

# Cassandra at Cloudkick

Dan Di Spaltro  
dan@cloudkick.com

Paul Querna  
paul@cloudkick.com

# Background - What is Cloudkick?

- Cloudkick is a Monitoring Platform
- Cloudkick stores lots of metrics per our data policy
- Previously, we used Postgres, but outgrew it fairly quickly
  - We had a high constant write load
  - Reads got slower and slower
- Alerting and Data collection
  - Cloud monitoring/management as a service
- Cassandra == Data Collection/Trending

# Data Model

- Archive CFs for raw data
- Rollups CFs for 5m, 30m, 1h, 4h, 1d
  - Think Lossless RRD
- Fat rows, relative low amount of keys vs columns
- Row cache doesn't work for us
  - Key cache does
- More details:

[https://www.cloudkick.](https://www.cloudkick.com/blog/2010/mar/02/4_months_with_cassandra/)

[com/blog/2010/mar/02/4\\_months\\_with\\_cassandra/](https://www.cloudkick.com/blog/2010/mar/02/4_months_with_cassandra/)

# Configuration

- Random Partitioner
- Higher memtable thresholds
- Default concurrent reads/writes
- Longer RPC timeouts
- No row cache
- 100% key cache
- Inserts are generally CL.ONE
- Running 0.6.0-RC1
  - Hey! It works...
  - rolling upgrades please!

# Client Code

- Custom Wrappers over Thrift, providing closer to ORMish functions
  - Written before things like LazyBoy
- We use it from:
  - Python (Django World)
  - Python (Twisted World)
  - Java
  - C++
- Don't write your own Thrift Wrapper unless you really need to. It sucks.
  - Many options now!

# Cassandra on Cloud Providers

- We run production in a cloud environment
  - FailureDetector configurable!
  - Watch your steal CPU / noisy neighbors
- Limited IO/disk options
  - Pretty pathetic
  - Careful about 50% disk capacity in  $\leq 0.6$ 
    - or you are screwed
  - Performance impact of Major Compaction
- Capacity planning is hard, so you should do it
- Trading agility for raw performance

# What we want to do but really don't want to do

Upgrade to 0.7

We're Hiring!