VIET NAM NATIONAL UNIVERSITY HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER AND ENGINEERING

Blockchain Storage Optimization

ASSISTOR: Nguyễn Thành Công
STUDENTS: Nguyễn Hữu Thiện - 2353133

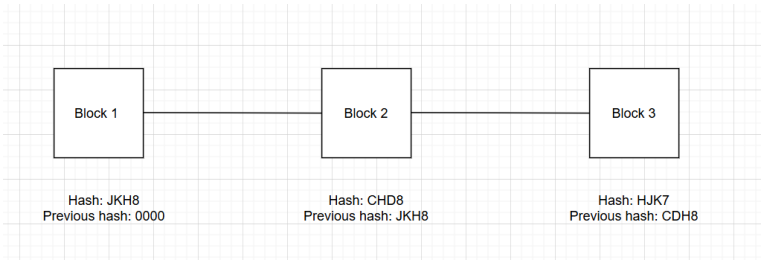
HO CHI MINH CITY, SEPTEMBER,2025

Contents

1 Understanding

Blockchain: Blockchain is a decentralized system that data is stored in each block. Each block is linked together securely, they are called chain. The first block is called *genesis block*
A block will include:

- Data is stored in each block
- Block's hash: Hash code of each block
- Previous block's hash: Store hash code of the previous block



Erasure coding: Erasure coding is a data protection method used in system design to ensure data reliability and availability. It works by dividing data into smaller chunks and then creating additional pieces of data called parity data using mathematical algorithms. This allows the system to recover the original data even if some chunks are lost or corrupted.
Erasure coding will include:

- Data Division: Original data is split into multiple chunks.
- Parity Creation: Additional parity chunks are created using algorithms like Reed-Solomon.
- Storage: Both data and parity chunks are distributed across different storage nodes or devices.
- Recovery: If some chunks are lost or damaged, the system can use the remaining chunks and the parity data to reconstruct the original data.

2 Core concept: Erasure coding (MDS, Reed–Solomon)

MDF (Maximum Distance Separable): In erasure coding data is split into K chunks and M parity chunks so the total chunks is N = K + M.
A code is called MDS if from any K out of N chunks, you can reconstruct the original data.

- The maximum number of loss trunks will be M
- The minimum required trunks for reconstructing data will be K

Reed - Solomon algorithm (RS): is a type of MDS code used for erasure coding
RS process:

- Data splitting
- Polynomial representation

- Encoding (Generating parity chunks)
- Storage/Transmission
- Decoding (Recovery)

3 Plan for research

1. Learn how to formulate Mixed - Integer Program (MIP).
2. Lookup some documents about Reed - Solomon algorithm to understand deeply each steps.
3. Read related document as references.
4. Try writing a Python program based on Reed - Solomon program.