

Scaling Dropbox

PRESLAV LE, NOVEMBER 7TH, 2016



breed 8:44 PM

Totally random question... but we've seen a few TOR's go down in US central... spontaneously reboot. Is anything interesting changed with MP code? or backfill, etc?



breed 9:09 PM

OK. We just lost 5 more switches. Currently I'm showing 7 racks down

@here

Zone
(west)

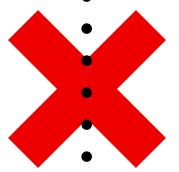
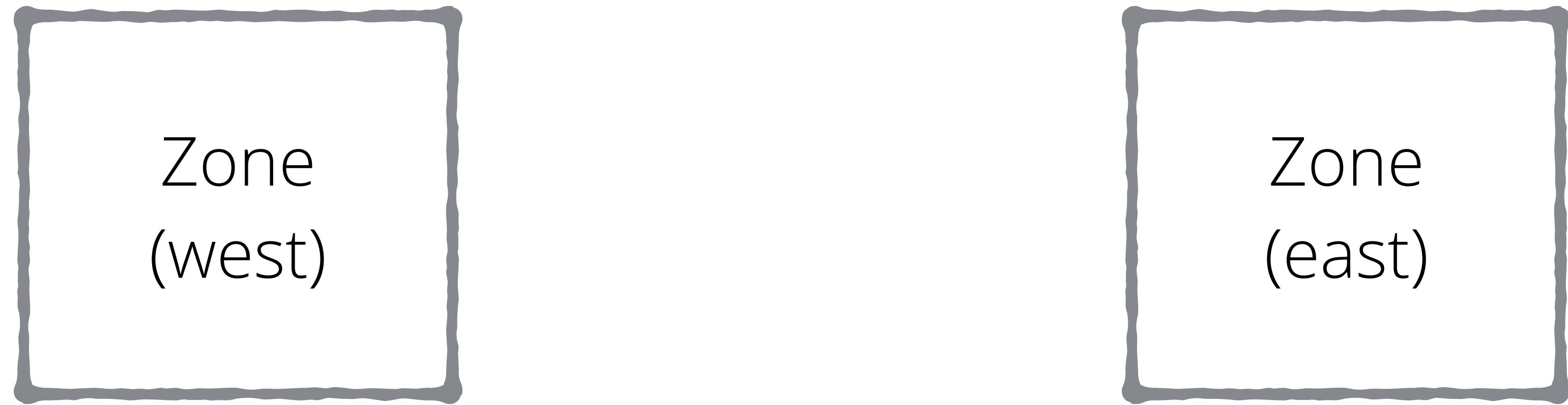
Zone
(east)

Zone
(central)

block.dropbox.com



block.dropbox.com



block.dropbox.com

Fear of the unknown

MEMORY LEAK



jason 9:27 AM

slow build up to 78%

then very quickly the switch goes down

SYNCHORNIZATION EVENT



mcot 9:41 AM

tor checker was enabled March 7 -- a couple people asked



mcot 9:41 AM

it does send the show mac address table command



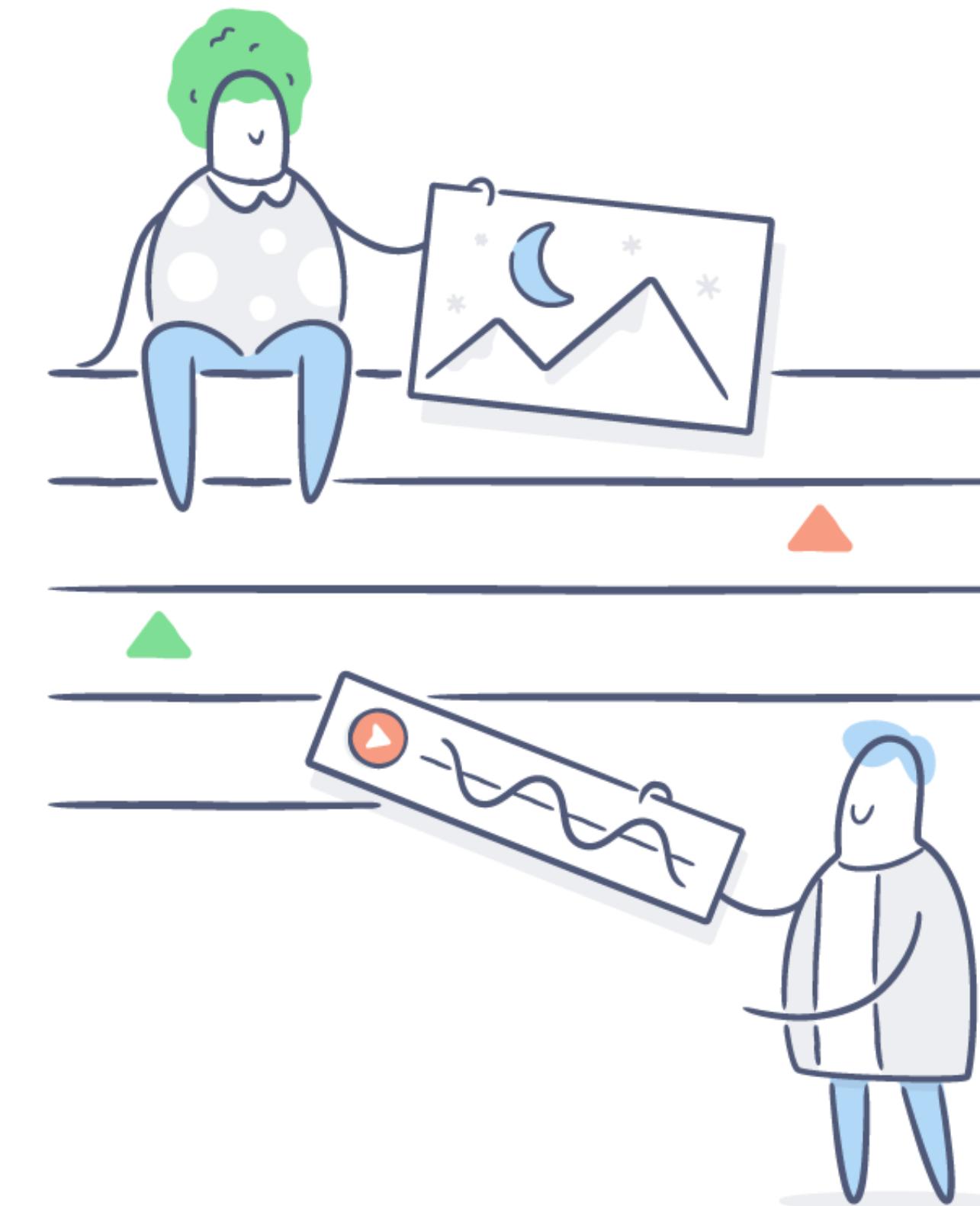
jason 9:43 AM

@presley: ^ that would explain the synchronization across racks with different uptimes

Success story

TODAY'S TALK

- 2012
- SCALING CHALLENGES
- 2016
- Q&A



PRESLAV LE

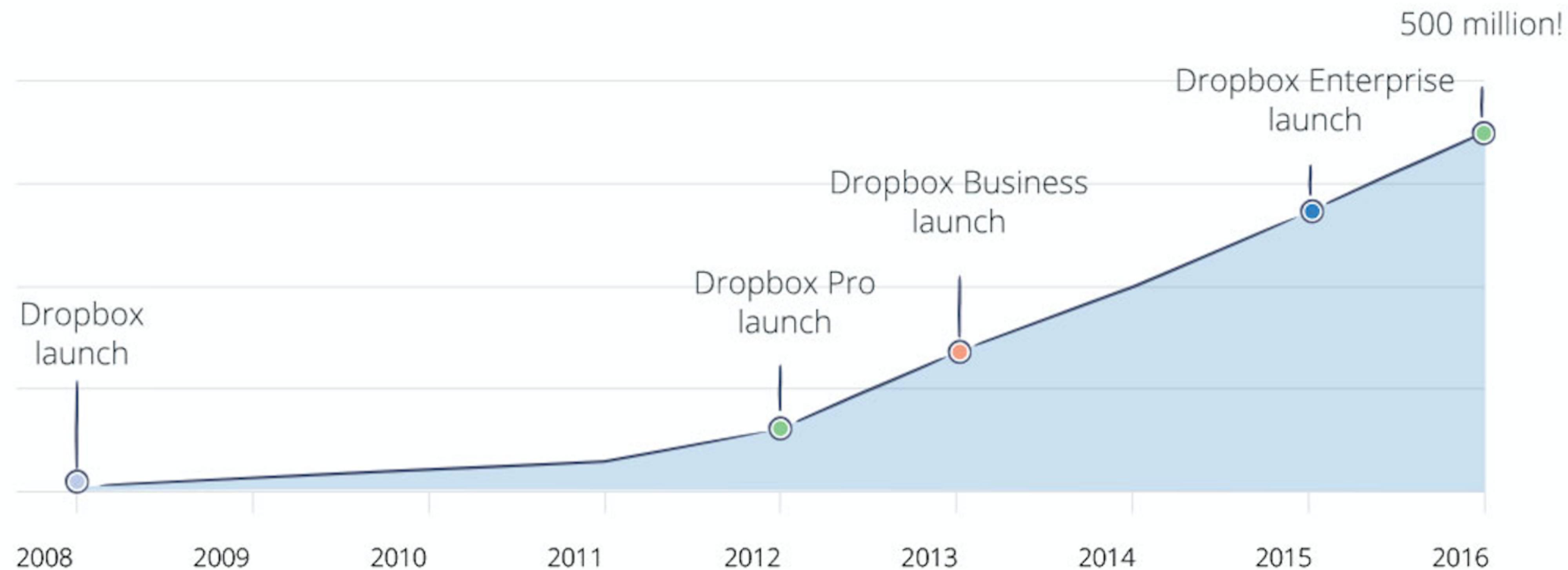


- At Dropbox since 2013
- **Projects:** Magic Pocket,
Infrastructure Performance,
Traffic team

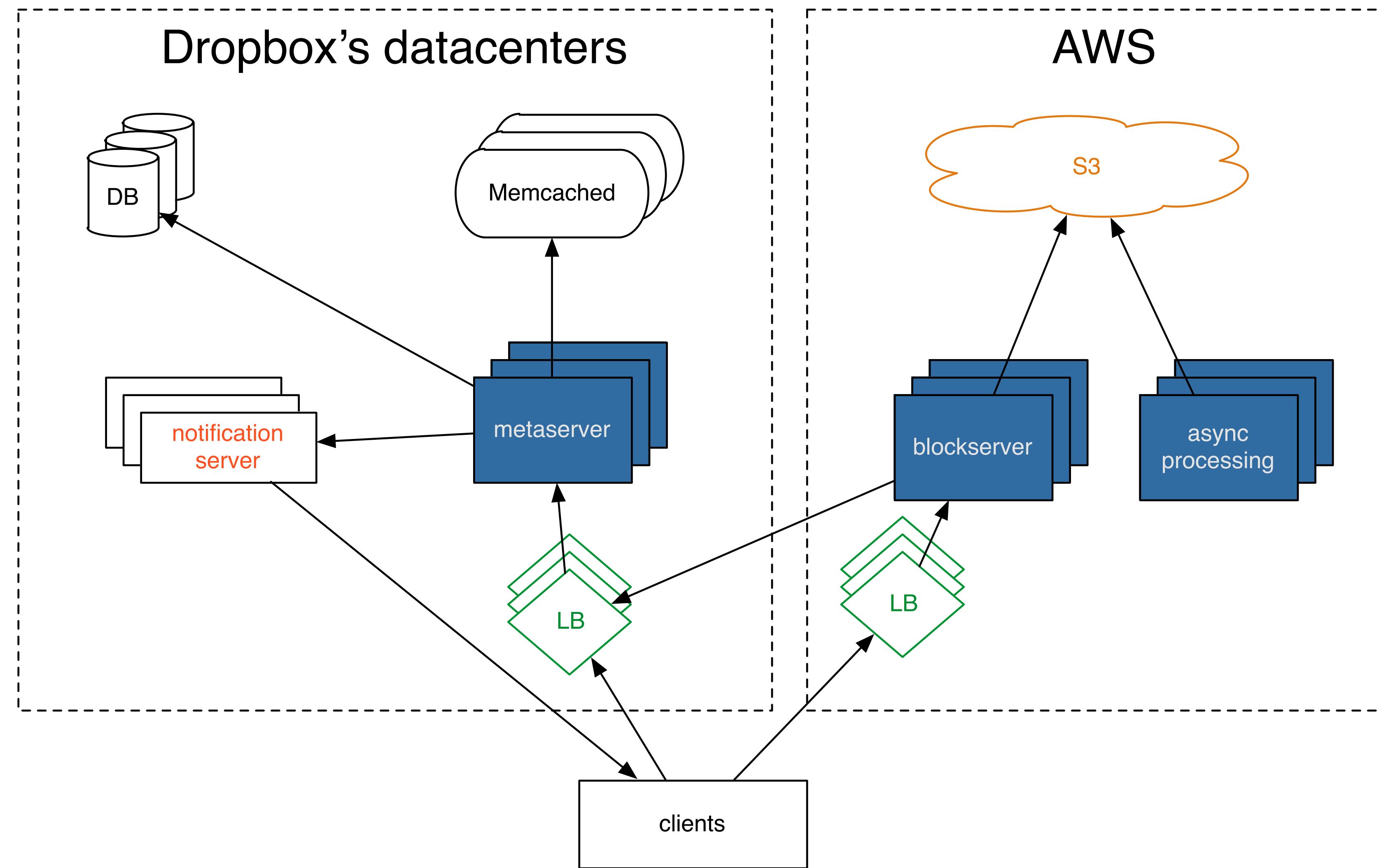
FILE, SYNC & SHARE



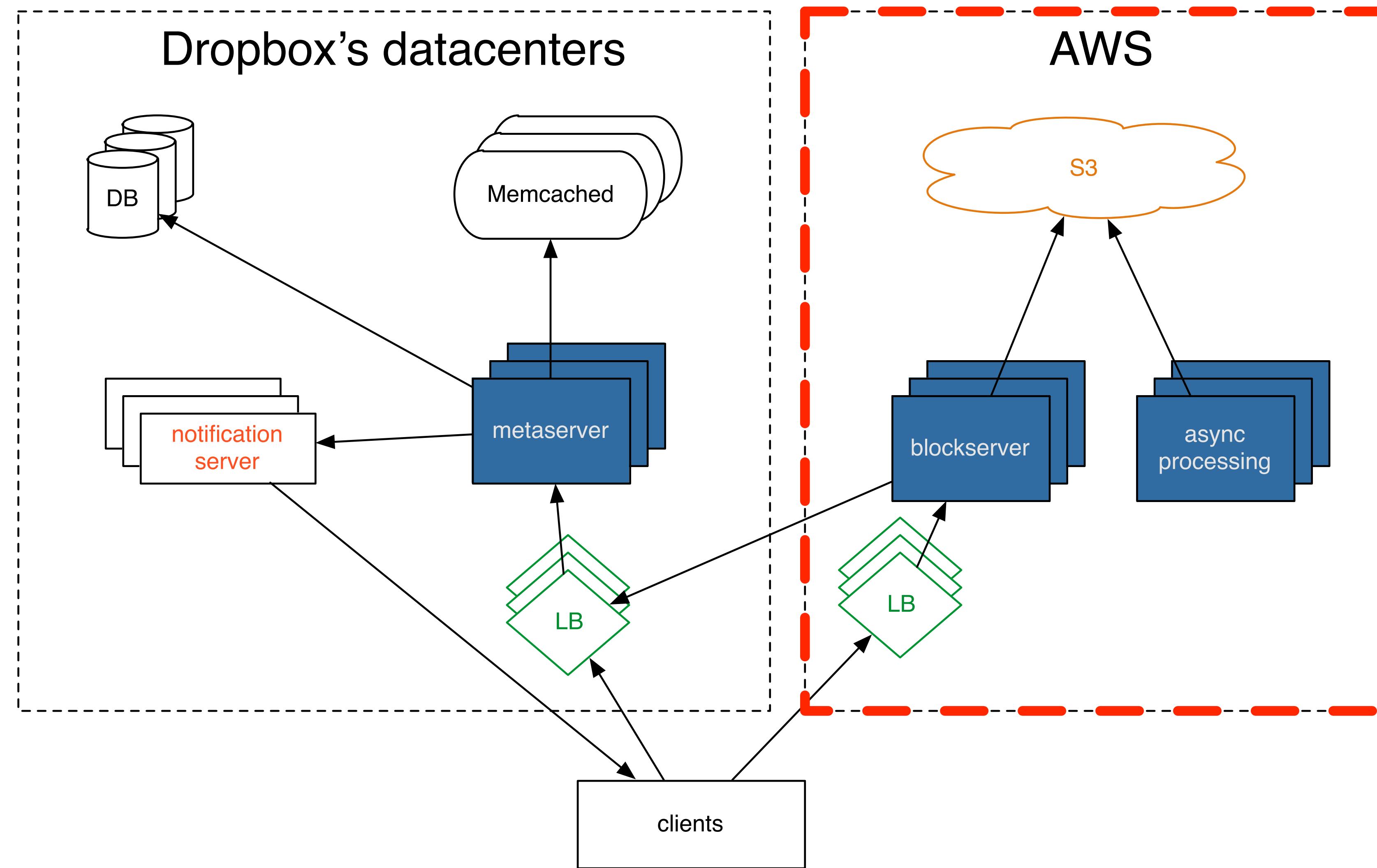
500 MILLION USERS



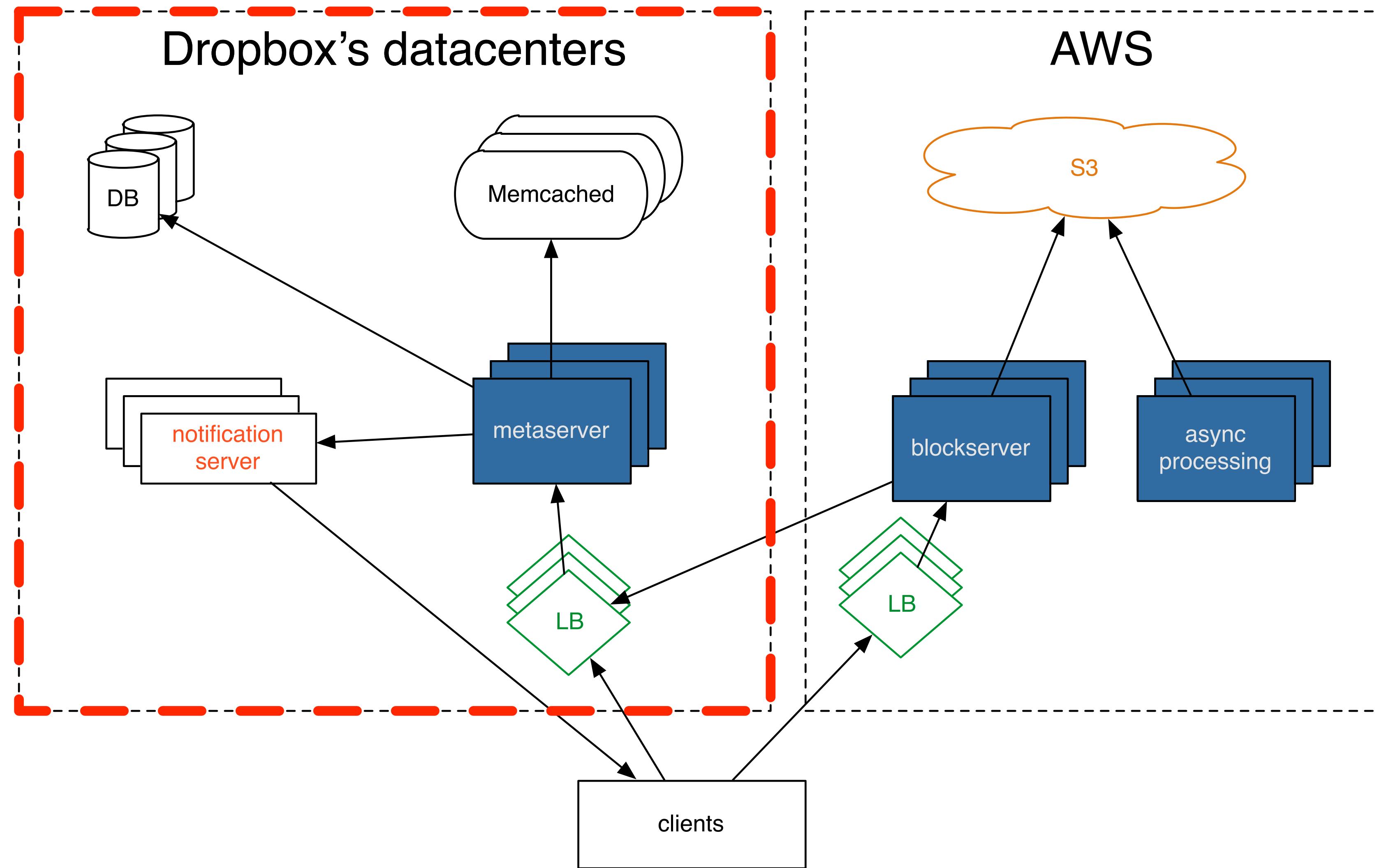
2012



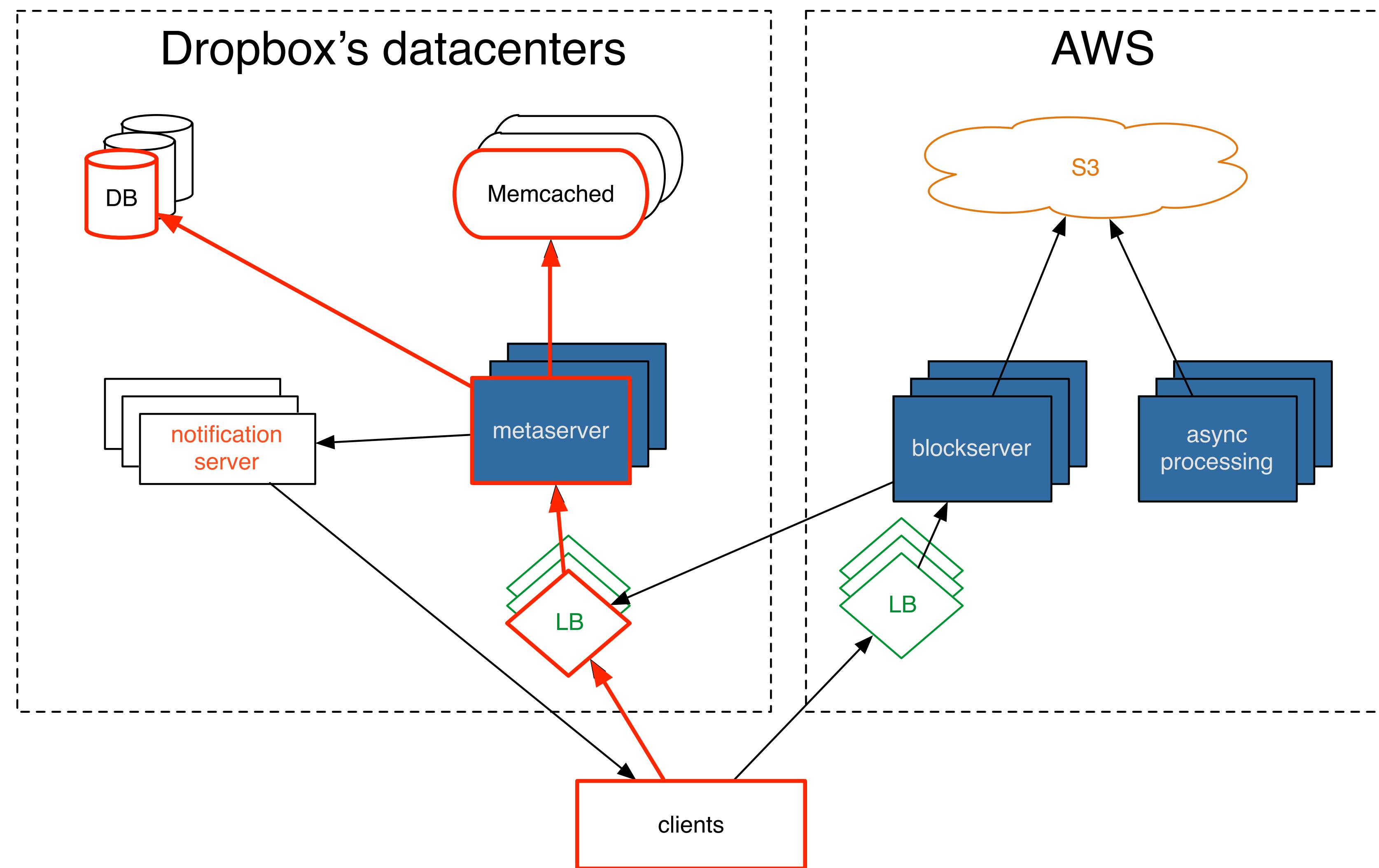
BLOCK DATA IN S3



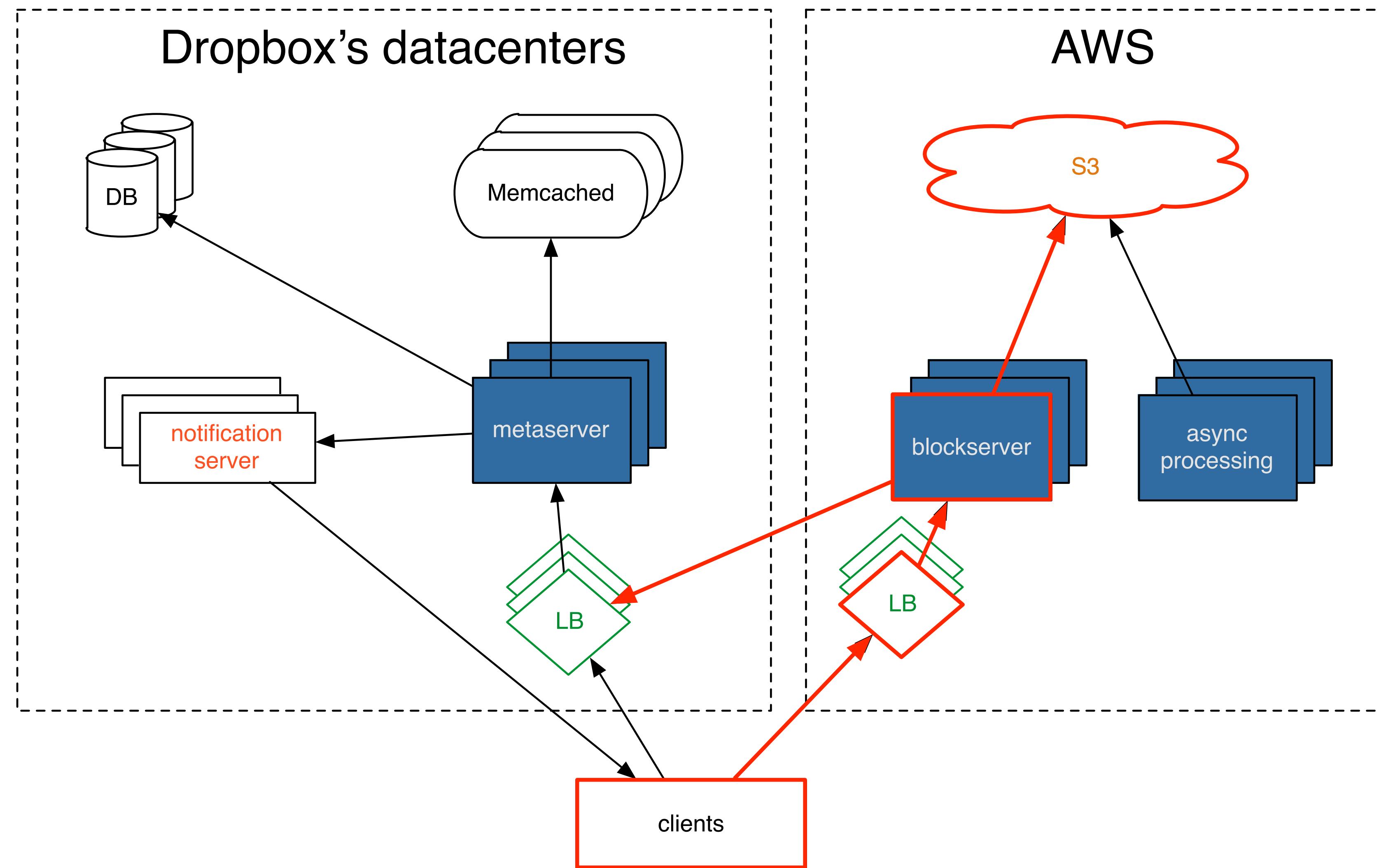
METADATA IN MYSQL



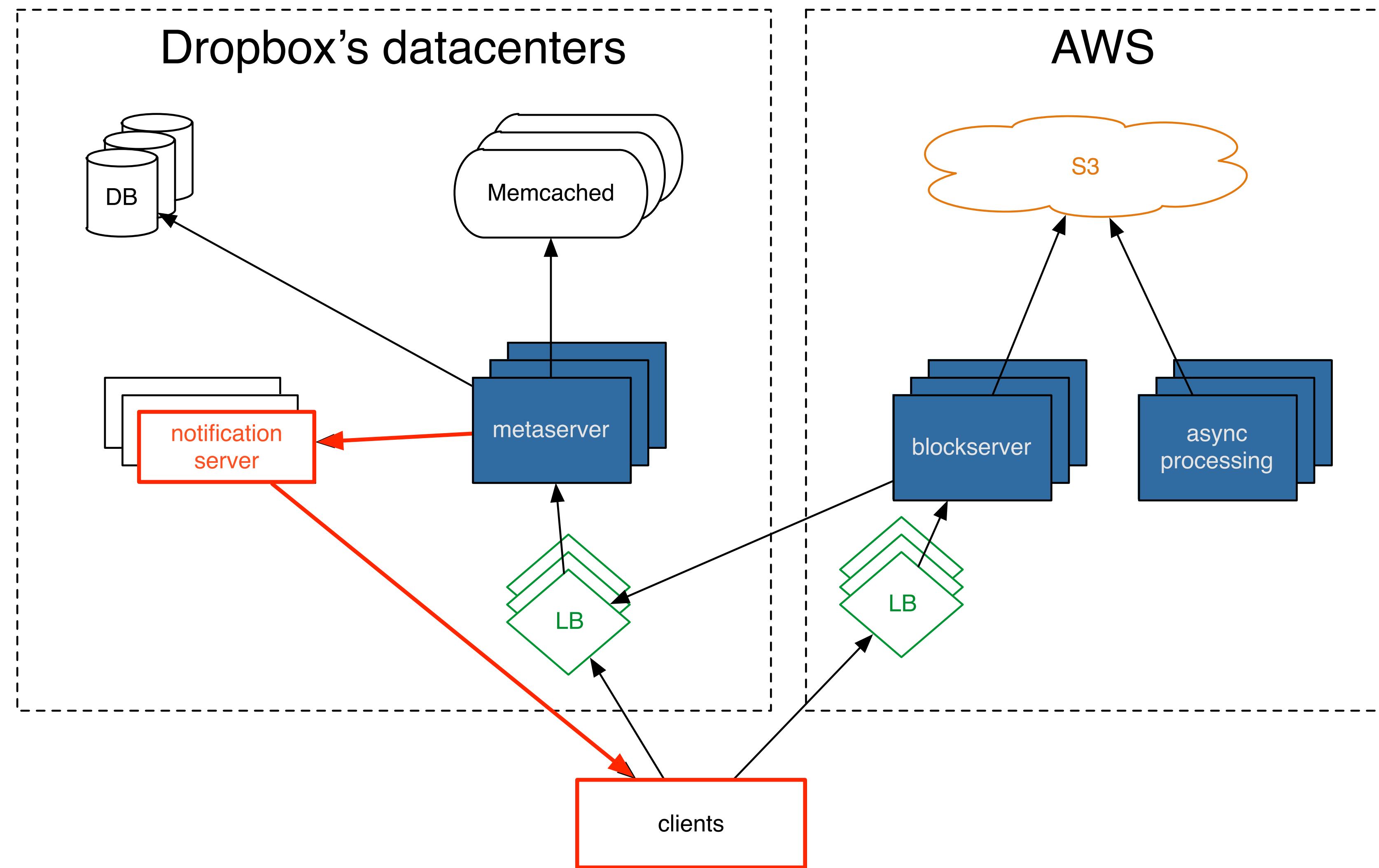
1. FETCH METADATA



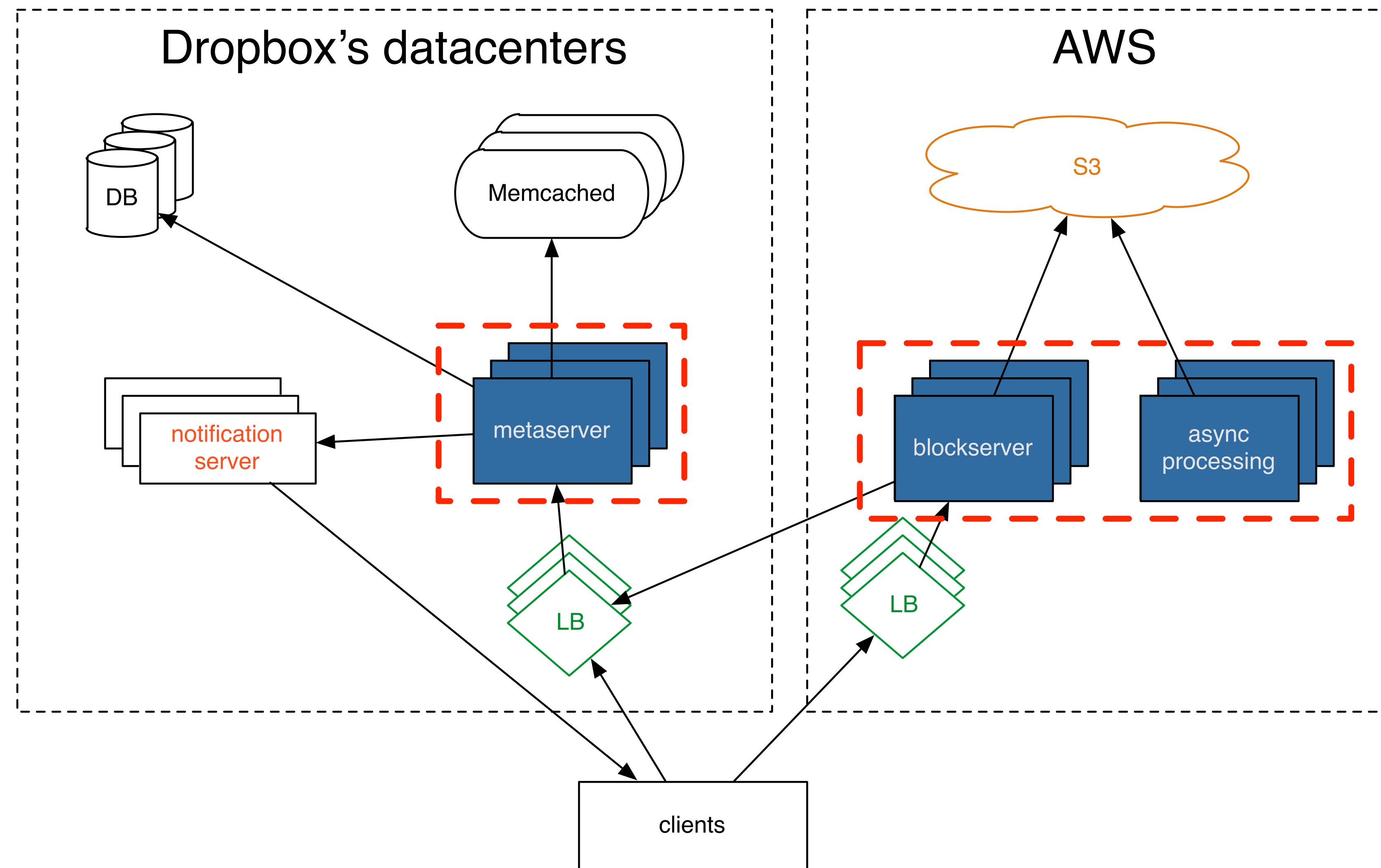
2. DOWNLOAD BLOCKS



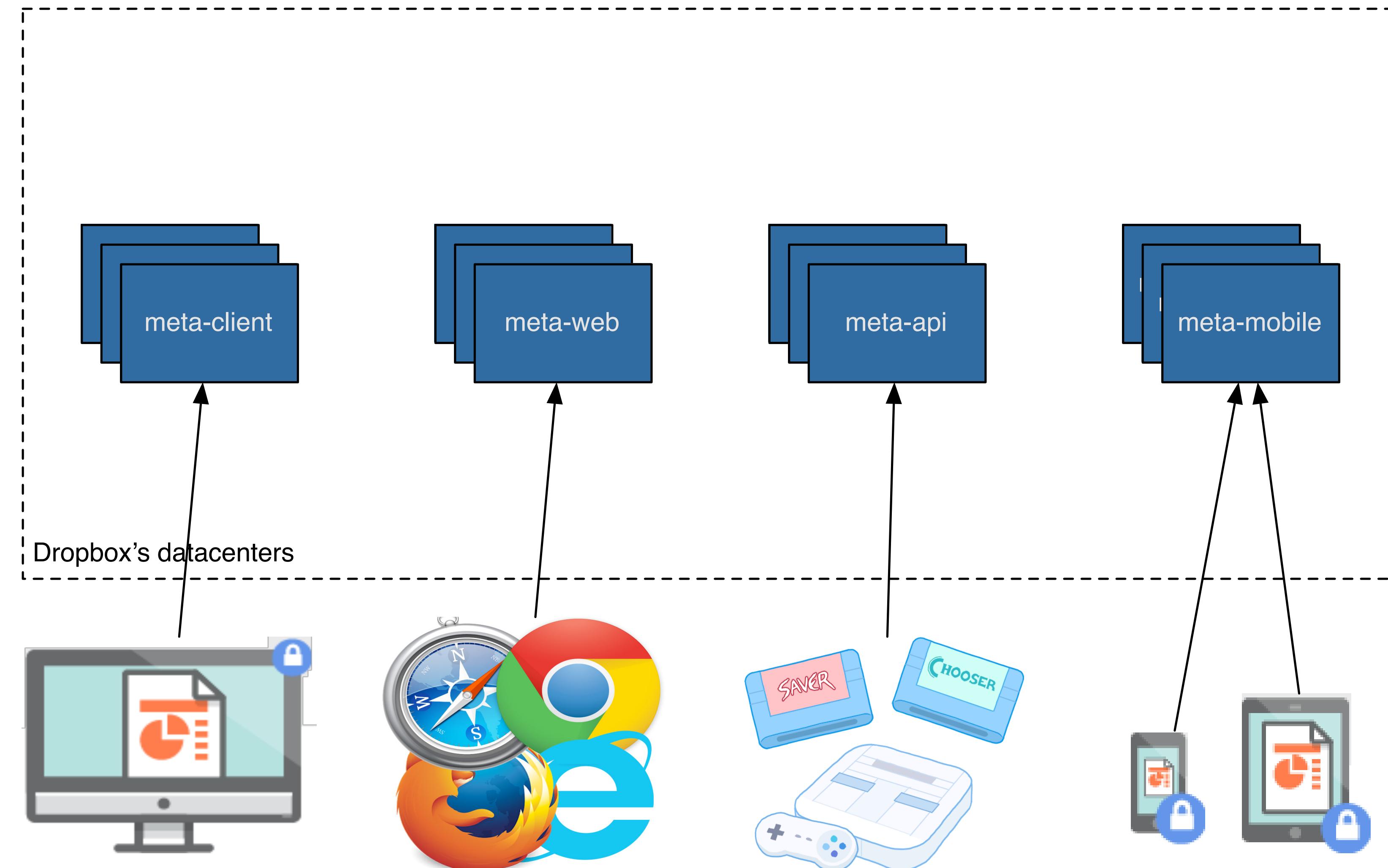
3. WAIT FOR NOTIFICATIONS



PYTHON EVERYWHERE



CLUSTER ISOLATION



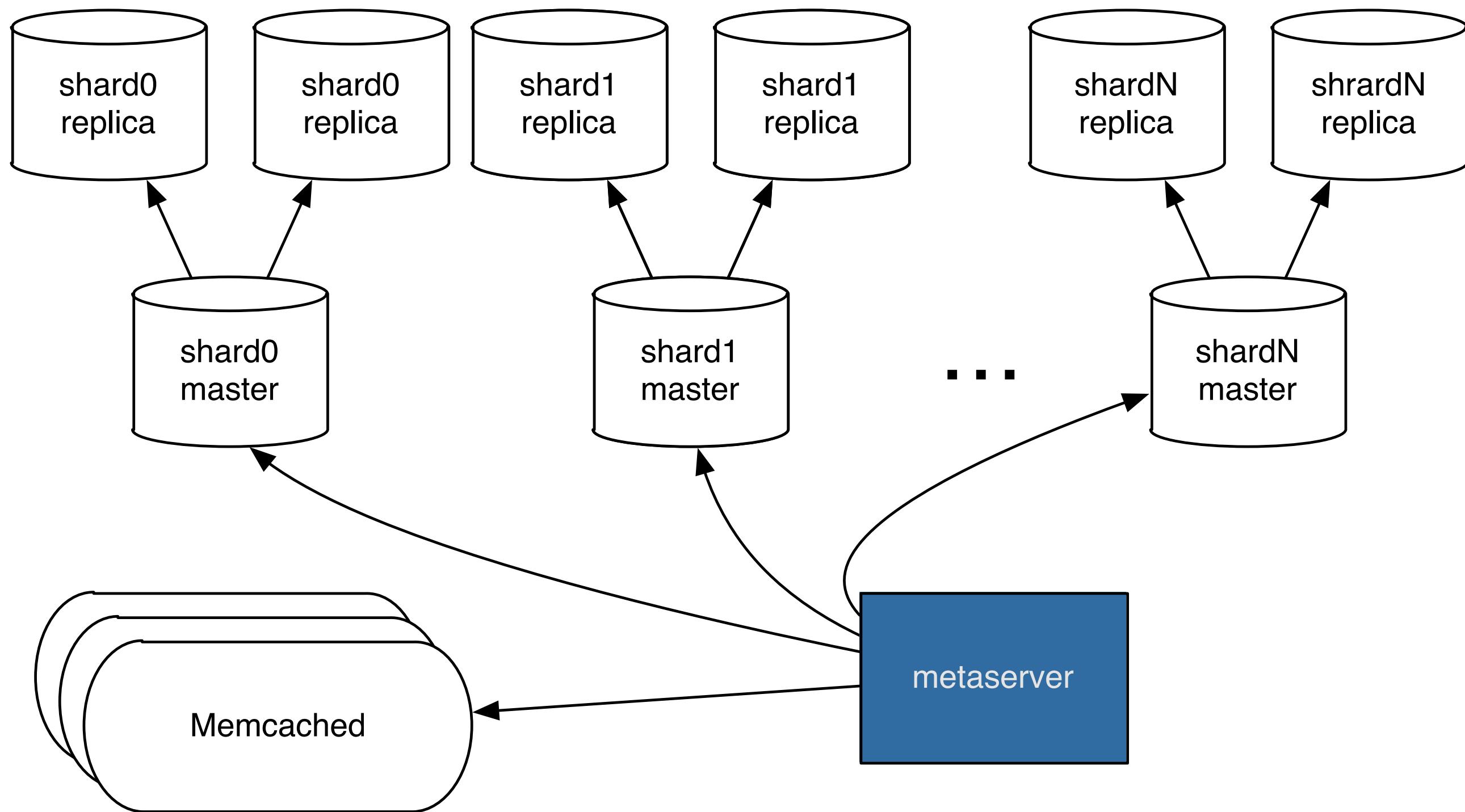
Scaling Databases

Scaling as Organization

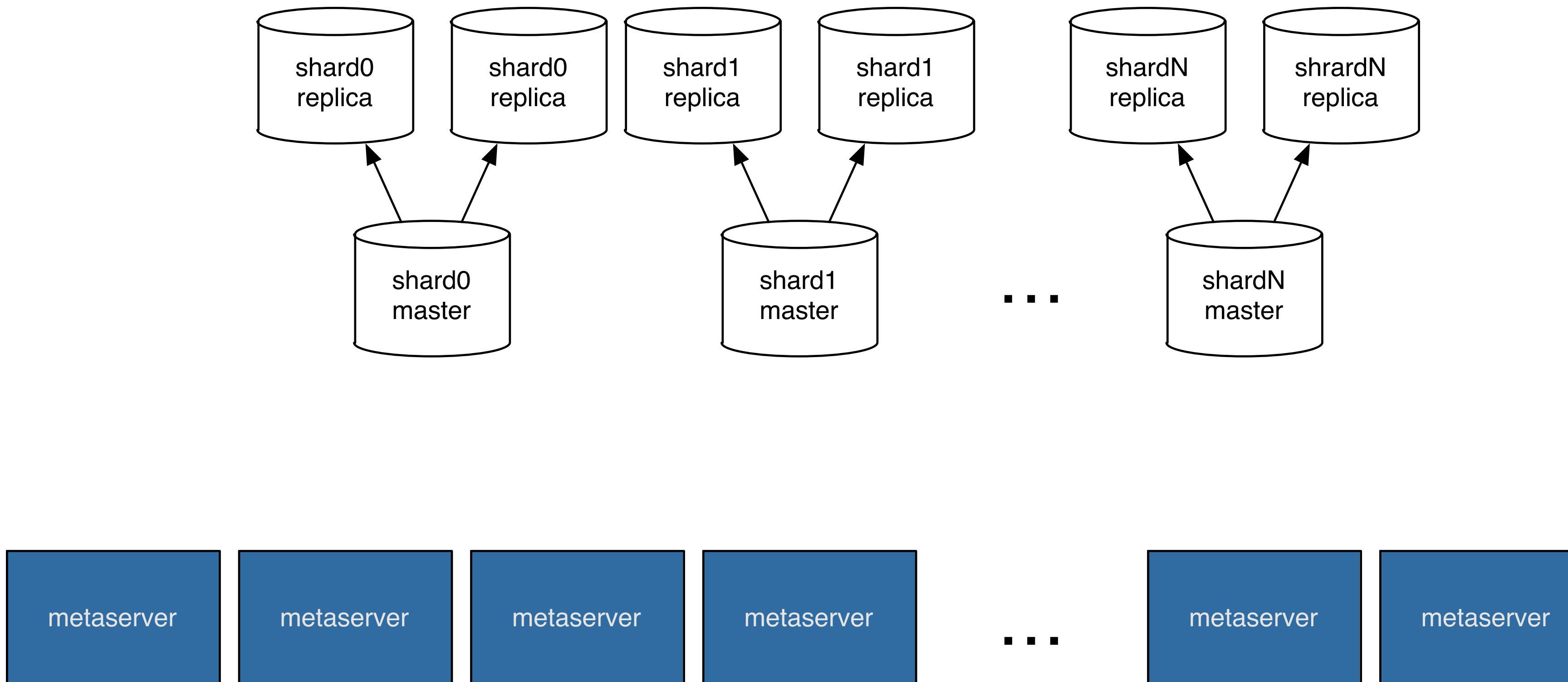
Scaling Software

Managing Complexity

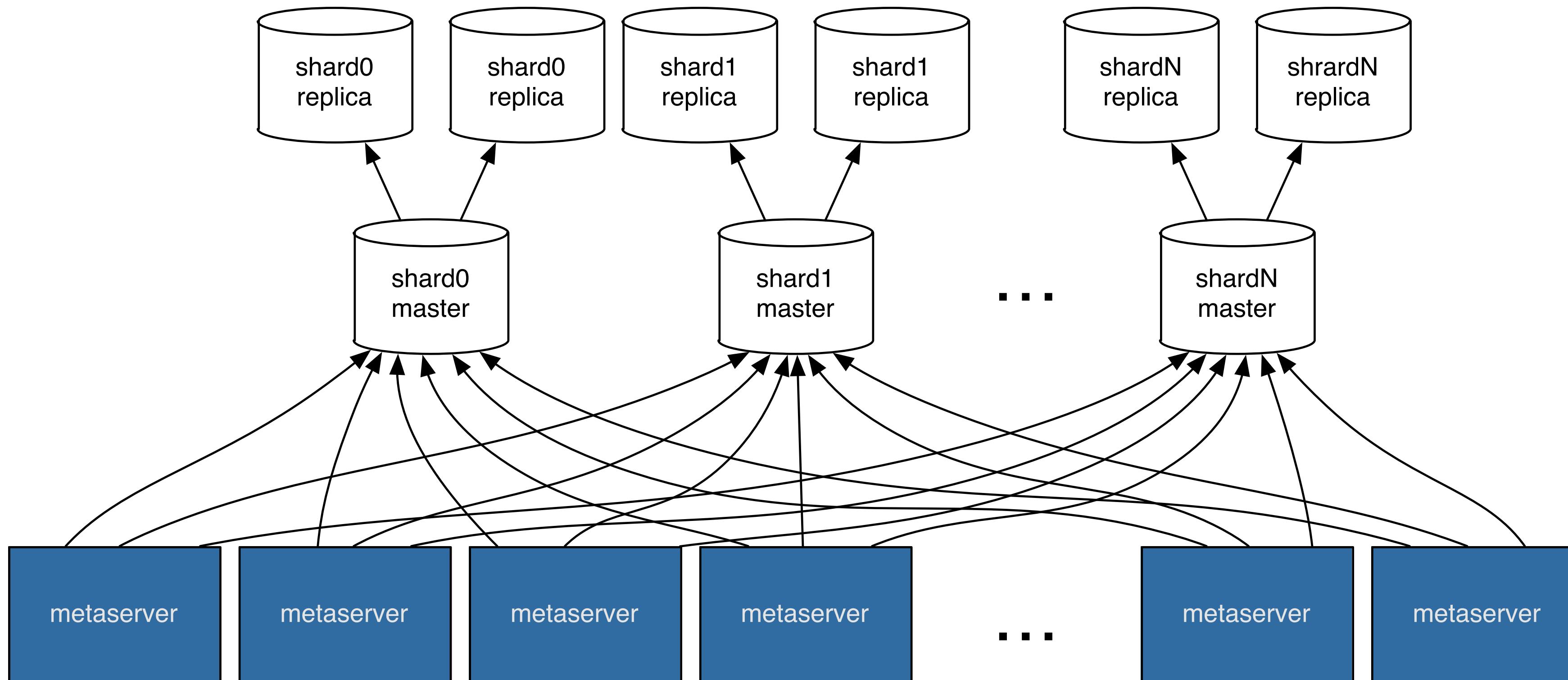
SCALING DATABASES



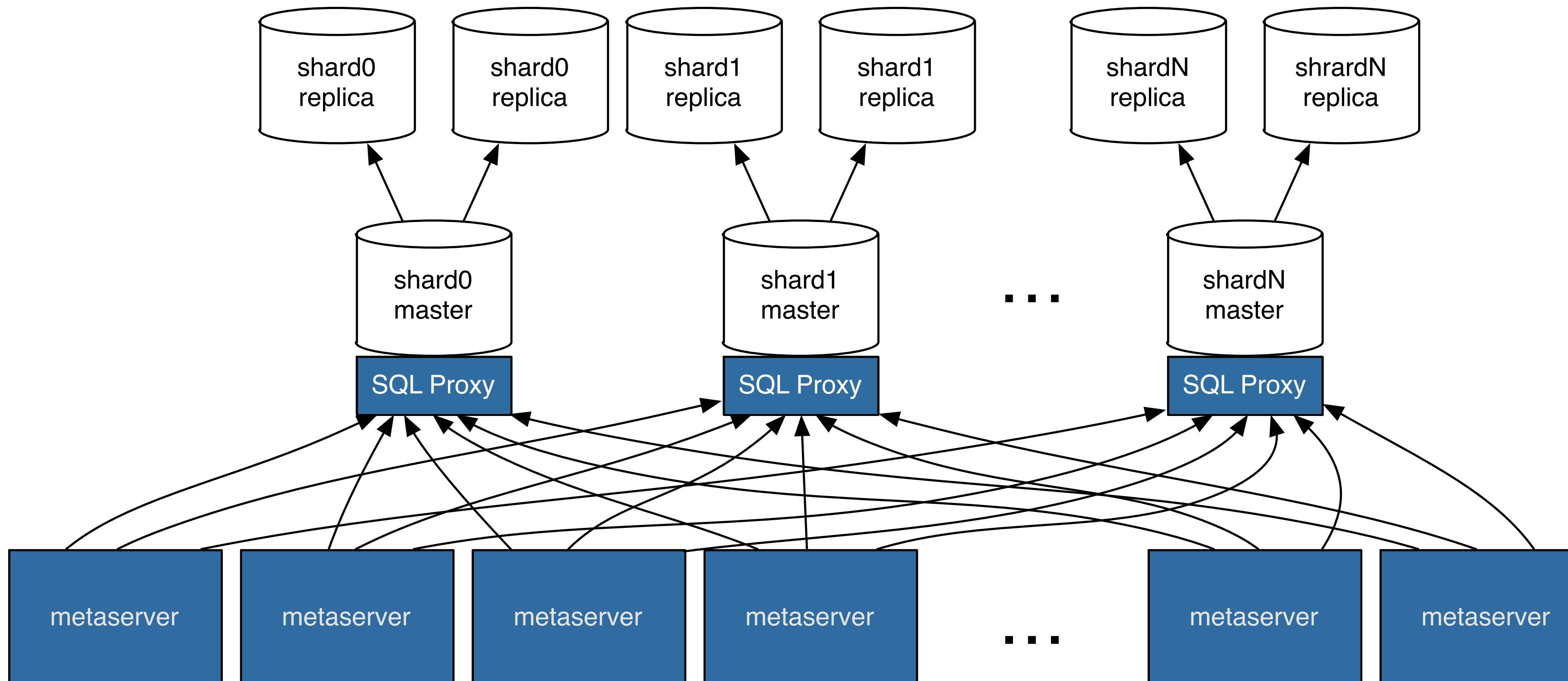
HORIZONTAL SCALING



CONNECTIONS



SQL PROXY



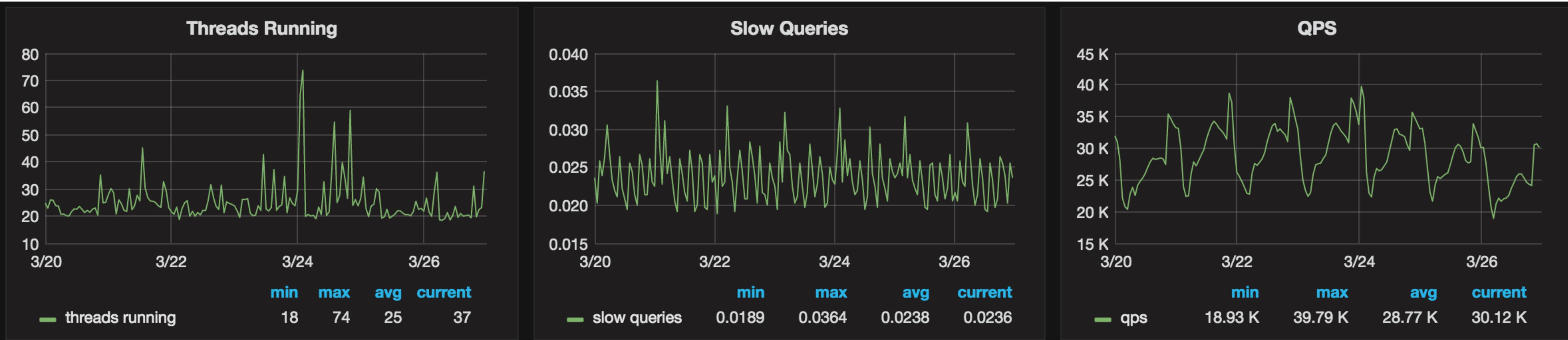
Scaling Databases

Scaling as Organization

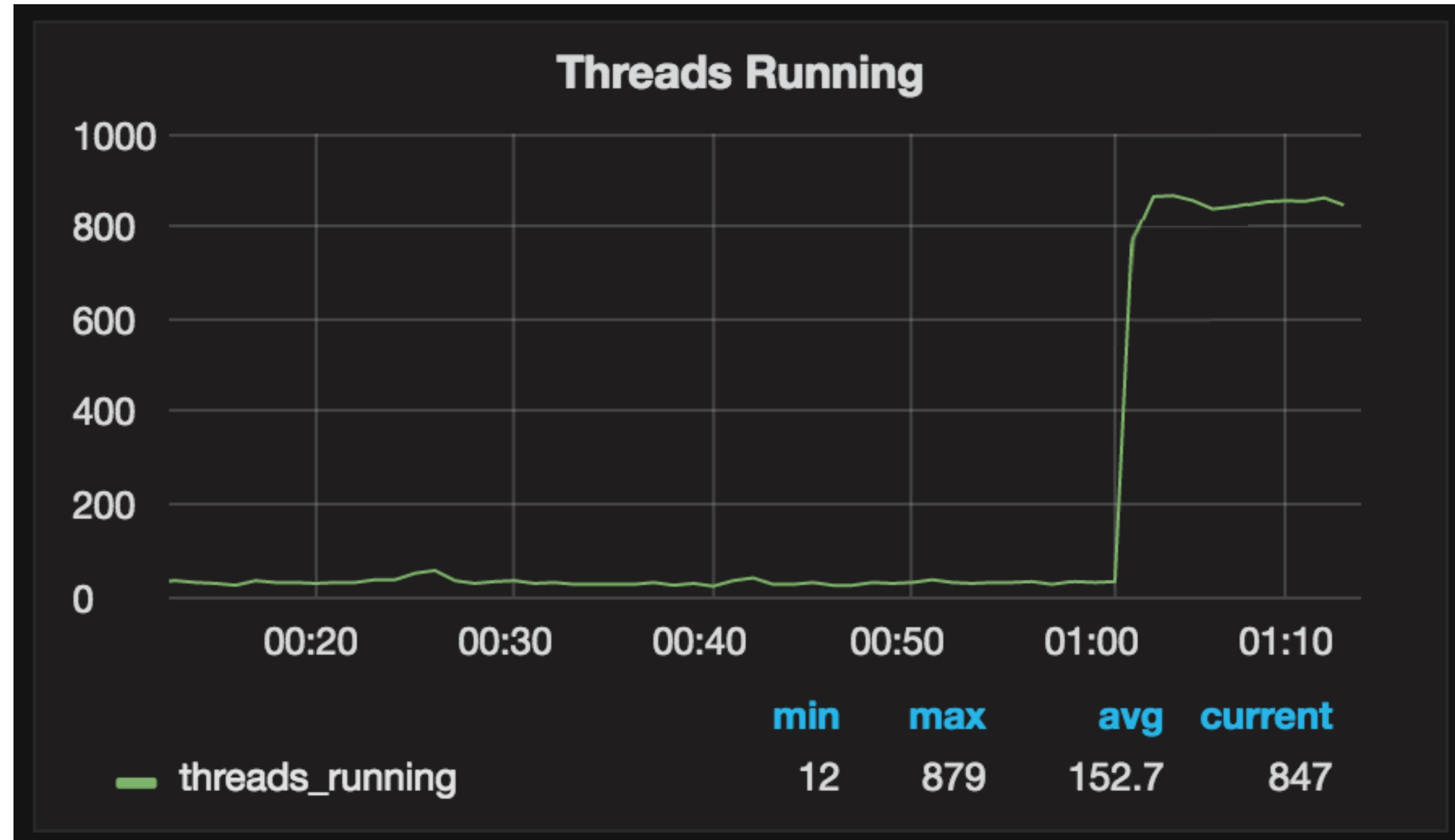
Scaling Software

Managing Complexity

GLOBAL DATABASE



AVAILABILITY ISSUES



PLAYBOOK

1. Check for ongoing deployments or newly enabled features

PLAYBOOK

1. Check for ongoing deployments or newly enabled features
2. Check for recently started background jobs

PLAYBOOK

1. Check for ongoing deployments or newly enabled features
2. Check for recently started background jobs
3. DBA oncall, please help!

Dropbox grew from 100 to 500 employees

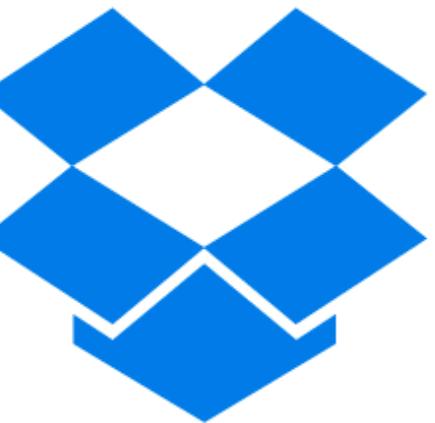
- Slow queries would adversely impact performance across the board

- Slow queries would adversely impact performance across the board
- More features => Managing more independent MySQL

- Slow queries would adversely impact performance across the board
- More features => Managing more independent MySQL
- Reactively (re)sharding individual databases as they hit capacity

- Slow queries would adversely impact performance across the board
- More features => Managing more independent MySQL
- Reactively (re)sharding individual databases as they hit capacity
- Impacted developer productivity

SCALABLE METADATA STORE
DESIGNED FOR MULTI-TENANCY

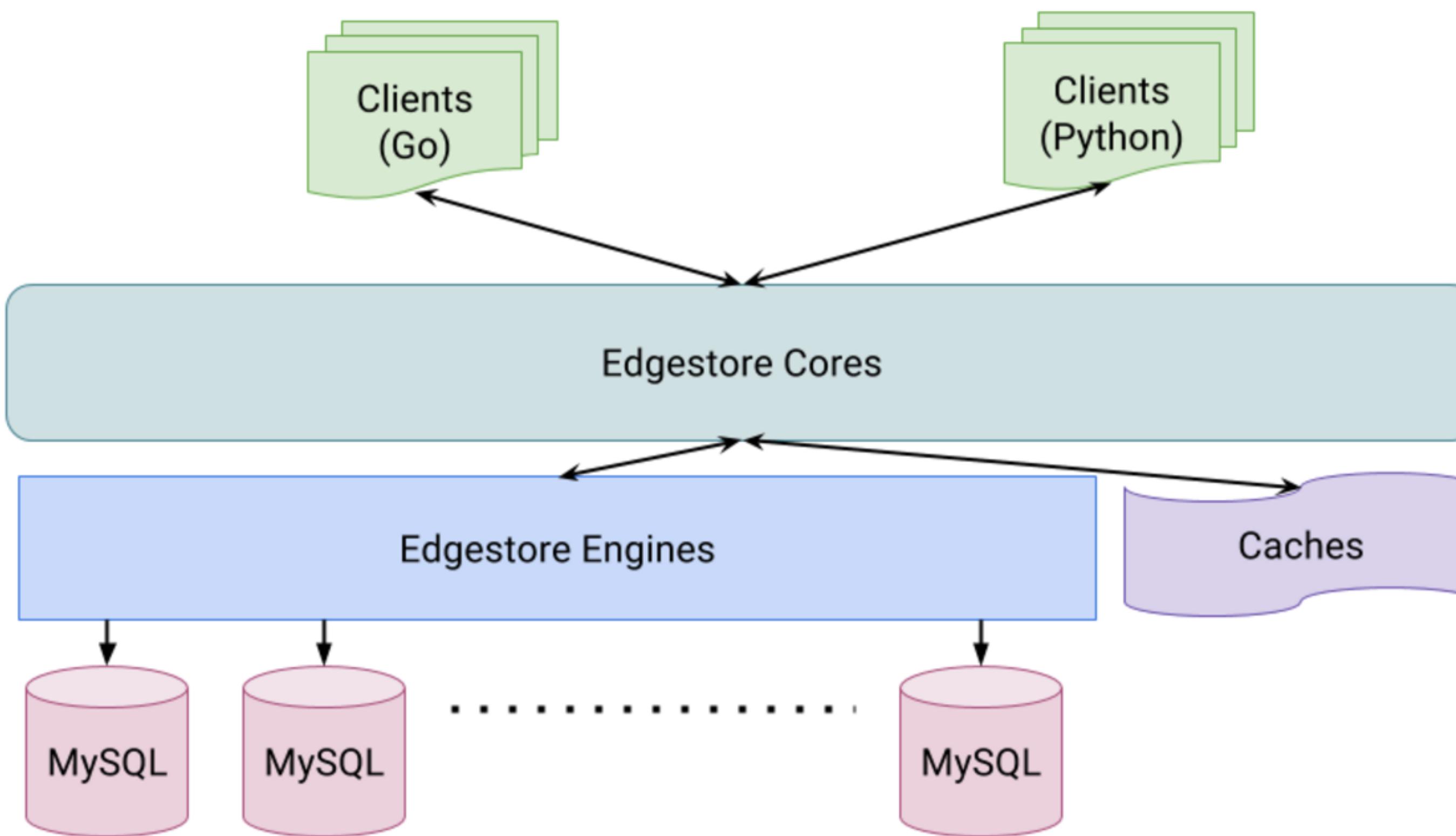


Edgestore

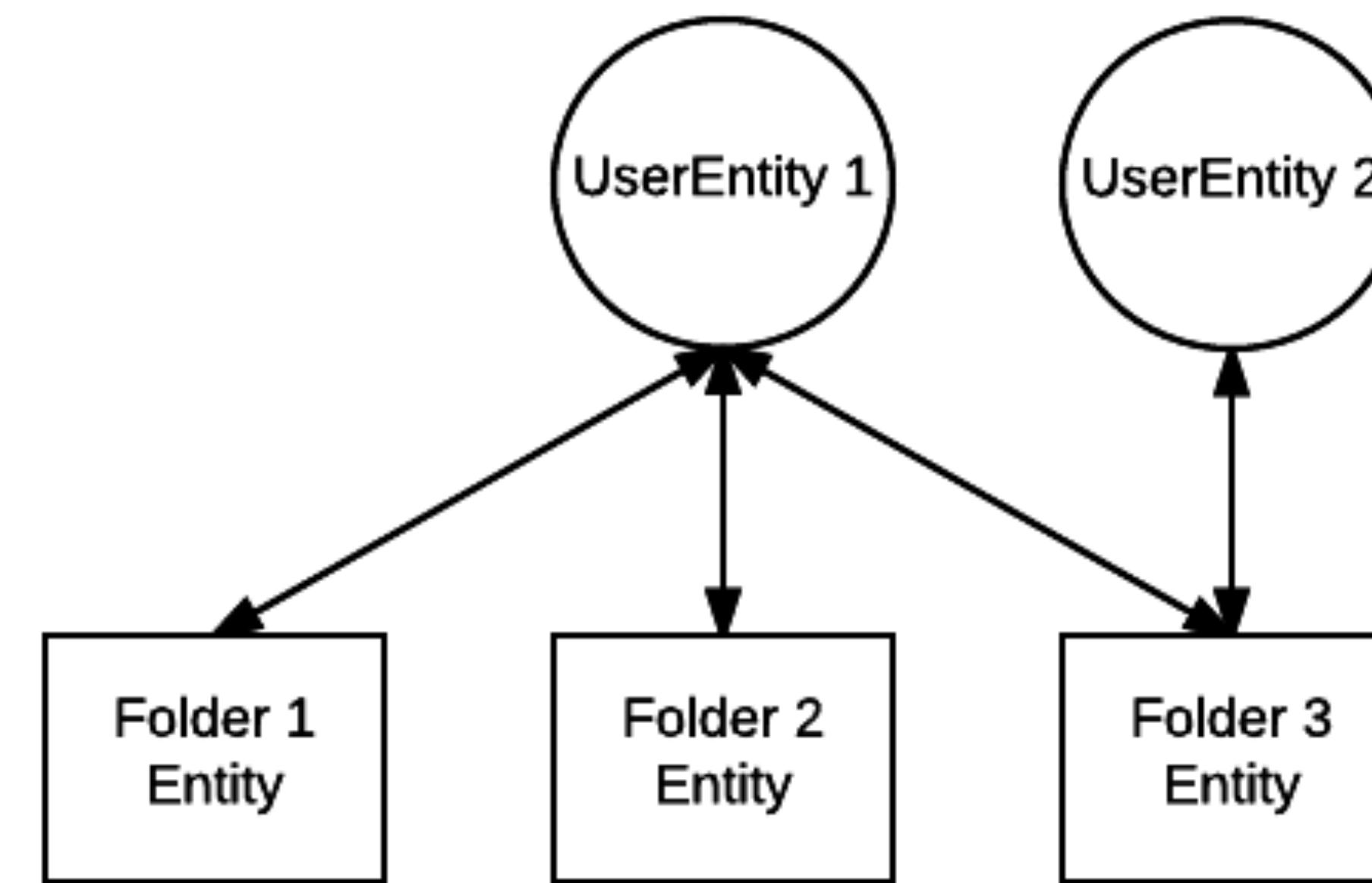
2013 — Present



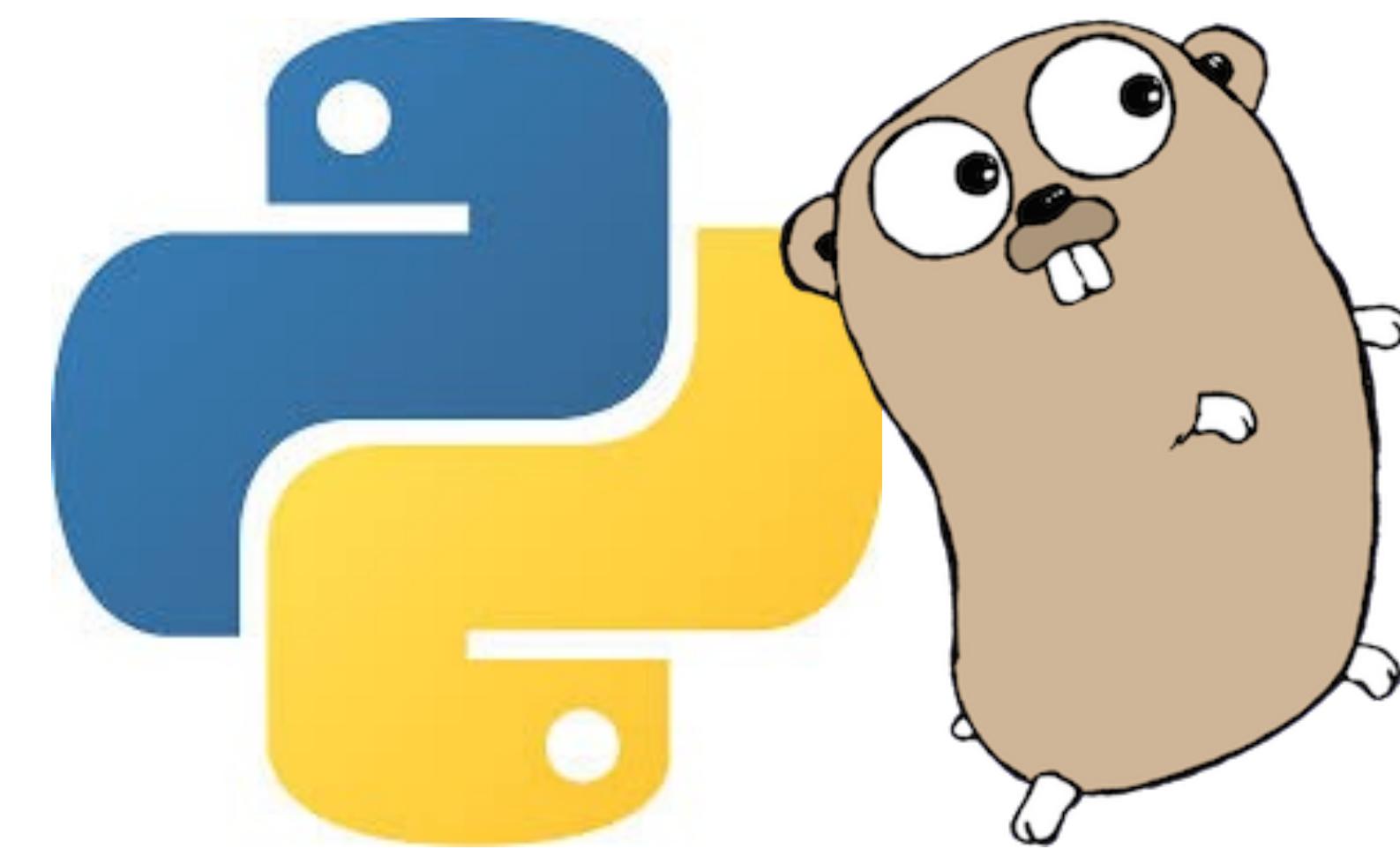
SHARDING AND CACHING BEHIND THE SCENES



ENTITIES AND ASSOCIATIONS



FIRST GO SERVICE

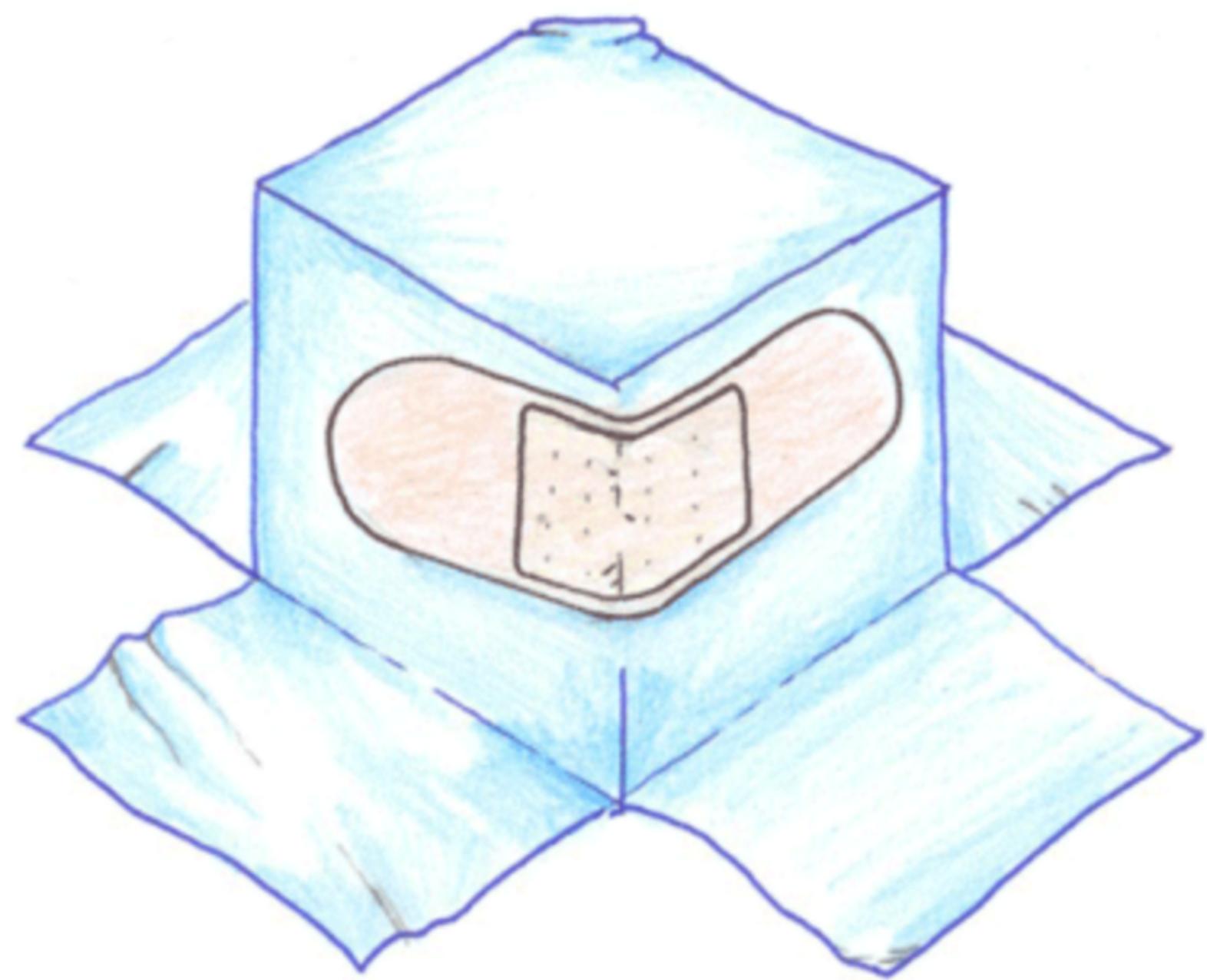


Scaling Databases

Scaling as Organization

Scaling Software

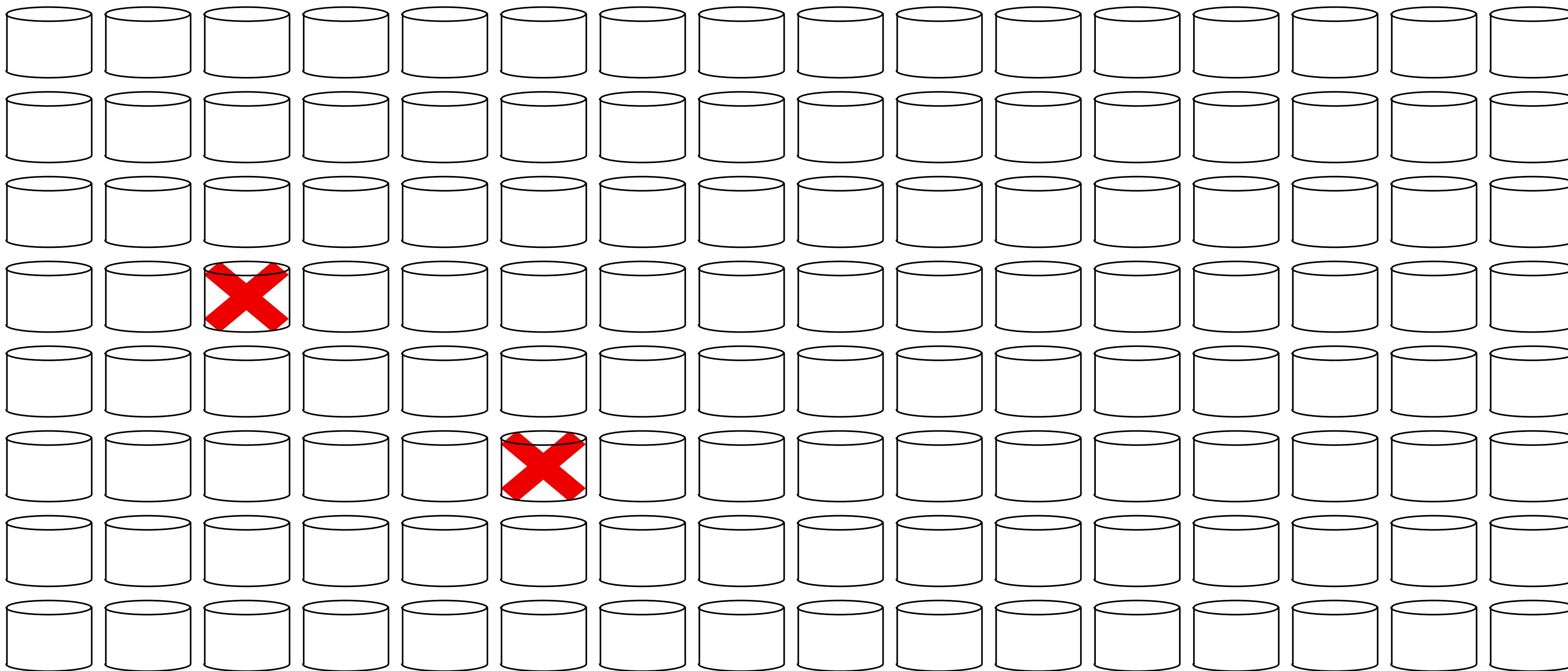
Managing Complexity



Error (5xx)

Something went wrong. Don't worry, your files are still safe and the Dropboxers have been notified. Check out our [Help Center](#) and [forums](#) for help, or head back to [home](#).

PERFECT STORM



SHARDING



PHOTO ALBUMS



Preslav Le ▾



Recents

Personal

Dropbox

Paper

Photos

Timeline

Albums

FEEDBACK

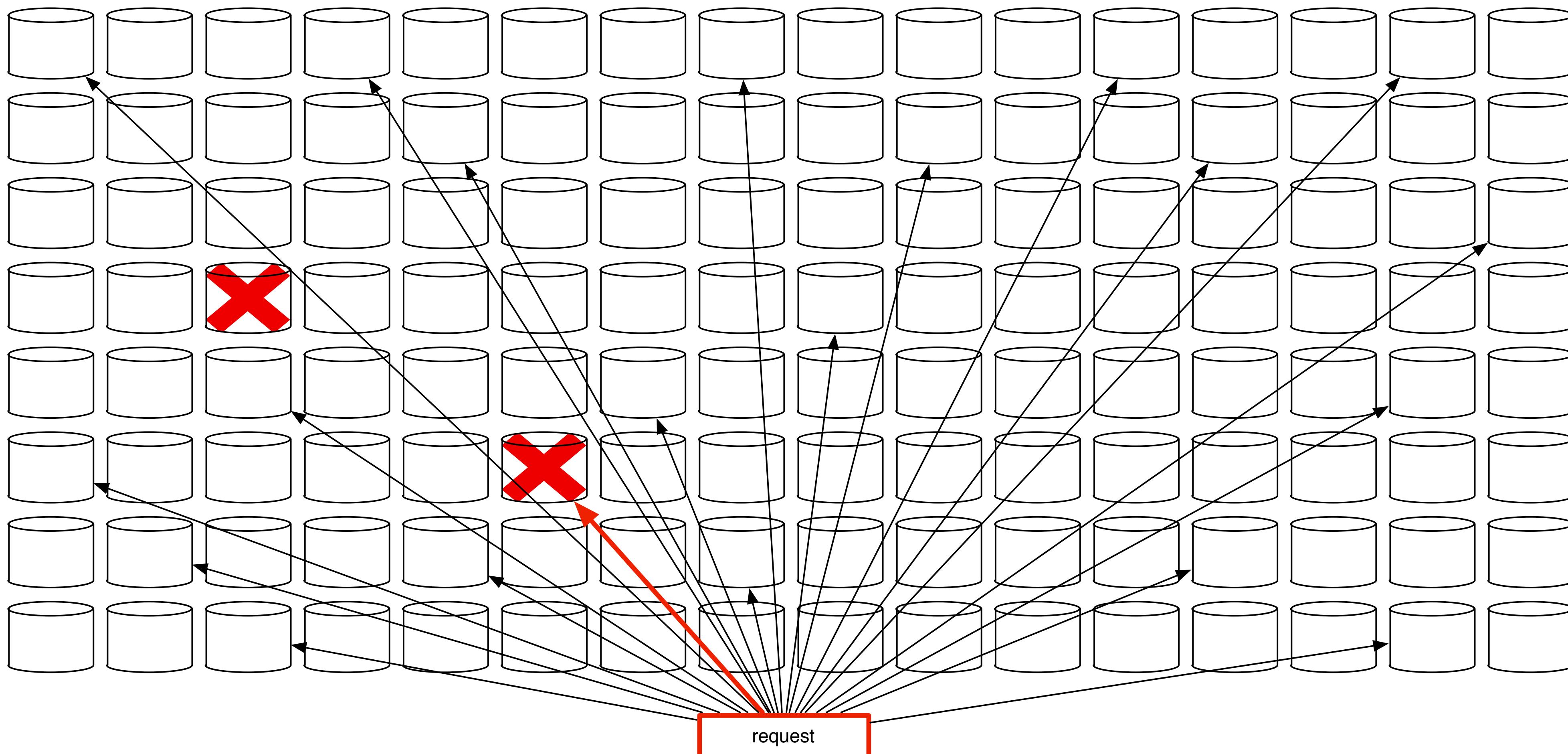


TEAM ADMIN CONSOLE

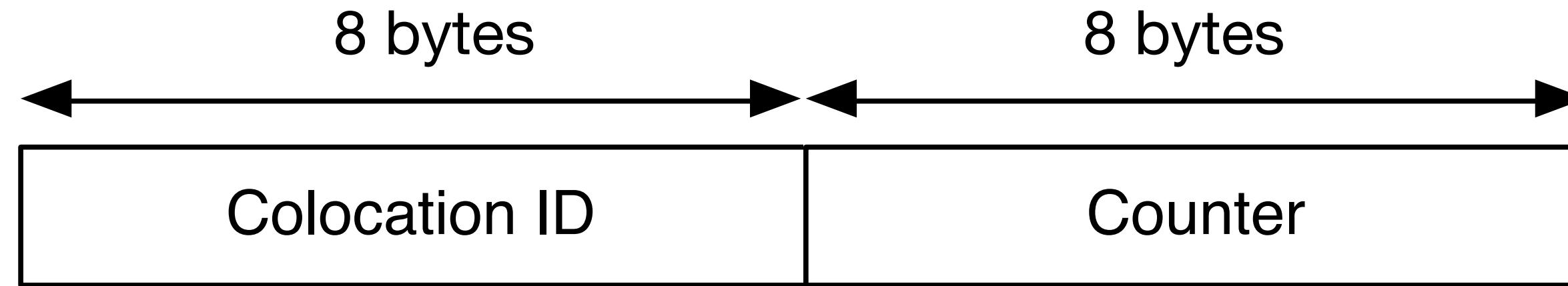
The screenshot shows the Dropbox Team Admin Console interface. At the top, there's a browser header with a lock icon and the URL <https://www.dropbox.com>. To the right of the URL are navigation icons (back, forward, search) and user information for "Janey Smith". Below the header, the main title "Members" is displayed next to the Dropbox logo. On the left, a sidebar lists various admin functions: Dashboard, Members (selected), Activity, Authentication, Sharing, Team folder, Groups, Account, and Help. A search bar at the top of the main content area says "Find a member of Pepo Inc". The main content area shows a table of current members. The table has columns for Name, User type, Usage, and Last web activity. It includes five rows for Annie Adams, Hannah Wu, Janey Smith, Stephanie Chen, and Will Butler, each with a gear icon for more options.

Current members (5)	Invited members (0)	Deleted members (34)	
Annie Adams (annie.adams@pepoinc.com)	TEAM ADMIN	1.55 GB	about 14 hours ...
Hannah Wu (hannah.wu@pepoinc.com)	USER MANAGEMENT ADMIN	204.03 MB	about a week ago
Janey Smith (janey.smith@pepoinc.com)	TEAM ADMIN	2.46 GB	in the last hour
Stephanie Chen (stephanie.chen@pepoinc.com)	SUPPORT ADMIN	171.89 MB	about a week ago
Will Butler (will.butler@pepoinc.com)	USER MANAGEMENT ADMIN	295.89 MB	about 2 days ago

REQUEST FANOUT



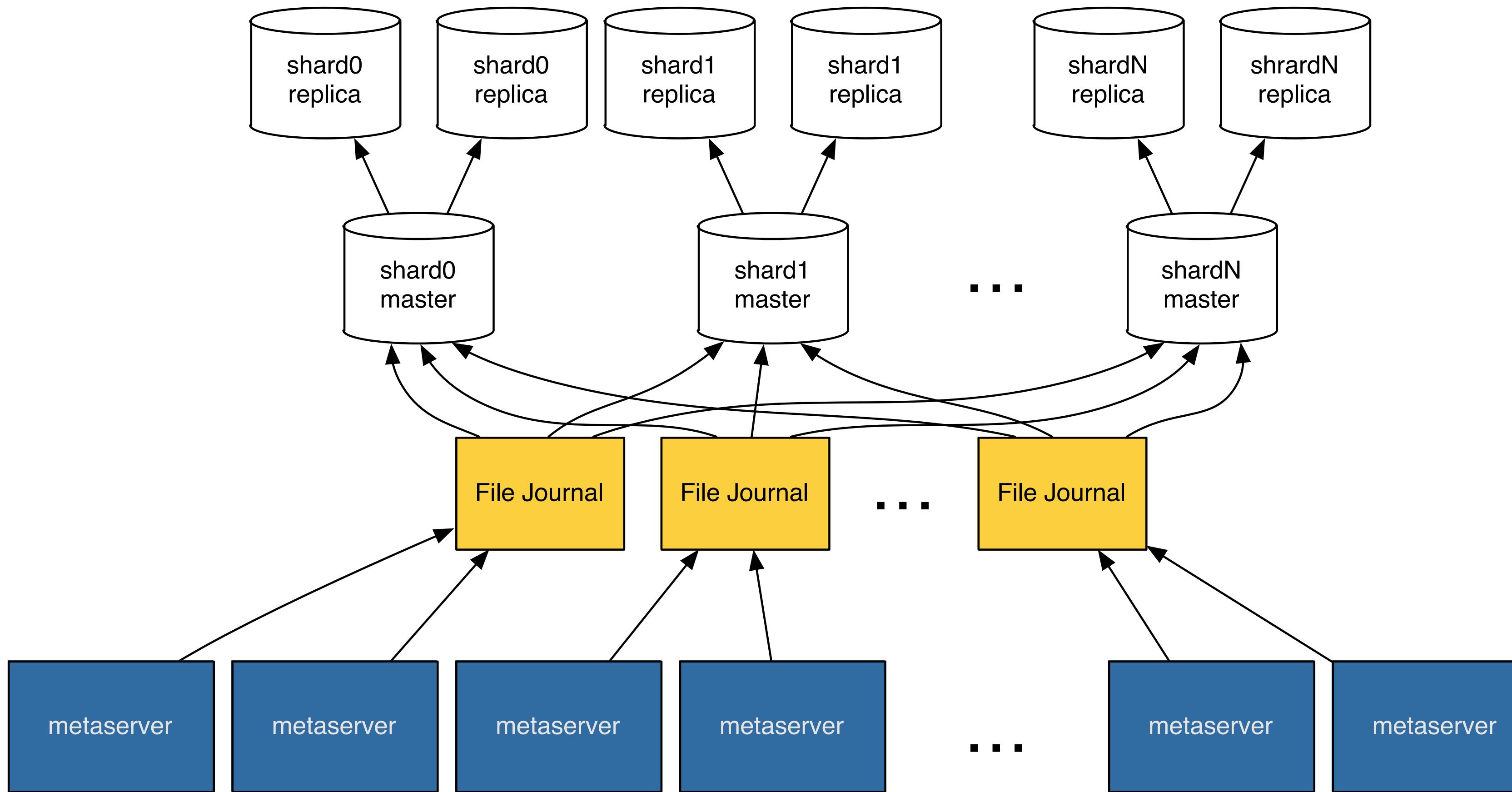
GLOBAL ID



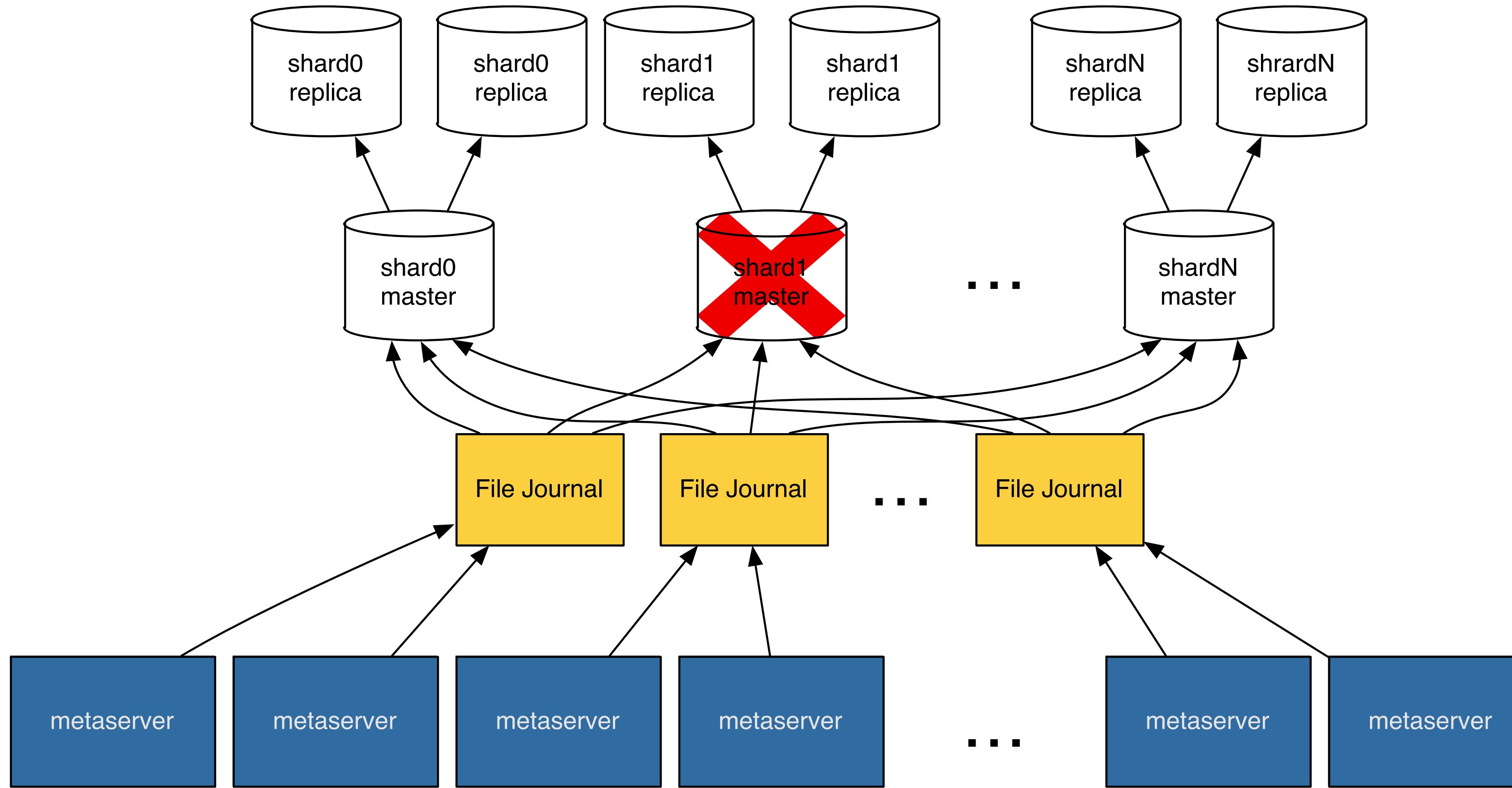
- Colocation ID: Identifies a shard
- Counter: Unique ID within the shard

Lack of colocation also hurts performance

NEW SERVICE: FILE JOURNAL



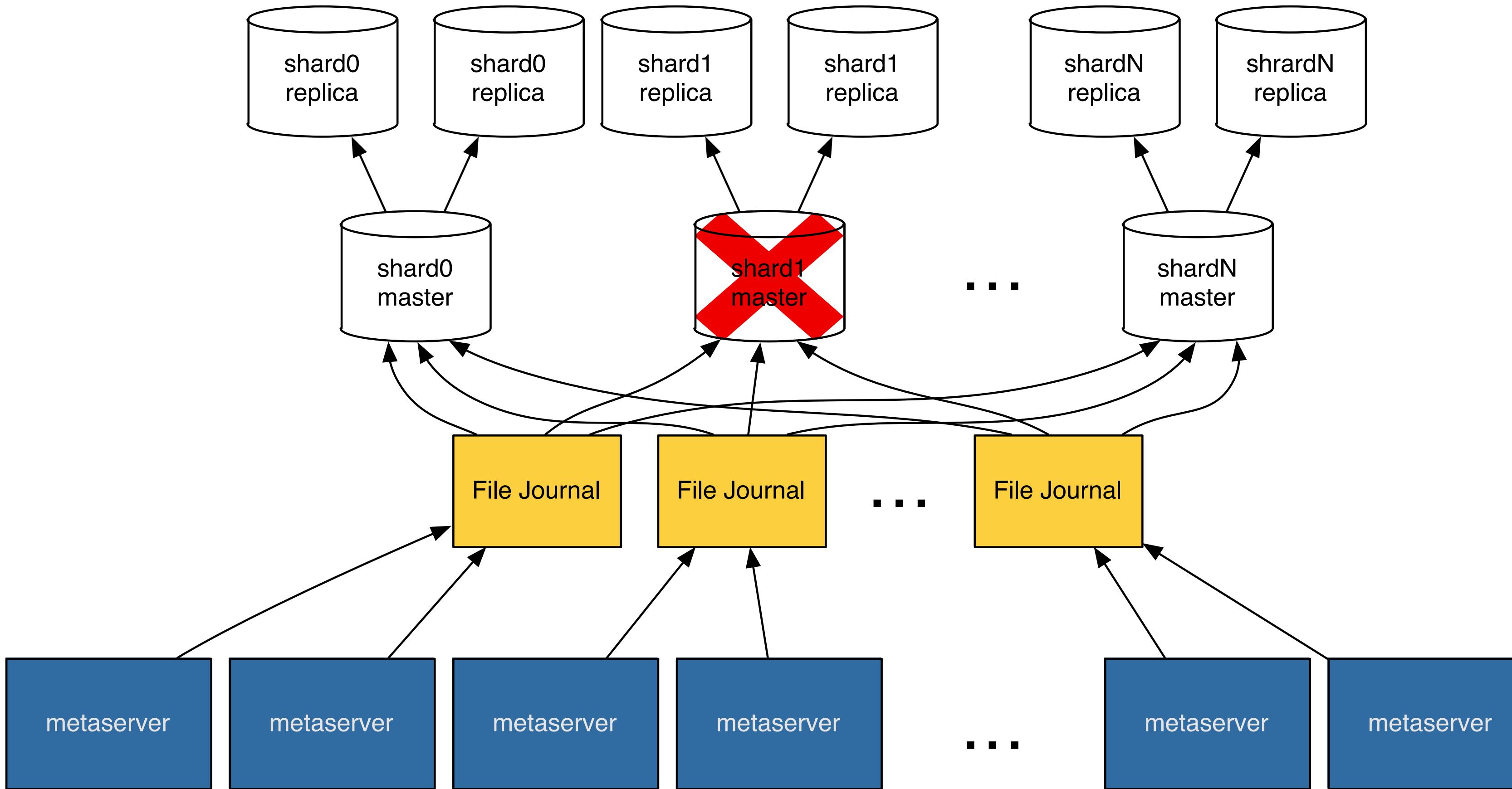
SHARD FAILURE



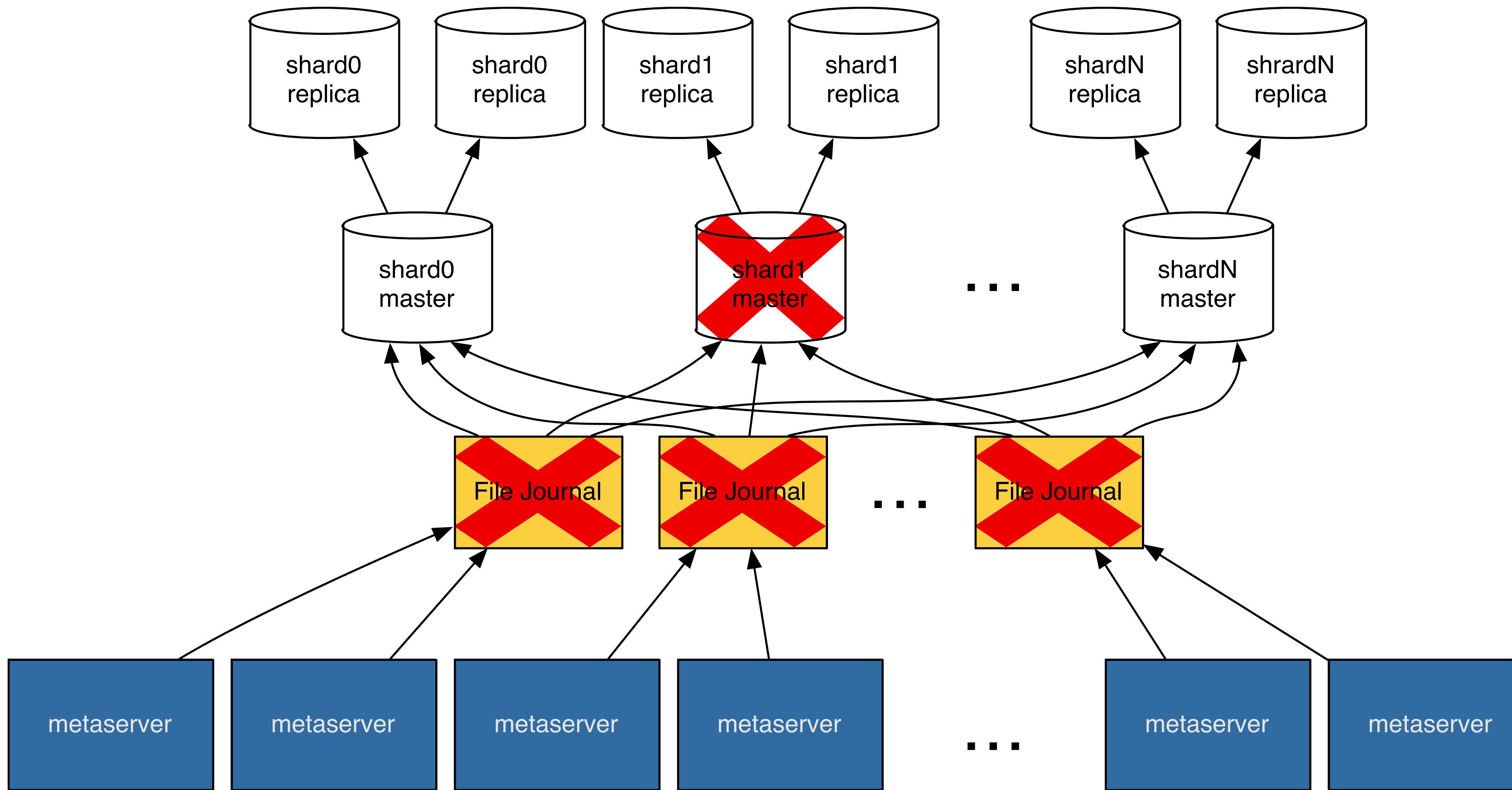
SHARDING (PART II)



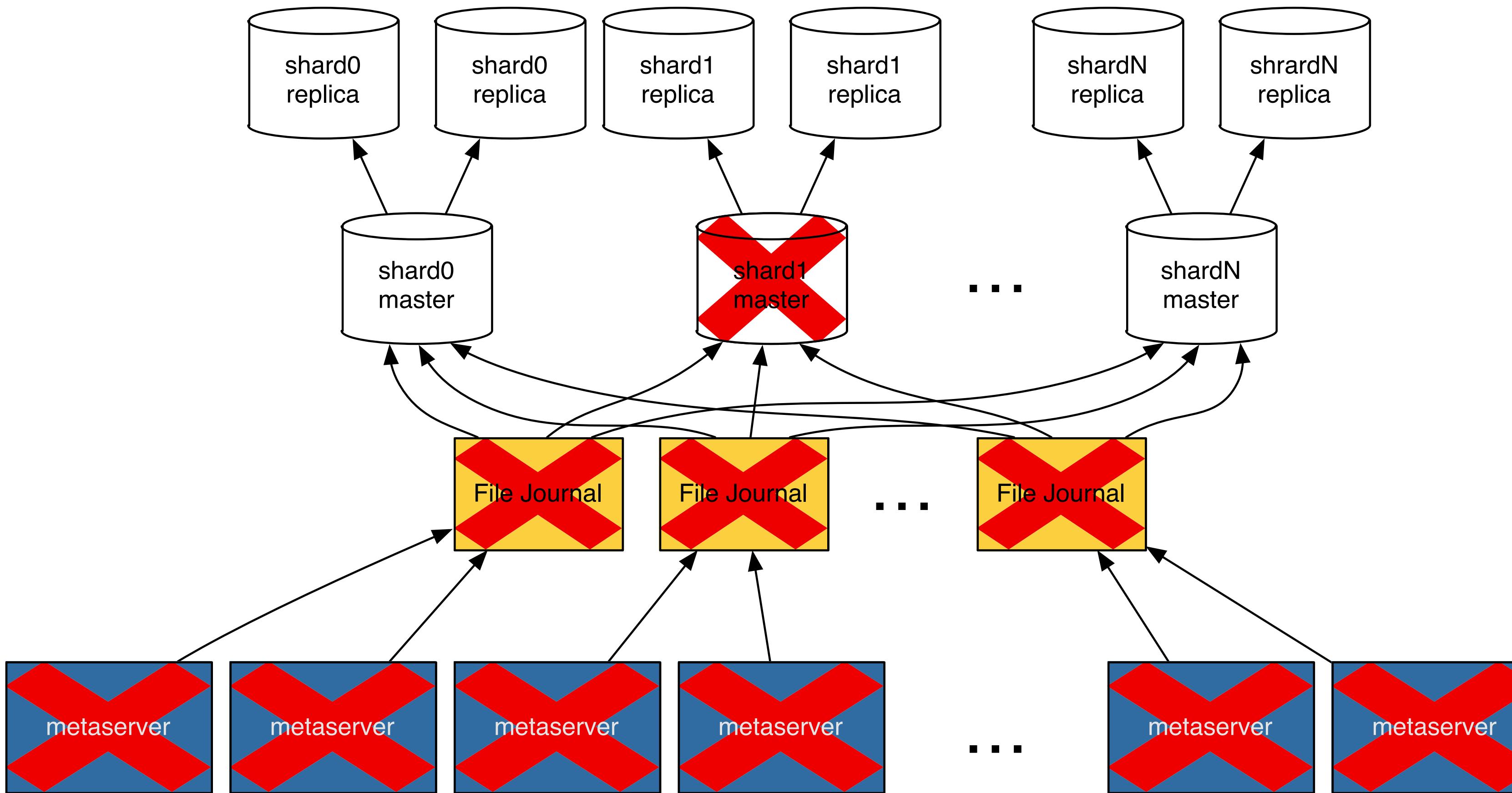
LONG TIMEOUTS



RUN OUT OF WORKERS



CASCADING FAILURE



SHARD ISOLATION

Limit resources dedicated to processing a single shard

Scaling Databases

Scaling as Organization

Scaling Software

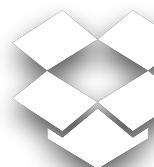
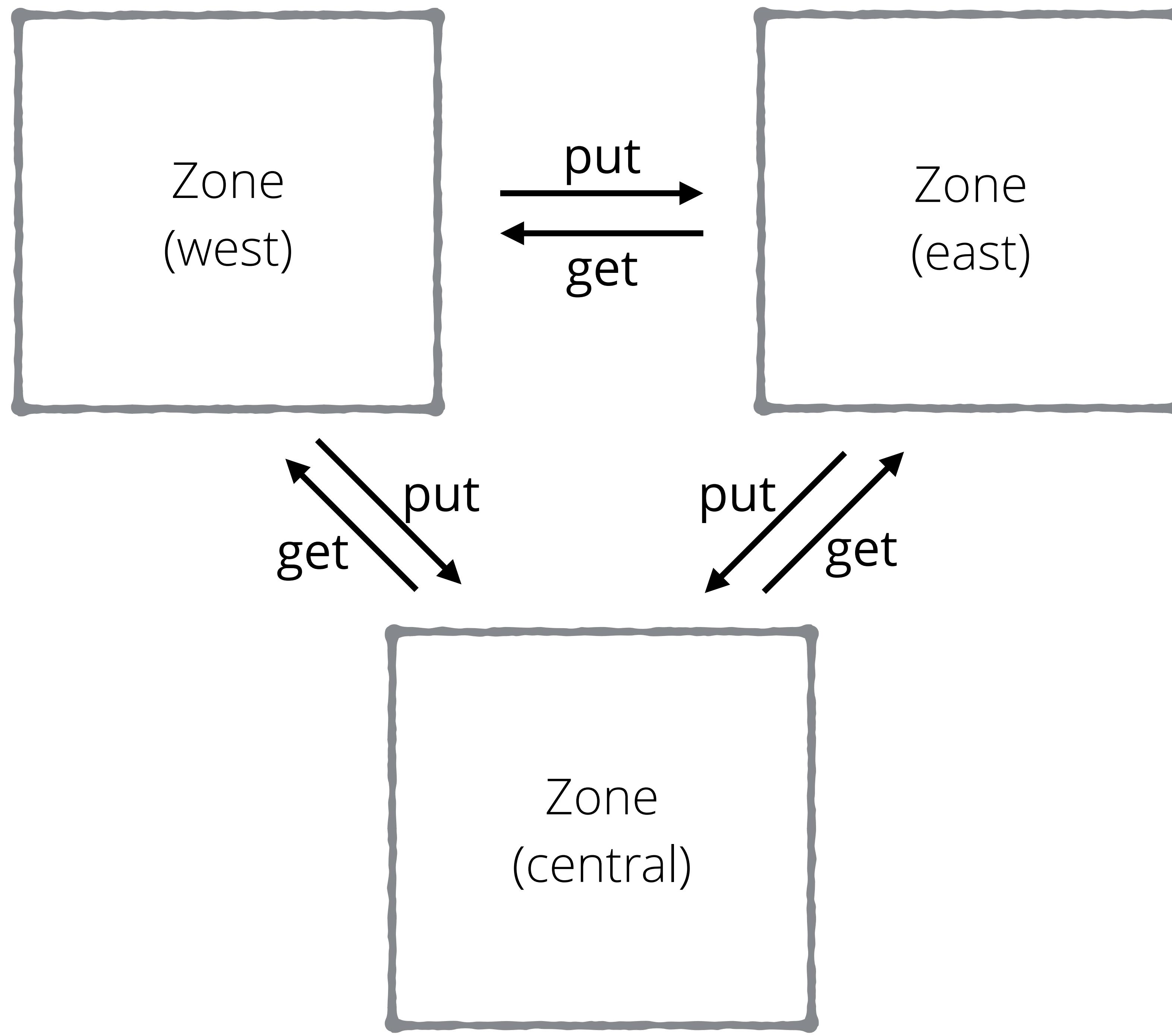
Managing Complexity

MAGIC POCKET BLOCK STORAGE SYSTEM

500PB+ user block data

3+ geographic regions

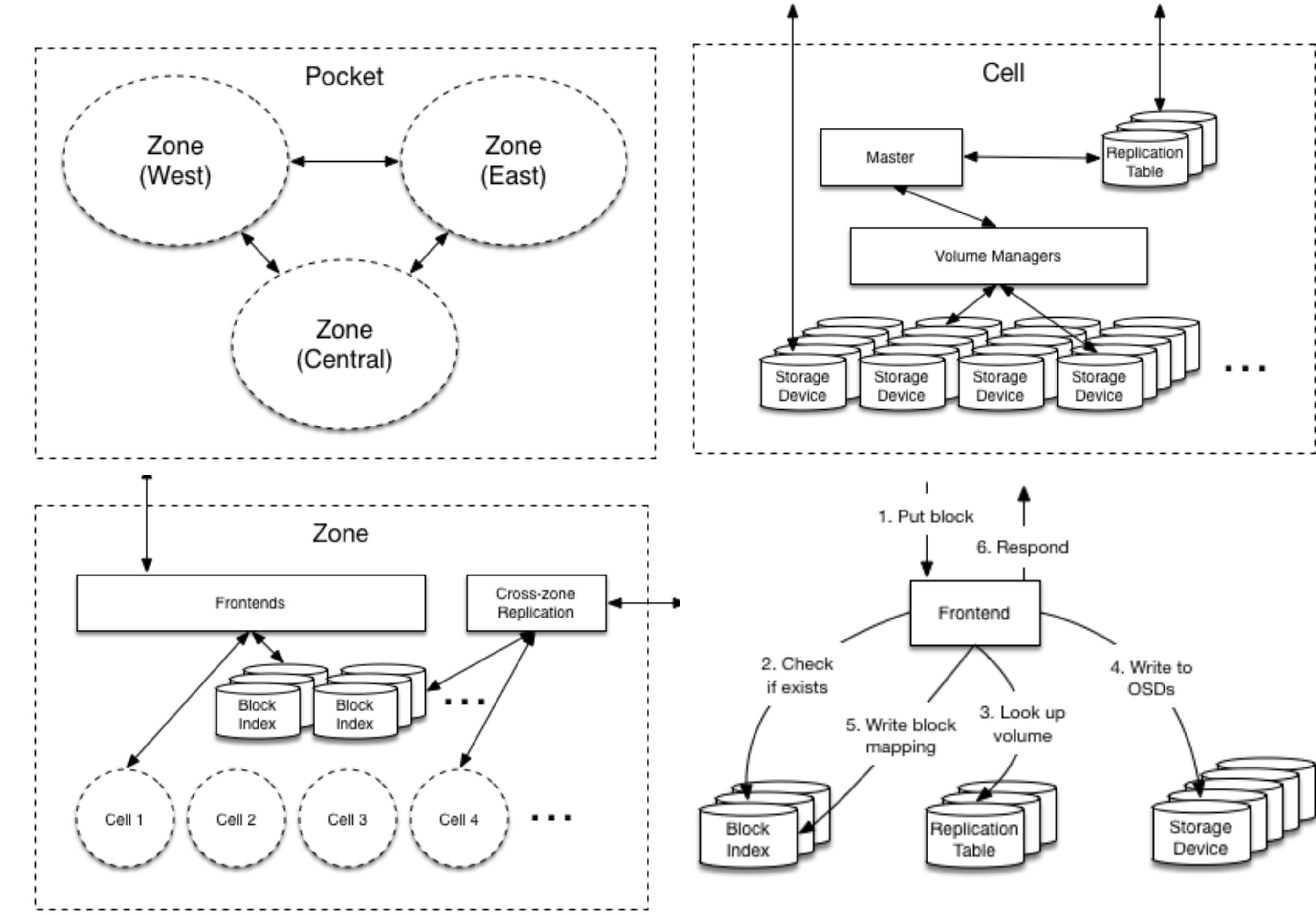
500+ million users





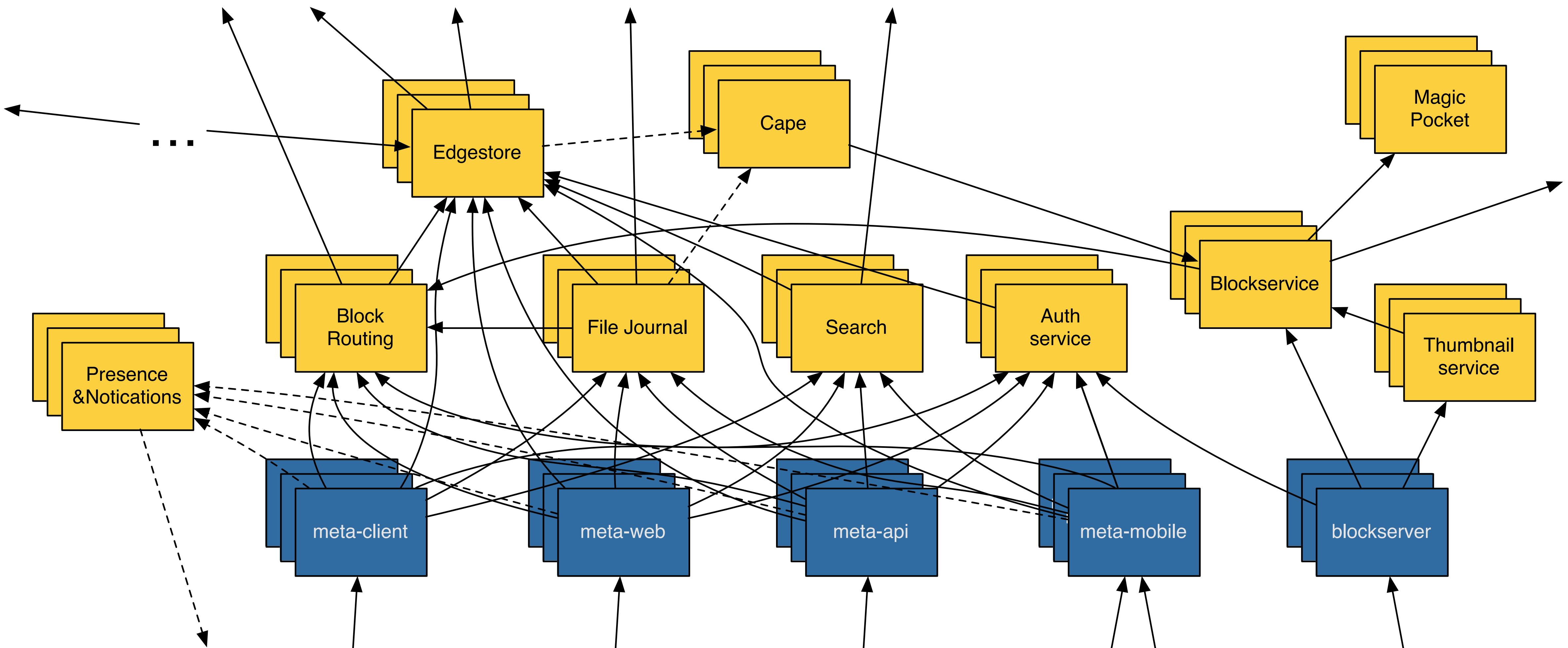
PYTHON, GO & RUST



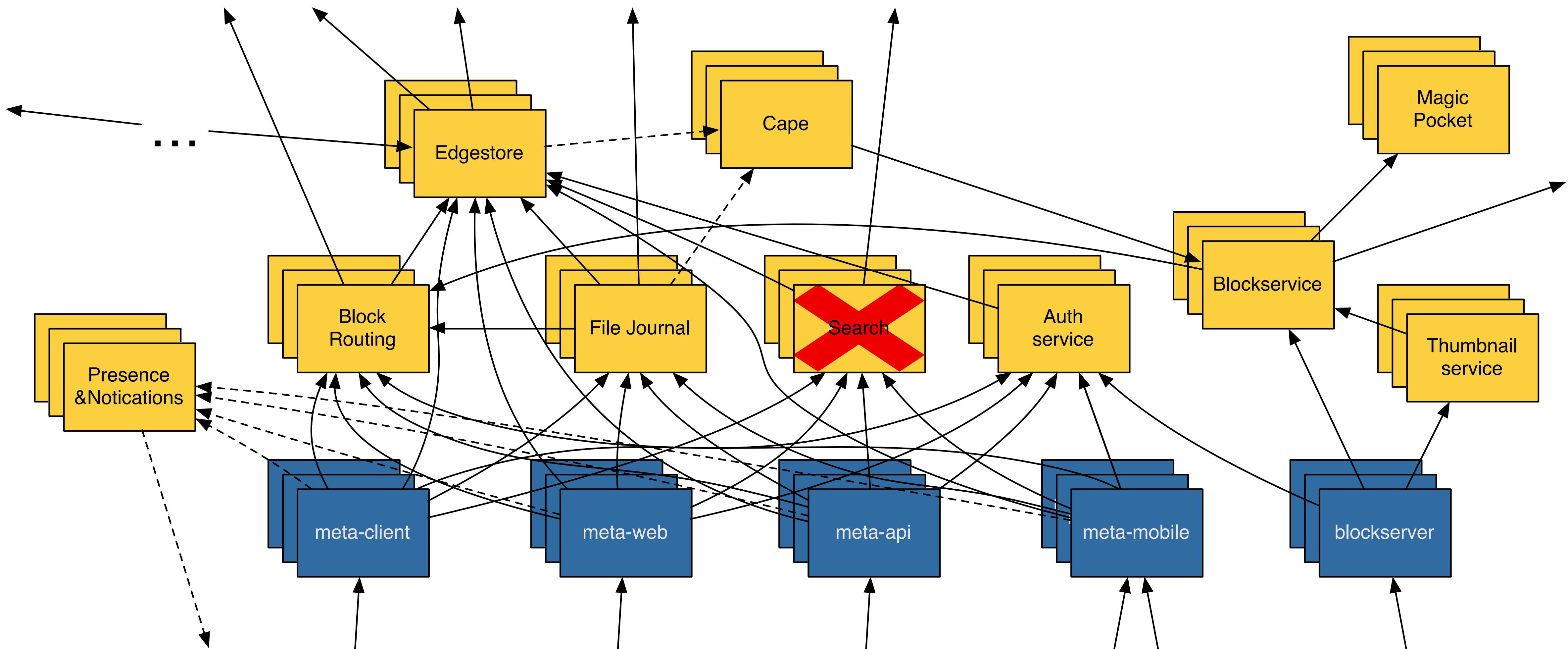


<https://blogs.dropbox.com/tech/>

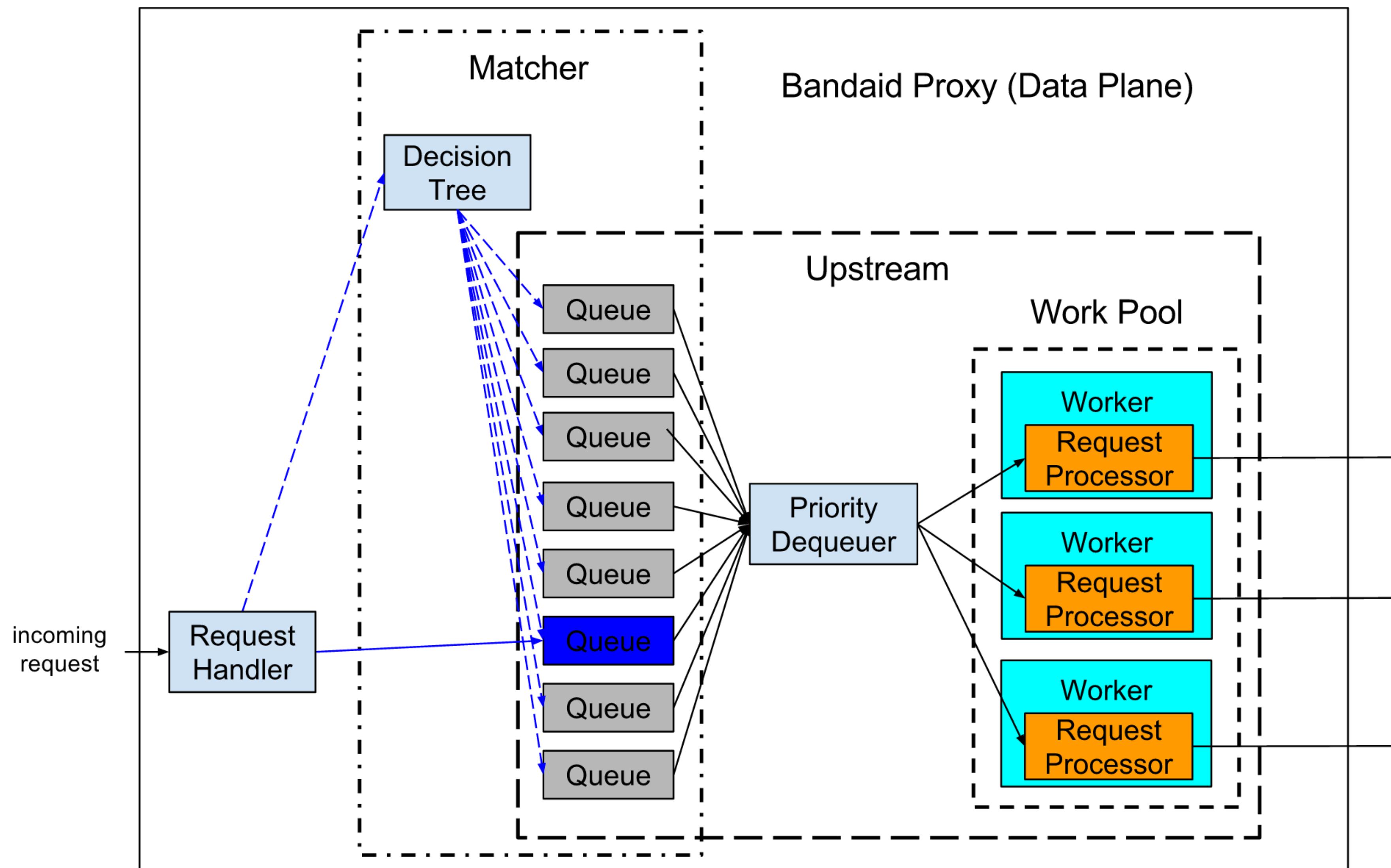
2016



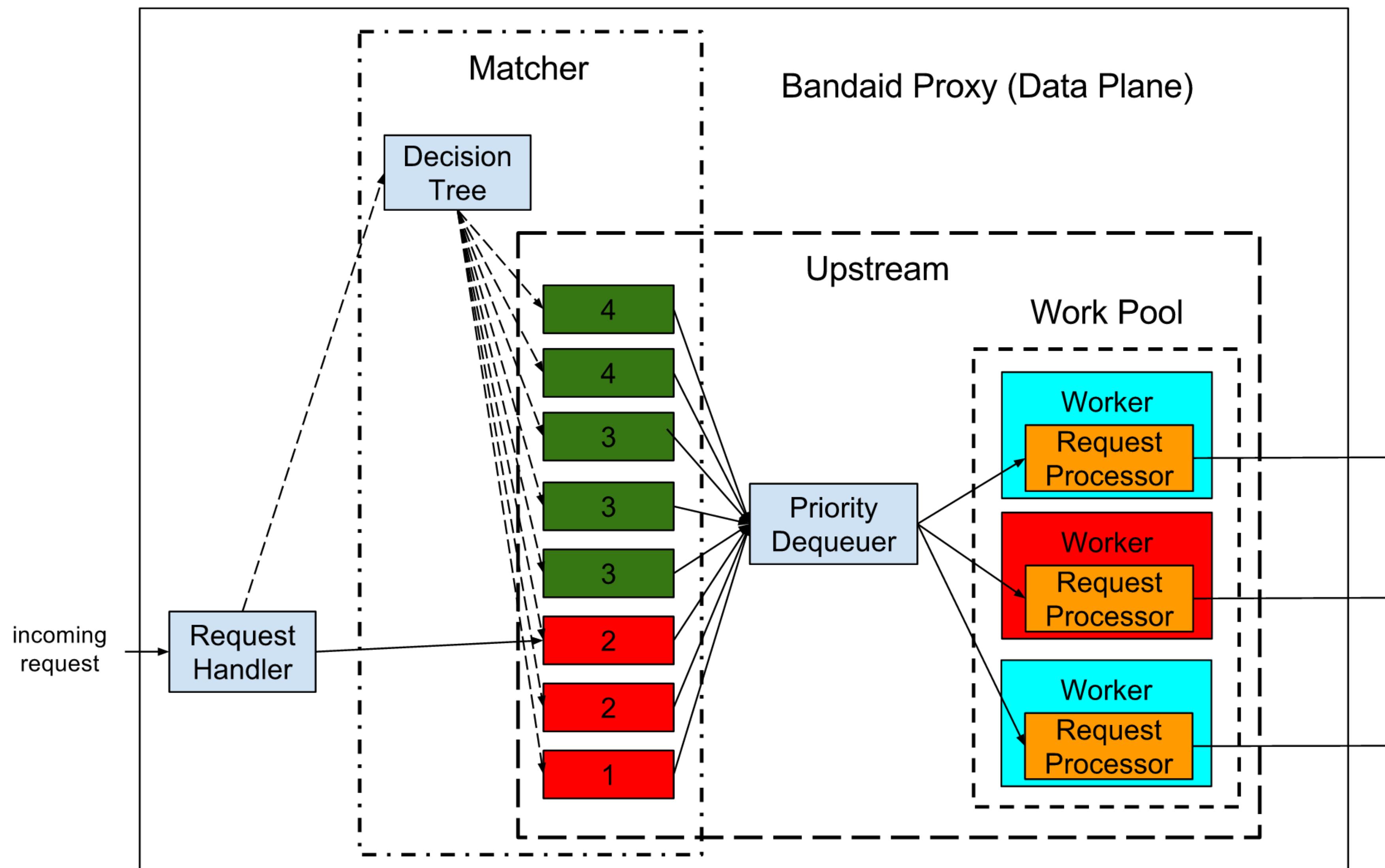
HOW TO PREVENT CASCADING FAILURE?



BANDAID: PER ROUTE ISOLATION



QUEUE PRIORITIZATION



Partition & Isolate
(data or services)

cluster isolation: Metaserver

data model isolation: Edgestore

shard isolation: File Journal

region isolation: Magic Pocket

route isolation: Bandaid

Isolation

