Replacing Cassandra: A Digital Transformation for the World’s Largest Digital Identity Network

Matthias Baumhof, Vice President Technology
Nick Blievers, Senior Director Engineering
ThreatMetrix, A LexisNexis Risk Solutions Company
Trusted User
or Cyber Threat?

"On the Internet, nobody knows you’re a dog."
Our intelligence is powered by 5,000 companies in our global network
Global Shared Intelligence:
130M transactions daily across 40,000 websites

The world’s largest Digital Identity Network
Allowing for quick and confident customer recognition

Source: Q1 2018 ThreatMetrix Cybercrime Report
Global Shared Intelligence at Scale

Transactions

• 3bn API transactions per month
• 55% are mobile

Active Entities

• Over 4.5bn active devices protected
• Over 1.4bn active accounts protected

~50 countries / second
~160 countries / minute
~240 countries / hour
Universal fraud & authentication decisioning in real time at global scale in the cloud

Customer calls API

< 100 ms

ThreatMetrix returns 100’s of attributes
Real-Time Decisioning Example: Identifying trusted returning customers

Account Registration and successful Login

After 12 months

User forgot password while trying to login, challenge?

No

Yes
Switching to Aerospike
General Requirements

• Predictive and fast performance with reads, max 1ms
• Read latency should not be affected with a heavy write workload
• I/O capacity for millions of reads/sec and writes/sec
• Predictive and fast access to 10’s of billions of keys and > 100 TB of raw data
• Node count in a manageable range: < 50 nodes
• Highly available, stores multiple copies of the data
• Multi Datacenter replication with minimal latency: < 500ms
• Easy to administer, operate and expand
• Clients for typical programming languages
Storage Requirements
Performance Requirements

• Slowdowns in user journey cost $$$
• Reportedly 100ms can cost Amazon 1% in revenue*
• Ever Increasing demand
What we were using

Custom SQLite clustered solution

Cassandra
The new storage layer

- “10x faster than what you use now”
- Sounds good, possibly too good to be true?
Migration

Custom SQLite → Cassandra → Aerospike
The Path To Better Performance
Is it really 10x?

*The new hardware is faster and has 3x the RAM, and better SSDs*
Unexpected benefits

- SSD write endurance
- Support
- User Defined Functions
What do we think now?

- World class products are a collection of trade-offs... but the right trade-offs.

- Aerospike does make concessions, but delivers in it’s promise of blindingly fast performance

- Can’t argue with the results!
Our Cluster Details
Cluster Growth
(\(\text{per Datacenter}\))

Now

28 Nodes
450 TB disk
20 TB RAM

End of 2018

48 Nodes
1,300 TB
35 TB RAM
ThreatMetrix real-time API architecture
Real-world performance

24h window, ~40,000 writes / sec, ~10,000 reads / sec

Average read latency: ~ 0.35ms measured at Analytics Engine
Thank you

Americas
+1 408 200 5700

Asia Pacific
+61 2 9411 4499

EMEA
+31 (0)20 800 0638

Sales
sales@threatmetrix.com

Partnerships
partners@threatmetrix.com

Web
threatmetrix.com