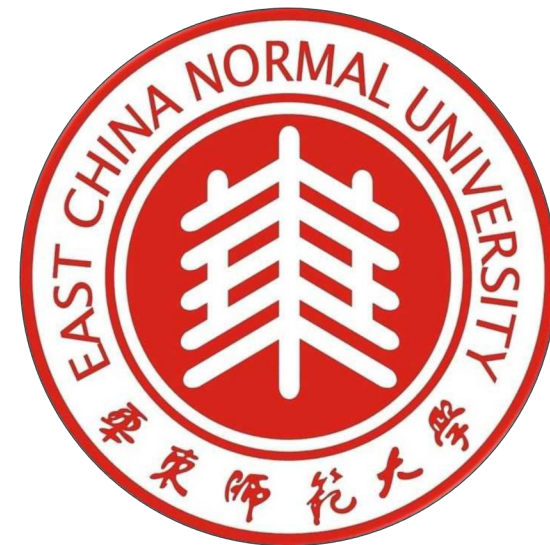


# SlimStore: A Cloud-based Deduplication System for Multi-version Backups

Zihao Zhang<sup>1</sup>, Huiqi Hu<sup>1</sup>, Zhihui Xue<sup>2</sup>, Changcheng Chen<sup>2</sup>, Yang Yu<sup>1</sup>,  
Cuiyun Fu<sup>2</sup>, Xuan Zhou<sup>1</sup>, Feifei Li<sup>2</sup>

<sup>1</sup> East China Normal University\*, China

<sup>2</sup> Alibaba Group, China



\* Shanghai Engineering Research Center of Big Data Management

# Background

## Data Backup

- Disaster recovery
- Regularly backup -> Multi-version backups

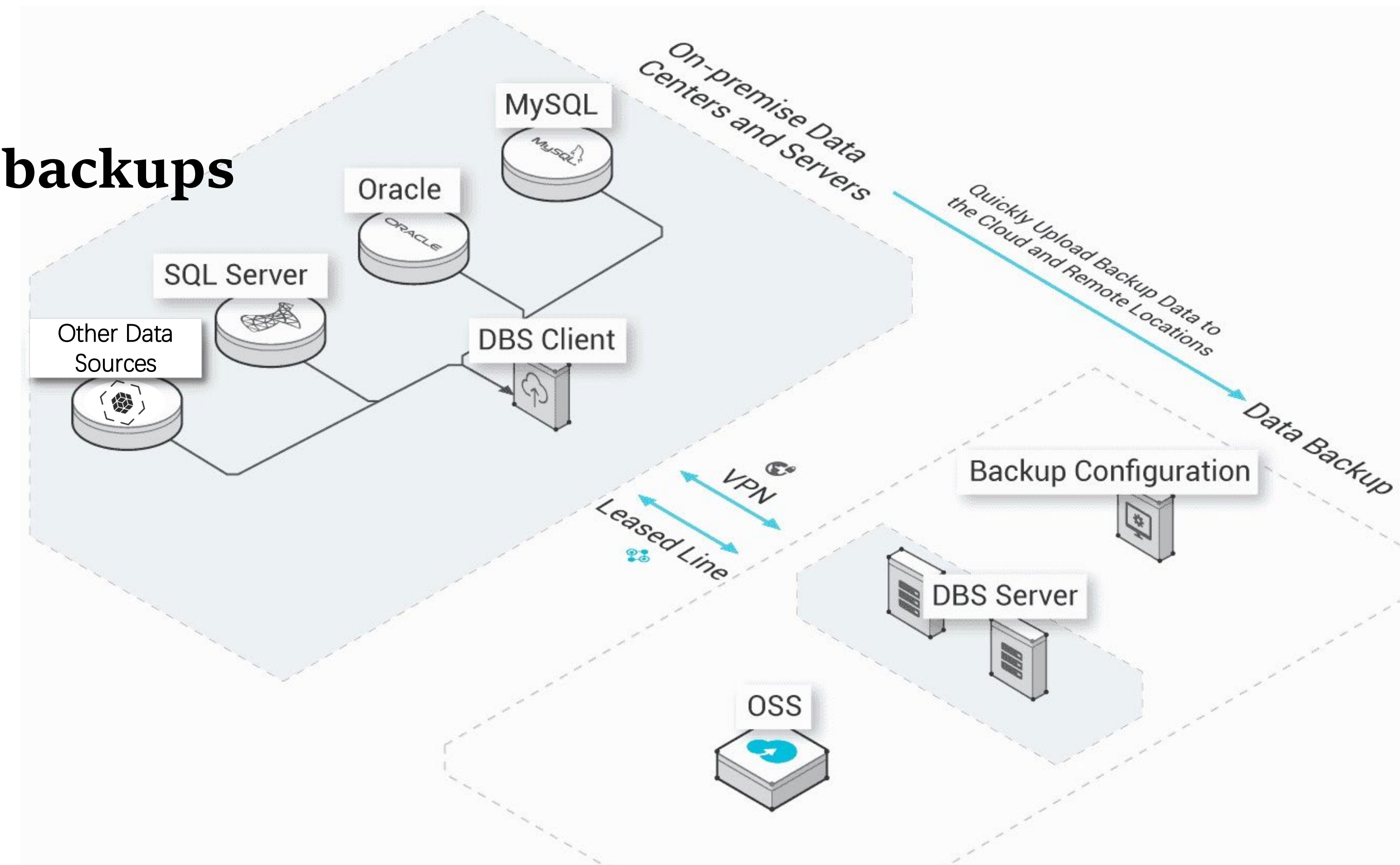
+

## Cloud Storage

- Transparently auto-scales
- Consumption-based pricing

||

## Cloud Backup





# Background

**How to reduce storage cost?**

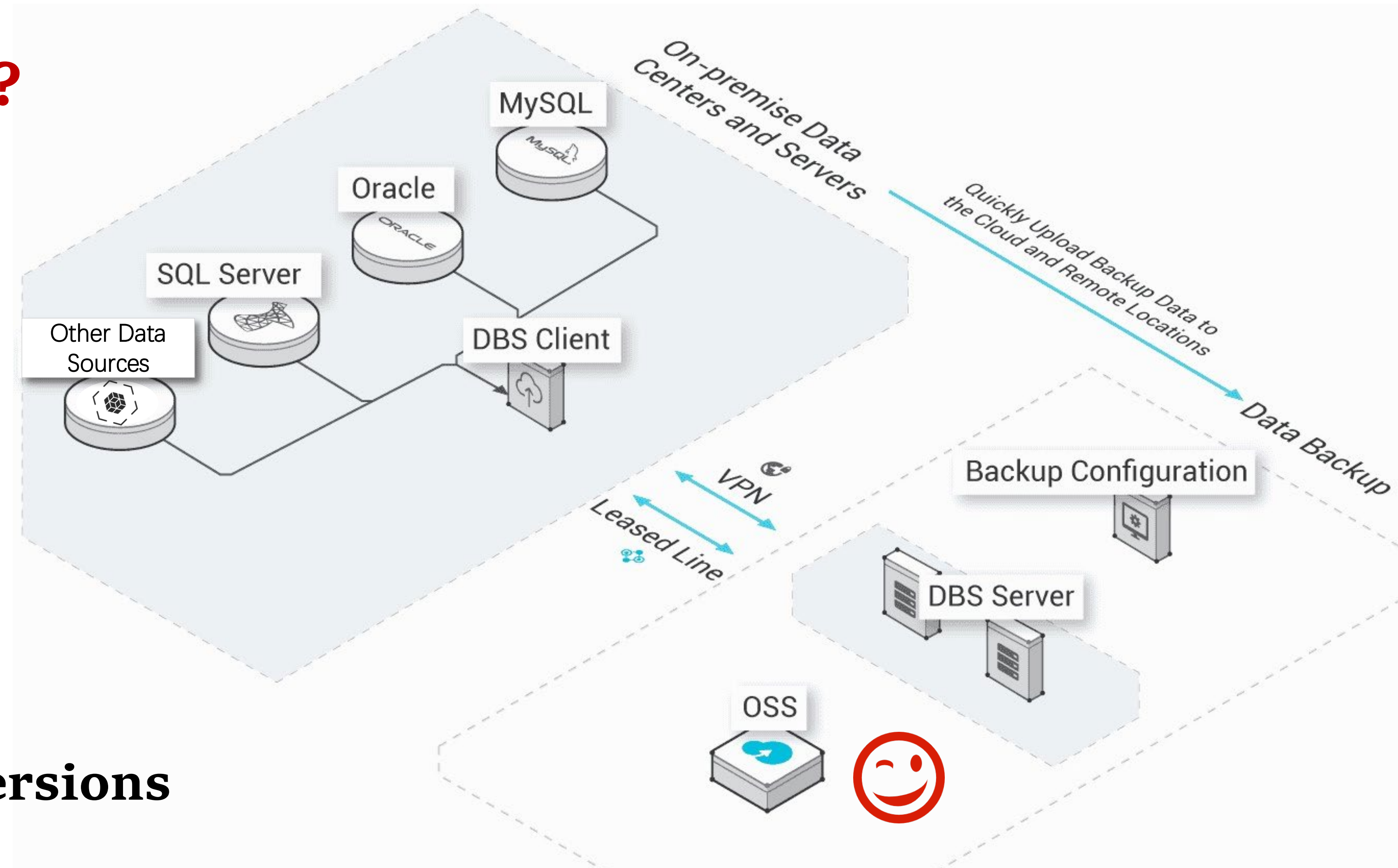
## Cheap Storage

- Object storage service  
e.g. Amazon S3, **Alibaba OSS**

## Data reduction algorithm

- Data deduplication

**Eliminating duplicates between versions**



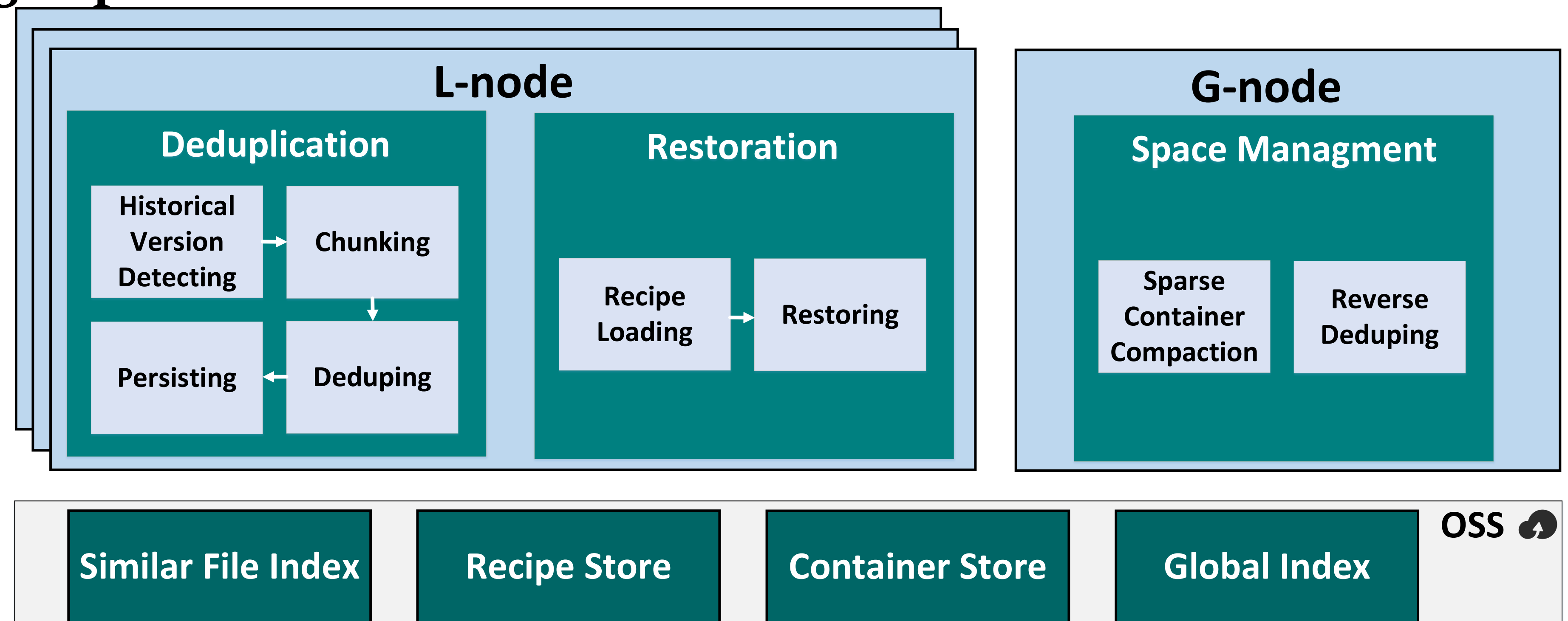
# Feature 1: Architecture - Separated storage and computation

**Storage Layer: Alibaba OSS**

**Computing Layer: Alibaba Cloud Elastic Compute Services**

- **L-Node: Online deduplication and restoration**
- **G-Node: Offline storage space reduction**

**Elastic expands  
according to the  
actual workload**



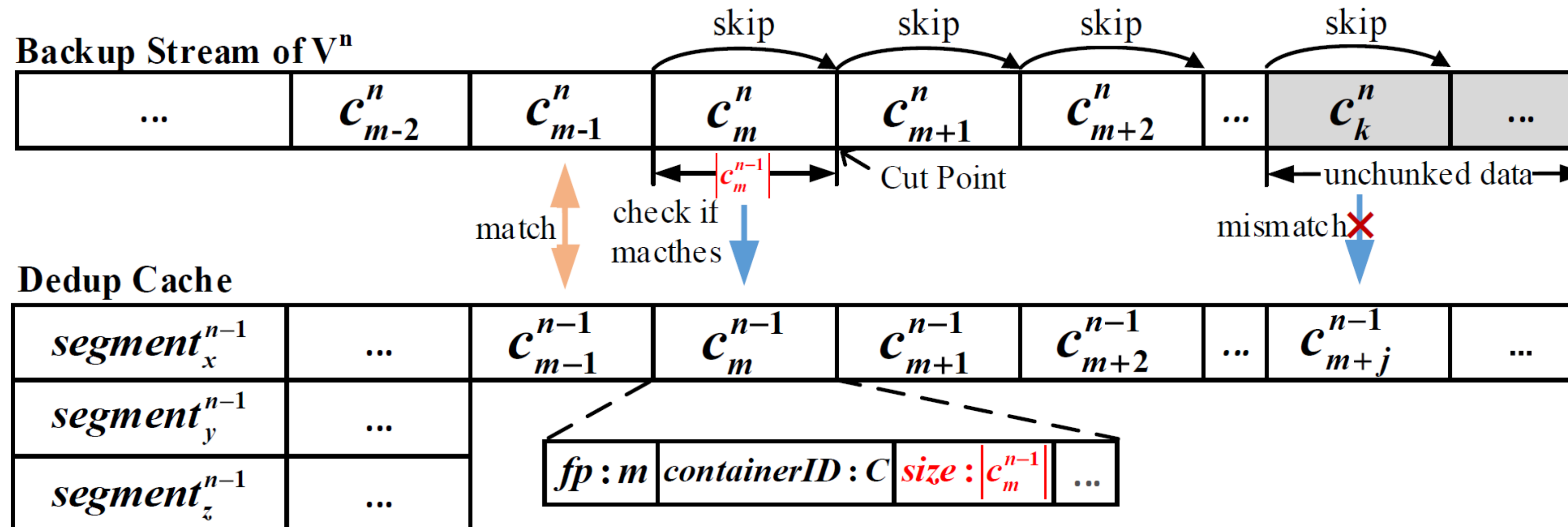
# Feature 2: Fast Online Deduplication

## Lightweight deduplication method

- Logical locality: chunk sequence in two versions are similar
- Prefetching chunks sequence of history version to quickly filter duplicate chunks

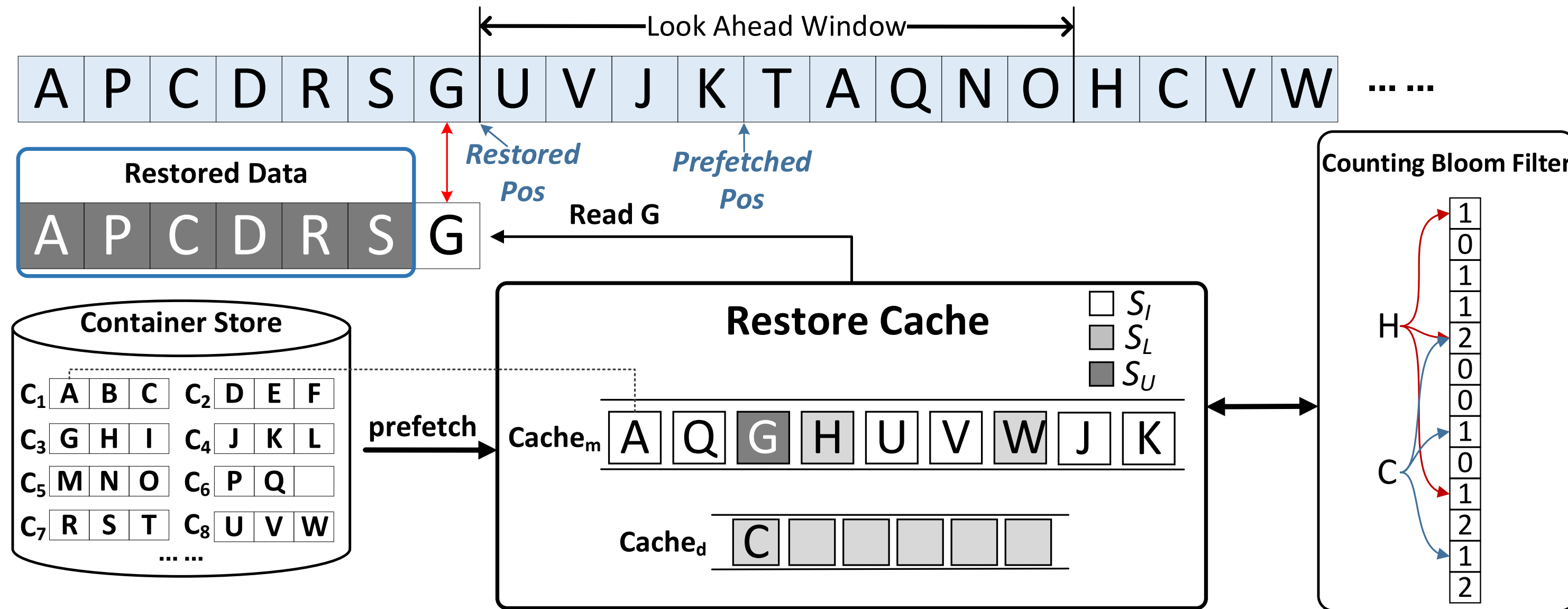
## History-aware skip chunking

- Using chunk size in the history version to skip to the next chunk point





# Feature 3: Fast Restoration



## Full-vision Restore Cache (FV cache)

- **Look-ahead window:**
  - Prefetching useful chunks in advance
- **Counting Bloom Filter:**
  - Preserving useful chunks in cache

## Sparse Container Compaction (SCC)

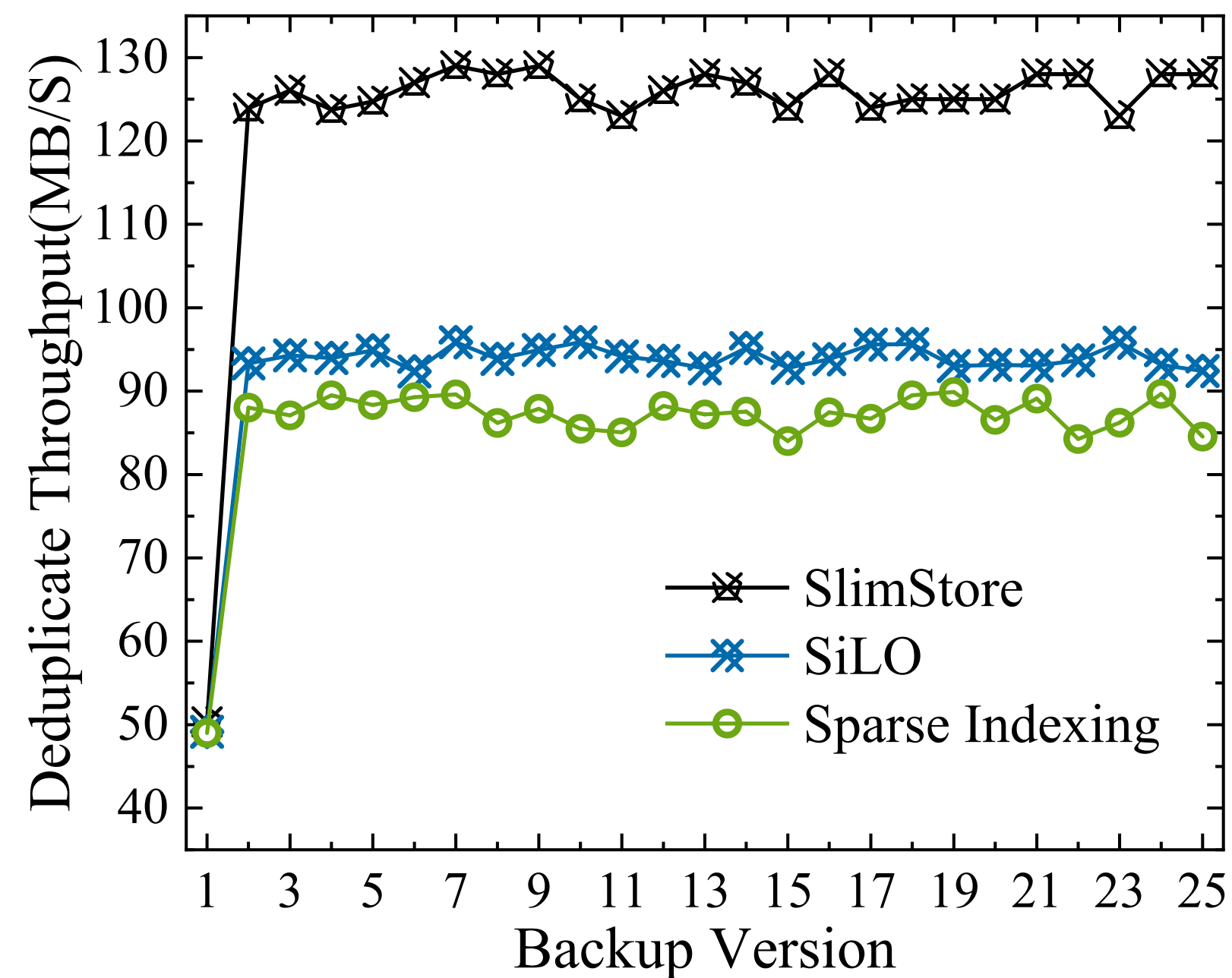
- **Working offline**
- **Adjusting the storage layout to obtain better physical locality for new versions**

# Experiments

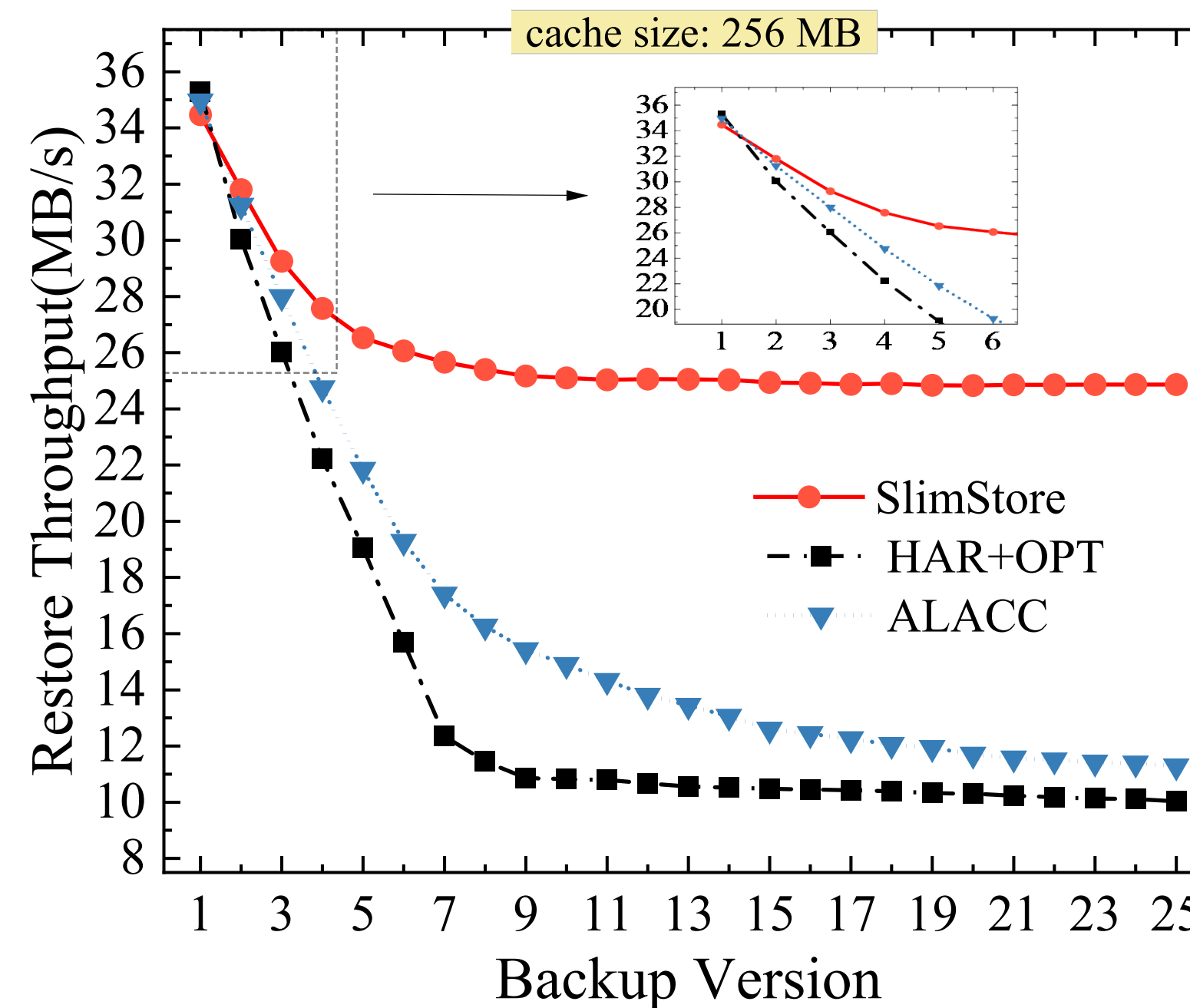
## Dataset

- **S-DB: 2.44TB database files with 25 backup versions**

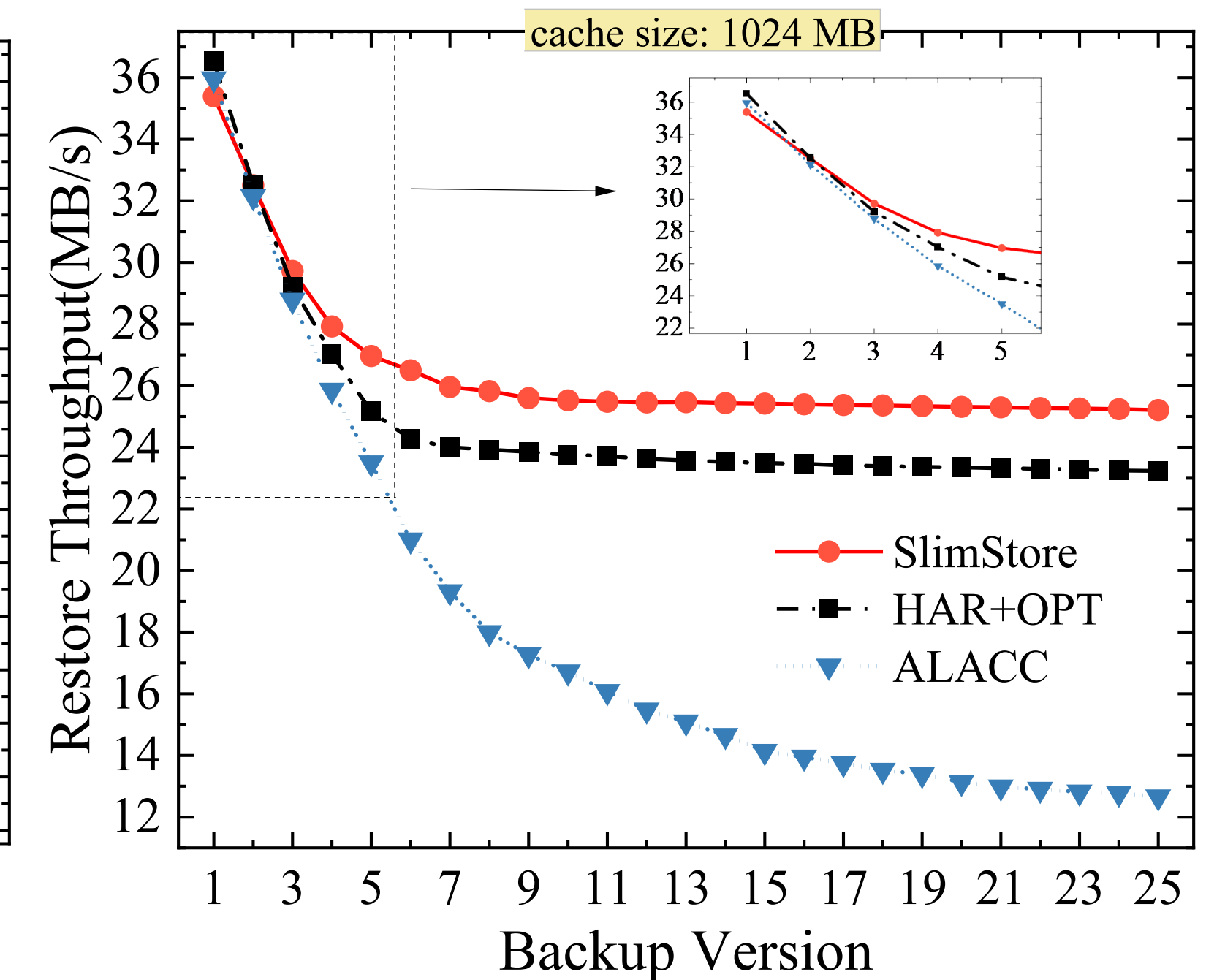
## Experimental Result



Deduplicate performance



Restore performance



**SlimStore outperforms state-of-art deduplicate and restore methods**

**THANKS!**

**Q&A**