RocksDB

Open Discussion on Roadmap and Vision

Dhruba Borthakur

Feb 3, 2016.

How many of you have hands-on experience with RocksDB?

Use RocksDB?

Contribute to RocksDB?

RocksDB is an open source-project

You get to decide and contribute how to take it forward

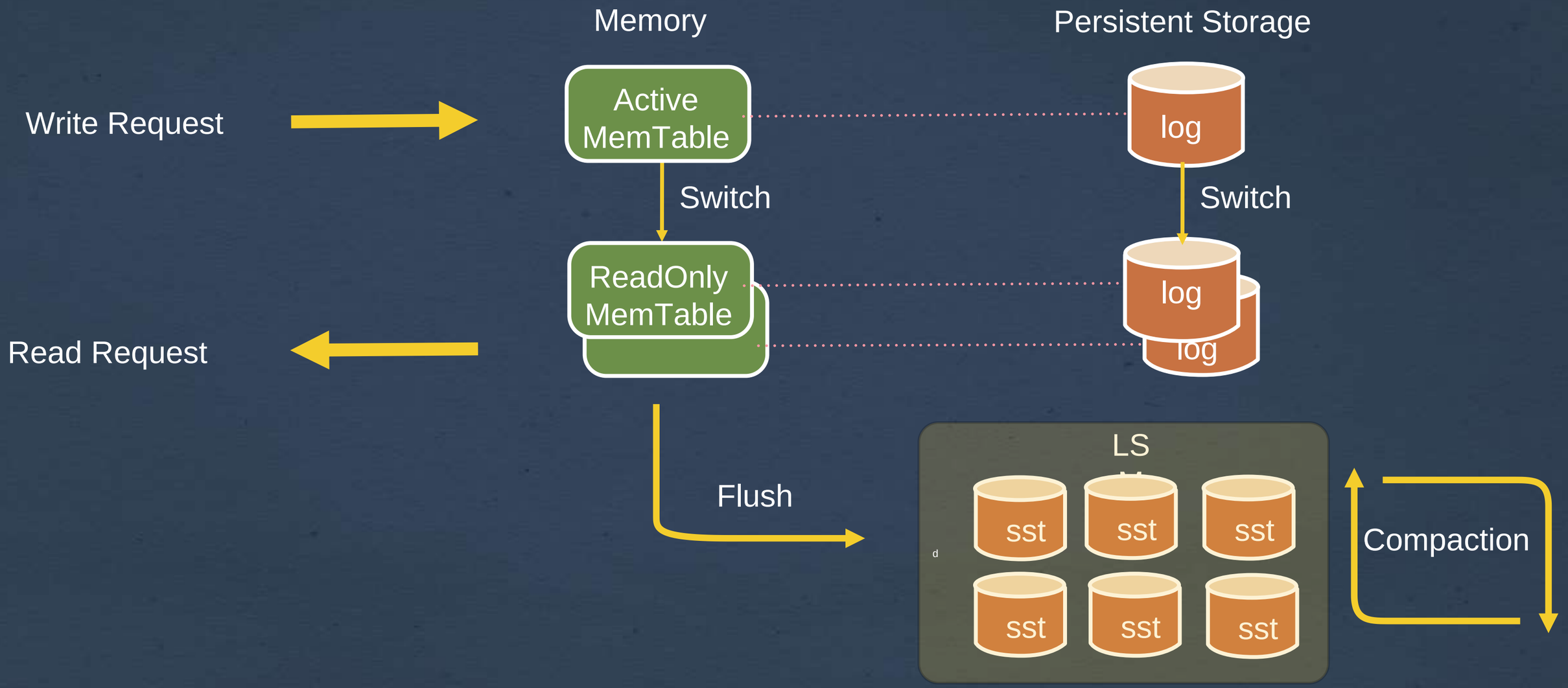
What is RocksDB?

- Key-Value persistent store
- Embedded
- Optimized for fast storage
- Server workloads

RocksDB API

- Keys and values are arbitrary byte arrays.
- Data are stored sorted by key.
- Update Operations: Put/Delete/Merge
- Queries: Get/Iterator

RocksDB Architecture

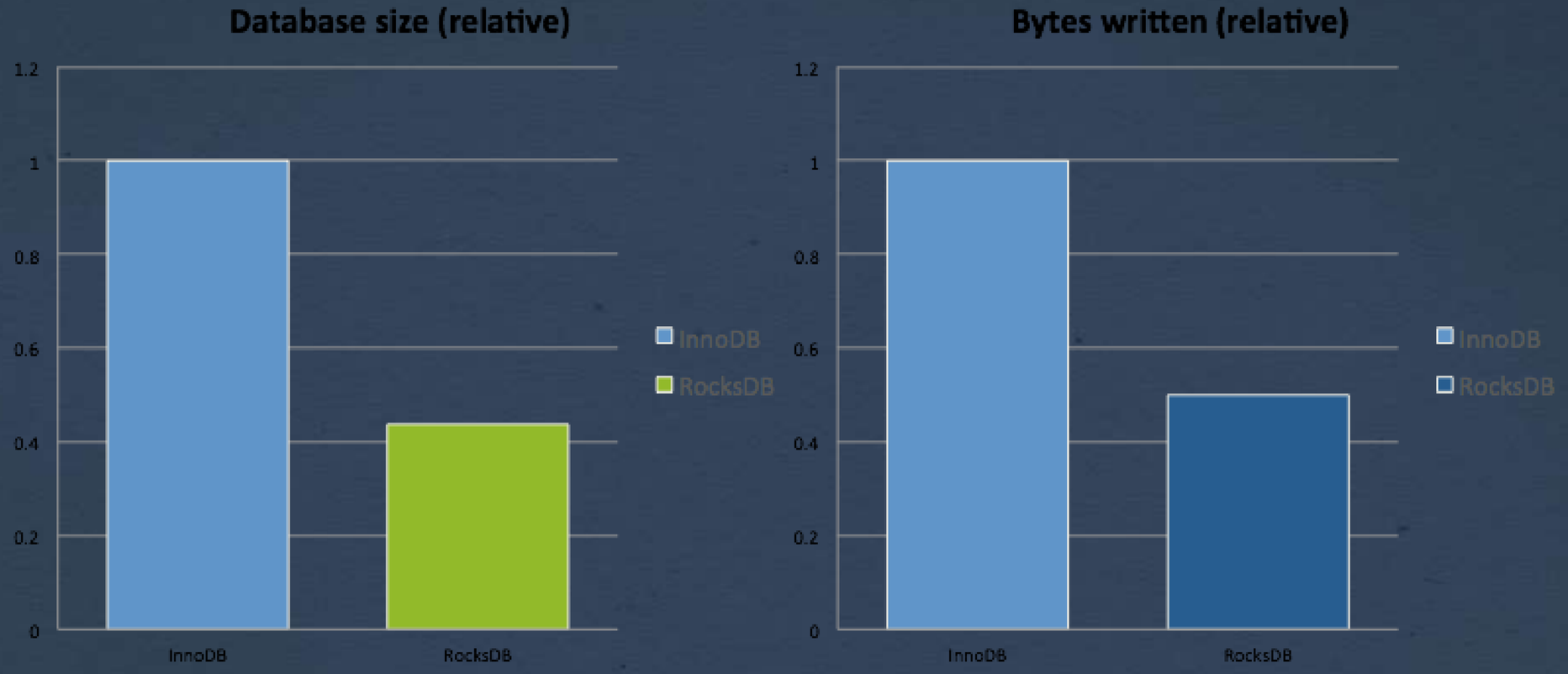


Why does FB develop RocksDB?

RocksDB in the mainstream

- Reduces a 5 TB MongoDB instance to 285 GB on MongoRocks

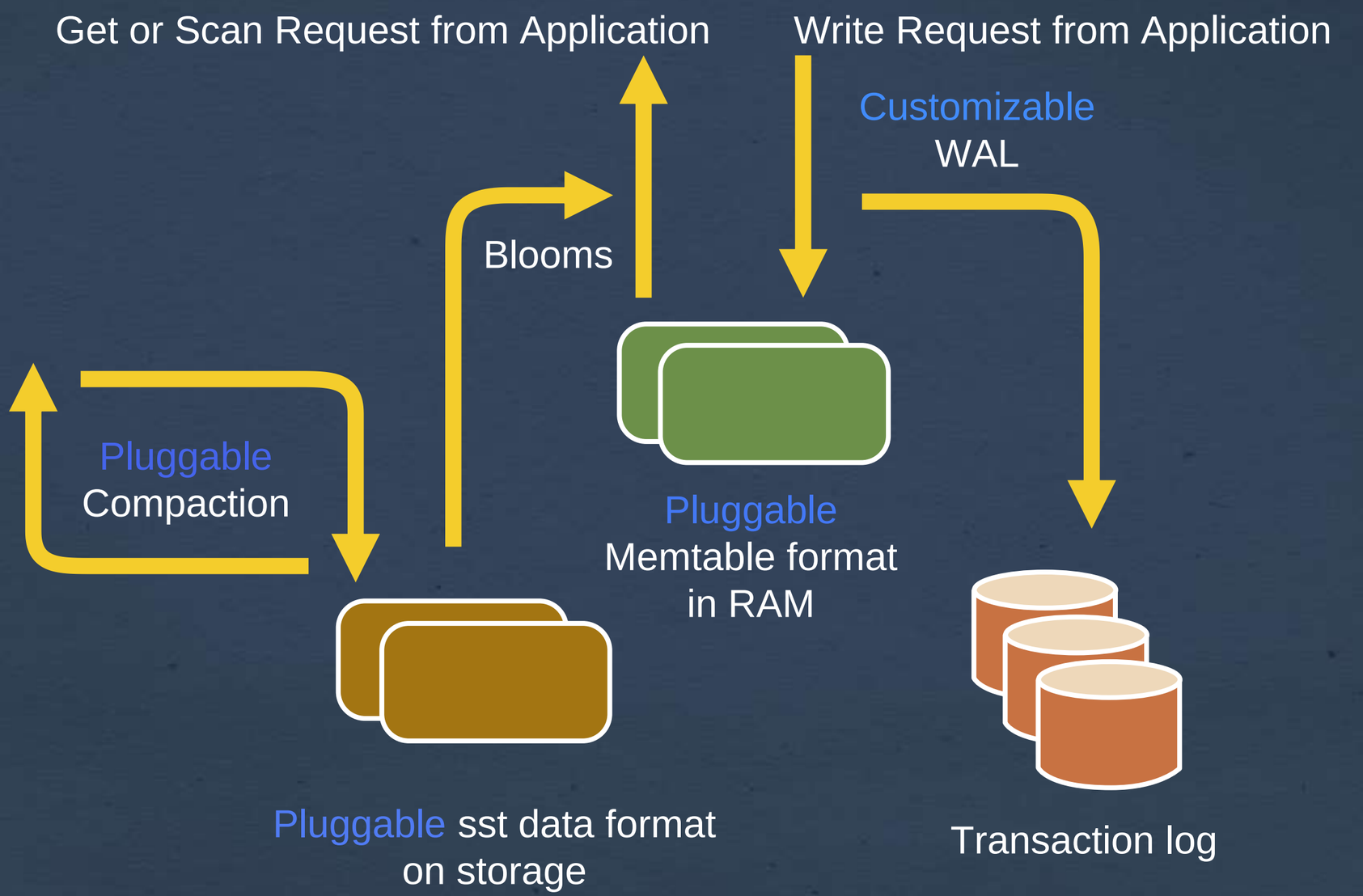
MySQL Innodb vs RocksDB



- LinkBench: open source benchmark for Facebook's workload
- Reduces MySQL flash storage space by 50% for LinkBench

It is easy to use RocksDB in your software

RocksDB: Open & Pluggable



RocksDB is a tool, not a solution by itself.

Embed it into your software solution.

RocksDB Integrations

- LinkedIn Feed , Yahoo Sherpa, Airbnb, Pinterest
- Microsoft Bing Platform (development)
- Apache Samza
- Rippled server (bitcoin style)
- RedHat CEPH
- Open Channel SSD

Any SSD Vendors out here today?

Make RocksDB use your SSD efficiently!

Any FileSystem vendors here today?

Make RocksDB use your filesystem efficiently!

Want to extend RocksDB functionality?

Use StackableDB

e.g. TTL Support, Geo-index, Redis-style

Vision for the future

- Most performant database engine on ram, SSD and disks
  - Optimize for next-generation storage hardware
- Flexibility to be deployed on varied environments
  - RocksDB’s components are pluggable
  - Enables software vendors to customize their solution

Roadmap for 2016

What do you want from RocksDB?

Here are some probable enhancements for RocksDB in 2016

Raw Devices

RocksDB would run on a device lun

RocksDB could use directIO and bypass OS cache

MergeOperator & CompactionFilter

RocksDB could support a lua-based Filter

RocksDB could support a Javascript-based Filter

Reduce storage footprint

Dictionary based compression

More tradeoffs between cpu and compression (zstd, etc)

Production support

Database repair

Standalone utilities to inspect files, blocks, MANIFEST

Tiered Storage

Single db can have RAM, SSD, disk & next-gen-hardware

Intelligent movement of data between storage tiers

Open Discussion