avg.
  - SLO (*Server Level Objective*), SLA (*Server Level Agreement*)
    - Example: 99% of requests must be below 100 ms delay

Reliability and Availability

- Reliability (MTTF) & Availability (MTTF/MTTF+MTTR) are very important, given the large scale
  - MTTF: mean time to failure
  - MTTR: mean time to repair
  - A server with MTTF of 25 years ➔ 50K servers would lead to 5 server failures a day
  - Annual disk failure rate of 2-10% ➔ 1 disk failure per hour

Outline

- 6.1 Introduction
- 6.2 Programming Models and Workloads
- 6.3 Architecture
- 6.4 Physical Infrastructure
- 6.5 Cloud Computing

What is Cloud Computing?

- Internet-based computing
  - a collection/group of integrated and networked hardware, software and Internet infrastructure (called a platform).
  - Using the Internet for communication and transport provides hardware, software and networking services to clients

Cloud Computing Characteristics

- **Remotely hosted**: Services or data are hosted on remote infrastructure.
- **Ubiquitous**: Services or data are available from anywhere.
- **Commodified**: The result is a utility computing model similar to traditional that of traditional utilities, like gas and electricity - you pay for what you would want!

Basic Cloud Characteristics

- Cloud are transparent to users and applications, they can be built in multiple ways
  - branded products, proprietary open source, hardware or software, or just off-the-shelf PCs.
- In general, they are built on clusters of PC servers and off-the-shelf components plus Open Source software combined with in-house applications and/or system software.

Cloud Computing

- Software as a Service (SaaS)
  - Instead of installing software on user's PC, software is run remotely on WSC ("the Cloud")
- SaaS customers are charged based on software use, not ownership

Cloud Computing Layers

Application Service (SaaS)	MS Live/ExchangeLabs, IBM, Google Apps; Salesforce.com Quicken Online, Zoho, Cisco
Application Platform	Google App Engine, Mosso, Force.com, Engine Yard, Facebook, Heroku, AWS
Server Platform	3Tera, EC2, SliceHost, GoGrid, RightScale, Linode
Storage Platform	Amazon S3, Dell, Apple, …

WSC - Redundancy

- Warehouse Scale Computers require high reliability
  - Storage reliability critical – can't afford to lose data
- Fault-tolerance – redundant systems
  - Storage – use RAID

RAID

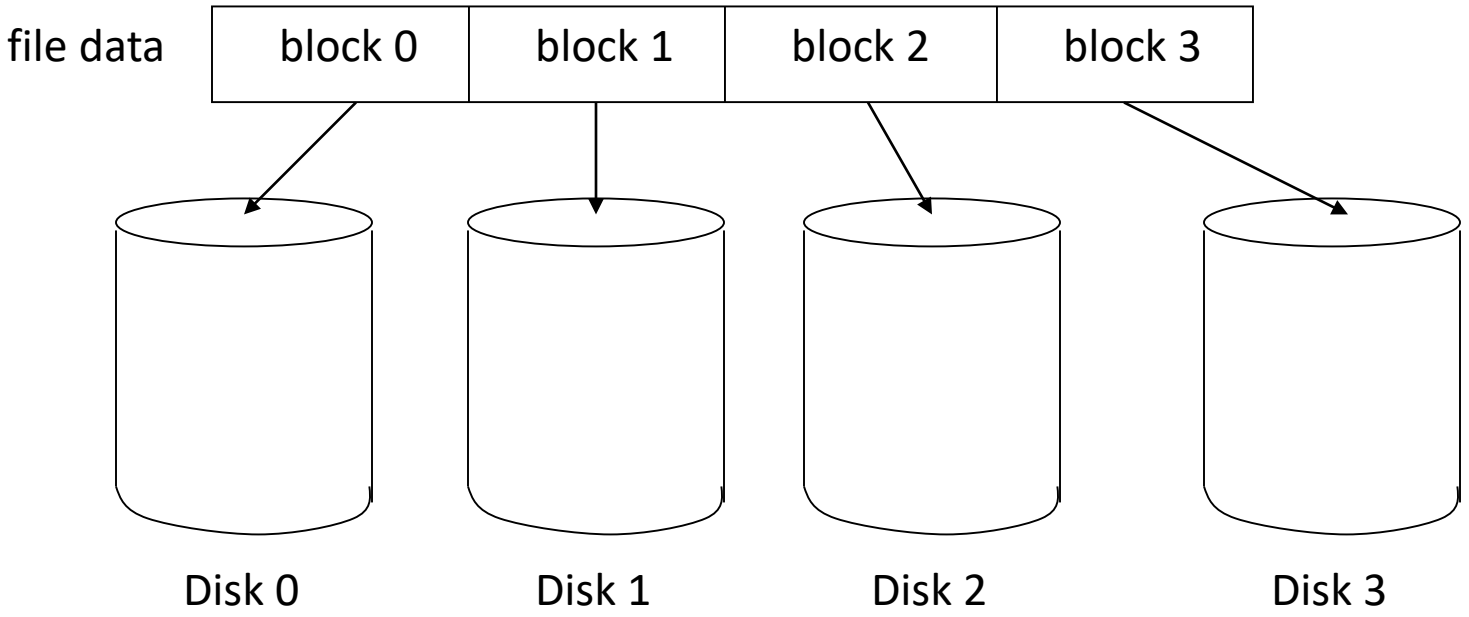
- Redundant Array of Independent Disks
  - Multiple drives, single host disk unit
- Basic idea is to connect multiple disks together to provide
  - large storage capacity
  - faster access to reading data
  - redundant data
- Many different levels of RAID systems
  - differing levels of redundancy, error checking, capacity, and cost

RAID Level 0

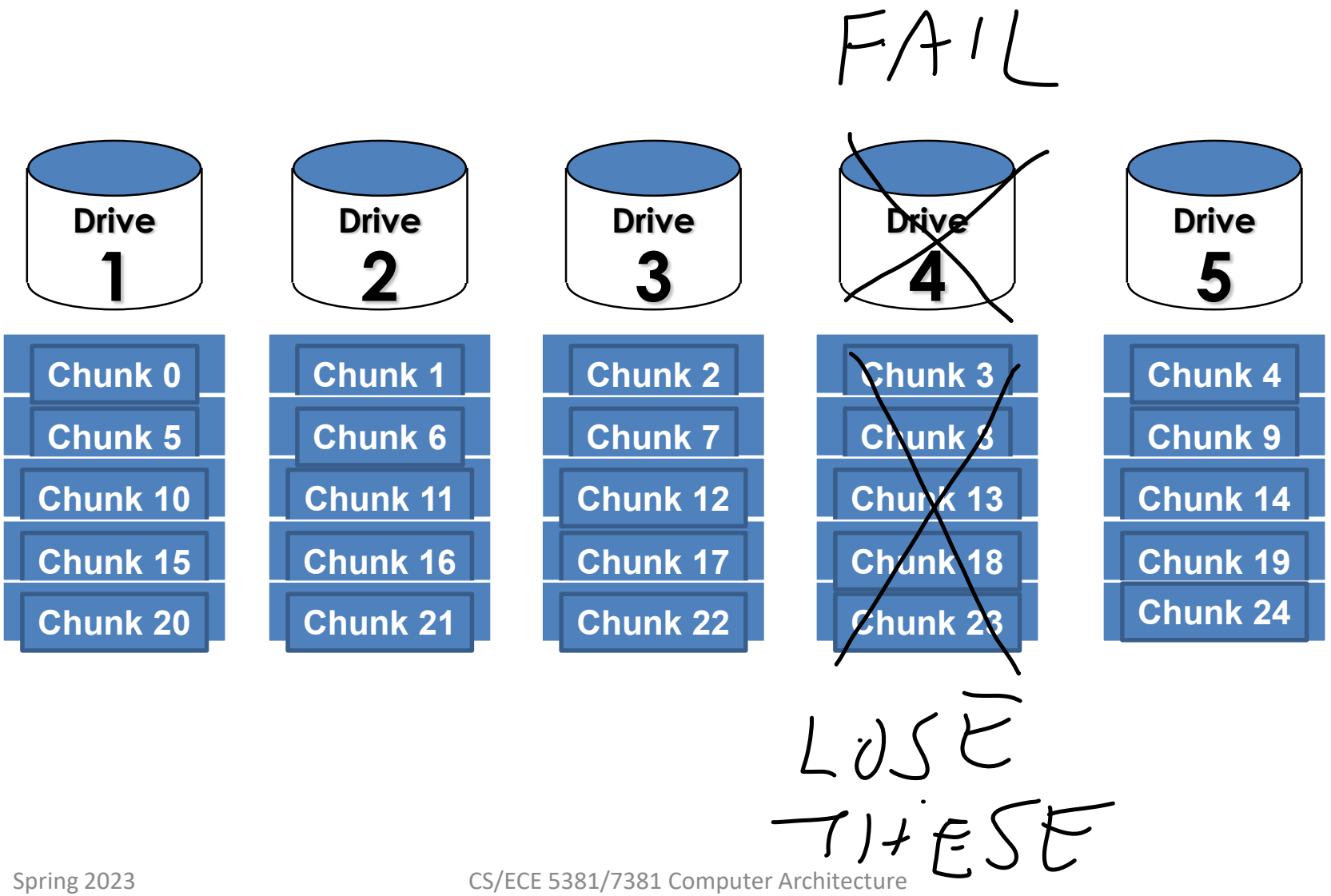
- Often called striping
- Break a file into blocks of data
- Stripe the blocks across disks in the system
- Simple to implement
  - disk = file block % number of disks
  - sector = file block / number of disks
- provides no redundancy or error detection
  - important to consider because lots of disks means low Mean Time To Failure (MTTF)

Striping

- Take file data and map it to different disks
- Allows for reading data in parallel



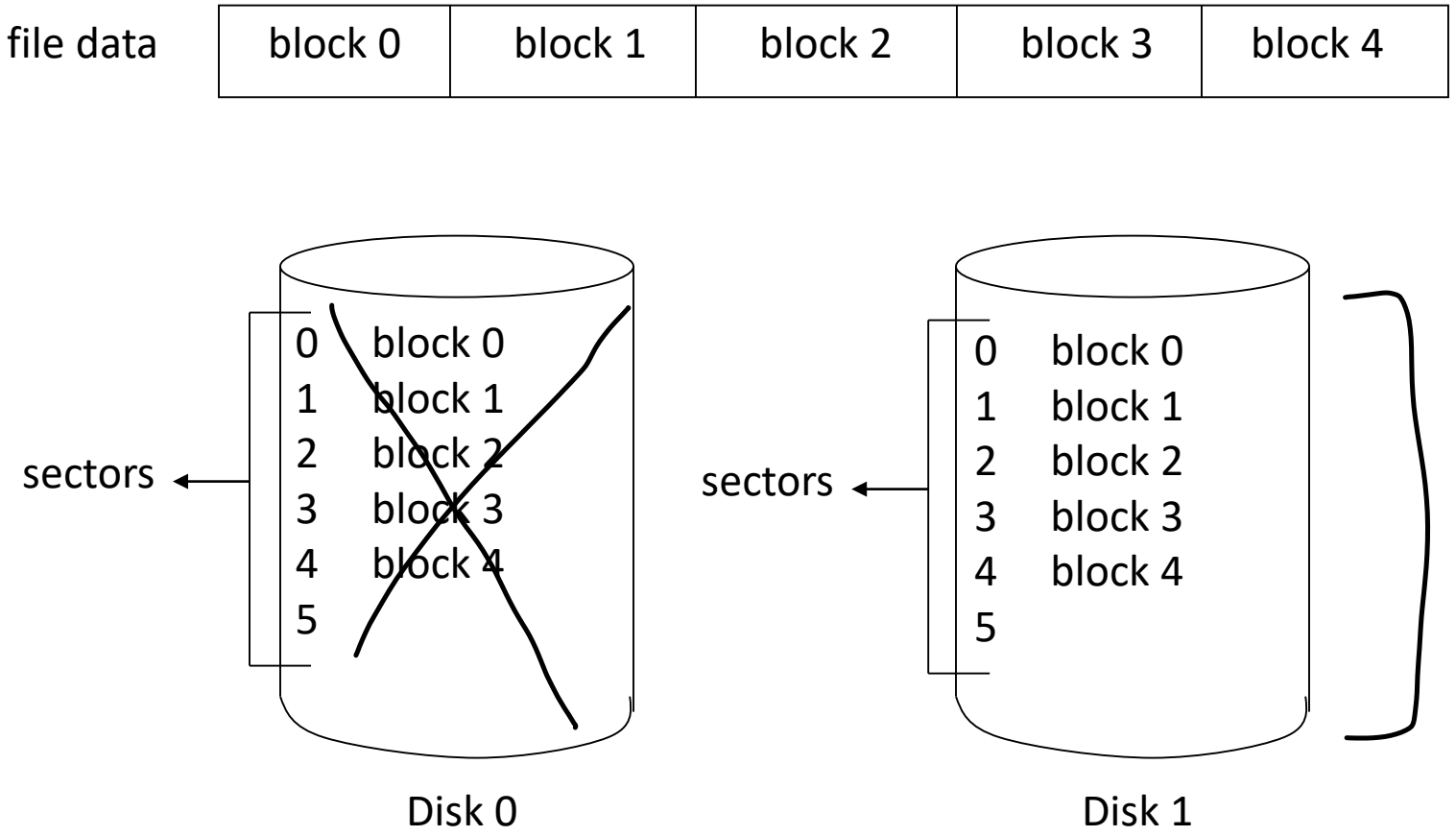
RAID 0 Striping



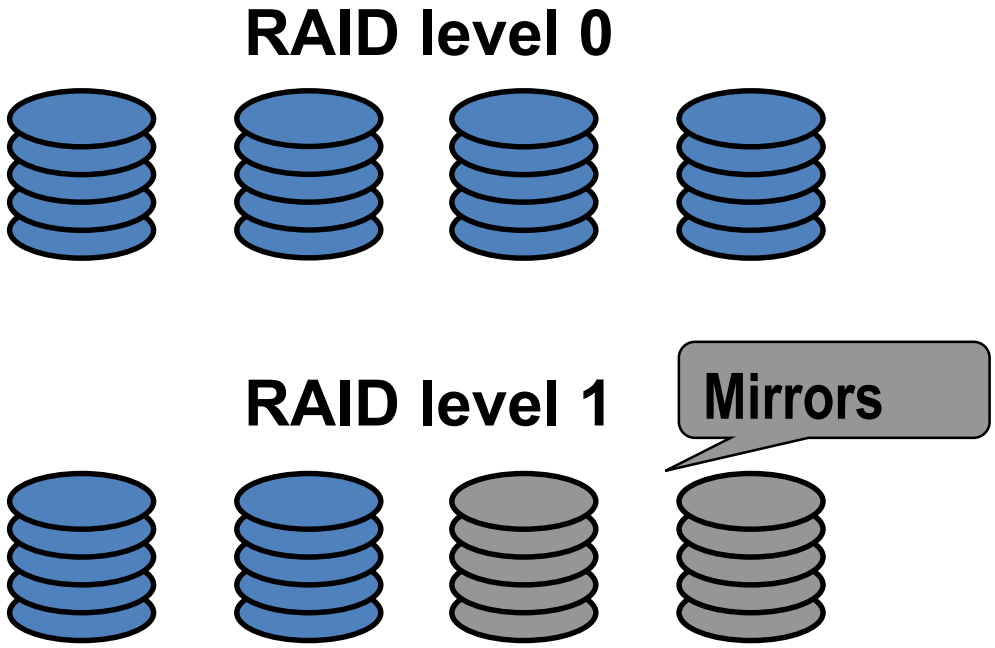
RAID Level 1

- A complete file is stored on a single disk
- A second disk contains an exact copy of the file
  - Mirroring
- Provides complete redundancy of data
- Read performance can be improved
  - file data can be read in parallel
- Write performance suffers
  - must write the data out twice
- Most expensive RAID implementation
  - requires twice as much storage space

Mirroring


file data
block 0
block 1
block 2
block 3
block 4
sectors
0
1
2
3
4
5
block 0
block 1
block 2
block 3
block 4
Disk 0
sectors
0
1
2
3
4
5
block 0
block 1
block 2
block 3
block 4
Disk 1
FAILS
COMPLETE COPY

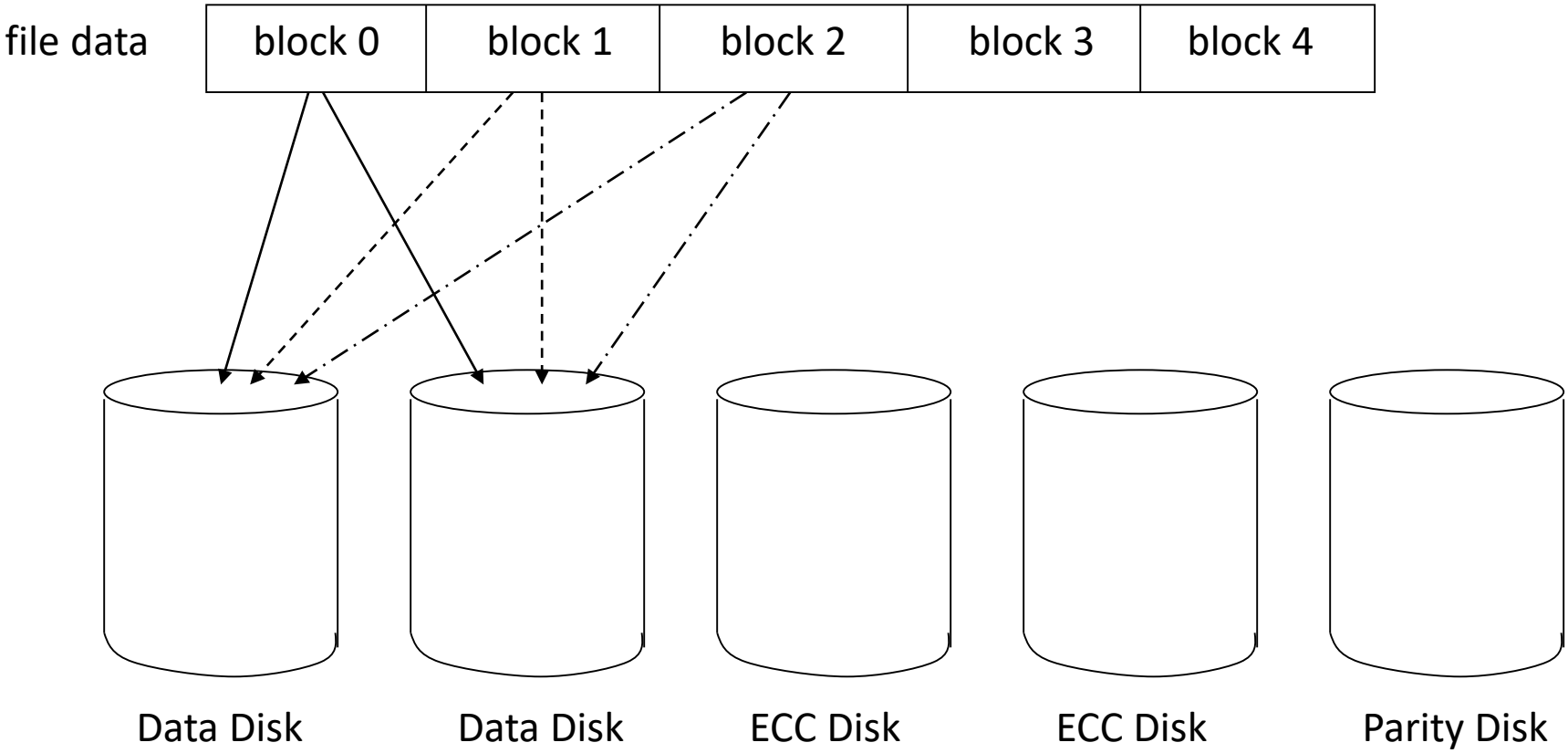
RAID levels 0 and 1



RAID Level 2

- Stripes data across disks similar to Level-0
  - difference is data is bit interleaved instead of block interleaved
- Uses ECC to monitor correctness of information on disk
  - Recall ECC: Error Correction Code
- Multiple disks record the ECC information to determine which disk is in fault
- A parity disk is then used to reconstruct corrupted or lost data

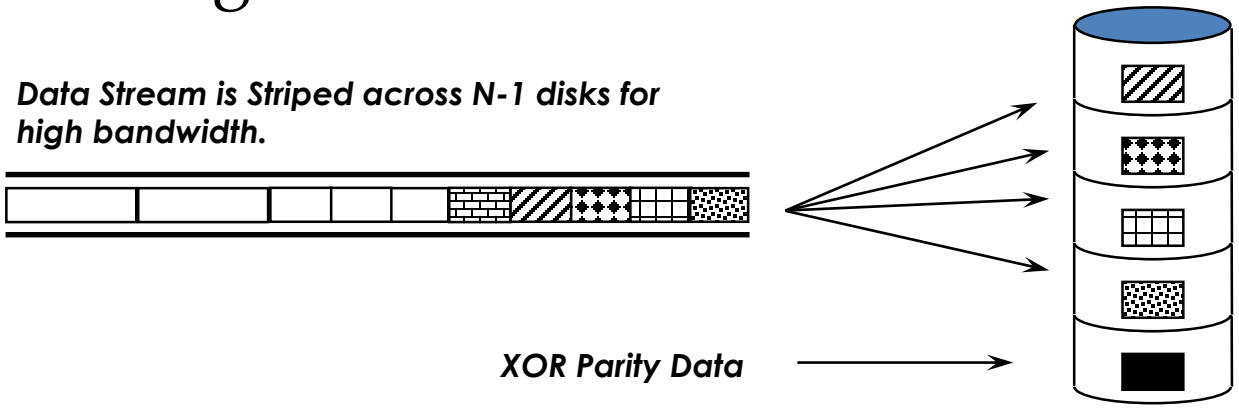
RAID 2


file data
block 0
block 1
block 2
block 3
block 4
Data Disk
Data Disk
ECC Disk
ECC Disk
Parity Disk

RAID 3

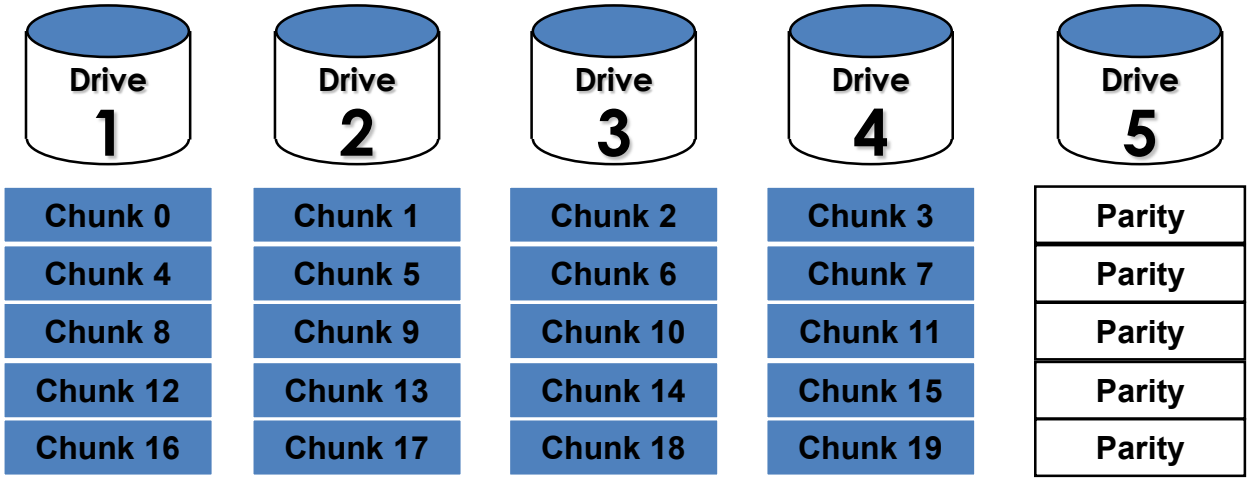
Disk Striping with dedicated parity drive

- High BW Performance; Cheap Availability
- Sector-granular data striping
- Single-threaded Access

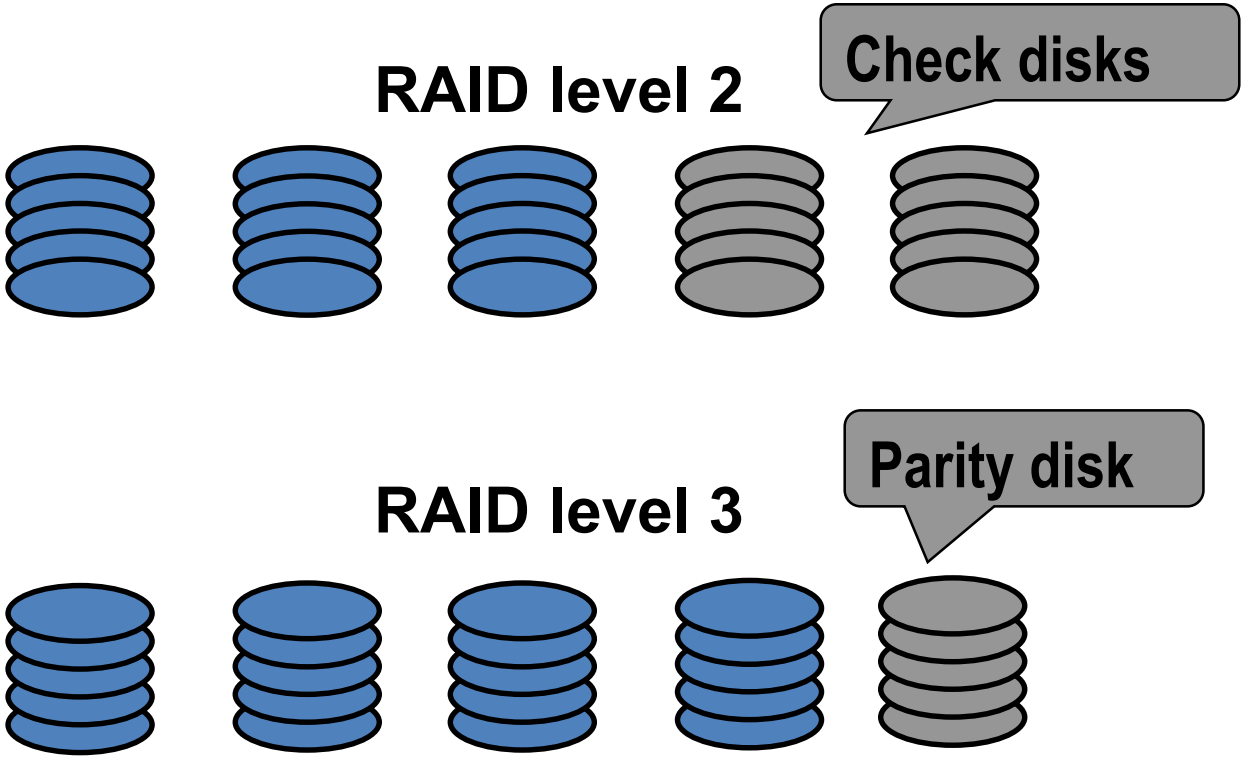


RAID 3 Striping

- ***Chunk size = single sector (pure RAID 3 would be single byte)***
- ***All parity data on same spindle***



RAID levels 2 and 3



How parity works?

- Truth table for XOR (same as parity)

A	B	A⊕B
0	0	0
0	1	1
1	0	1
1	1	0

Recovering from a disk failure

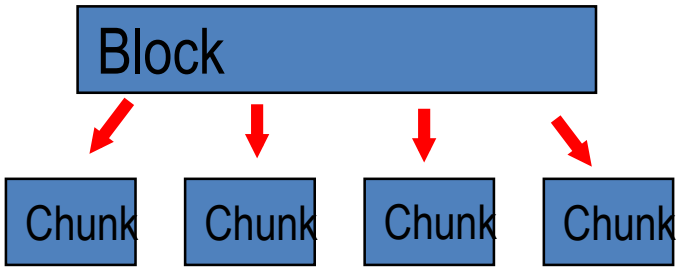
- Small RAID level 3 array with data disks D0 and D1 and parity disk P can tolerate failure of either D0 or D1

D0	D1	P
0	0	0
0	1	1
1	0	1
1	1	0

D1⊕P=D0	D0⊕P=D1
0	0
0	1
1	0
1	1

How RAID level 3 works (I)

- Assume we have $N + 1$ disks
- Each block is partitioned into N equal chunks

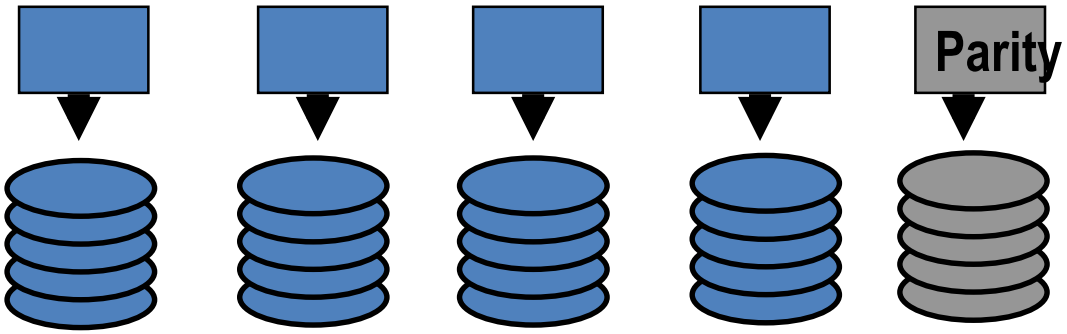


$N = 4$ in example

How RAID level 3 works (II)

- XOR data chunks to compute the parity chunk

- Each chunk is written into a ***separate disk***



Domain-Specific Architectures

(Chapter 7, Hennessy and Patterson)

Note: some course slides adopted from publisher-provided material