Using Hints to Improve Inline Block-Layer Deduplication

14th USENIX Conference on File and Storage Technologies 2016

Sonam Mandal,[1] Geoff Kuenning,[3] Dongju Ok,[1] Varun Shastry,[1]
Philip Shilane,[4] Sun Zhen,[1,5] Vasily Tarasov,[2] Erez Zadok[1]

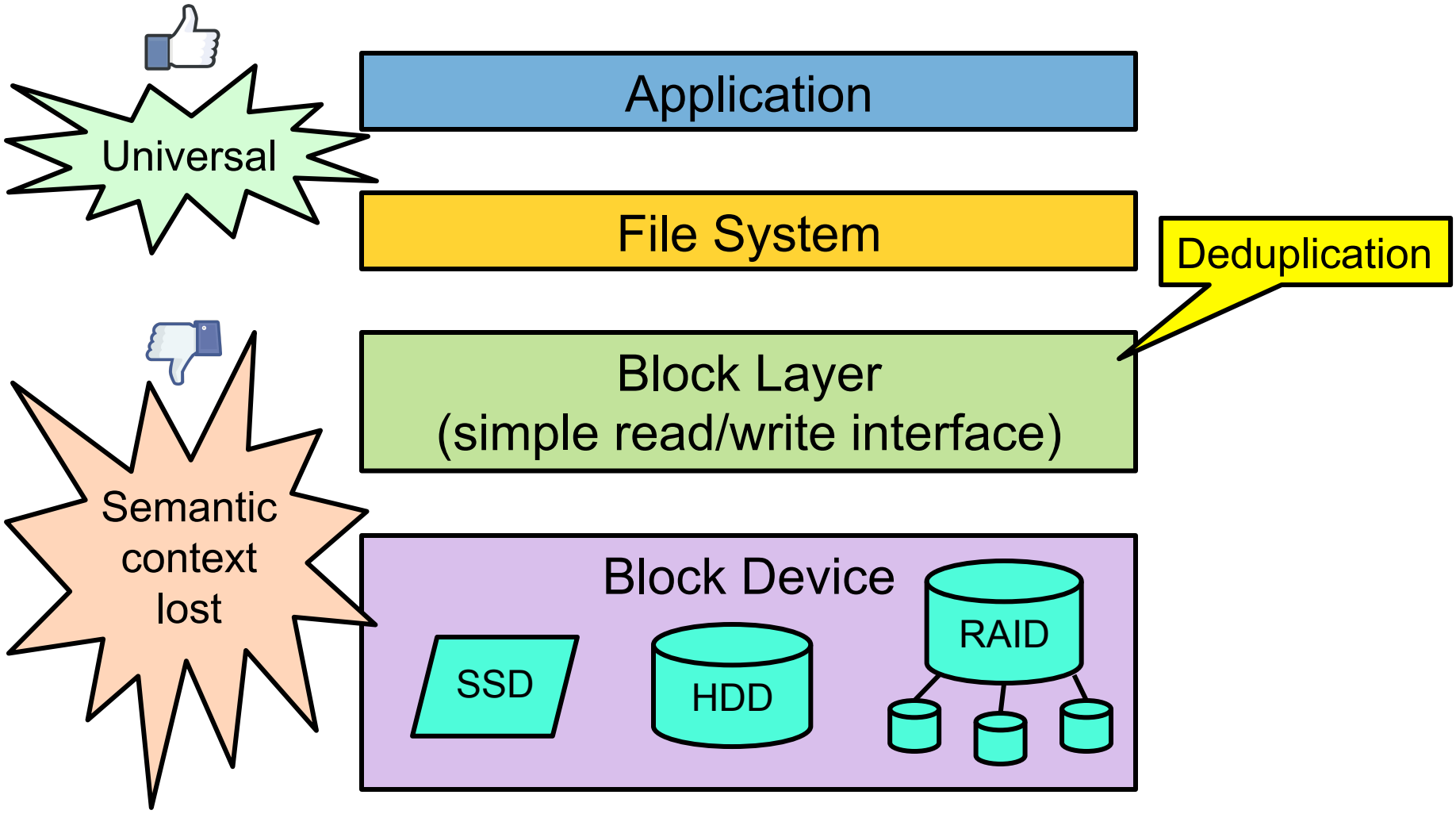
[1]Stony Brook University; [2]IBM Research – Almaden; [3]Harvey Mudd College; [4]EMC Corporation; [5]HPCL, NUDT, China

Stony Brook University HARVEY MUDD COLLEGE IBM Research EMC²

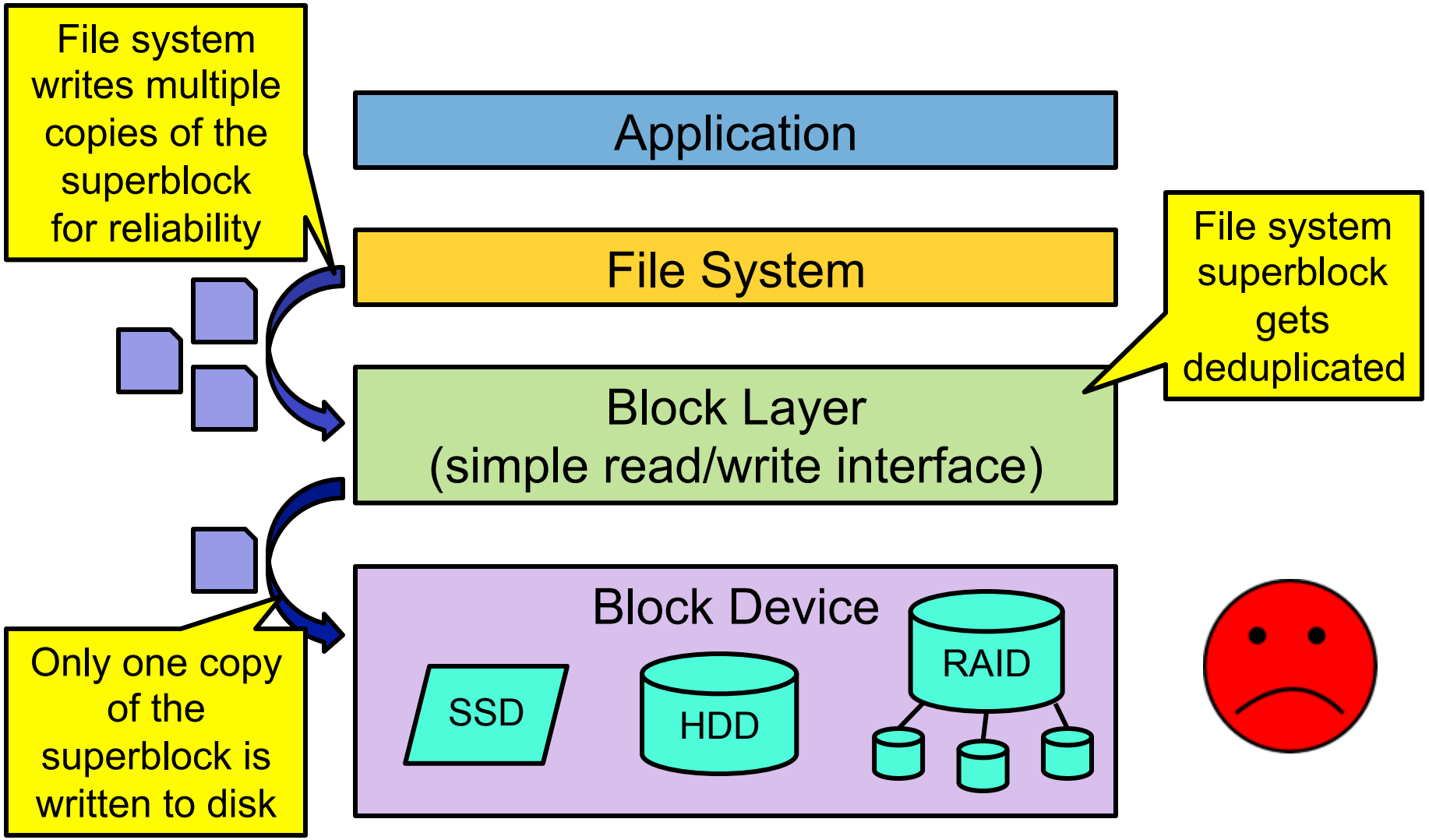
Outline

- Introduction
- Dmdedup Overview
- Hints
- Evaluation
- Conclusion and Future Work

Introduction: Storage Stack

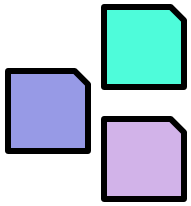


Introduction: Reliability



Introduction: Efficient Resource Utilization

Application writes out **unique** data blocks (does not deduplicate)





Application

File System

Block Layer
(simple read/write interface)

Block Device

SSD

HDD

RAID

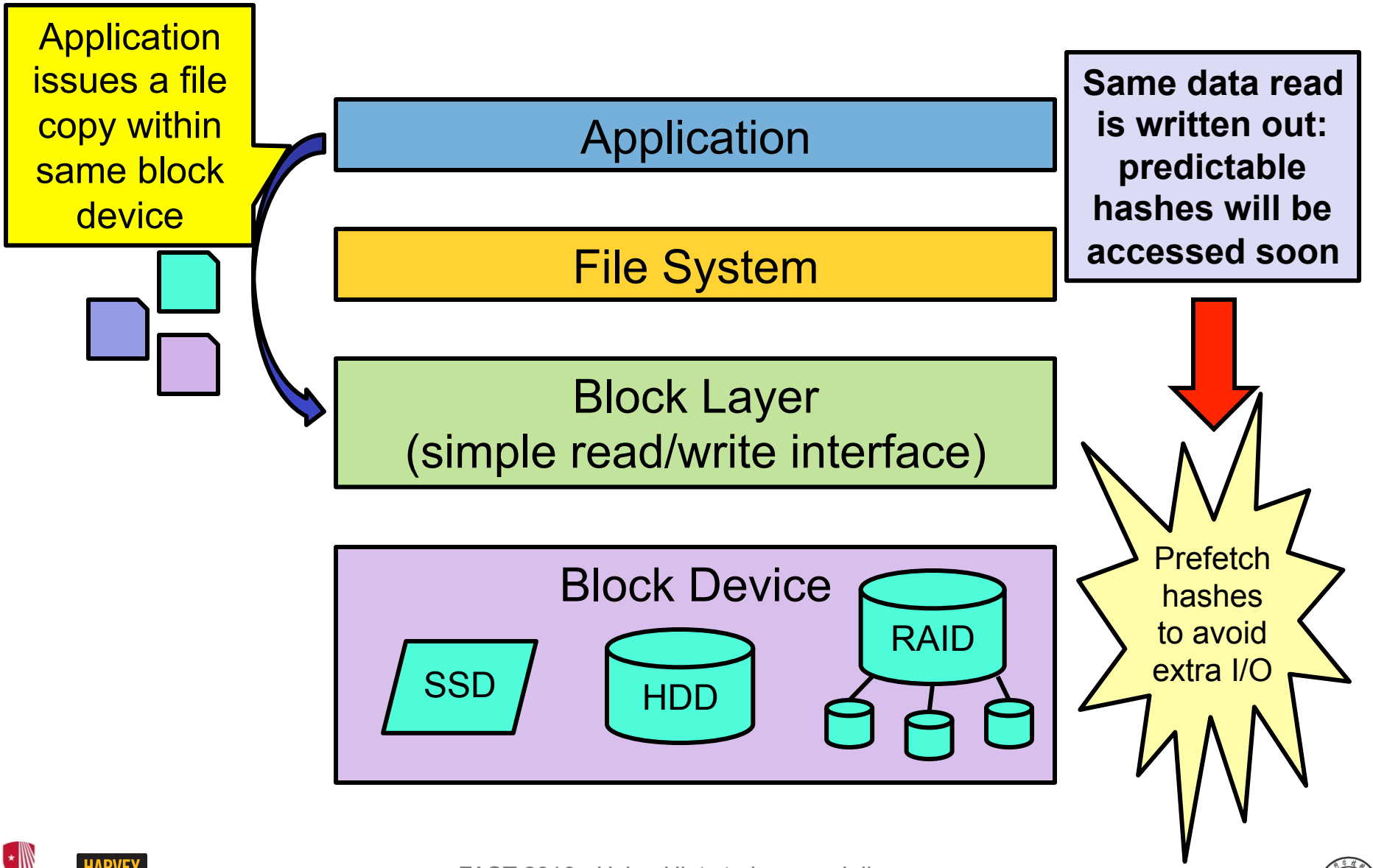
Deduplication Overhead

CPU

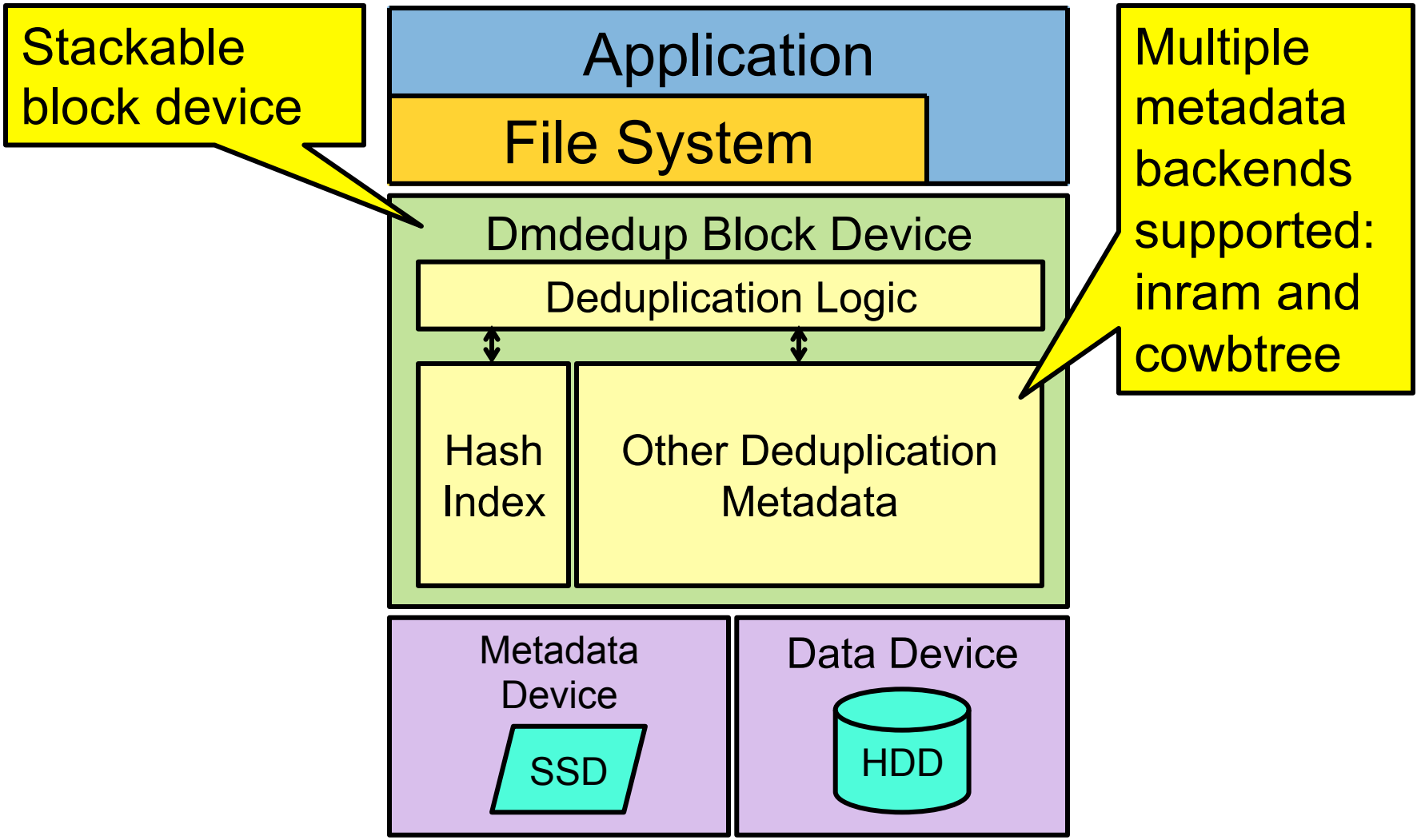
Memory

I/O to disk

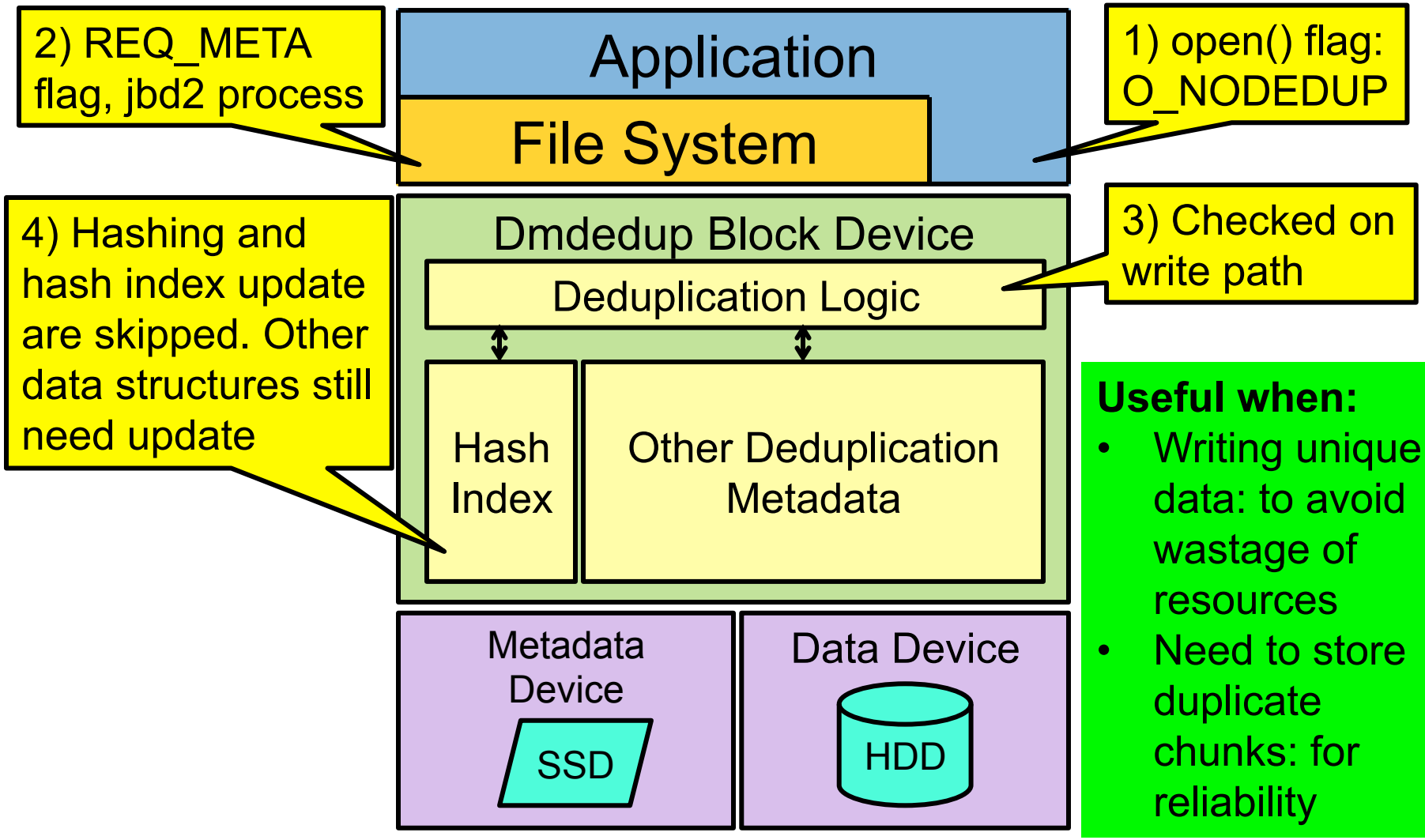
Introduction: Performance



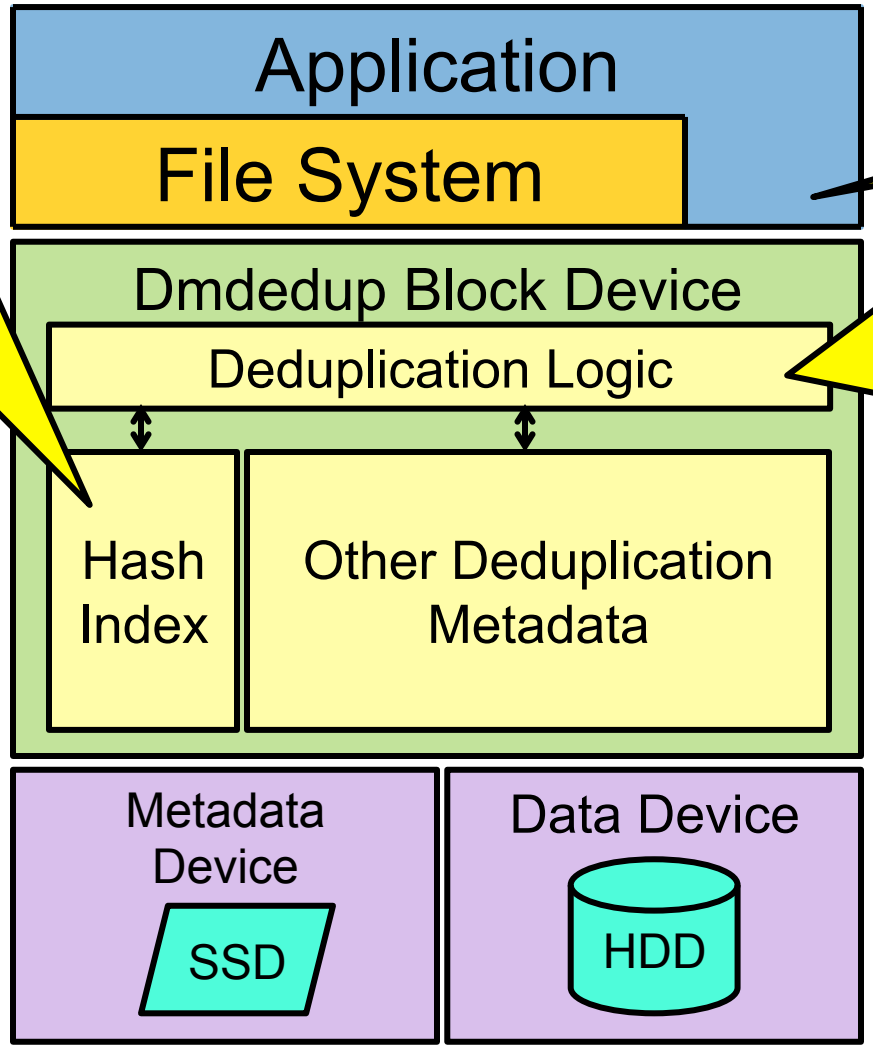
Dmdedup Components



NODEDUP Hint



PREFETCH Hint



3) Hash index lookup is skipped if the hash is found in prefetch cache on write. Other data structures still need lookup

1) open() flag: O_PREFETCH

2) We don't usually add the hash on read path. On hint, we add hash to prefetch cache on read path.

Useful when: Soon to be used hashes are already known: hash index lookup is an expensive operation.

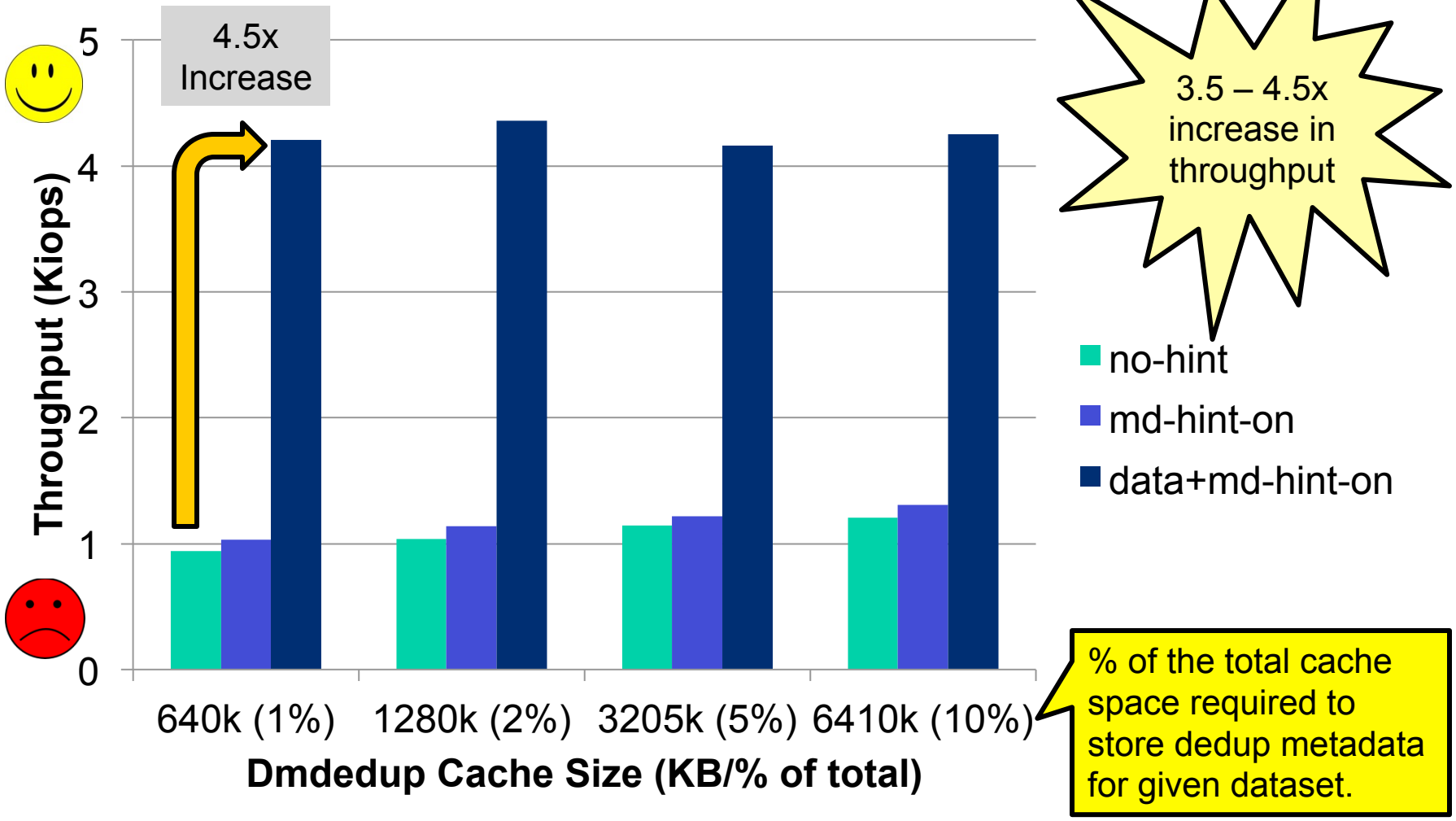
Evaluation: Workloads

- NODEDUP hint:
  - Filebench's Fileserver workload
    - Data and metadata ops emulating a file server
  - Configuration
    - No NODEDUP hint
    - Metadata only marked with NODEDUP hint
    - Metadata and data marked with NODEDUP hint
- PREFETCH hint:
  - File copy workload modified to pass PREFETCH hint
- Dmdedup's Cowbtree backend used for all experiments

NODEDUP Hint: Fileserver Workload

- Modified Filebench
  - Write data based on a given duplicate distribution – unique data for results shown
- Modified Fileserver workload definition
  - 4KB blocks instead of 1KB blocks.
- Experimented with different file systems

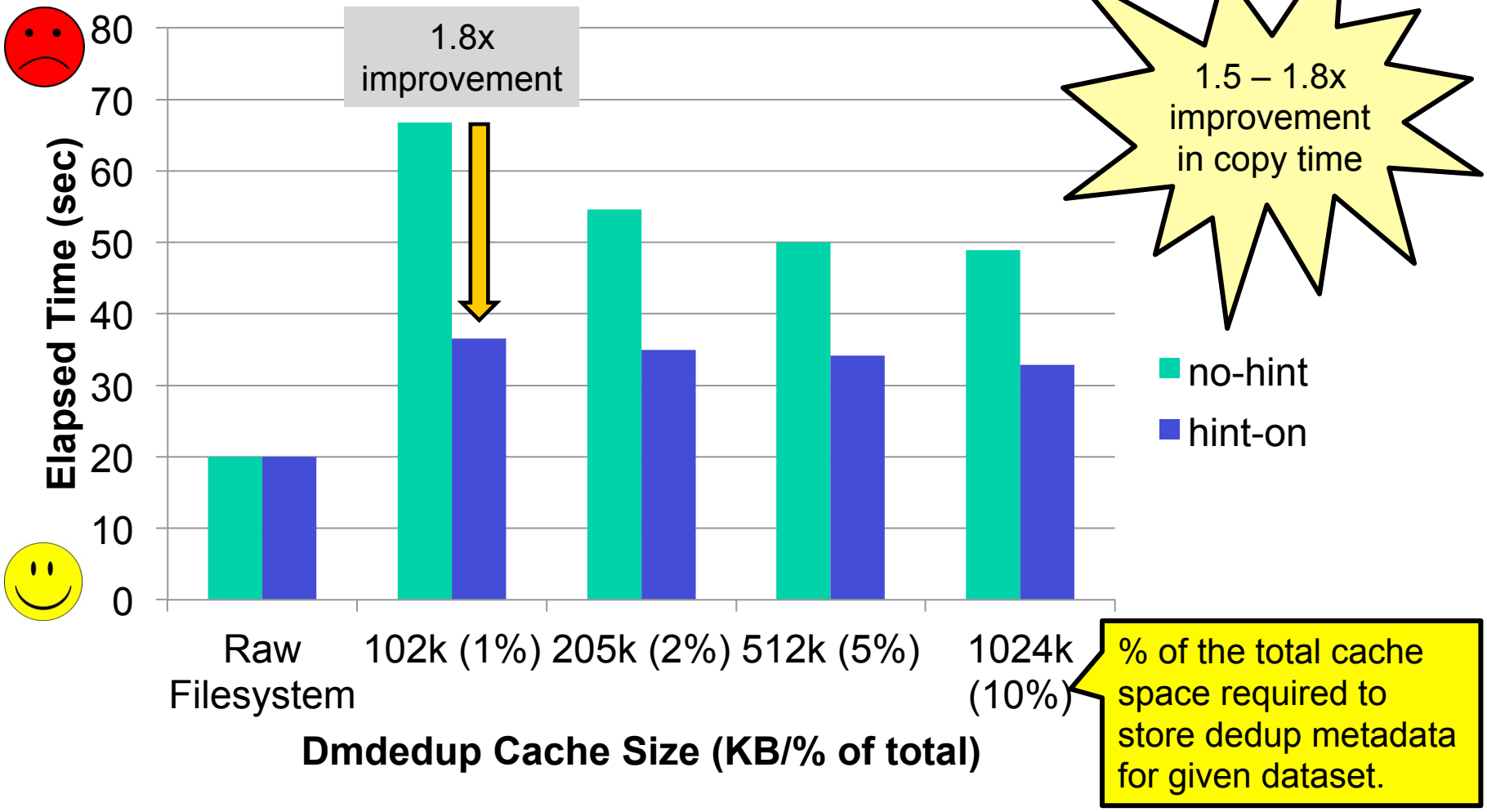
NODEDUP Hint: Fileserver Workload on Nilfs2



PREFETCH Hint: File Copy Workload

- Copy 1GB file with unique 4KB blocks
  - Modified dd – O_PREFETCH open() flag on read path
  - Unmodified dd
- Called sync and umount to make sure data reached block layer.
- Experimented with different file systems

PREFETCH Hint: File Copy Workload on Nilfs2



Conclusion and Future Work

- Conclusion – Hints help
  - Adds reliability
    - Duplicate file system metadata blocks
  - Improves performance
    - Prefetches hashes
  - Uses resources efficiently
    - Avoids deduplication overhead for unique data
- Future Work
  - Hint support for other file systems
  - PREFETCH Hint for Nilfs2 segment cleaning
  - Cross-block device support for hash prefetching

Using Hints to Improve Inline Block-Layer Deduplication



Sonam Mandal,[1] Geoff Kuenning,[3] Dongju Ok,[1] Varun Shastry,[1]
Philip Shilane,[4] Sun Zhen,[1,5] Vasily Tarasov,[2] Erez Zadok[1]

[1]Stony Brook University; [2]IBM Research – Almaden; [3]Harvey Mudd College; [4]EMC Corporation; [5]HPCL, NUDT, China

Git Repository for Dmdedup: git://git.fsl.cs.sunysb.edu/linux-dmdedup.git

Stony Brook University · HARVEY MUDD COLLEGE · IBM Research · EMC²