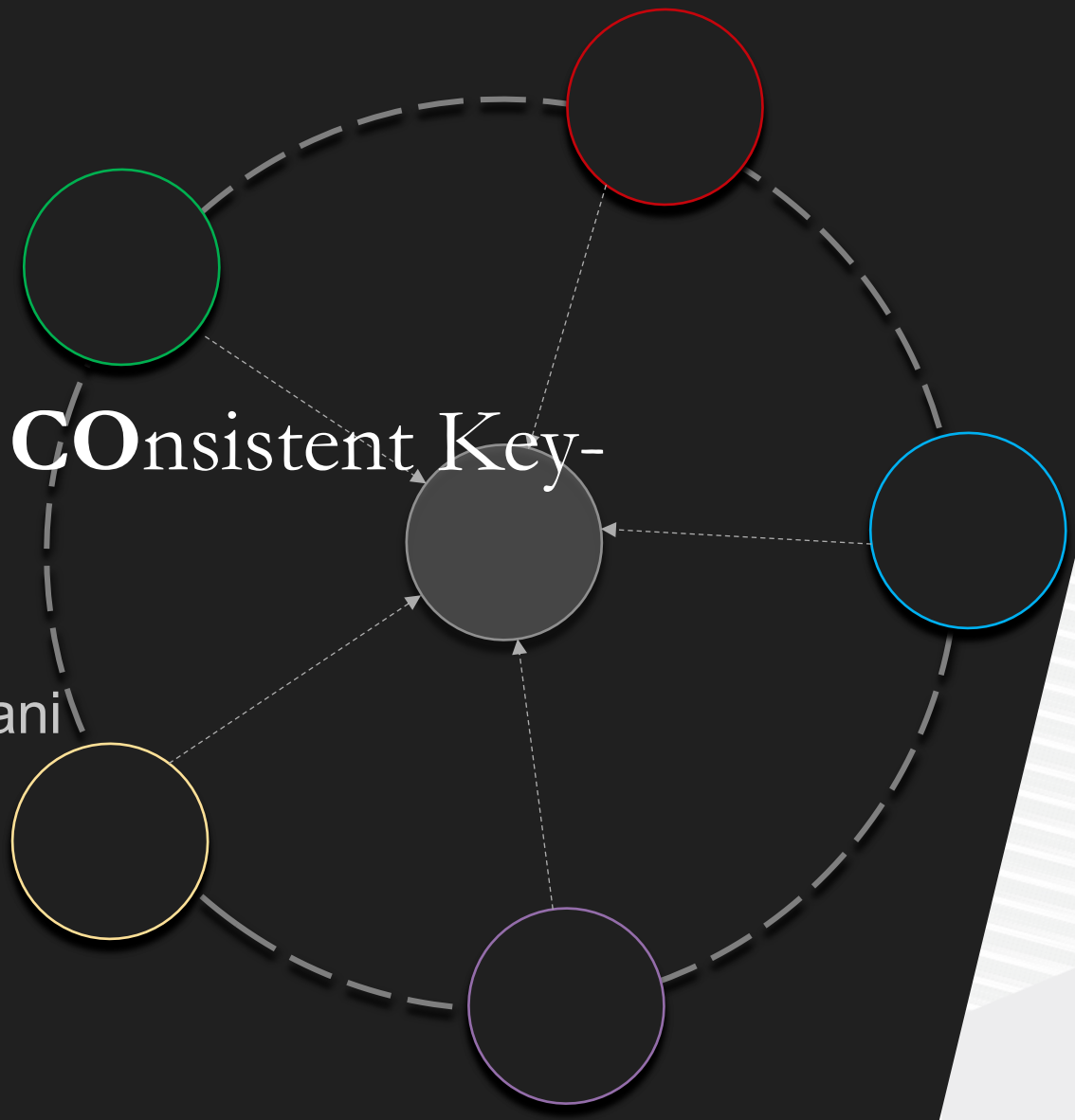
DISCO:

DIstributed **S**trongly- **CO**nsistent Key-Value Store

Adil, Basava, Deepti and Kalyani

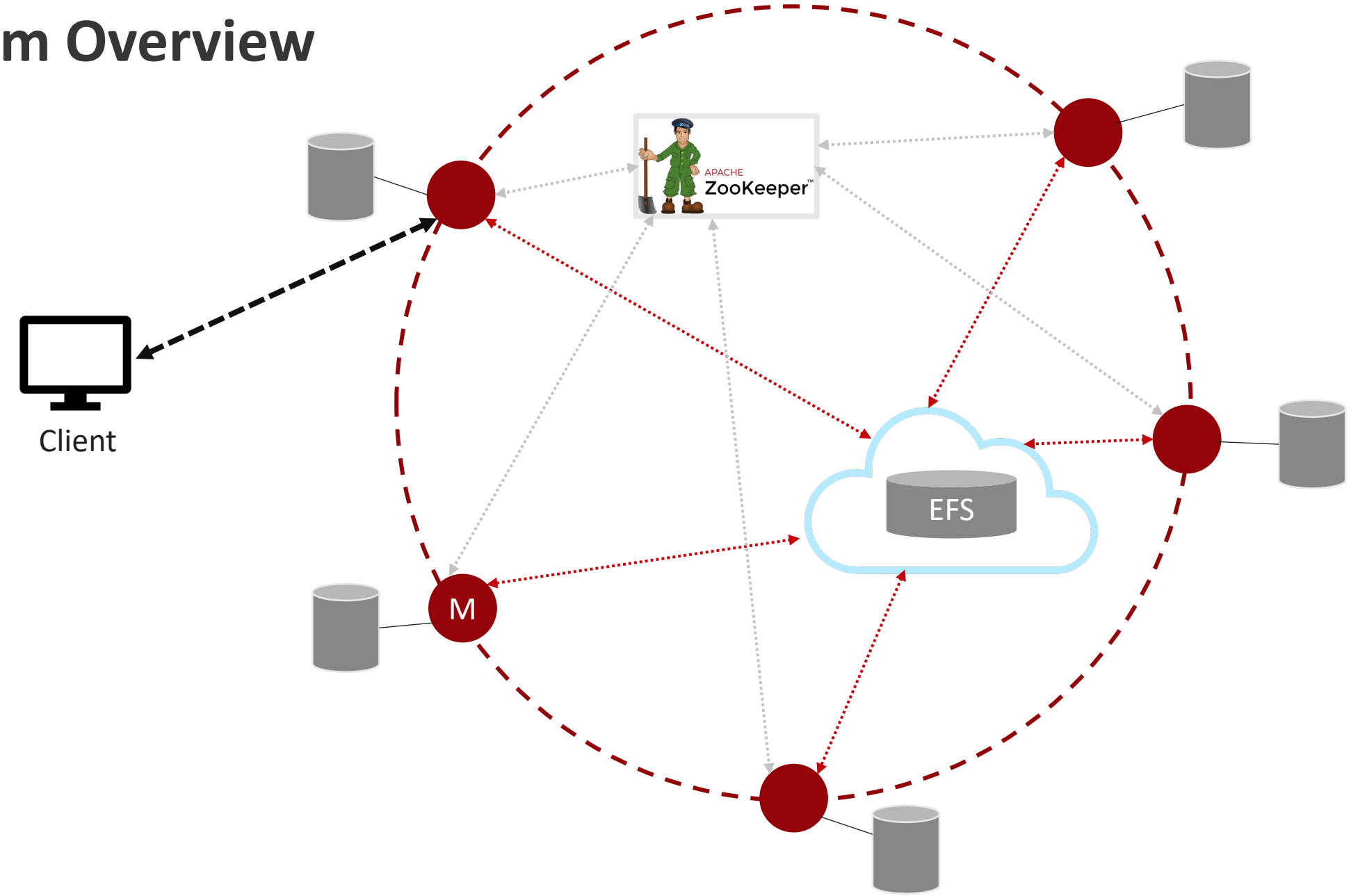


API

- PUT(key, value)
- GET(key)
- DELETE(key)
- GETRANGE(prefix)*

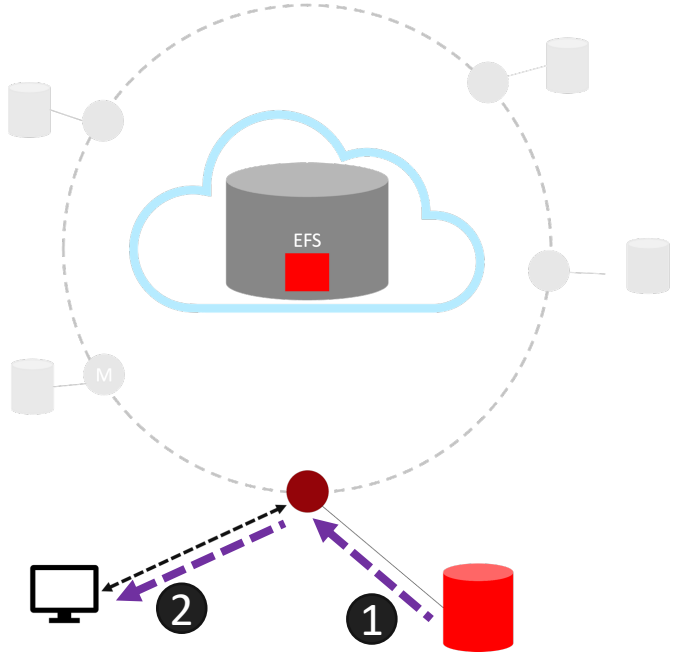
* Only available in locality mode

System Overview

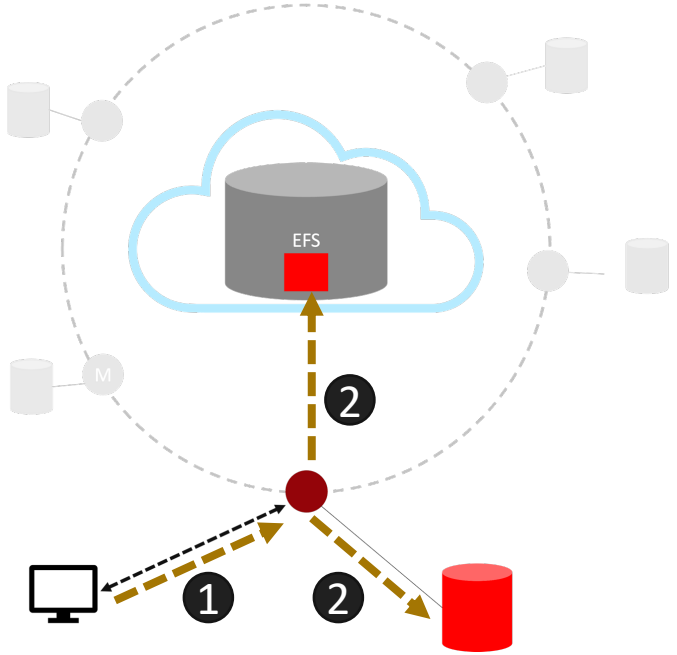


Levels of LevelDBs

- Each node maintains a local copy of its EFS levelDB instance



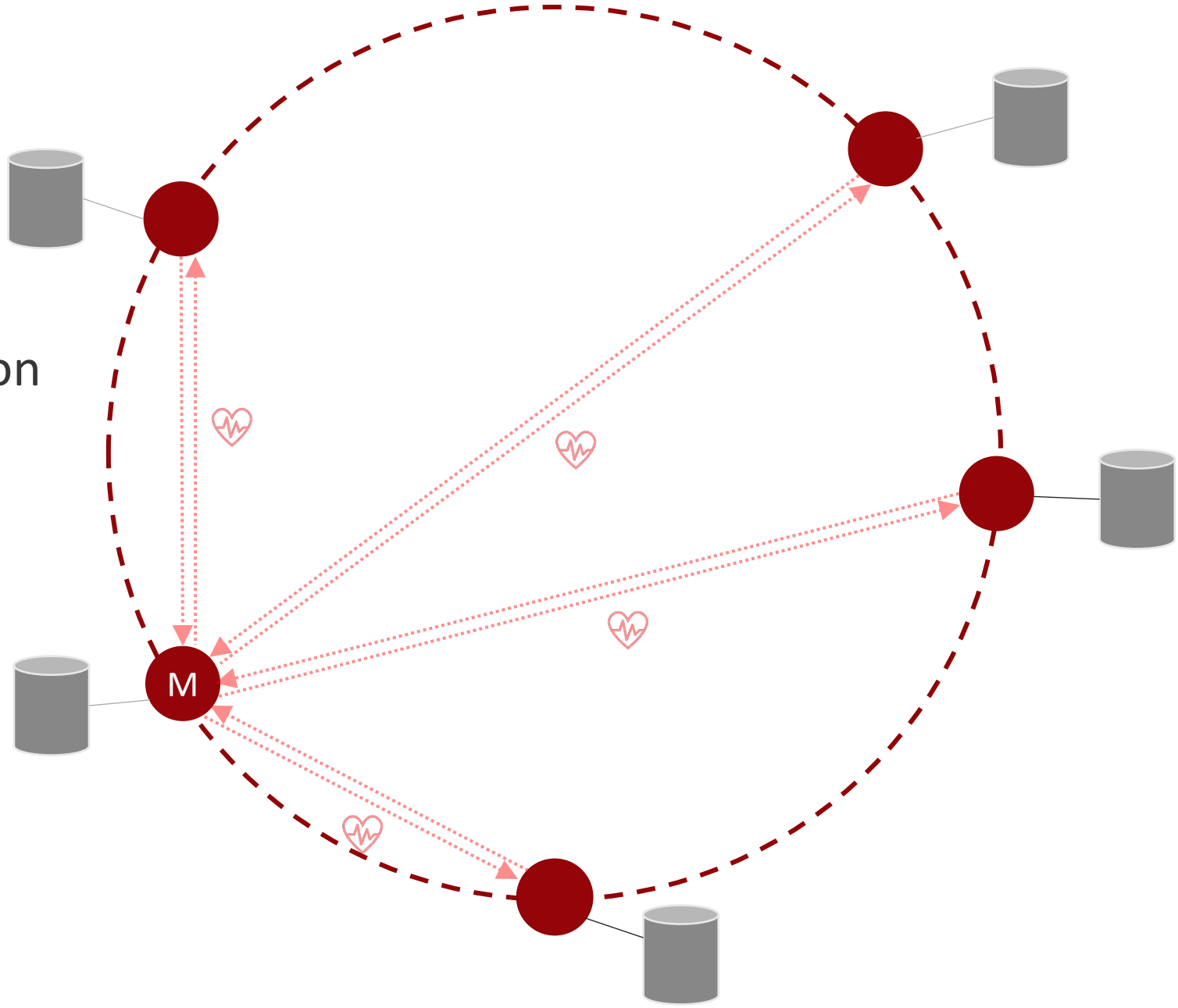
GET and GETRANGE served from local copy



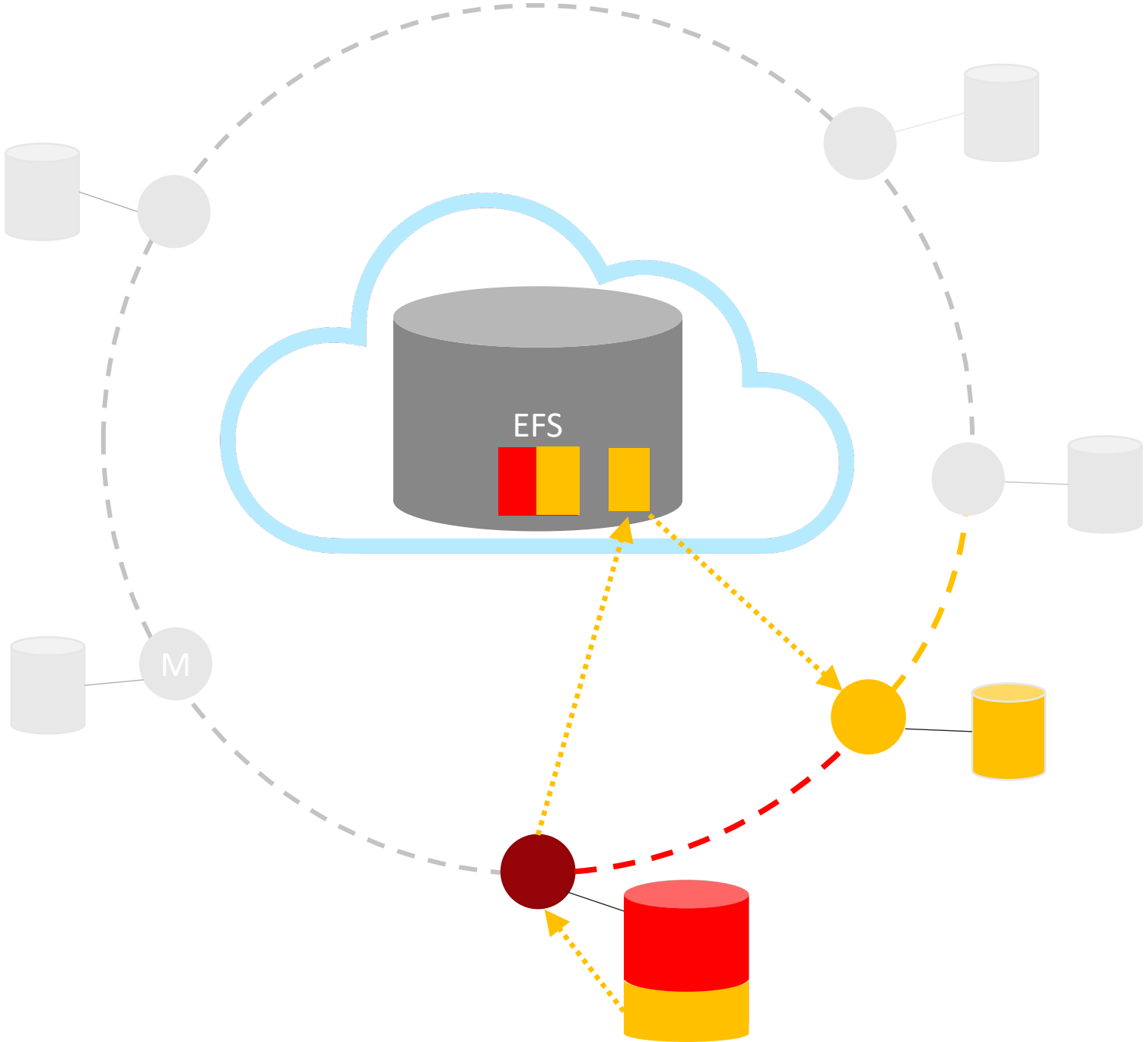
PUTs sent asynchronously to both EFS and local copy.

Master

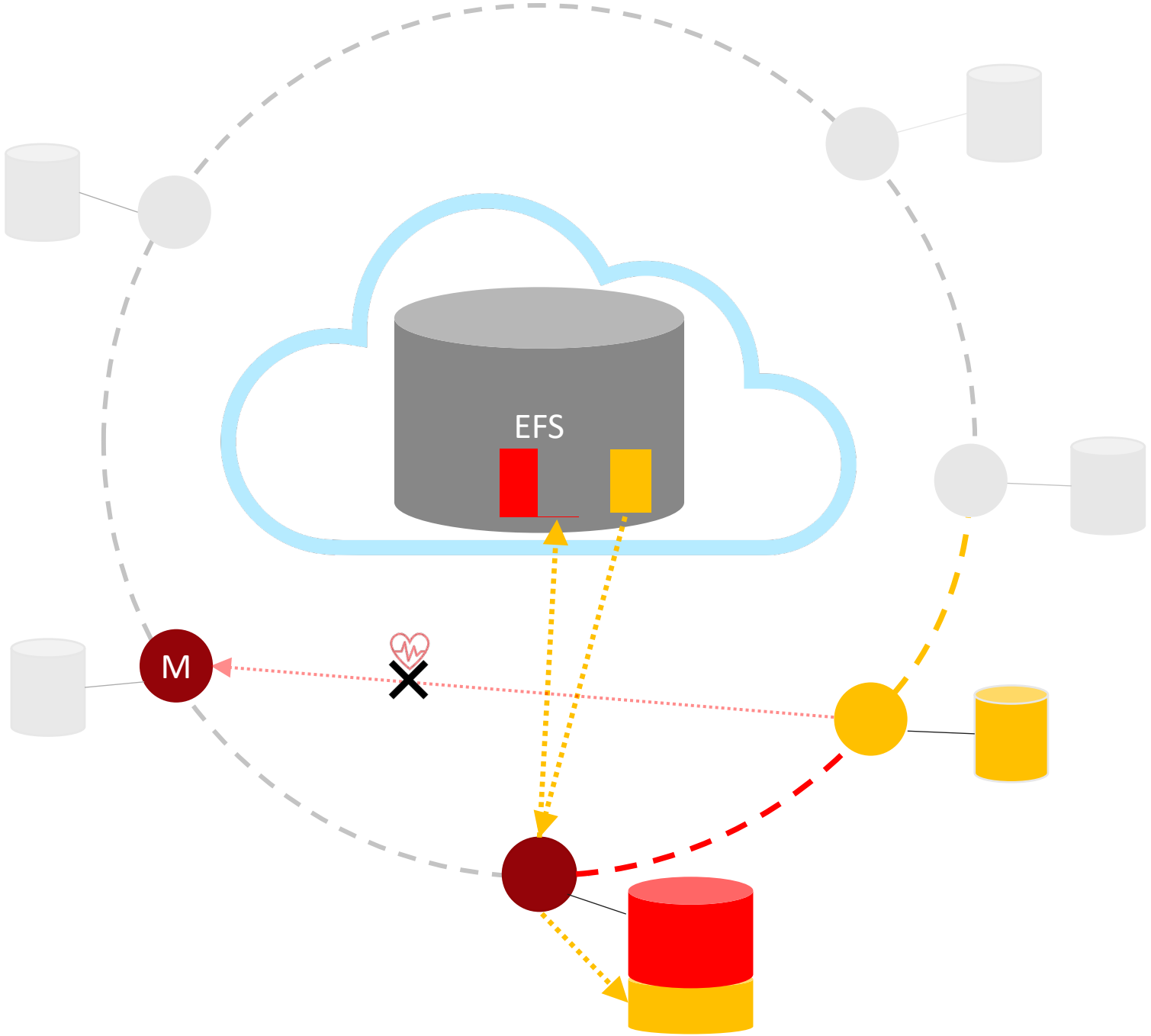
- Heartbeats
- Broadcast Hash Table
- Failure triggers election



Node Joins



Node Exits



Modes

- Default – All writes async. Reads done from local copy.
- Locality – Keys hashed based only on prefix. Supports reading ranges of keys in one RPC call.
- Write – Writes are buffered and written in one batch.
  - Buffer is committed before serving any incoming GET call
  - Minimizes calls to EFS
  - Possibly lose data if a crash occurs!

Crash Consistency

	INIT	GET	PUT	GETRANGE	SPLIT	MERGE
MASTER	✓	✓	✓*	✓	✓	✓
FOLLOWER	✓	✓	✓	✓	✓	✗

A crash during PUT may result in data loss in Write-Batch mode

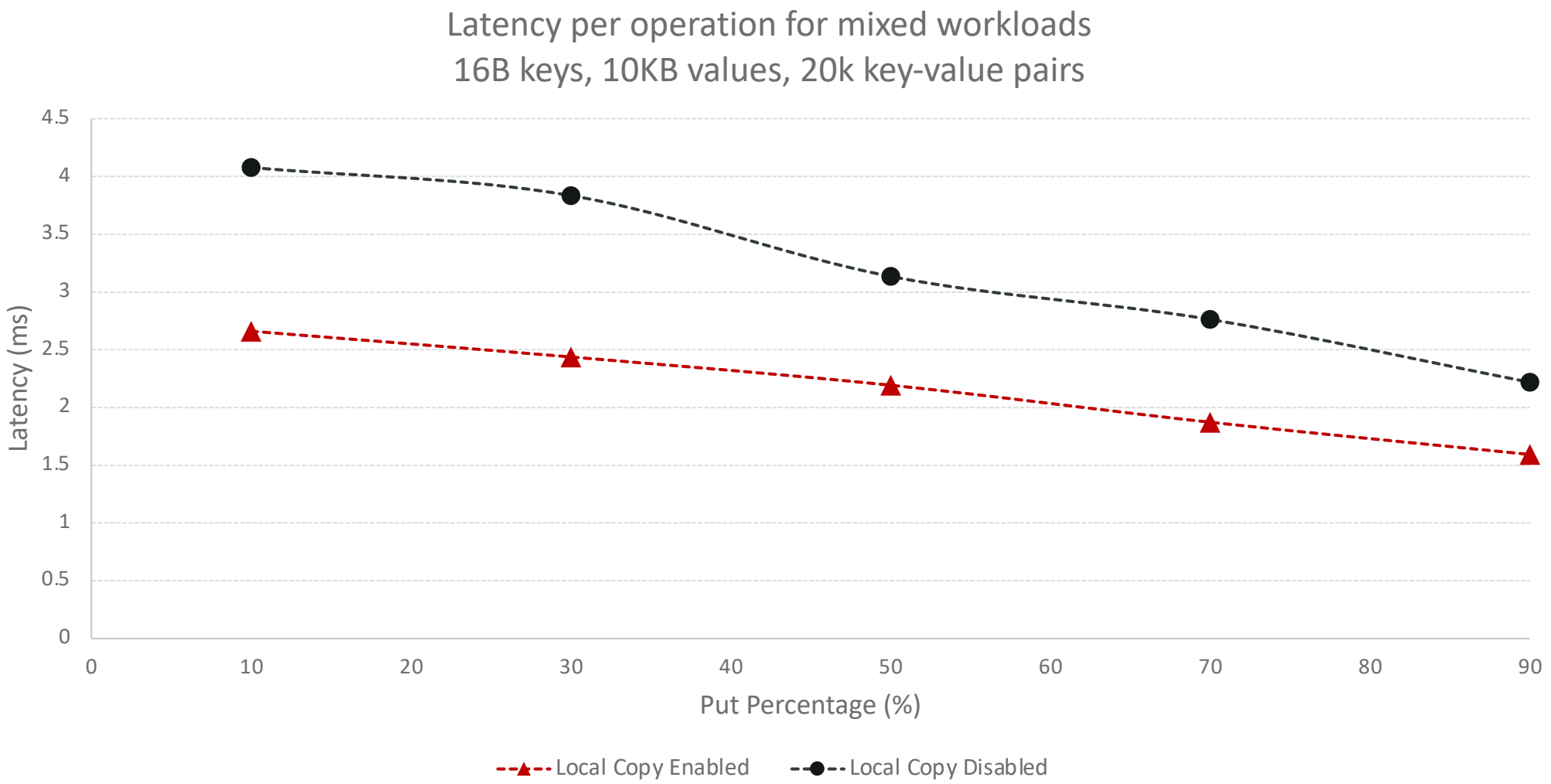
Demo 1

https://drive.google.com/file/d/1XHOCRqLAf8LfT1DVzHAxz977yB6rHs0b/view?usp=sharing

Demo 2

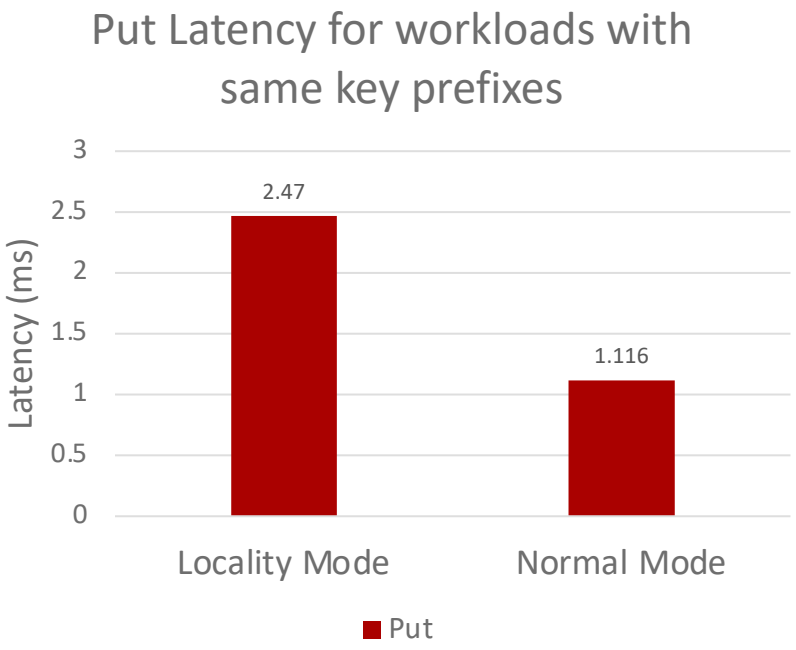
https://drive.google.com/file/d/1Xv_8dBew6OmgO20-uc16iWtsQFluqJ_Z/view?usp=sharing

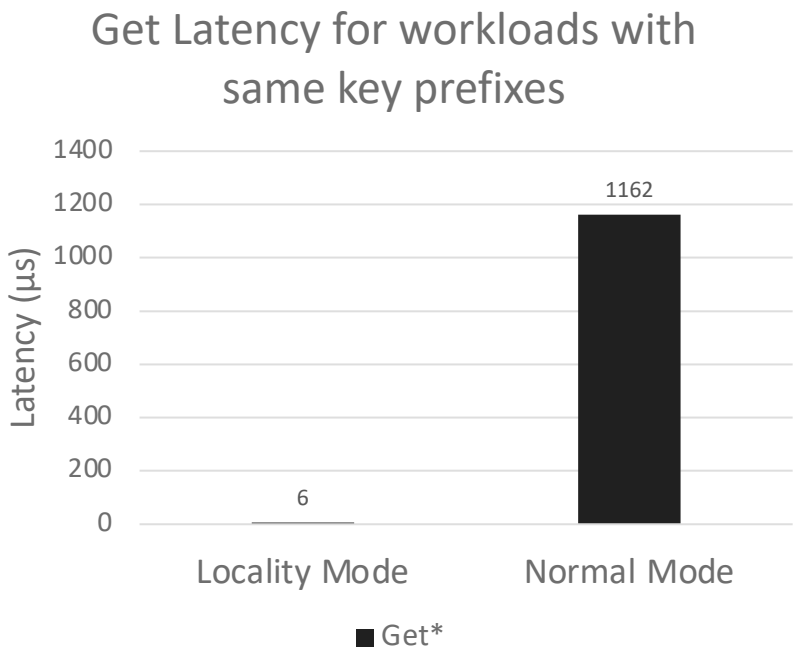
Impact of Local Copy



Prefix GETs in LOCALITY mode

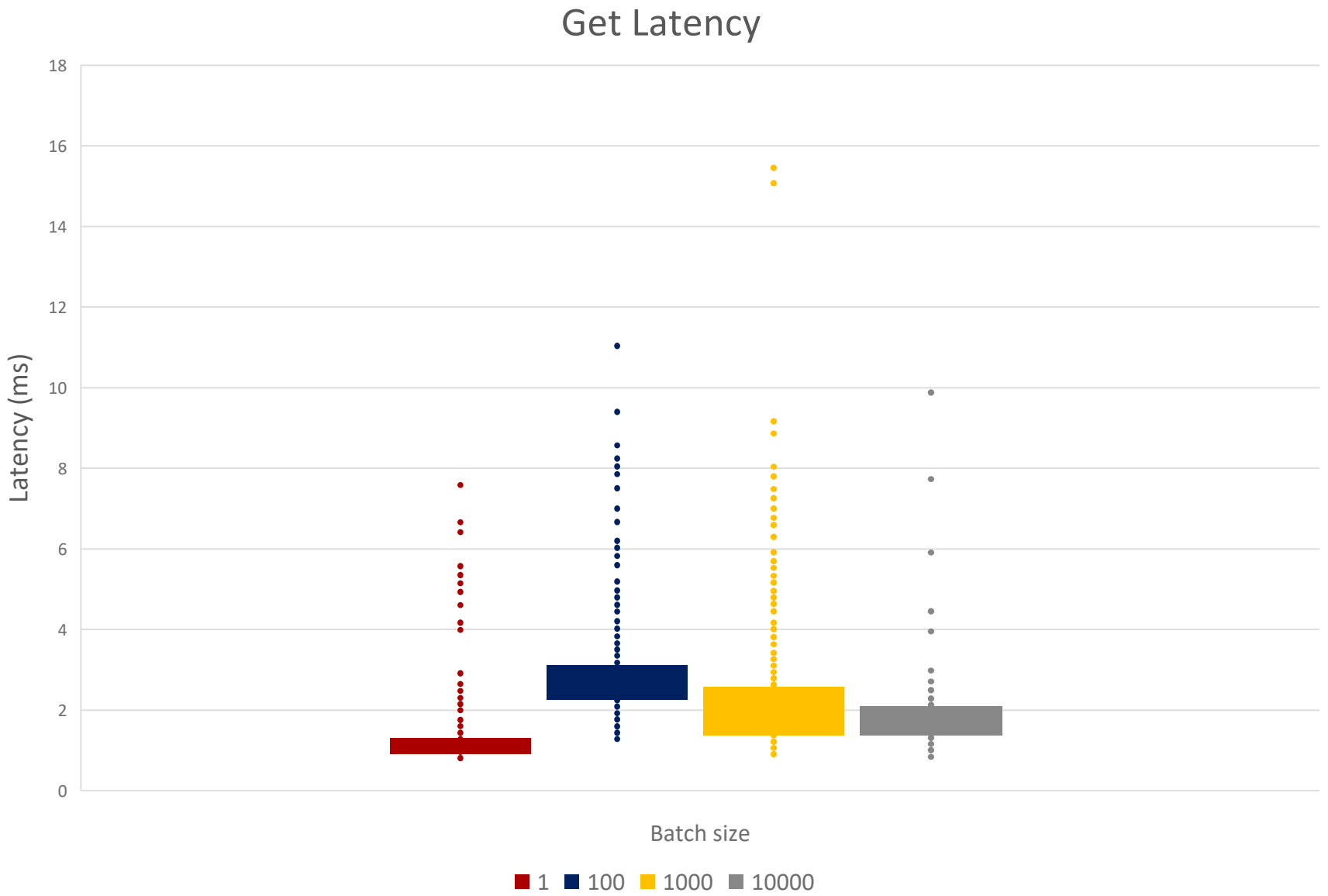
Key: 16B Value:100B 10,000 Key-Value pairs





* GETRANGE in Locality Mode

Cost of Write-optimized Operation



Cost of Write-optimized Operation