Department of Information Technology A.Y. 2024-2025

Class: TY Btech. IT **Semester: VI**

Subject: Big Data Lab **Batch: I1-1**

Name: Dipti Agarwal **SAP: 60003220189**

Experiment – 2: Study and analyze RAID for any real world problem statements

Aim: To study and analyze RAID for any real world problem statements.

Requirements: PC and Internet

Brief Theory:

- Overview of RAID.
- RAID Levels.
- Comparison of RAID levels.
- Give case studies where these RAID levels can be applied for each level.

For Example:

Case Study: University Data Center

A large university runs a data center to host virtual machines (VMs) for faculty and students.

Current Situation/Issues:

- The VM storage is based on a RAID 0 array with 8 disks, each 2TB capacity. ○ A single disk failure has previously resulted in loss of student projects and research data.
- Performance is not an issue, but data redundancy is critical.
- The university cannot afford new disks, but restructuring is possible.

Task:

- Recommend a RAID level that prevents data loss while using the existing hardware.
- Explain how the new configuration balances performance and redundancy.
- Suggest additional low-cost backup strategies to protect student work.

Solution:
Recommended RAID Level: RAID 10 (1+0)

Justification:
Since data redundancy is critical, and no new disks can be purchased, we must reconfigure the existing 8-disk RAID 0 setup into a RAID 10 array.

Observation:

Case Study 1: Hospital Data Storage

- **Current Situation/Issues**:
  - The hospital currently uses a **RAID 0** array to store patient records, medical images, and critical data.
  - **Data loss risk** is high because RAID 0 provides no redundancy.
  - A failure of a single disk could result in the **loss of critical patient data**, potentially affecting patient care and violating legal regulations (e.g., HIPAA).
  - There is a need to **balance performance** and **data protection** without significant cost.
- **Task**:
  - Recommend a RAID level that offers **redundancy** while maintaining **performance** and **cost efficiency**.
  - Ensure that the solution meets the hospital's data protection and availability needs.
- **Solution**:
  - **Recommended RAID Level**: **RAID 5**
- **Justification**:
  - **RAID 5** offers **single-disk redundancy** through parity, ensuring that if one disk fails, the data can still be reconstructed from the remaining disks.
  - It balances **performance**, **storage efficiency**, and **data protection**—the storage space required for redundancy is minimal (only the equivalent of one disk).
  - RAID 5 is a cost-effective solution for providing **data protection** without the need for doubling the storage capacity (as in RAID 1), which makes it ideal for the hospital's needs.

Case Study 2: Video Editing Studio

- **Current Situation/Issues**:
  - The video editing studio uses a **RAID 1** configuration for storing video files and projects.
  - While **RAID 1** provides **data redundancy**, it lacks the **performance** required for handling large video files and rendering high-resolution videos.
  - The system is **sluggish** during large file processing, which is hindering productivity.
  - The studio needs a solution that provides both **fast read/write speeds** and **data redundancy** for its editing projects.
- **Task**:
  - Recommend a RAID level that enhances **performance** while still providing **data redundancy** to protect critical video files.
- **Solution**:
  - **Recommended RAID Level**: **RAID 10 (1+0)**
- **Justification**:
  - **RAID 10** combines **RAID 0 (striping)** for **performance** and **RAID 1 (mirroring)** for **redundancy**.
  - It offers **high read/write speeds** due to the striping, making it ideal for high-performance applications like video editing.
  - The mirroring ensures **data protection**, allowing the studio to maintain **redundancy** without compromising on speed or capacity.
  - RAID 10 is well-suited for environments with high-volume data processing needs, such as video rendering.

Case Study 3: E-commerce Website

- **Current Situation/Issues**:
  - The e-commerce website uses a **RAID 5** array for its web hosting, handling critical product data, customer orders, and transaction records.
  - While RAID 5 offers good **data protection** and reasonable **performance**, the website experiences **slower response times** during high-traffic sales events.
  - There is a concern about **potential downtime** or **data loss** if two disks fail simultaneously, especially during peak traffic periods.
- **Task**:
  - Recommend a RAID configuration that ensures **data protection**, **reliability**, and **better performance** during periods of high web traffic.
- **Solution**:
  - **Recommended RAID Level**: **RAID 6**
- **Justification**:
  - **RAID 6** provides **dual parity**, allowing the array to tolerate the failure of **two disks** simultaneously without data loss.
  - This makes it more **robust** than RAID 5, offering greater protection during **high-traffic periods**, when the likelihood of multiple disk failures increases due to higher workload.
  - RAID 6 is suitable for environments like e-commerce websites that need **high availability**, **data integrity**, and **reliability** while minimizing downtime, especially during peak traffic events like sales or promotions.

Conclusion:

This experiment analyzed different RAID levels and their applications in real-world scenarios, focusing on balancing performance, redundancy, and cost. By reviewing case studies from a hospital, video editing studio, and e-commerce website, we identified the most suitable RAID configurations for each situation—RAID 5 for data protection in a hospital, RAID 10 for performance and redundancy in a video studio, and RAID 6 for high availability in an e-commerce environment. The experiment highlights how RAID can be optimized to meet specific data storage and performance needs.