Accelerating the Data Deduplication Performance with GPU in Hybrid Storage Systems

Prince Hamandawana, Awais Khan, **Changgyu Lee**, Sungyong Park, Youngjae Kim
Department of Computer Science and Engineering
Sogang University, Seoul, Republic of Korea

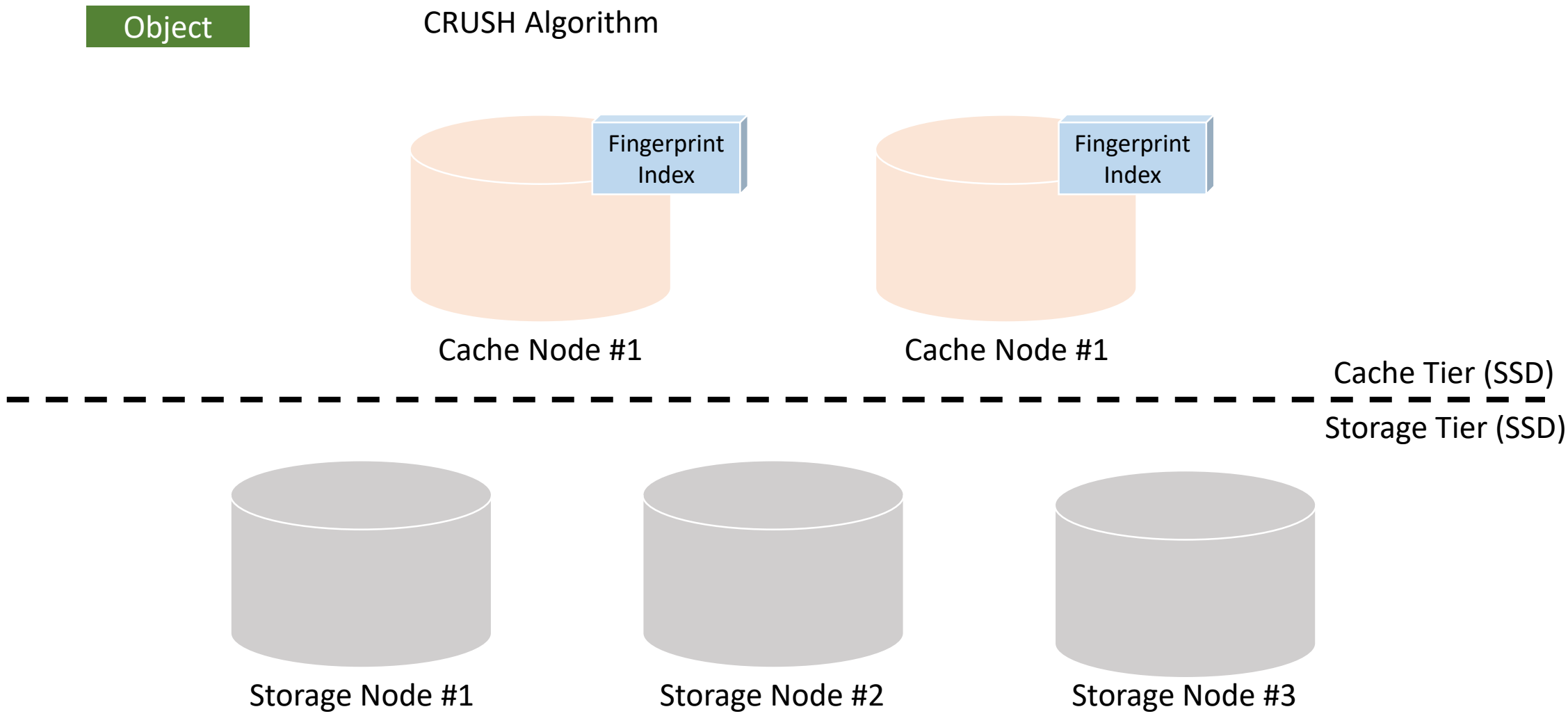
PDSW-DISCS 17 WIP session
November 13, 2017, Denver, USA

Inline Deduplication in Cloud Storage System

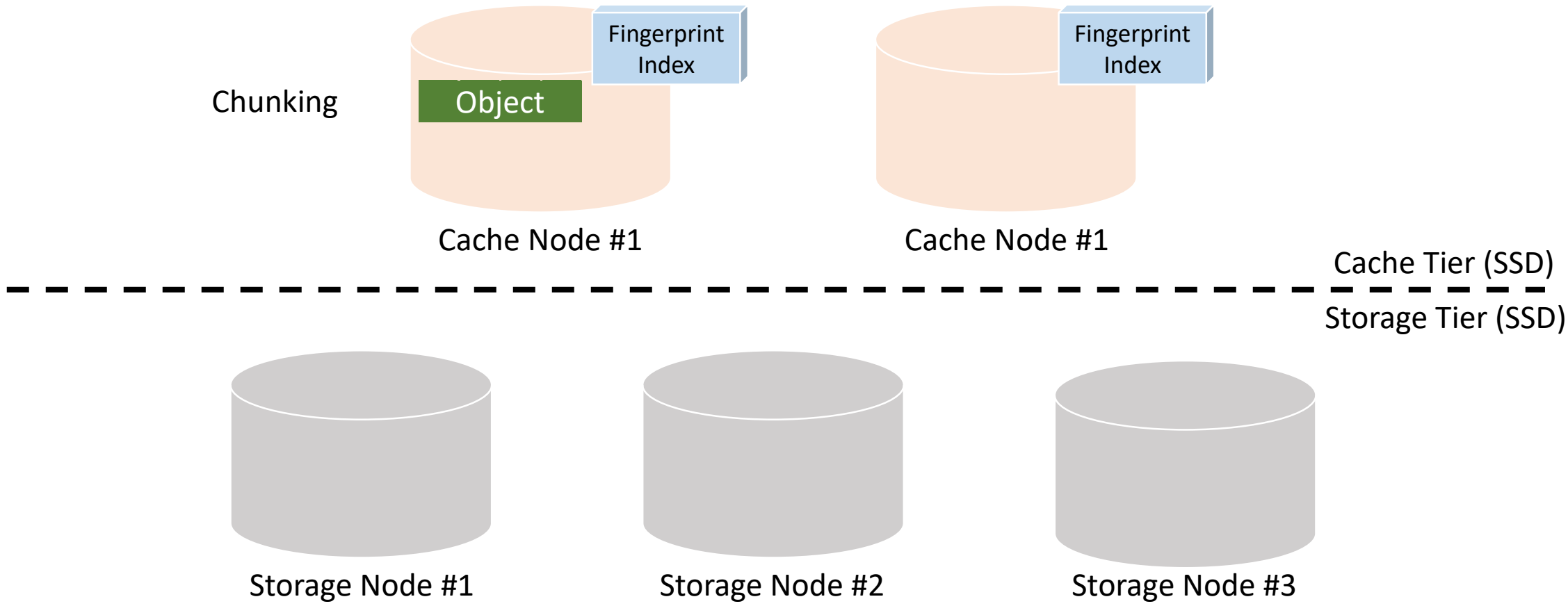
- To achieve high space utilization in Tiered Cloud Storage System, following techniques are discussed in community
  1. Compression
  2. Erasure Coding
     - Can’t remove replicated data across cluster
     - Difficult to deploy inline mode
  3. **Inline Data Deduplication**
     - Higher Storage Efficiency by removing replicated data across cluster
     - Eliminating duplicated data in Cache tier
     - But, overhead of inline deduplication directly affects to performance.
- In Hybrid Storage system, Cache-tier nodes equip SSDs and inline deduplication can reduce amount of writes to SSD.

  → Lower Write Amplification, Longer Lifetime

Inline Deduplication Framework on Ceph

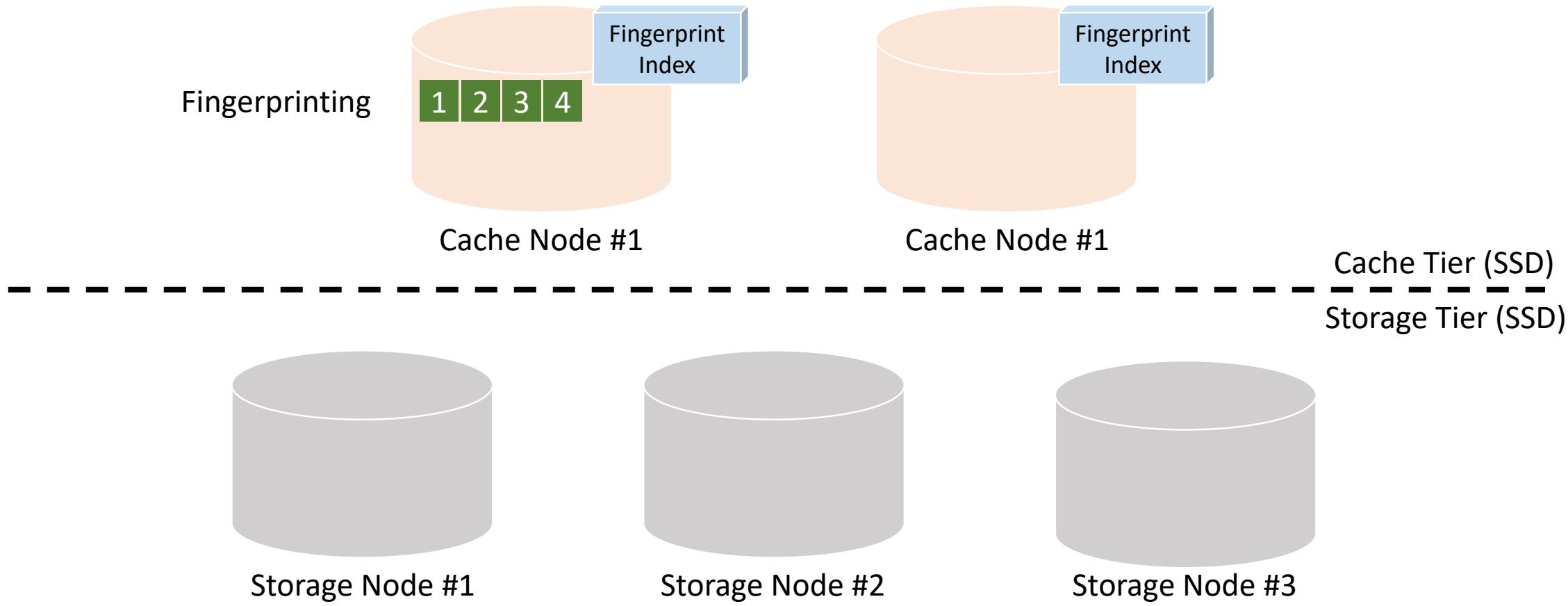


Inline Deduplication Framework on Ceph

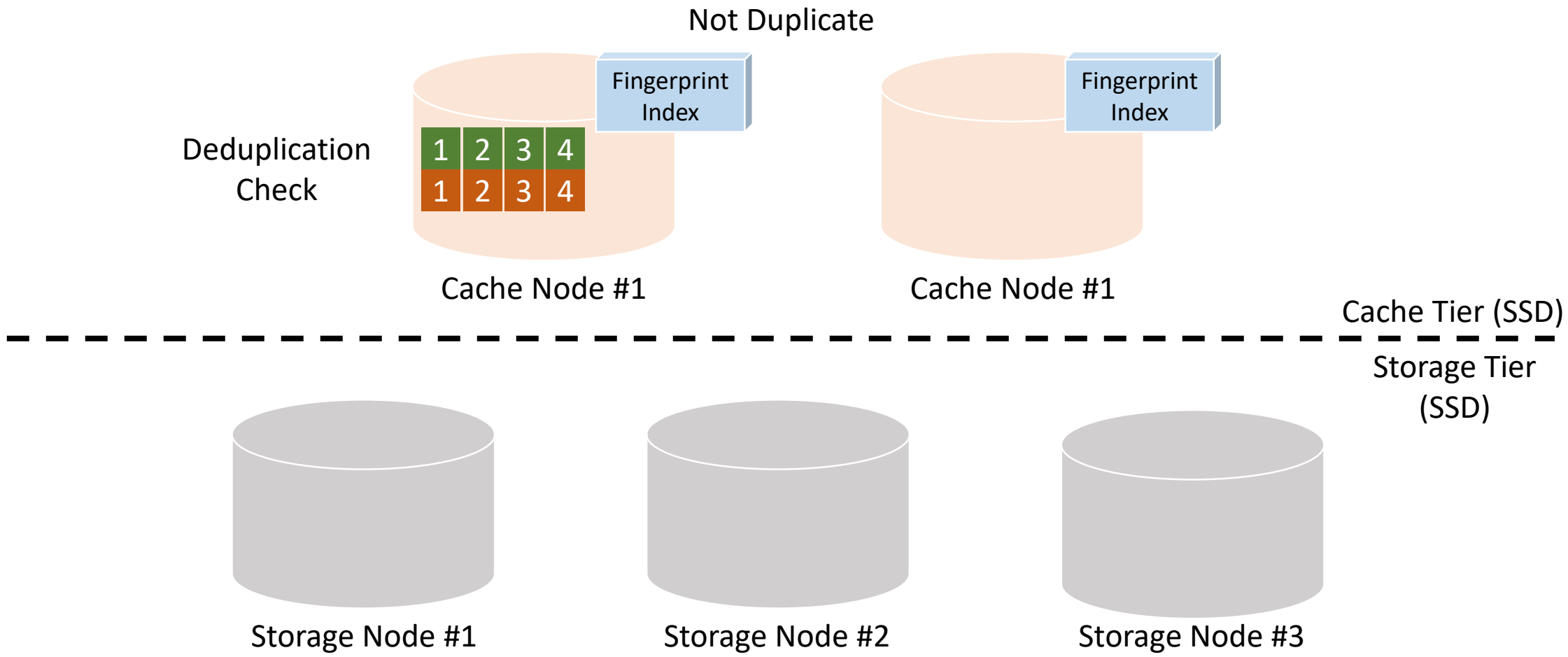


SOGANG UNIVERSITY

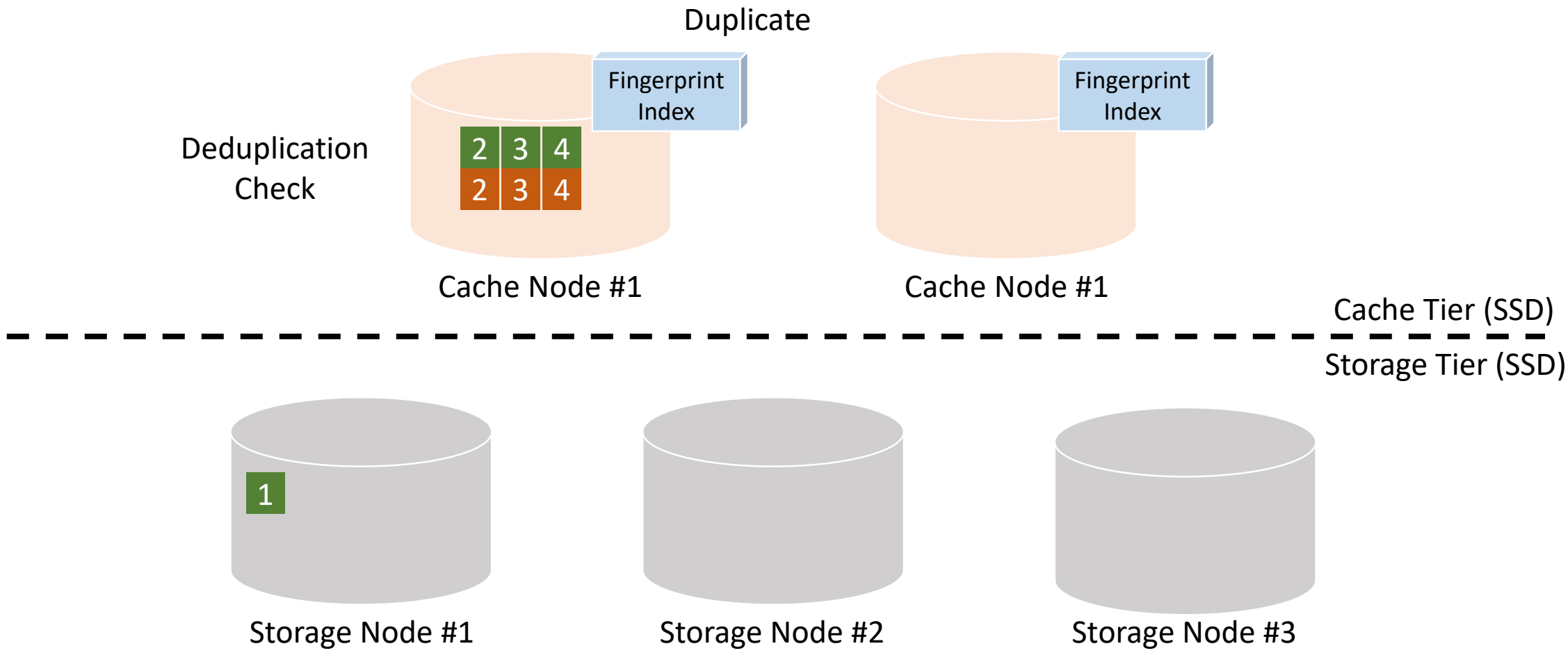
Inline Deduplication Framework on Ceph



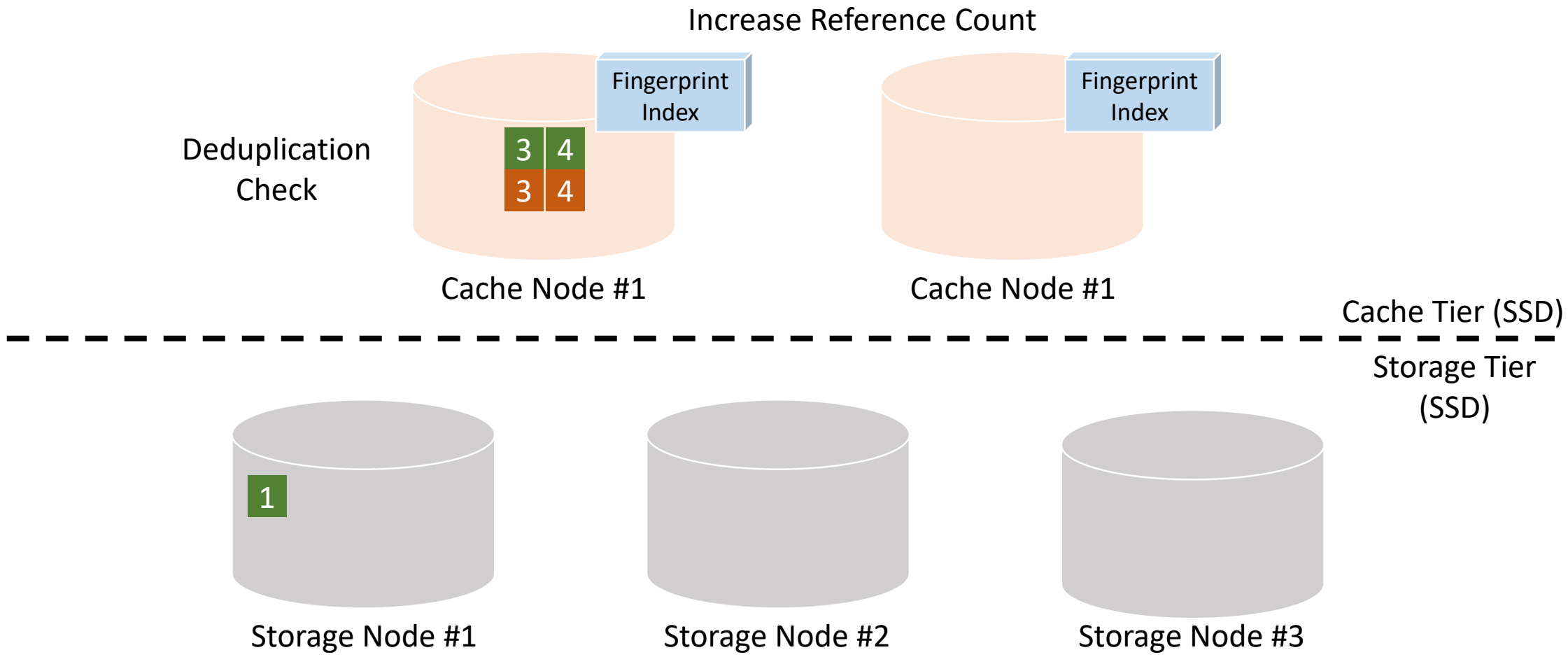
Inline Deduplication Framework on Ceph



Inline Deduplication Framework on Ceph



Inline Deduplication Framework on Ceph



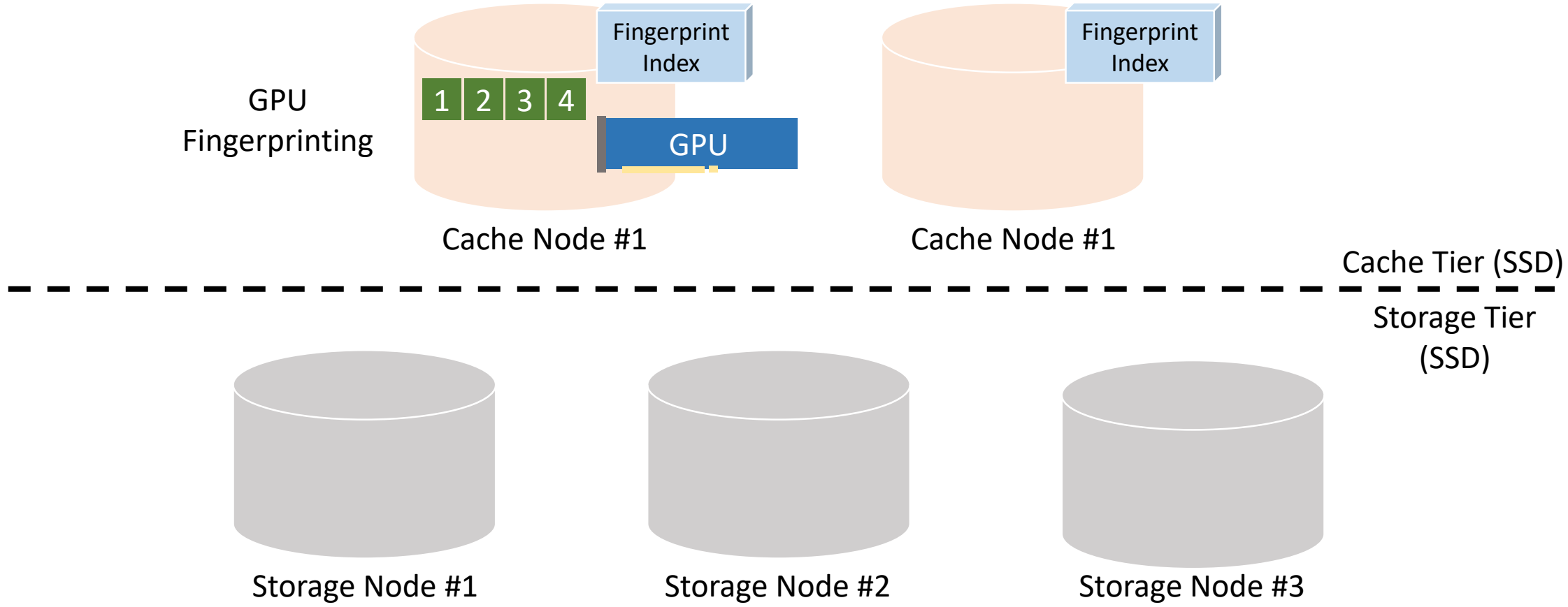
Fingerprint Overhead and GPU Acceleration

- Deduplication overhead consists of
  - Chunking
  - Calculating Fingerprint
  - Fingerprint Query

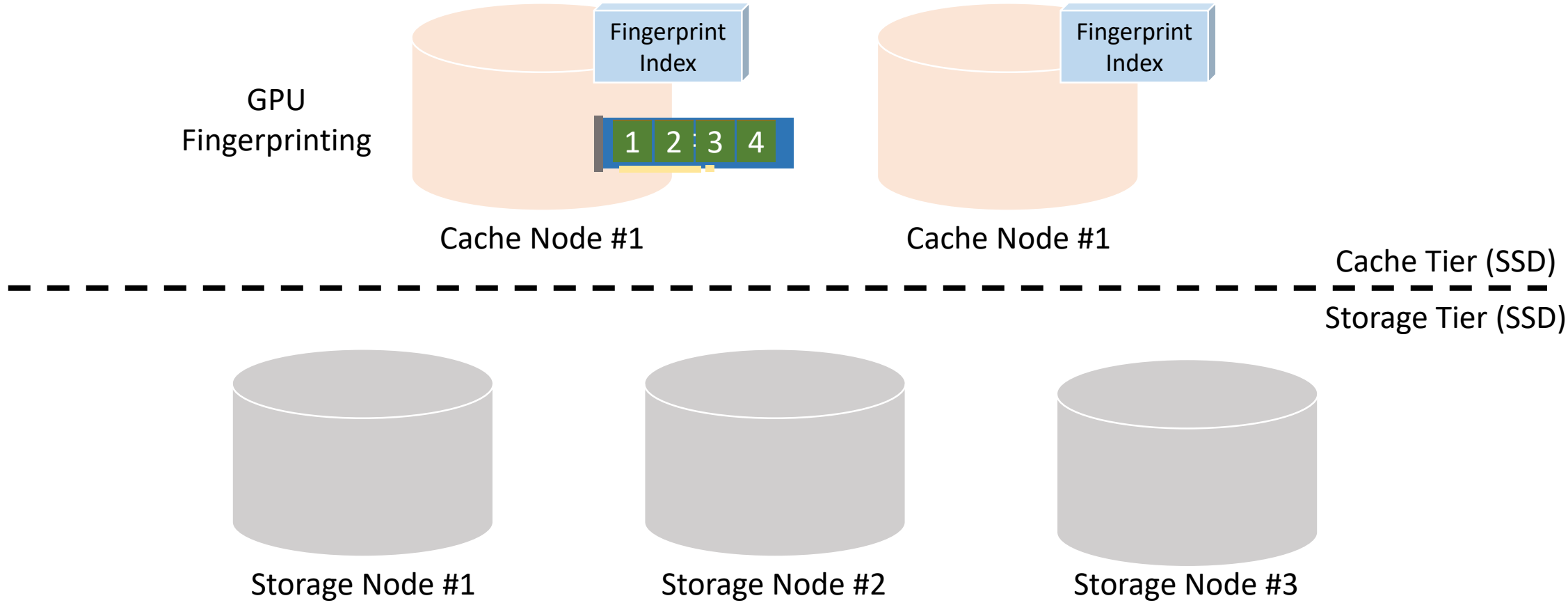
- We observed fingerprint overhead is more than 70% in total deduplication overhead.

- To reduce fingerprinting overhead, we propose to use GPU Acceleration for fingerprinting.

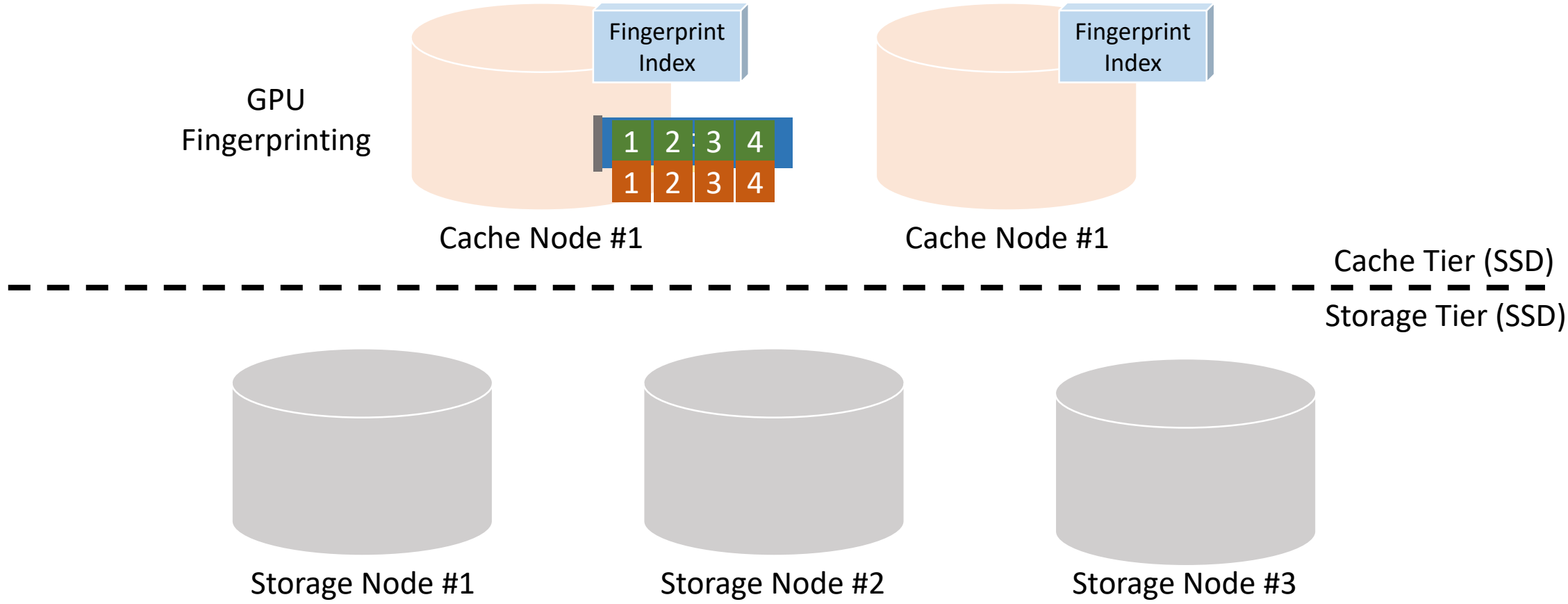
Accelerating Fingerprint Calculation with GPU



Accelerating Fingerprint Calculation with GPU



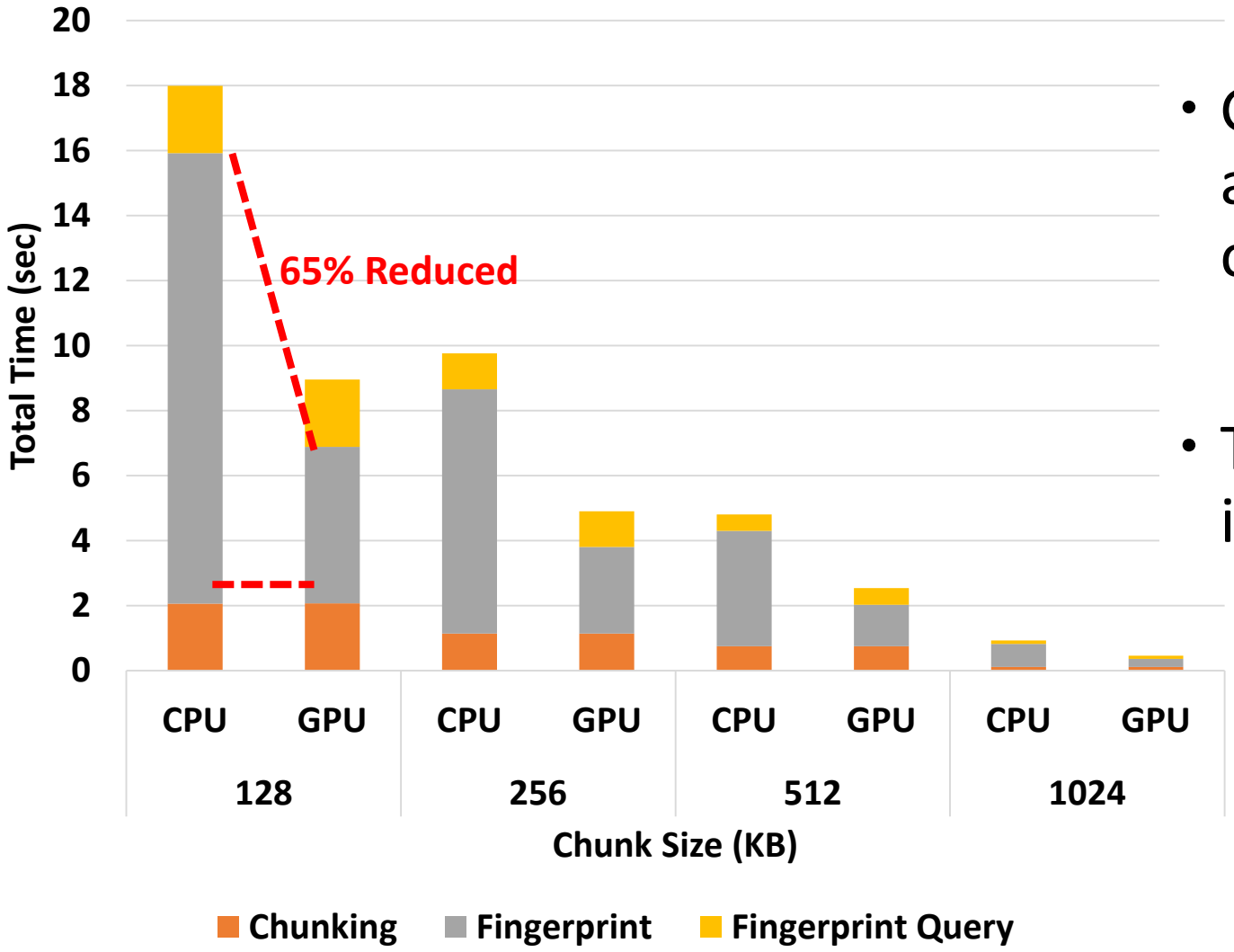
Accelerating Fingerprint Calculation with GPU



Experiment Setup

- Ceph Jewel v10.2.5
- CUDA Toolkit 8.0
- 4 OSD server
  - Intel Xeon ES-2640 v3 @ 2.60GHz
  - 32GB memory
  - 12GB NVIDIA Tesla K80 GPU
  - 2 SSDs (Cache Tier), 4 HDD (Storage Tier)
- Ceph RBD Client
- Total 1GB size random 4MB writes using fio benchmark

Preliminary Results



- GPU Fingerprinting reduced about 65% of fingerprint overhead.
- Total Deduplication overhead is reduced to 52%.

SOGANG UNIVERSITY

Q&A

- Contact: **Changgyu Lee** (changgyu@sogang.ac.kr)
  Department of Computer Science and Engineering
  Sogang University, Seoul, Republic of Korea