

Riak Use Cases: Dissecting the Solutions to Hard Problems

Andy Gross <@argv0>

Chief Architect, Social Media Liability

Basho Technologies

Riak

- ✦ Dynamo-inspired key value database
 - ✦ with full text search, mapreduce, secondary indices, link traversal, commit hooks, HTTP and binary interfaces, pluggable backends
- ✦ Written in Erlang and C/C++
- ✦ Open Source, Apache 2 licensed
- ✦ Enterprise features (multi-datacenter replication) and support available from Basho

Choosing a NoSQL Database

- ✦ At small scale, everything works.
- ✦ NoSQL DBs trade off traditional features to better support new and emerging use cases
- ✦ Knowledge of the underlying system is essential
- ✦ A lot of NoSQL marketing is bullshit

Tradeoffs

- ✦ If you're evaluating Mongo vs. Riak, or Couch vs. Cassandra, you don't understand your problem
- ✦ By choosing Riak, you've already made tradeoffs:
 - ✦ Consistency for availability in failure scenarios
 - ✦ A rich data/query model for a simple, scalable one
 - ✦ A mature technology for a young one

Distributed Systems: Desirable Properties

- ✦ Highly Available
- ✦ Low Latency
- ✦ Scalable
- ✦ Fault Tolerant
- ✦ Ops-Friendly
- ✦ Predictable

1000s of Deployments



User/Metadata Store Comcast



User profile storage for xfinityTV mobile application

Storage of metadata on content providers, and content licensing info

Strict latency requirements

Notification Service

Yammer

A screenshot of the Yammer notification interface for a community named 'FOUR LEAF CONSULTING'. The interface is divided into three main sections: a left sidebar, a central notifications area, and a right sidebar. The left sidebar contains navigation options like 'My Feed', 'Direct Messages', 'Notifications', 'Community Feed', and 'More'. The central notifications area shows a list of notifications, including mentions, replies, and likes. The right sidebar contains 'Community' information, 'Following Suggestions', 'Group Suggestions', 'Related Networks', and an 'Invite' section with an email input field. The notifications are as follows:

- You were mentioned in a thread:** Sarah Schwartz: @Jessica Halper when will the powerpoint be ready for our meeting on Friday? 11 minutes ago. View thread >
- Phil Spitzer replied to your message:** Phil Spitzer in reply to Jessica Halper: I think this is an excellent idea! 12 minutes ago. View thread >
- Phil Spitzer likes your message:** Jessica Halper in reply to Jesse Wilkinson: Personally, I think producing new product lines is the best strategy because it will help us expand our offering and makes us more competitive. 3 months ago. Liked by Phil Spitzer. View thread >
- Sarah Schwartz likes your message:** Jessica Halper Marketing: Heading down to Peppendine University tomorrow morning to film a video and attend the Social Media Garage meeting. Looking forward to the trip! 4 months ago. Liked by Sarah Schwartz. View thread >

Yammer notification module powered by Riak

Session Store

Mochi Media



First Basho Customer (late 2009)

Every hit to a Mochi web property = 1 read,
maybe one write to Riak

Unavailability, high latency = lost ad revenue

Document Store

Github Pages / Git.io

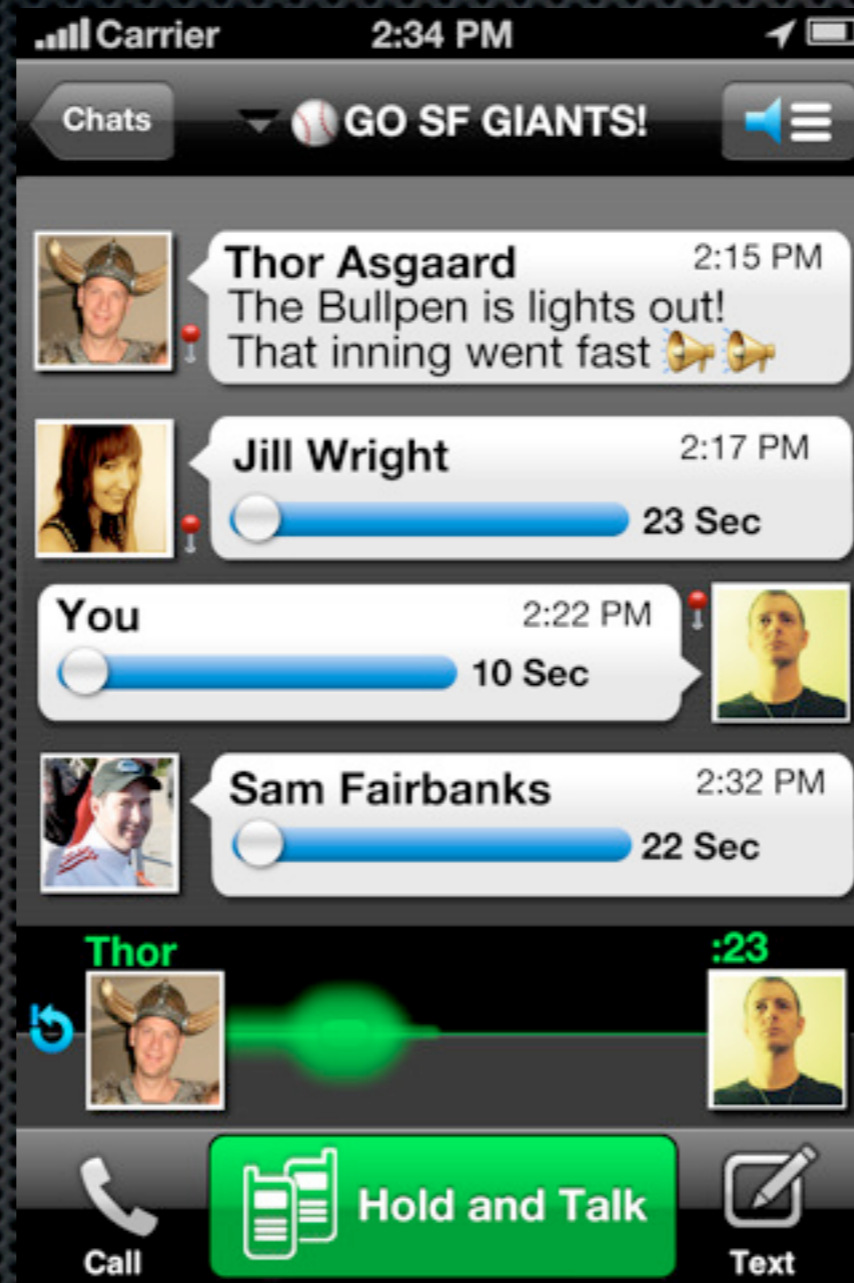


Riak as a web server for Github Pages

Webmachine is an awesome HTTP server!

Git.io URL shortener

Walkie Talkie Voxer



Voxer - Initial Stats

- ✦ 11 Riak Nodes
- ✦ ~500GB dataset
- ✦ ~20k peak concurrent users
- ✦ ~4MM daily requests

Then something happened...

Walkie Talkie App Voxer Is Going Viral On iPhones And Androids, Trending On Twitter



A screenshot of a Twitter post from the account **sodmg.com** (@souljaboy). The tweet text is "Voxer. Soulja Boy." and includes a verified badge that says "They SODMG". The tweet has 50+ retweets and 8 favorites. Below the tweet are icons for Reply, Retweet, and Favorite. The user's profile picture shows a man in a hat, and the bio lists "sodmg.com" and "@souljaboy".



Voxer - Current Stats

- ✦ > 100 nodes
- ✦ ~1TB data incoming / day
- ✦ > 200k concurrent users
- ✦ > 2 billion requests / day
- ✦ Grew from 11 to 80 nodes Dec - Jan

riakmedia v2 free disk

PLAY

SHARE GRAPH

EDIT

EXPORT

Annotations ▾



View: 2012/02/12 00:00 – 2012/03/19 23:55

Past:

2d

1w

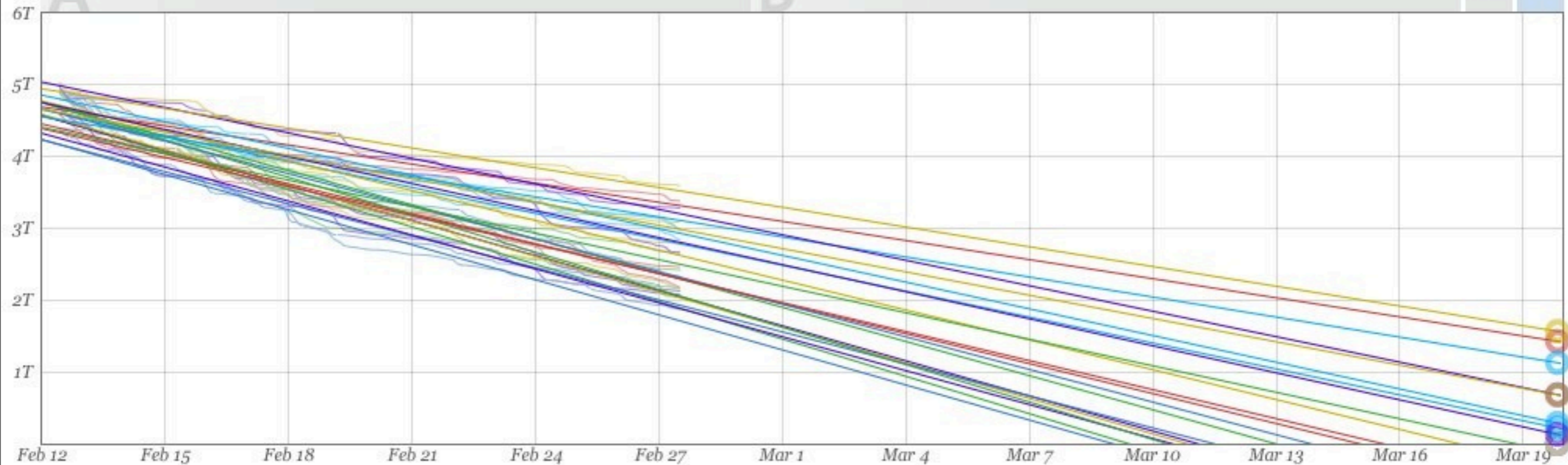
2w

4w

1y



A Original Graph **B** Linear Regression





Distributed Systems: Desirable Properties

- ✦ High Availability
- ✦ Low Latency
- ✦ Horizontal Scalability
- ✦ Fault Tolerance
- ✦ Ops-Friendliness
- ✦ Predictability

High Availability

- ✦ Failure to accept a read/write results in:
 - ✦ lost revenue
 - ✦ lost users
- ✦ Availability and latency are intertwined

Low Latency

- Sometimes late answer is useless or wrong
- Users perceive slow sites as unavailable
- SLA violations
- SOA approaches magnify SLA failures

Who cares about latency?

The screenshot shows a Bloomberg news article. At the top, there is a 'MARKET SNAPSHOT' section with market data for U.S., Europe, and Asia. Below that is a navigation bar with 'QUICK NEWS', 'VIEW', 'MARKETS', 'PERSONAL FINANCE', 'SUSTAINABILITY', 'TV', and 'RADIO'. The main article title is 'Secret Fed Loans Gave Banks Undisclosed \$13B' by Bob Ivry, Bradley Keoun, and Phil Kuntz, dated Nov 27, 2011. The article text discusses the Federal Reserve's secret bailout of banks in 2008. On the right side, there is a 'More Stories' section with links to 'U.S. Stock-Index Futures Gain After Report on IMF Planning Loan to Italy', 'Texas to Ask Supreme Court for Stay of Maps', 'Euro Advances After Report of IMF Italy Loan Plan: Aussie, Kiwi Strengthen', and 'Asia Stocks, U.S. Futures Rally on Italy'. Below the main article is a social media sharing section with 'Recommend', 'Tweet' (956), 'Share' (74), and 'Print' options. At the bottom right, there is an 'Advertisement' placeholder. Several black arrows originate from a central point on the right and point to various elements: the 'U.S. Bonds' link in the navigation bar, the 'U.S. Stock-Index Futures Gain' link, the 'Euro Advances' link, the 'Asia Stocks' link, the main article title, the 'Tweet' button, the 'Share' button, and the 'Advertisement' placeholder.

SOA

Who cares about latency?



Sometimes high latency looks like an outage to the end user.

Fault Tolerance

- ✦ Everything fails
 - ✦ Especially in the cloud
- ✦ When a host/disk/network fails, what is the impact on
 - ✦ Availability
 - ✦ Latency
 - ✦ Operations staff

Predictability

“It’s a piece of plumbing; it has never been a root cause of any of our problems.”

Coda Hale, Yammer

Cost



[@moonpolysoft](#)

Cliff Moon

Amortize the cost of an database across its entire life. Turns out the only thing that matters is operational cost.

6 Nov via [TweetDeck](#) ☆ [Favorite](#) ↻ [Retweet](#) ↩ [Reply](#)

Retweeted by [murf](#) and 22 others



Operational Costs

- ✦ Sound familiar?
 - ✦ “we chose a bad shard key...”
 - ✦ “the master node went down”
 - ✦ “the failover script did not run as expected...”
 - ✦ “the root cause was traced to a configuration error...”
- ✦ ***Staying up all night fighting your database does not make you a hero.***

High Availability: Erlang

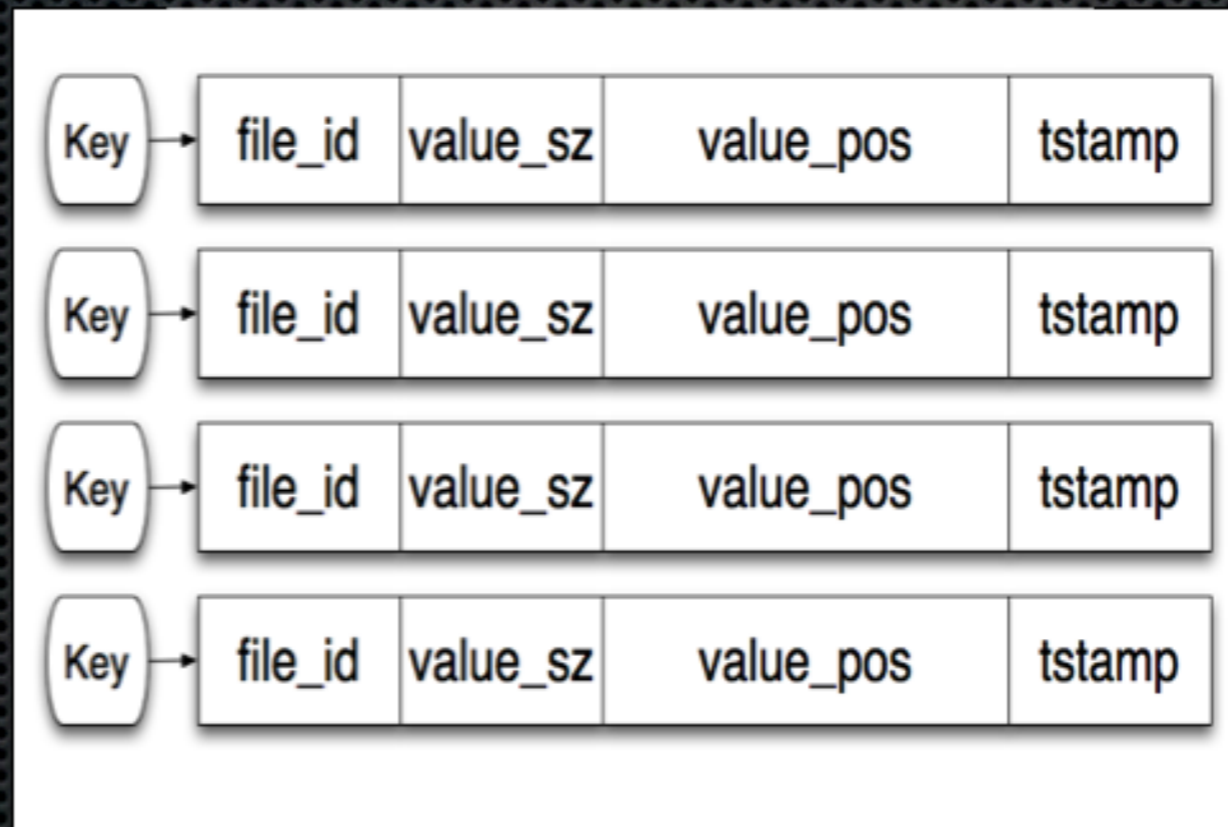
- ✦ Ericsson AXD-301: 99.99999999% uptime (31ms/year)
- ✦ Shared-nothing, immutable, message-passing, functional, concurrent
- ✦ Distributed systems primitives in core language
- ✦ OTP (Open Telecom Platform)

High Availability: Riak Core

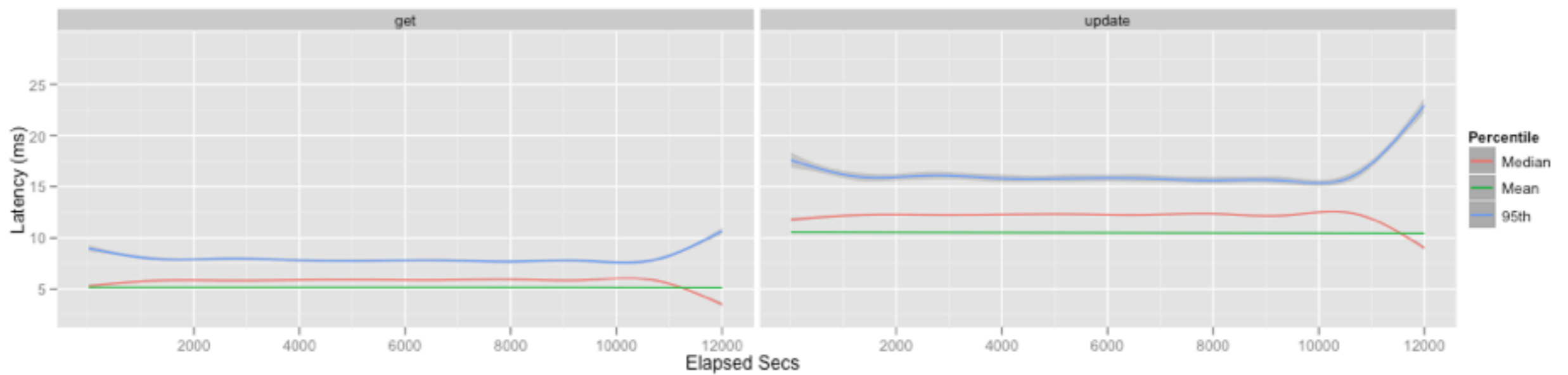
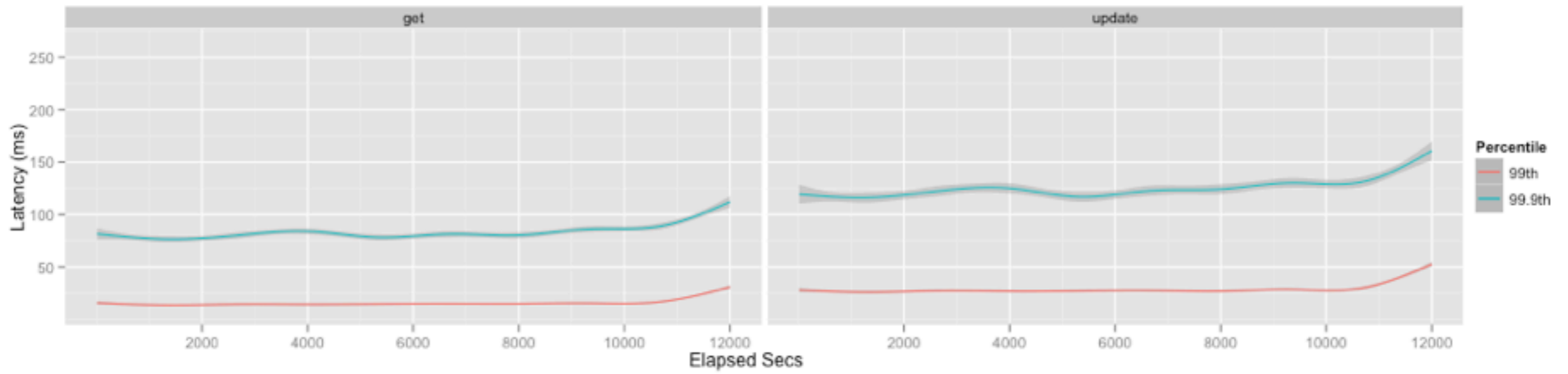
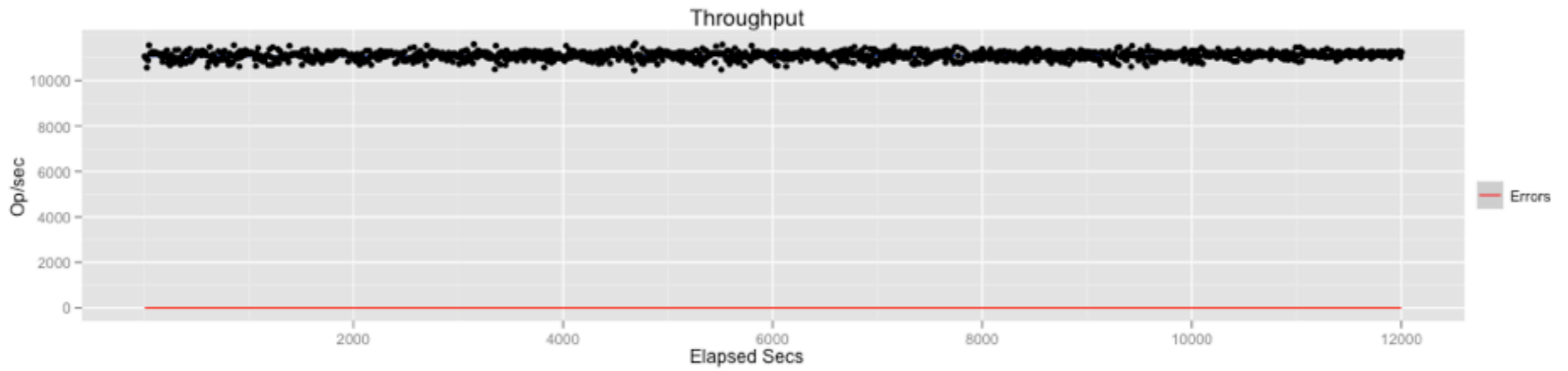
- ✦ Dynamo abstracted: distributed systems toolkit
- ✦ Exhaustively tested
- ✦ In production use at AOL, Yahoo, others
- ✦ Insulates local storage and client API code from the hard problems

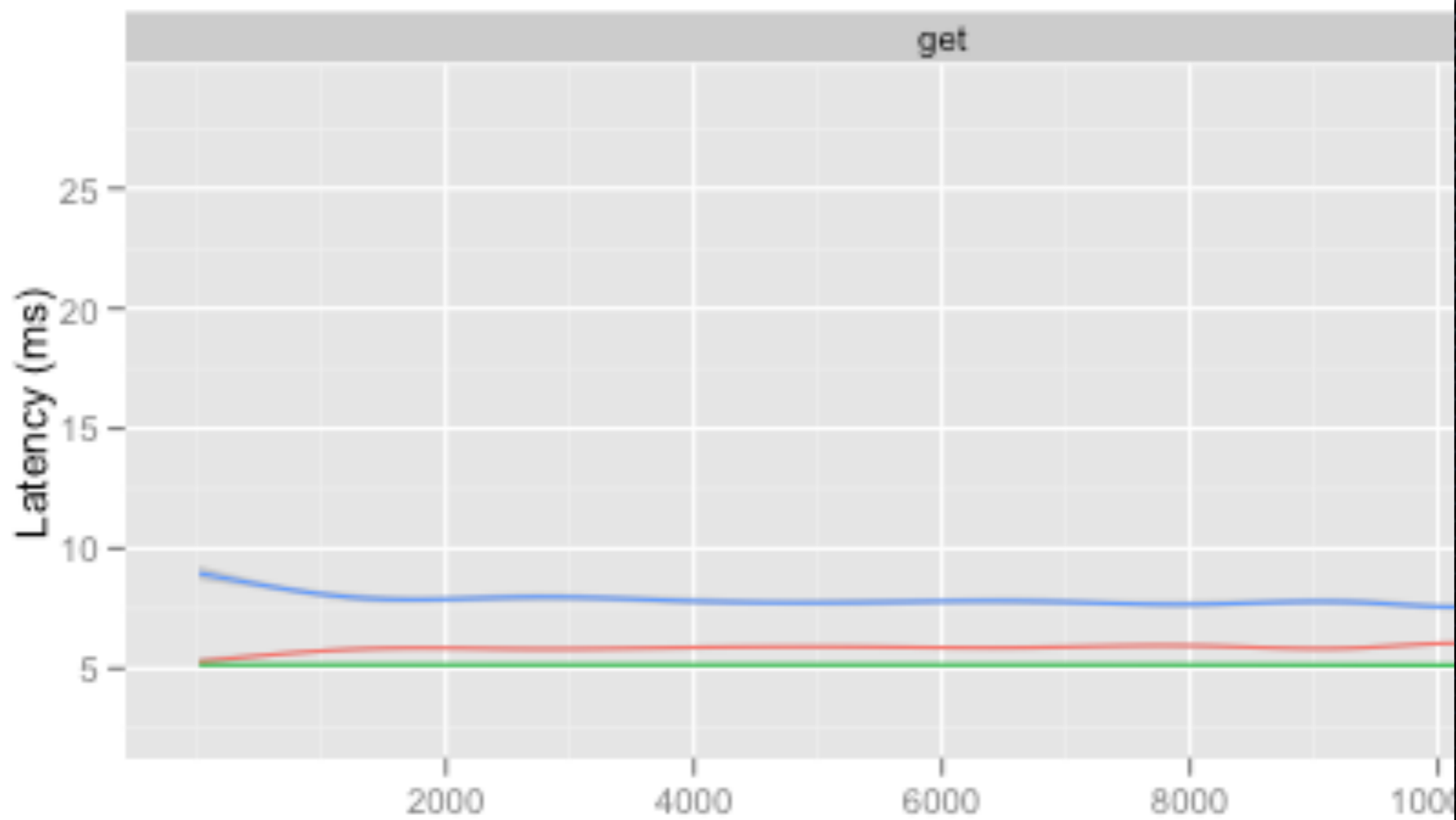
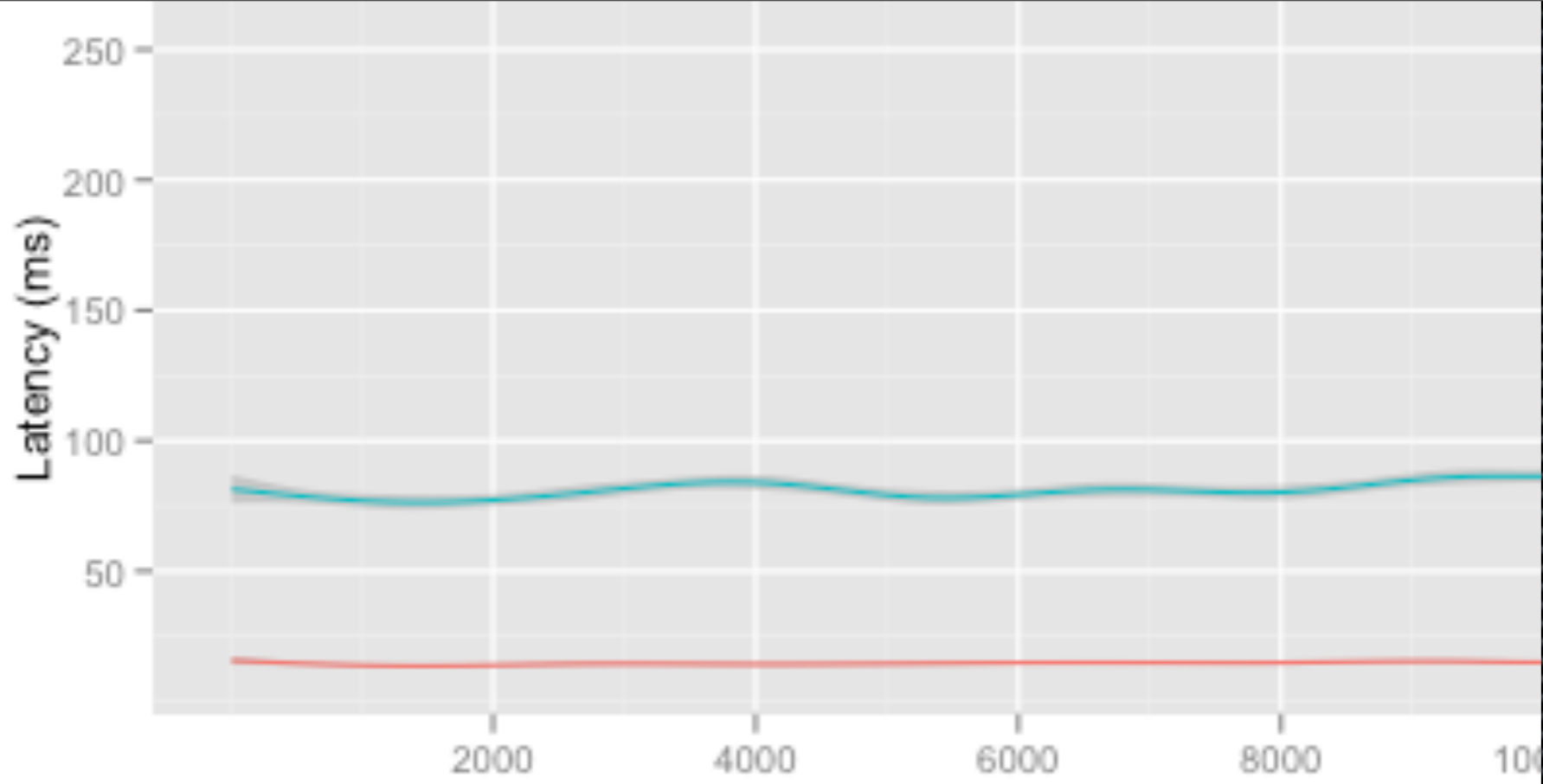
Low Latency: Bitcask

Low Latency: All reads = hash lookup + 1 seek



Tradeoff: Index must fit in memory





Low Latency: Erlang VM

- ✦ Erlang VM was designed for soft-realtime apps
 - ✦ Preemptively scheduled lightweight threads
 - ✦ GC is per-thread, not stop-the-world
- ✦ Sophisticated scheduler + message passing = effective use of multicore machines.

Questions?