Fault Tolerance and Availability:

Building resilient systems that keep your business running when failures occur

What are Fault Tolerance and Availability?

Fault Tolerance

The ability of a system to continue operating correctly despite the presence of hardware, software, or network faults. It's about graceful degradation rather than complete failure.

Availability

The percentage of time a system is operational and accessible to users. Measured as uptime divided by total time, typically expressed in "nines" of reliability.

Why are They Critical for Modern Business?

Minimize Downtime Impact

Prevent costly service interruptions and protect against data loss that can damage reputation and revenue streams.

Maintain Business Continuity

Ensure critical operations continue seamlessly even during system failures, keeping customers satisfied and business flowing.

Enhance User Experience

Provide consistent, reliable service that builds trust and maintains customer loyalty in competitive markets.

Understanding Different Types of System Faults

Hardware Faults

Physical component failures including disk crashes, memory errors, server malfunctions, and power supply issues that can bring systems offline instantly.

Software Faults

Programming errors, configuration mistakes, corrupted data, and design flaws that cause applications to crash or behave unpredictably.

Network Faults

Communication disruptions including packet loss, bandwidth congestion, router failures, and connectivity issues that isolate system components.

Strategies for Achieving High Availability

01

Implement Redundancy

Deploy backup systems, duplicate components, and multiple data centers to ensure continuity when primary systems fail.

02

Eliminate Single Points of Failure

Identify and redesign critical components that could bring down the entire system if they fail.

03

Choose Deployment Modes

Active-active systems run simultaneously for load sharing, while active-passive keeps standby systems ready for failover.

Essential Fault Tolerance Techniques

1

Data Replication

Create multiple synchronized copies of critical data across different locations and storage systems to prevent data loss during failures.

- Real-time synchronization
- Geographic distribution
- Automatic failover

2

System Checkpointing

Periodically save complete system states to enable rapid recovery from the last known good configuration when errors occur.

- Scheduled snapshots
- Incremental backups
- Recovery point objectives

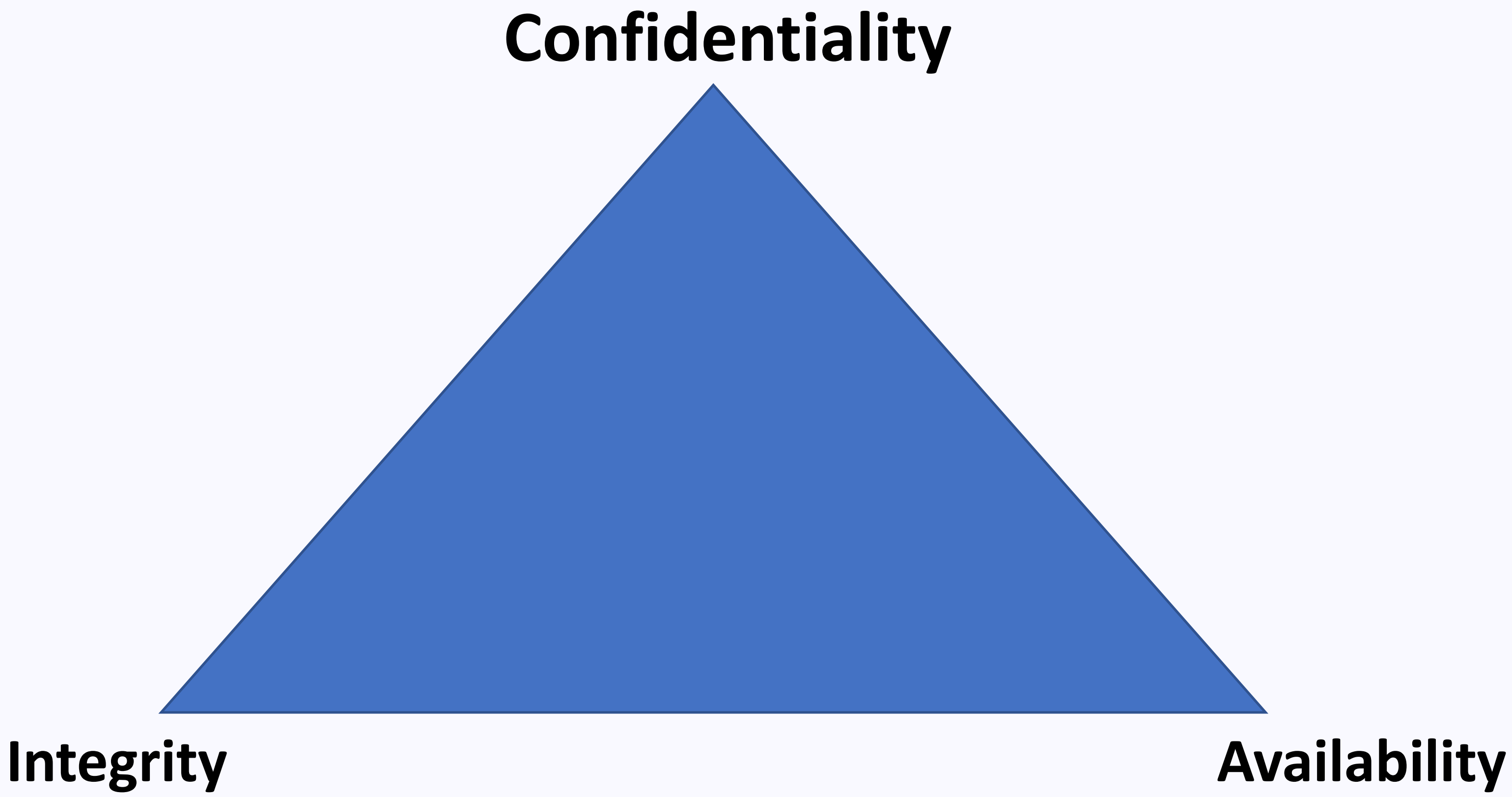
3

Error Detection and Correction

Implement monitoring systems and automated correction mechanisms to identify and resolve issues before they impact users.

- Proactive monitoring
- Self-healing systems
- Automated remediation

Cybersecurity architecture is the structured design of security controls, technologies, processes, and policies that protect an organization's IT infrastructure, data, and applications from cyber threats. Think of it as a blueprint for securing digital systems, much like how building architecture defines how to make a structure safe and functional.



The CIA Triad

1. **Confidentiality – *keeping information secret***
    1. **Ensures only authorized people/systems can access data.**
    2. **Achieved by:**
        1. **Encryption**
        2. **Access controls (IAM, RBAC, MFA)**
        3. **Network segmentation**
    3. **Example: Credit card details in an e-commerce database should only be seen by authorized systems, not everyone.**

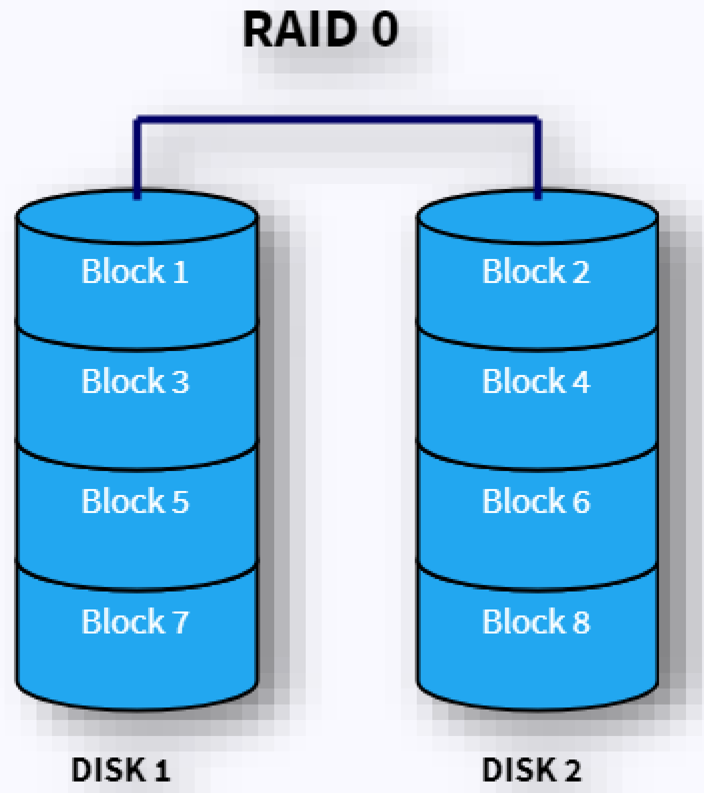
2. **Integrity – *keeping information accurate and unaltered***
    2. **Protects data from being modified, corrupted, or tampered with.**
    3. **Achieved by:**
        2. **Hashing (e.g., SHA-256)**
        3. **Digital signatures**
        4. **Checksums**
        5. **Database integrity controls**
    4. **Example: A financial transaction must not be changed during transmission.**

3. **Availability – *keeping information and systems accessible when needed***
    3. **Ensures systems and data are usable by authorized users at the right time.**
    4. **Achieved by:**
        3. **Redundancy (backup servers, cloud failover)**
        4. **Load balancing**
        5. **DDoS protection**
        6. **Regular system maintenance**
    5. **Example: An online banking app must be available 24/7; downtime is a failure of availability.**

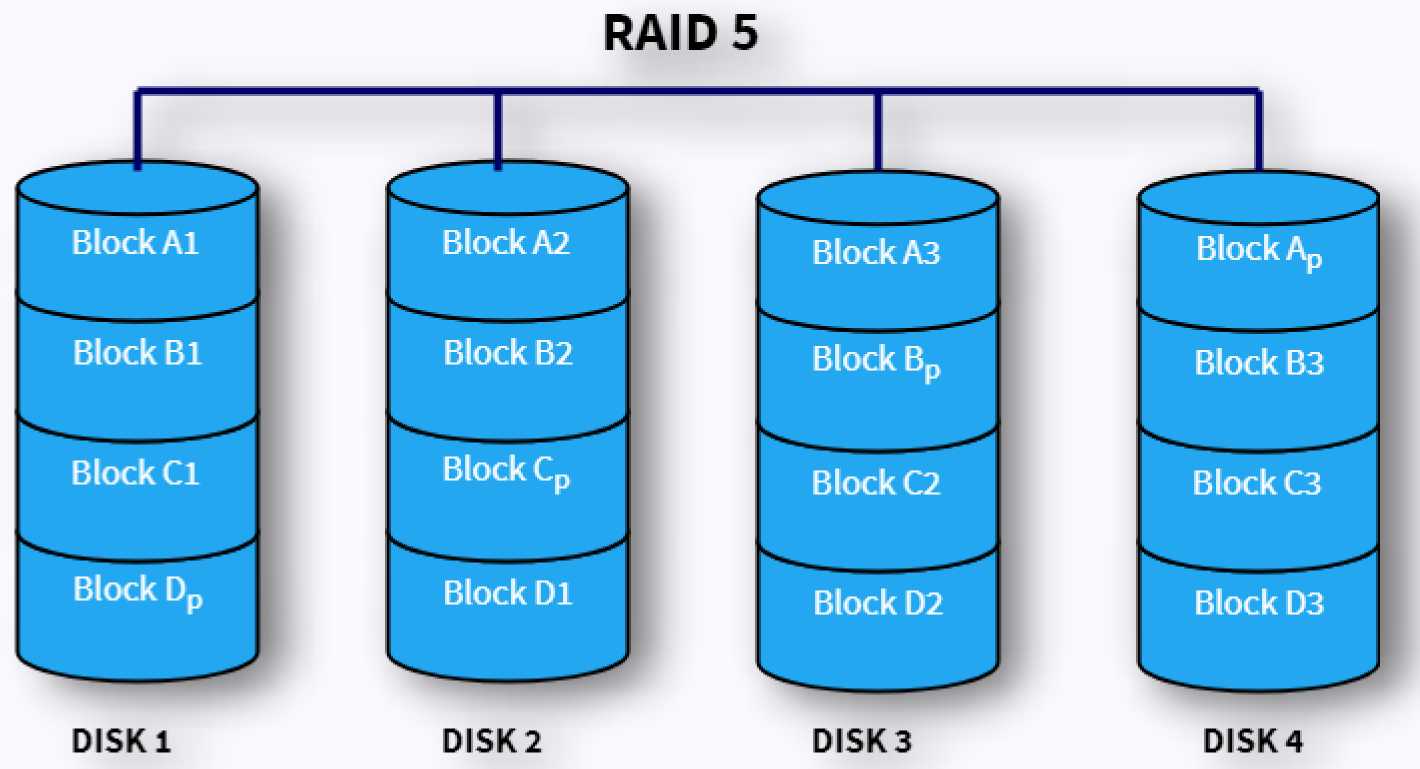
RAID (Redundant Array of Independent Disks) is a storage technology that combines multiple physical hard drives or solid-state drives (SSDs) into one logical unit to improve data performance, protection, or both

- ❑ **RAID 0**

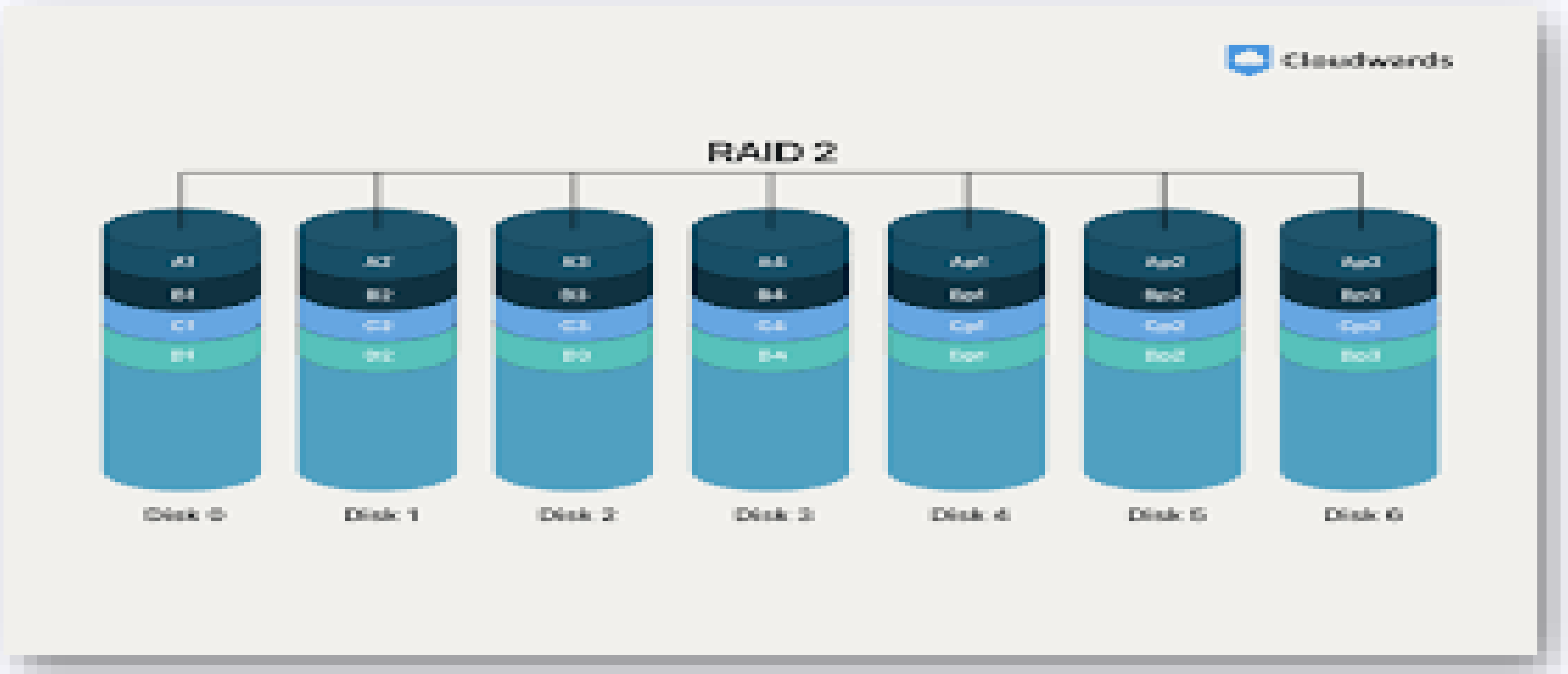
RAID 0 is based on data striping. A stream of data is divided into multiple segments or blocks and each of those blocks is stored on different disks. So, when the system wants to read that data, it can do so simultaneously from all the disks and join them together to reconstruct the entire data stream. The benefit of this is that the speed increases drastically for read and write operations. It is great for situations where performance is a priority over other aspects. Also, the total capacity of the entire volume is the sum of the capacities of the individual disks. The downside, as you may have already guessed it is that there is almost no redundancy. If one of the disks fails, the entire data becomes corrupt and worthless since it cannot be recreated anymore.



- **RAID 5 is very similar to RAID 4, but here the parity information is distributed over all the disks instead of storing them in a dedicated disk. This has two benefits — First, there is no more a bottleneck as the parity stress evens out by using all the disks to store parity information and second, there is no possibility of losing data redundancy since one disk does not store all the parity information.**



- **RAID 2** is a storage technology that uses **bit-level striping with error correction based on Hamming codes**. In this setup, data is divided at the bit level and spread across multiple hard drives, while additional drives are dedicated to storing error-correcting code. When data is read or written, the system calculates and stores error correction information, which can later be used to detect and correct single-bit errors or reconstruct data if a drive fails. This design ensures a high level of **data integrity and fault tolerance**, contributing to system reliability and availability. However, RAID 2 is rarely used in modern systems because it requires a large number of disks and is less efficient compared to more practical RAID levels such as RAID 5 or RAID 10


Cloudwards
RAID 2
A1
B1
C1
D1
A2
B2
C2
D2
A3
B3
C3
D3
A4
B4
C4
D4
Ap1
Bp1
Cp1
Dp1
Ap2
Bp2
Cp2
Dp2
Ap3
Bp3
Cp3
Dp3
Disk 0
Disk 1
Disk 2
Disk 3
Disk 4
Disk 5
Disk 6