

diggg

Owen Byrne, Co-Founder/Sr. Software
Engineer

*IEEE San Francisco Communications
Society*

Digging Into the Technology Behind the Development of Digg

- Introduction
- How we Scaled
- Conclusion
- Q&A

What is Digg?

The Idea - Oct. 2004

- Slashdot - User created content
- Del.icio.us - Bookmark / Share / Save
- Friendster - Social hookups, large groups, no task
- Macrumors.com - Page 2
- Great content that didn't make it to the home page of many news sites

Digg: Democratizing Media For All

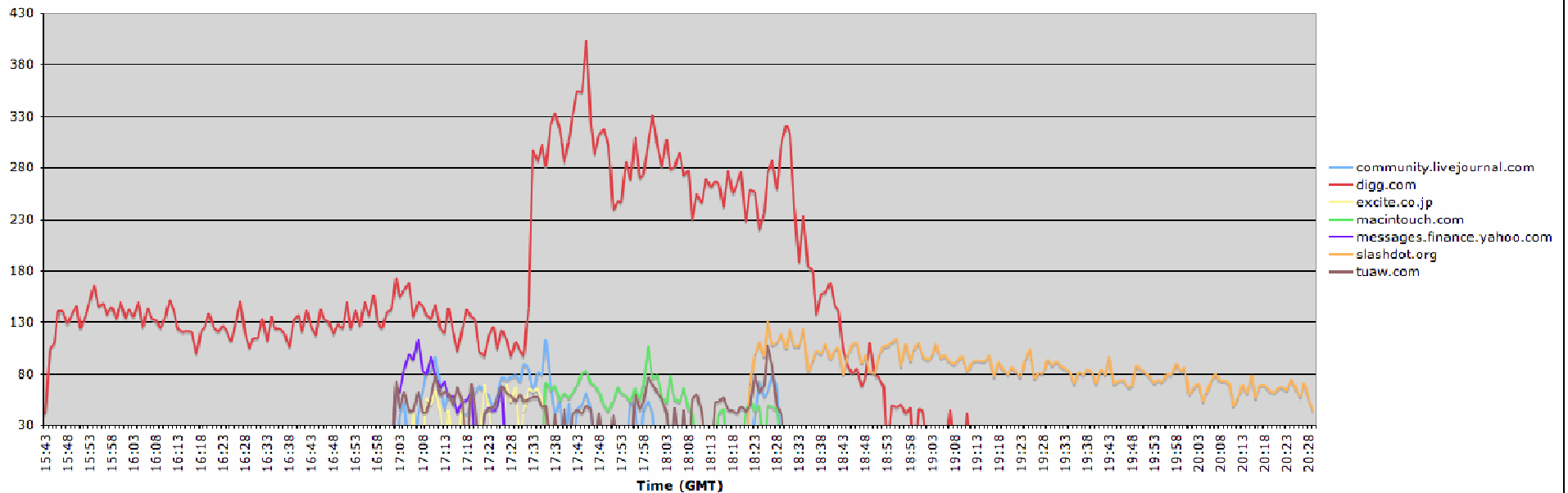
- Passionate community - individuals have driven our success (e.g. AlbertPacino, digitalgopher)
- Editors and reporters sit on the Digg homepage (including my former employer)
- Readily adaptable to other media - e.g. videos, podcasts

Stats

- 1M+ Registered users
- 10M+ Pages per day
- 15M+ Unique Visitors/Month
- 10% monthly growth (users and pageviews)
- 100+ Linux boxes
 - Web Servers
 - Memcache Servers
 - Mysql Slaves
 - Development
 - Monitoring, spares, office support, etc.

The “Digg Effect”

Incoming referrers per minute from large sites



\$2,000 to Launch

- Developed a project spec - Elance
- Open source technologies - LAMP
- Basic utilitarian design
- \$99 a mo. hosting

Feature Decisions

- Innovation – Avoid me too features
- Simple and rewarding - one click (Digg, bury, #1)
- Used “ajax” where it made sense
- Tools to connect to other sites (blog this, javascript widget)
- Experiment - spy, “cloud view,” visualizations (stack, swarm)

Pre-Funding (Dec '04-Mar '05)

- Monitoring is “someone is working on the site most of the time”
- Standard LAMP architecture
- Growth constrained by hardware
 - Optimization
 - Denormalization
 - Thanks to Jeremy Zawodny

Growth

- Paris Hilton
- Thomas Hawk / Price Rite Photo
- Word of mouth, PR, minimal advertising
- New categories added in June 2006
- Steady growth with occasional insanity

Problems

- Log files
- myIsam bad, InnoDB good
- Mysql full-text search doesn't scale
- Javascript compression

Seed Funding Investment

- Small Number of Servers (May 2004-Sep 2004) - web server, mysql master, mysql slave
- Ad-hoc monitoring
- One Dedicated operations person
- Silverorange (Digg gets a professional design) -- Daniel Burka
- Growth continues to outstrip hardware, though now more constrained by developer resources

Series A Investment

- \$100k on Servers. Spares!
- Everyone in the same location
- Operations Department (currently four people)
- Hire senior developers

Description of Digg Architecture

- LAMP + memcached
- MySQL 5, innodb
- Lucene full-text search
- Digg 1.0 (Dec. 2004)
- Digg 2.0 (July 2005)
- Digg 3.0 (July 2006)

Open Source

- Monitoring - Nagios, Cacti
- Javascript - Prototype, Scriptaculous
- Backend - Memcached

Development Process

- Tools

- Subversion

- Bugzilla

- Wiki

- Virtual Servers

- Growing the Team. 1/7 local

- <http://digg.com/jobs>

Conclusion

- Engage with your users
- Don't forget about the business model
- Be frugal
- Think about scaling early

Q&A