AEROSPIKE USER SUMMIT 2018

AEROSPIKE

When Five Nines Is Not Enough: What 100% Uptime Looks Like

Timothy G Smith
Senior Vice President & General Manager
Global Technical Infrastructure & Operations
AppNexus

Who is AppNexus?

- **AppNexus is an internet technology company that powers the real-time sale and purchase of digital advertising**
- **We help publishers sell advertising space**
- **We help buyers purchase effective advertising**
- **Before you ask...**
- **Our platform supports billions and billions of transactions each day**

 All rights reserved. © 2018 Aerospike Inc

Billions and Billions of Transactions Per Day?

- **AppNexus conducts 10B ± 1.5B transactions per day**
  - NYSE 4.2M transactions per day
  - Nasdaq 11.2M transactions per day
  - Visa 150M transactions per day
  - All of the above are average daily transactions in January 2018
  - Combined NYSA, Nasdaq, and Visa had about 5B transactions in the month of January
- **Several million dollars of "Gross Transaction Value" per day**
- **Each transaction is for very small (tiny even) amounts of money**
- **Lots of opportunities to transact every day: in excess of 300B most days**

 All rights reserved. © 2018 Aerospike Inc

AppNexus Operational Environment

- **AppNexus does not close. Ever!**
- **AppNexus has no natural or scheduled maintenance windows where ad serving can be unavailable**
- **Aerospike must be available every second of every day because it is in the real time path**

 All rights reserved. © 2018 Aerospike Inc

How do you achieve 100% uptime?

- **Use Internet Routing (BGP) and DNS to flexibly route traffic to data centers**
  - Route around failures
  - BGP anycast techniques are well understood
  - Simple load balancing using “smart” DNS server
  - More sophisticated load balancing using stateful proxies like NGINX
- **Everything must scale horizontally**
- **Everything must survive individual server failures**
- **If a whole data center goes down send the traffic to another data center**
- **Simplifying assumptions: Each individual transaction by itself is very low value so a handful of failures in 5 million requests per second is a rounding and/or measurement error**

 All rights reserved. © 2018 Aerospike Inc

Basic Real Time Ad Serving Traffic Flow

- BGP and DNS get the traffic to a data center
- One of many proxies gets the inbound HTTP connection
- Proxy passes it through to AppNexus Impression Bus server node
- Impbus requests data from numerous data sources including Aerospike
- Impbus hands request to AppNexus Bidder or customer's bidder
- Business logic applied
- Transaction occurs or does not occur depending on business logic
- Generate records about the transaction and hand off to non-real time data pipeline which keeps track of various things like who spent what where
- Repeat up to 5m times per second

 All rights reserved. © 2018 Aerospike Inc

Aerospike at AppNexus

History

- **Aerospike has been in a critical path at AppNexus for 7+ years**
  - Since before it was called Aerospike
  - It replaced a memcached product from Schooner that no longer exists
  - Since before Aerospike had any Enterprise-level features
    - No Quick Restarts
    - No XDR
    - No Backups
  - Support used to be calling Srini or Brian directly
- **Aerospike has successfully grown with AppNexus**

 All rights reserved. © 2018 Aerospike Inc

Aerospike at AppNexus

Current Deployment

- 18 clusters spread across 6 data centers on 3 continents
- 200 total Aerospike nodes
- Aerospike holds 100TB of data
- Largest SSD based cluster holds 28TB of data
- Largest memory based cluster holds 5TB of data
- Cluster sizes are currently between 2 and 28 nodes

 All rights reserved. © 2018 Aerospike Inc

Details on one cluster

- **24 node SSD-based cluster**
- **Each node**
  - 2 x 12 core E5-2670 v3 @ 2.30GHz CPUs
  - 512G RAM
  - 24 x 400GB Intel DC S3700 SSDs
  - 4 x 10G NICs
- **Why 2.5 inch SATA SSD instead of NVMe?**
  - 3+ years ago there were far fewer NVMe offerings
  - Testing of NVMe drives indicated that they were not ready for prime time (bandwidth and firmware limitations)
  - The Intel DC S3700 SATA SSDs tested very well and have worn very well

 All rights reserved. © 2018 Aerospike Inc

More details on a cluster

- **Cluster contents**
  - 9+ Billion objects
  - 28TB of data
  - Object size range is roughly 45% < 512 bytes and 55% > 512 bytes
- **Transaction rates**
  - 1M reads/s sustained for 12+ hours a day
  - Between 200k and 300k writes/s 24 hours a day
  - 10Gb/s of outbound network traffic for 12+ hours a day per node
- **Response times**
  - Read: < 8ms is 99.8%
  - Write: < 8ms is 99.9%

 All rights reserved. © 2018 Aerospike Inc

Operational Experiences

- **We replace every server every 3 years**
- **Upgrades are fun!**
  - They did not used to be though
- **We perform upgrades/restarts at all hours**
  - Middle of the day? Yes indeed!
  - During peak request times? Absolutely positively yes.
- **Why?**
  - Because we can!
  - The impact is very very minimal since restarts are very fast
  - Ops staff and engineers are around just in case something does happen

 All rights reserved. © 2018 Aerospike Inc

Operational Experiences

- **Hardware fails. It happens.**
  - Best case
    - Hardware fails completely
    - Aerospike automatically rebalances
  - Worse case
    - Hardware failure causes a single node to be partially up
    - User intervention required to turn down Aerospike on that node
- **Aerospike Cluster response**
  - Possible degradation in service until the broken node is removed
  - Cluster automatically rebalances to maintain object redundancy
  - Clients immediately react appropriately to cluster membership changes

 All rights reserved. © 2018 Aerospike Inc

Operational Experiences

Capture all instrumentation available

- **Aerospike has a ton of good statistics**
  - Capture them
  - Store them
  - Look at them and think about them
  - Alert on them
- **Keep an eye on, at minimum**
  - Cluster Size
  - System Utilization
  - Available %
  - Client read/write error rates
  - Device/Memory free percentages

 All rights reserved. © 2018 Aerospike Inc

Operational Experiences

Capacity Issues – Your business growth is good!

- **What changed?**
  - Request rates?
  - Object sizes?
  - Write-cache hit rate?
- **Adjust configurations**
  - Can we tune our way out of this?
    - Are there available resources we are not utilizing?
  - Can/Should we increase/decrease any settings?
  - Evaluate and adjust settings with help from documentation and/or Aerospike Support

 All rights reserved. © 2018 Aerospike Inc

Operational Experiences

Capacity Issues Continued

- **Grow horizontally**
  - Add more nodes.
    - Buy... Beg... Borrow... Whatever...
  - Aerospike Cluster automatically rebalances
  - Aerospike Clients automatically utilize the new nodes
- **Grow “vertically”**
  - Evaluate new hardware
    - Discuss with Aerospike support
  - Use ACT to validate SSDs
  - Spec newer generation of CPU, memory, NICs
  - Test Test Test

 All rights reserved. © 2018 Aerospike Inc

Summary

- **Use Aerospike**
- **Do not be afraid**
- **Do not treat it like Oracle or MySQL or SQLserver**
- **We run it to redline sometimes so maybe you can too. YMMV**
- **If something can go wrong with Aerospike core key/value store features there is a good chance we have seen it over the years and already opened a ticket with Aerospike and Aerospike already fixed it**
- **We have not had an actual outage since <INSERT DATE HERE> and that one was at least half our fault**

 All rights reserved. © 2018 Aerospike Inc