membase.org: The Simple, Fast, Elastic NoSQL Database
NorthScale
Matt Ingenthron
OSCON 2010
Membase is an Open Source distributed, key-value database management system optimized for storing data behind interactive web applications.

All aspects of membase are *simple, fast and elastic* by design.
Value

image courtesy http://www.flickr.com/photos/vintagedept/3617706196/
Simple

(with a replica)
Fast

- Original use case: speed up access to authoritative data as a distributed hashtable
- Must be at *at least as fast* as a highly tuned DBMS
- Designed for modern datacenter substrate
  - Designed for VM and cloud deployments
  - Designed with modern systems and flash storage devices in mind

image courtesy [http://commons.wikimedia.org/wiki/File:Disassembled_HDD_and_SSD.JPG](http://commons.wikimedia.org/wiki/File:Disassembled_HDD_and_SSD.JPG)
Elastic

- Add nodes without losing access to data
- Maintain consistency when accessing data
  - membase is a CA type system
- Scale linearly by just adding more nodes
Who?
Fault-tolerant memcached Cluster

ARCUS cache cloud

at 

the biggest web portal in Korea
What is Project Arcus?

- Memcached
  - Common protocol across PHP, Java, C applications
- Moxi (Memcached proxy) based
- In-house automatic fault-detection and failover solution
- Collectd-based monitoring
- Proxy and cache server administration UI
- Private cloud service
Previous Deployments

- A few individual memcached installations

Problems
- No fault-tolerance
  - Hardware failures are common (heat, network switch failure, etc)
- No automatic scalability
  - To add / remove a memcached server, they need to rebuild code, distribute, and restart all clients
Today

- Memcached clusters
  - Fault-tolerance transparent to clients
    - Consistent hashing in moxi (memcached proxy)
  - Cache As A Service (CaaS)
    - All major services in NHN started using cache
    - Multitenancy across cache services
Tomorrow

- Membase deployments being developed
  - Persistence (Memory, SSD, HDD)
  - *No proxy* – no additional network latency
    - Smart-clients
    - Failover at the client built upon libvbucket
- Server provision and monitoring UI
Performance impact

**Performance**
- Response Time: X 10
- Throughput: X 16.6

**DB Load**
- 50%
- 34%
NHN work in membase.org

- Provided architectural direction
- Provided real-world use cases and specific requirement
- Bug fixes, code review
- Feature contributions
  - spymemcached
  - moxi enhancements
- Sponsors membase open-source
The Guts

Photo Courtesy http://www.flickr.com/photos/pellia/76804760/
Clustering

- Underlying cluster functionality based on erlang OTP
- Have a custom, vector clock based way of storing and propagating...
  - Cluster topology
  - vBucket mapping
- Collect statistics from many nodes of the cluster
  - Identify hot keys, resource utilization
vBuckets

- Hash keys to vbuckets not to servers
- Shared, clusterwide configuration mapping vBuckets to servers
  - Any given vbucket is only enabled in one location at a given time
  - Replication locations specified as part of the vBucket mapping
- moxi and vbucket aware clients receive updated configurations
  - A “not my vbucket” reply is an indication that the client vbucket map is wrong, needs update
- See Dustin’s blog for more information
TAP

- A generic, scalable method of streaming mutations from a given server
  - As data operations arrive, they can be sent to arbitrary TAP receivers
- Leverages the existing memcached engine interface, and the non-blocking IO interfaces to send data
- Three modes of operation
Clients, nodes and other nodes

- Client
  - port 11211 memcached operations
- moxi + Client
  - port 11210 memcached operations
  - REST/comet cluster topology and vbucket map

- moxi
- membase (memcached + membase engine)
- ns_server

- TAP
- vbucketmigrator

memcached operations with tap commands
Get Involved

- **Open Source**
  - Nearly all Apache 2.0 licensed
  - Some components BSD based, owing to history
- **Try using it!**
  - NorthScale packaged, tested, integrated installs
    - RPM for RHEL/CentOS 5.2 and later
    - .deb for Ubuntu 9.04
  - Let us know what you need
    - Already questions about iterating
- **Looking for platform porting**
  - Have some RedHat/CentOS, Ubuntu covered
    - Could certainly integrate a bit better
    - Especially needed with moxi
Roadmap

- Beta 2: shipped July 21, 2010
  - Feature: disk > memory
  - Feature: standalone moxi
  - Feature: binary protocol
  - Enhanced: Web UI and management

- Beta 3, within this quarter
  - Planned: Web UI, analytics and visualization
  - Planned: Multi-tenancy, bucket management rules
  - Planned: more TAP examples

- General Availability follows
Finding us

- membase.org
- http://wiki.membase.org
  - Has some docs, ramping up more docs, examples
- http://groups.google.com/group/membase
- Chat options
  - irc.freenode.net #membase
  - XMPP MUC: membase@membase.org
Q&A

Matt Ingenthron
matt@northscale.com