

Background

Key-value stores

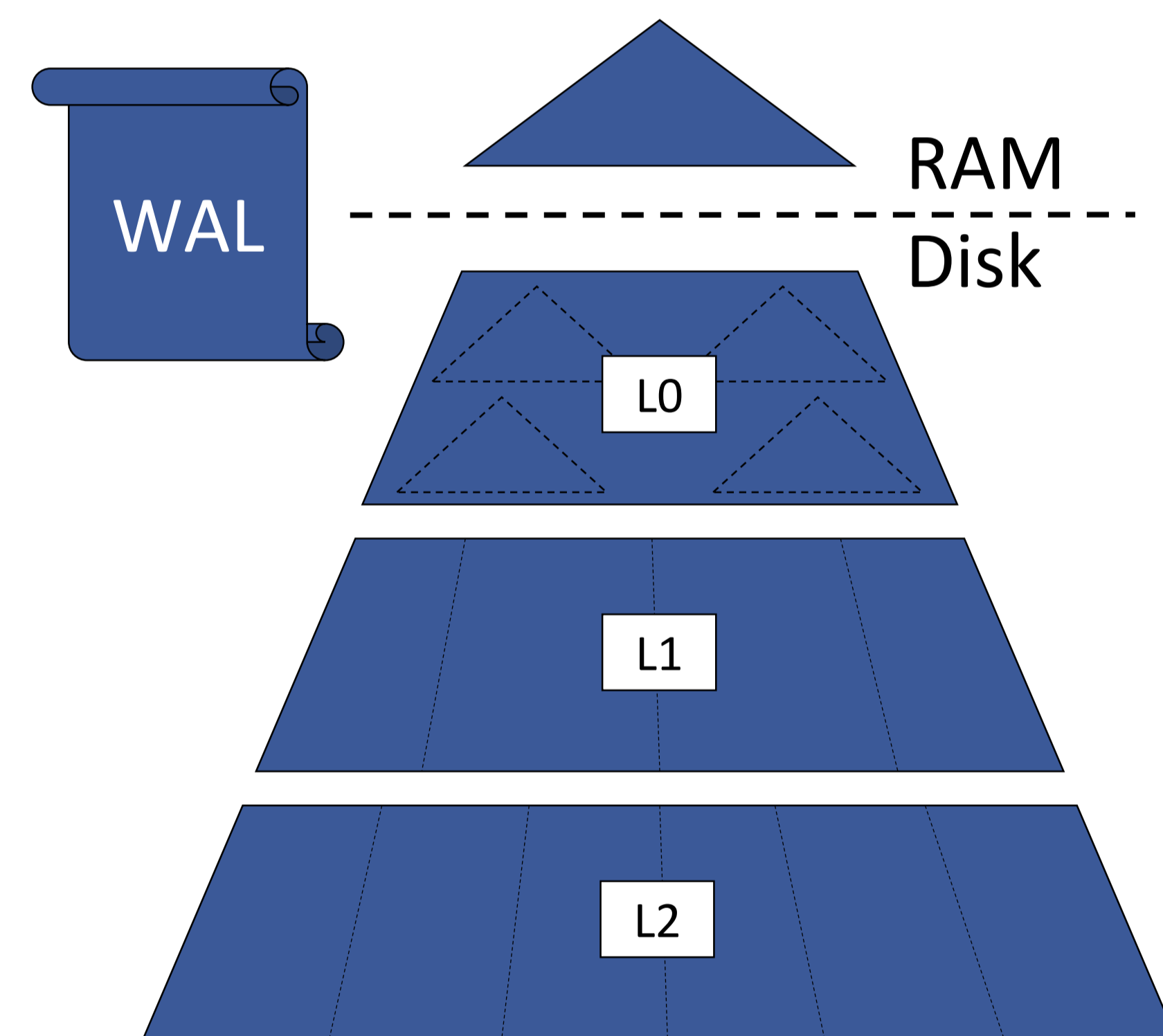
- Key → value
- Put, Get, Delete, atomic Scan

Persistence motivation

- Survive crashes
- Hold data beyond RAM size

Main challenges

- Consistency
- Performance



Log-Structured Merge (LSM)

- In-memory “cache” of latest puts
- Persistent log file
- Multiple storage levels

Issues

- Single log file
- Single in-memory data structure
- Storage files might overlap
- Painful rebalancing

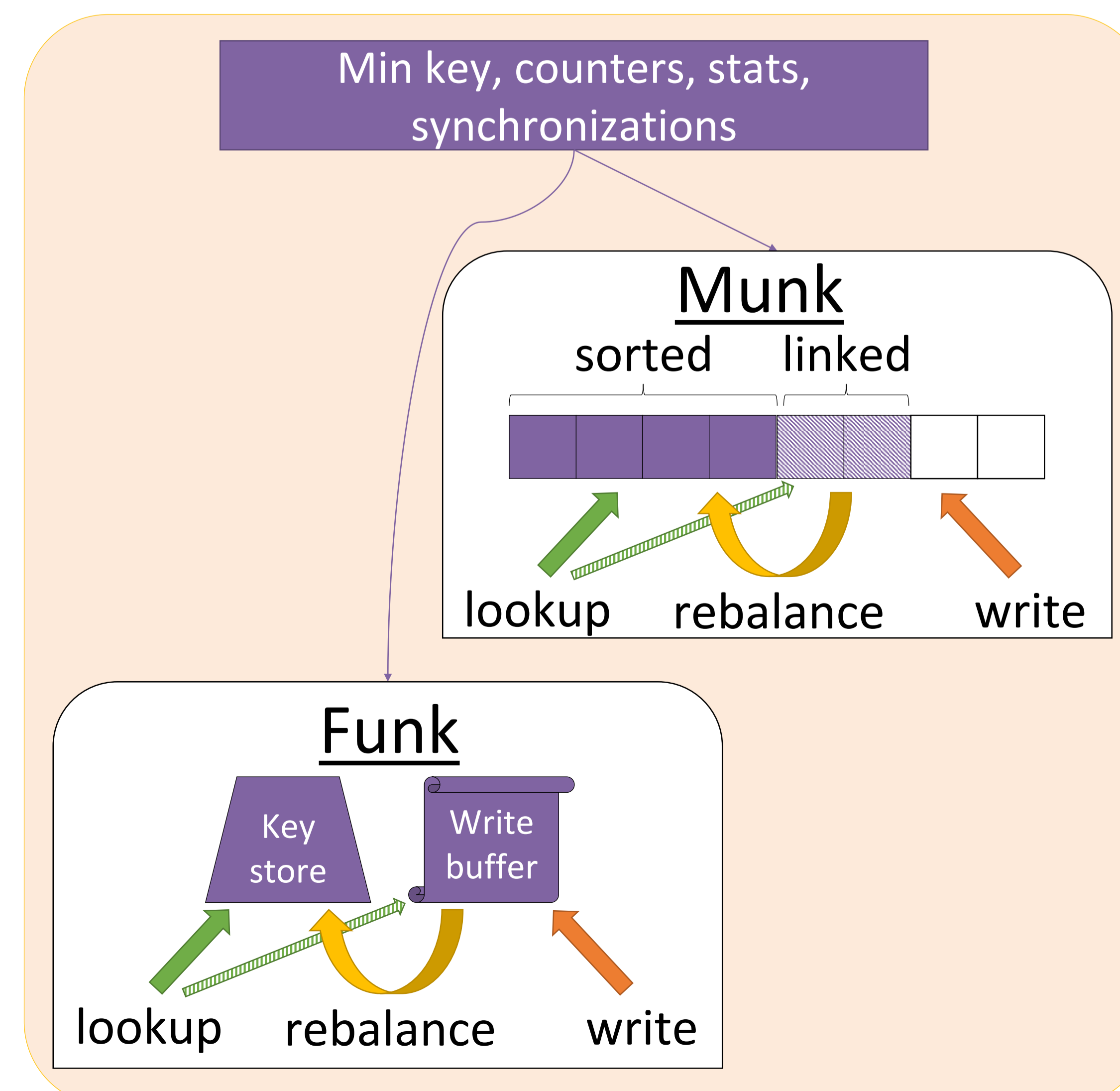
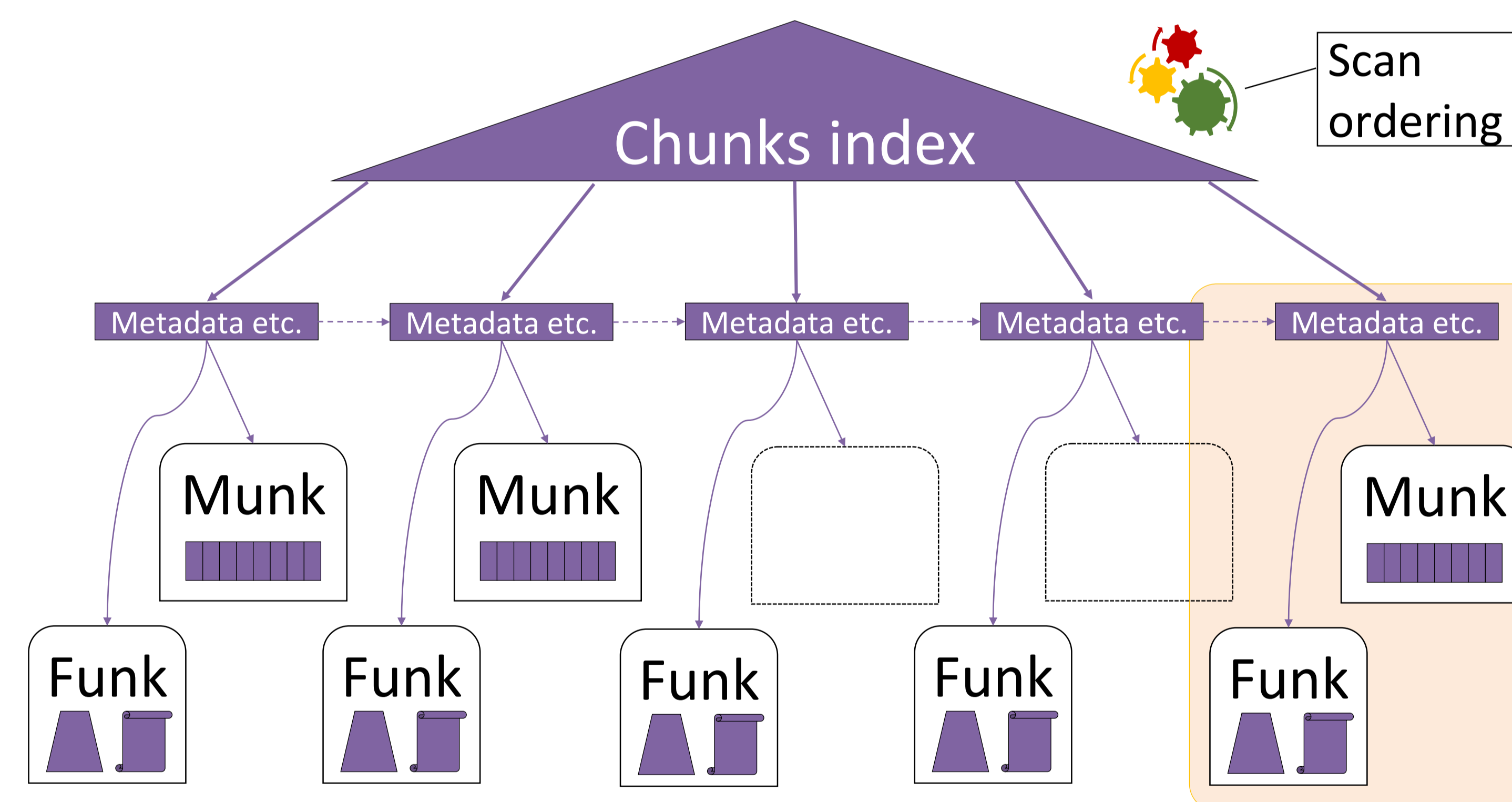
PiWi – Persistent KiWi

Ancestry: KiWi [PPoPP17]

- Chunk-based in-memory structure
- Balanced, cache friendly
- Highly concurrent, wait/lock-free

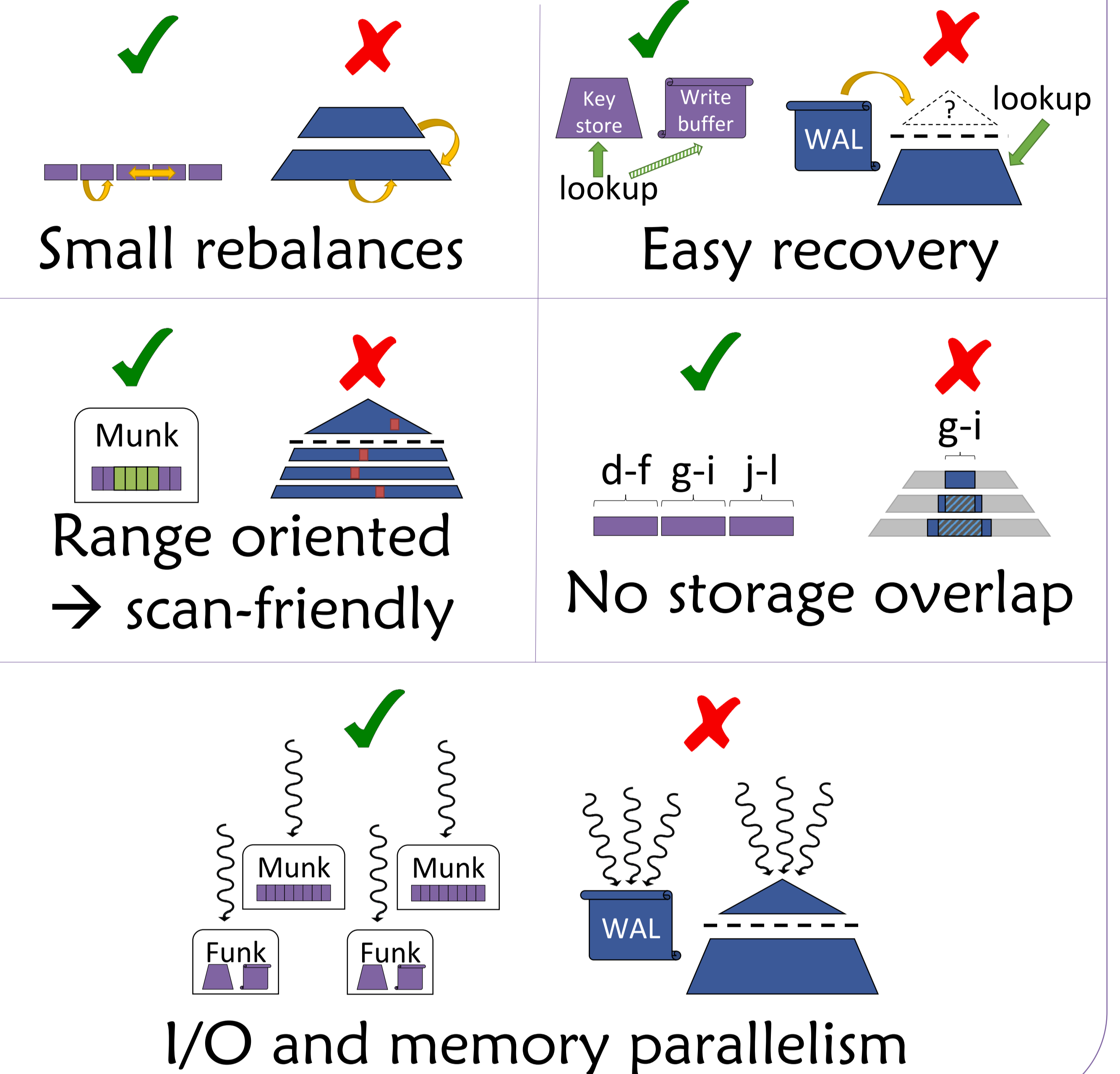
Adding persistency

- All data stored in files (Funk)
- Hot data stored in memory (Munk)
- Not lock-free, scans still atomic

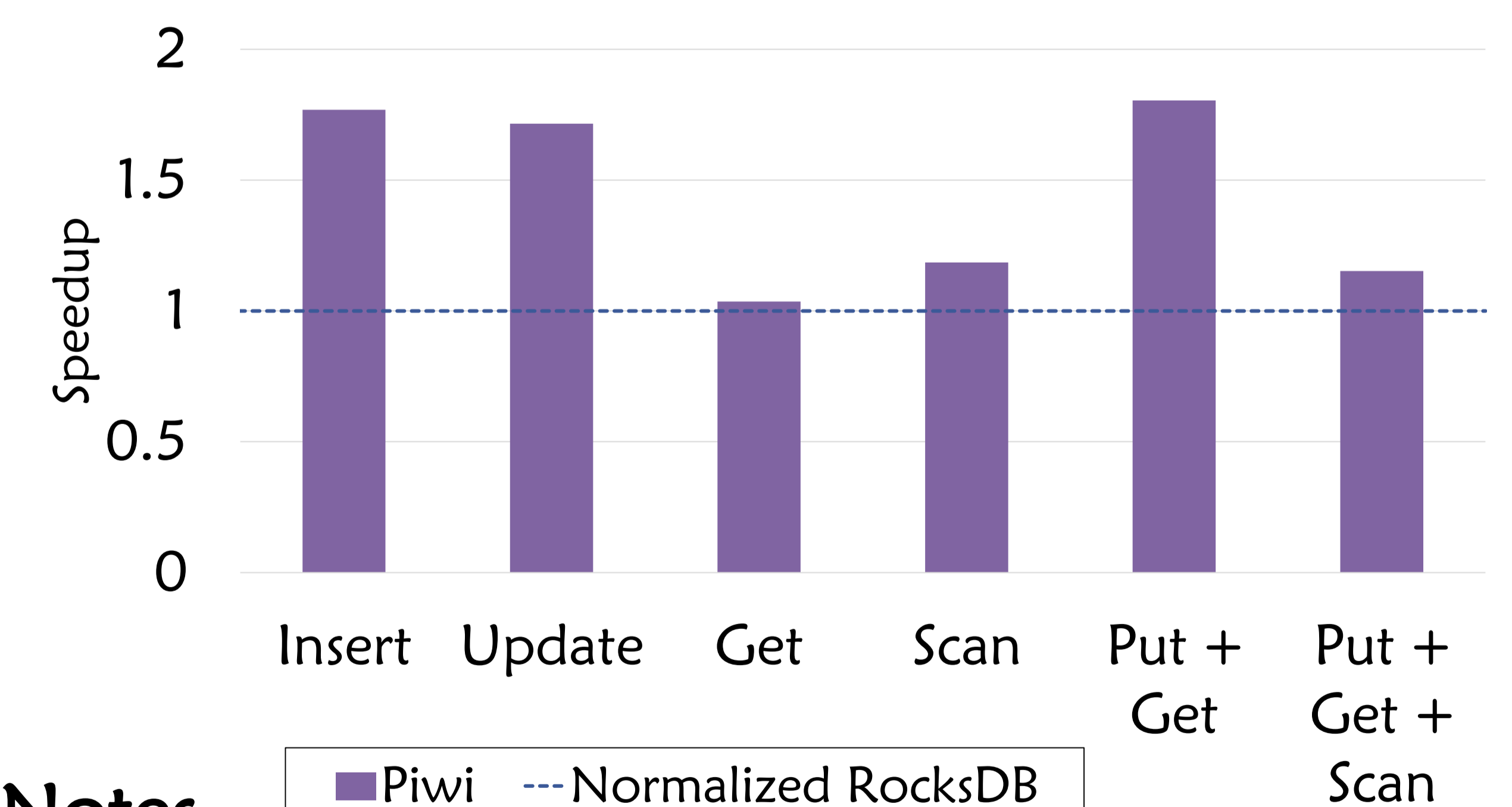


Results

Design advantages



Piwi vs. RocksDB (Facebook)



Notes

- 100M (unique) keys, ~80GB DB size
- 16 application threads, SSD
- Piwi has slightly better (avg.) read and write amplification