Serving Analytics in Real-time to 10000s of Servers

Rushan Chen
Analytics Platform
rchen@zynga.com

XLDB 2011, Menlo Park, CA
Access to User Insights

Game Web Servers

Aggregates(R)
Other real-time data(R/W)

Data Services Membases

Aggregates

Vertica Data Warehouse

Tracking Data
Built Using Membase

• Membase
  — <K,V> store like Memcached
  — with persistence(sqlite) & replication
  — by Couchbase with contribution from Zynga

• Zynga Data Services
  — one of the bigger deployments of Membase
  — one master cluster + multiple slave clusters
  — only write to master
Data and API

• 5B items in each Membase cluster
  — daily aggregates + other real-time data
  — info on each user + indexed lists of users

• API
  — direct access library
  — REST

• Usage
  — in many services and most Zynga games
  — to quantify, filter and compare
Operation

- Server configuration tool: zookeeper
- Connection pooling: mcmux
- Monitoring and alerting: munin + nagios
  - replication queue
  - memory usage
  - request and connection rate
- Backup and Recovery
  - backup sqlite db files on schedule
  - restore from defragmented backups.
Next Steps

• Improved Replication
• Server Side Clustering
• API Performance