The FuzzyLog: A Partially Ordered Shared Log

Joshua Lockerman, Jose M. Faleiro, Juno Kim, Soham Sankaran
Daniel J. Abadi, James Aspnes, Siddhartha Sen, Mahesh Balakrishnan

Yale University      UC Berkeley      UC San Diego      Cornell University
University of Maryland College Park    Microsoft Research    Facebook
(work done at Yale University)
distributing state

- consistency
- durability
- scalability
- availability

... complex systems!
shared log → simple systems

sequencer

replicated state

updates in shared log
shared log → simple systems

✓ consistency
✓ durability
✗ scalability
✗ availability (network partitions)

sequencer

replicated state

updates in shared log
applications are partially ordered

Database

Alice
applications are partially ordered
applications are partially ordered

**Database**

- USA
- Asia

*can we provide the simplicity of a shared log without imposing a total order?*
two sources of partial order

- data shards
- geo-replication
FuzzyLog – a partially ordered log

Database

Alice

Bob

Bob*

A → G → J

B → D → I

C → E → J → K
outline

• abstraction
• applications
• implementation
• evaluation
one color per data shard

Database

Bob

reading

appending
one color per data shard

Database

Alice

Bob

H → I → J → K

A → B → C → D

Appending multi-color append
each color is in multiple regions

USA

geo-replication

Asia
void append(colors, data);

USA

{ A → G → H → I }

{ B → D }

Asia

{ C → E }

{ dashed line connecting I to C/E }
void sync(callback);
recap: the FuzzyLog API

• append a node to one or more colors
  • each color is totally ordered within a region
  • each color is causally ordered across regions.

• sync nodes in a single color
  • nodes are seen in log-order
outline

• abstraction
• applications
• implementation
• evaluation
applications

• transactions: AtomicMap

• best-effort consistency: CAPMap
AtomicMap: write-only tx

TX:
Alice = $5
Bob = $0
Bob has $5

Alice has $0

AtomicMap: read-write tx

Transfer Alice’s money to Bob

Alice had $5

Bob has $5

Alice has $0
applications

• transactions: AtomicMap

• best-effort consistency: CAPMap
CAPMap: best-effort consistency
CAPMap: best-effort consistency
CAPMap: best-effort consistency
outline

• abstraction
• applications
• implementation
• evaluation
implementation

- single region
- multiple regions
FuzzyLog implementation

Chain Replication

FuzzyLog storage servers

application servers
FuzzyLog mechanism: multi-appends

FuzzyLog storage servers

application servers

Chain Replication
FuzzyLog mechanism: multi-appends
FuzzyLog mechanism: multi-appends

Phase 1

Client A

Server 1

Server 2

Client B
FuzzyLog mechanism: multi-appends

Phase 1
FuzzyLog mechanism: multi-appends

Phase 1
FuzzyLog mechanism: multi-appends

Phase 1

Max = 12

Server 1

Server 2

Client A

Client B

Max = 13
FuzzyLog mechanism: multi-appends

multi-appends are assigned unique timestamps.
FuzzyLog mechanism: multi-appends

the FuzzyLog supports atomic multi-appends without a centralized sequencer.
implementation

- single region
- multiple regions
FuzzyLog mechanism: shadow logs

USA

Asia

A

B

...
FuzzyLog mechanism: shadow logs
outline

• abstraction
• applications
• implementation
• evaluation
AtomicMap performance

Ms of puts/sec

Workload (% of multi-shard puts)

# of clients

AtomicMap/ Dapple

1/8
2/8
4/8
8/8
16/16
32/16

0%
0.1%
1%
10%
100%
CAPMap performance

Latency (ms) vs. Time Elapsed (ms)

- Primary puts
- Secondary puts

CAPMap performance
other applications

- transactions: AtomicMap
- best-effort consistency: CAPMap
- RedBlue
- CRDTs
- transactional CRDTs
- FuzzyZookeeper
conclusion

• a shared log simplifies distributed systems but imposes a total order

• the FuzzyLog is a partially ordered shared log:
  • data sharding
  • geo-replication

• we obtain the simplicity of the shared log approach without its limitations!
Thank You!

Mossy Log by Shawn Harquail used under CC BY-NC

Globes from:
https://visualpharm.com/free-icons/globe-595b40b65ba036ed117d2d7e
https://visualpharm.com/free-icons/globe-asia-595b40b65ba036ed117d2d7f