

Social Media Analytics Using Greenplum's Data Computing Appliance

Johann Schleier-Smith
Co-founder & CTO
Tagged Inc.

@jssmith

February 2, 2011

What is Tagged?

- 3rd largest social network in US (minutes spent)
- We enable anyone to meet and socialize with new people
- 100 million member accounts
- Longest sessions of any social network



Social Media Challenges

- Competing products are always a click away
- 1 or 2 years to obsolescence
- 50+ things on every page
- Collective behaviors dominate

Must have world-class analytics for big data

The screenshot shows a Facebook profile for 'Johann'. The profile picture is a man wearing sunglasses. The cover photo features Adobe and Photoshop logos. The page is filled with various posts, including text updates, photos, and video thumbnails. A 'What's New?' section at the top right shows a grid of items from friends and groups. The left sidebar contains a list of user statistics and activity. The bottom of the page has a comment input field.



Analytics with Greenplum

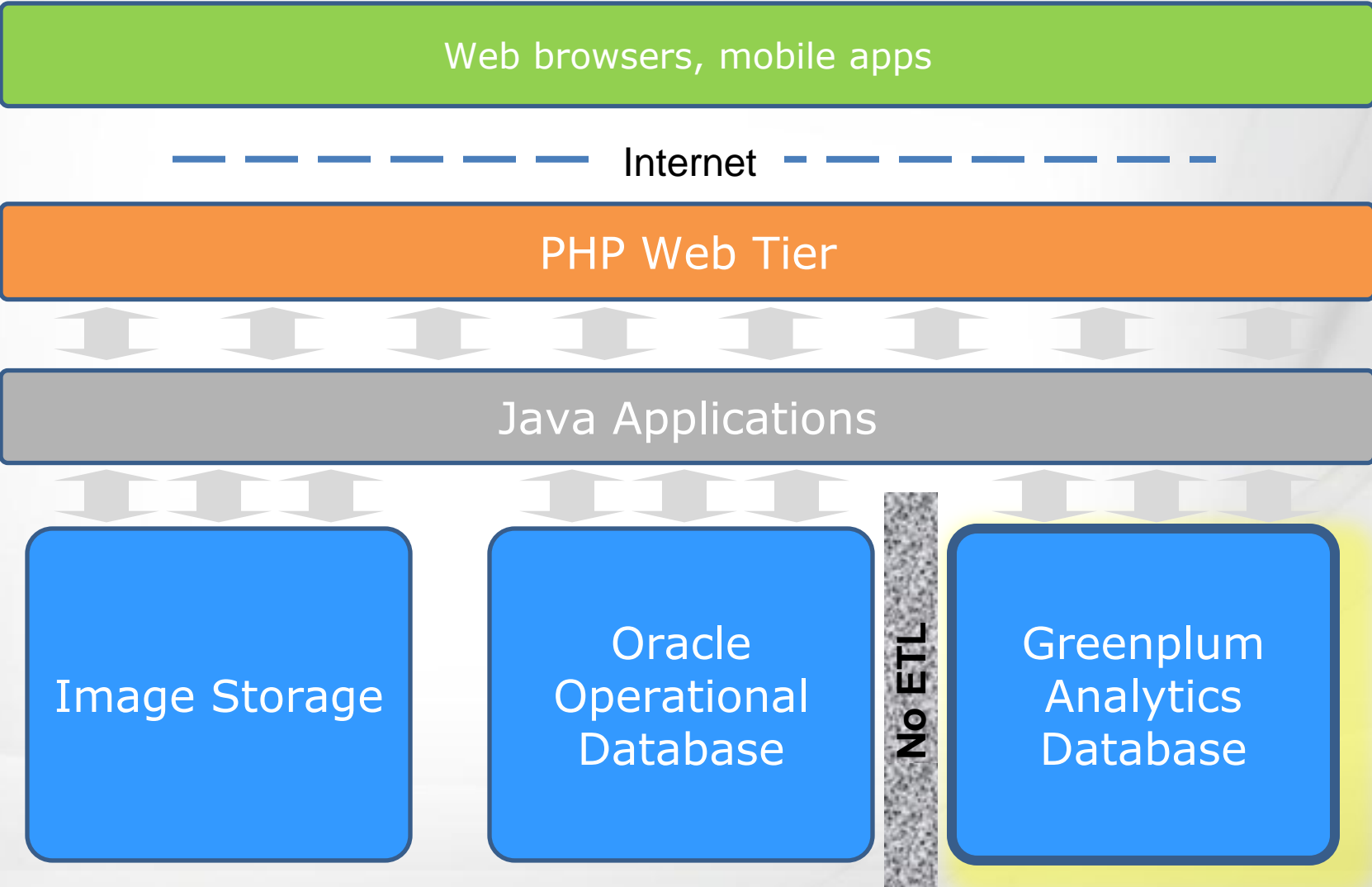
Case Studies

Best Practices

Some Numbers

- Monthly Metrics
 - 5 billion page views
 - 20 million visitors
 - 3 PB downloaded
 - 10 billion ad impressions
- Monthly Logging
 - 50 billion log entries
 - 120 log streams
 - 10 TB log data

Tagged Infrastructure



Greenplum Data Computing Appliance



- Tagged deployment since 2008
- 2 cabinets
- 8 x data nodes
 - 8 CPU cores
 - 32G memory
 - 48 x 500G disk
- Compute machine
 - 64 CPU
 - 256G memory
 - 192TB disk
 - 384 spindles
 - 32 gbit network throughput
- Greenplum professional services determined appropriate sizing

Greenplum's compute and data architecture

- Same scalability principles used for big web services
- No data too big
- Always fast since CPU and storage scale together
- RAID + replication for data protection and availability

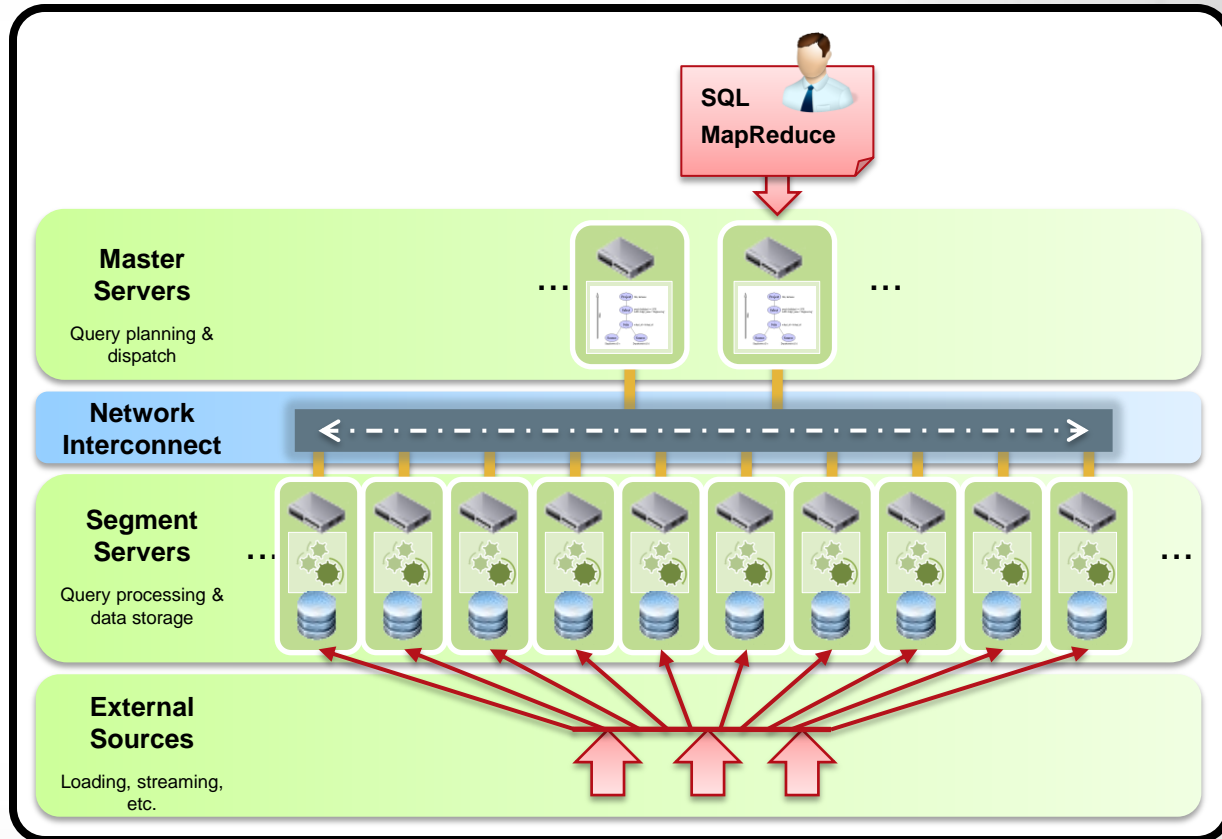


Diagram courtesy of EMC

Logging Approach

- Events for everything
 - Login
 - Page view
 - Friend request sent, confirmed, rejected
 - Message sent, read, deleted
 - Virtual gift given
 - Pets game trade
 - Farm game crop planted
 - Mafia player attacked
- No dependence on operational database
- Log more rather than less
 - Save data as best suited for analysis
 - Include extra “state of the world” information
 - Don’t think about space or cost constraints
- Keep data for 6-24 months

Analytics Approaches

- Dashboards / reports
- A/B testing
- Ad-hoc study
- Predictive modeling

Dashboards

- Java + JavaScript graphics libraries are our platforms
 - YUI tables and forms
 - Google visualization API
 - Our business is software so we build it ourselves
- Easy for everyone to check metrics, see what may have changed
- Visualizations quick to scroll through, over 100 graphs
- Working on automated anomaly detection



A/B Testing

Report ID: [67962](#), [67961](#), [66521](#), [66484](#), [66368](#), [66367](#), [65962](#), [65929](#), [65021](#), [65002](#), [65001](#), [64991](#), [64985](#), [64504](#), [64503](#), [64502](#), [64501](#), [64364](#), [64363](#), [64362](#), [64361](#), [64344](#), [64343](#), [63937](#), [63936](#), [63935](#), [63783](#), [63632](#)

Run Dates: From: 01/31

Report Name:

Bug Number: (delimit with comma) TA:

Description:

Options: Email No

Slices: gender, cc_iso, age, pets_net, locale, last_login

Metrics: Cafe Met, 7 day, Cafe, Comment, Photo, Test, Video, Connect, FB P

<< first <

Test Result: 59127

new_home_whats_new

Test ID: 29211

From: 10/15/2010 18:56:54

To: 11/01/2010 13:42:38

Bugs

November 2010

Su	Mo	Tu	We	Th	Fr	Sa
31	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	1	2	3	4
5	6	7	8	9	10	11

Versions and Summary Data:

Version	Weight	Description	Friend Requests
1	1	"Default"	--
2	1	"MeetMe static box"	+2.45% (±0.62%)
3	1	"No suggestion box"	+1.67% (±0.6%)
4	1	"No suggestion box, lower top ad"	+1.88% (±0.6%)

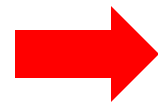
Profile

Version	Profile Views Received				Num Users	To
	Num Users	Total Views Received	Avg Views Received	Err Views Received		

SQL for Ad-hoc Analysis

- Business and Product have questions, we have answers
- Complex, deep analysis for games
- Motivate new feature development
- Declarative language >10x more productive than Java
- Greenplum supports SQL 2003 OLAP extensions

Analytics with Greenplum



Case Studies

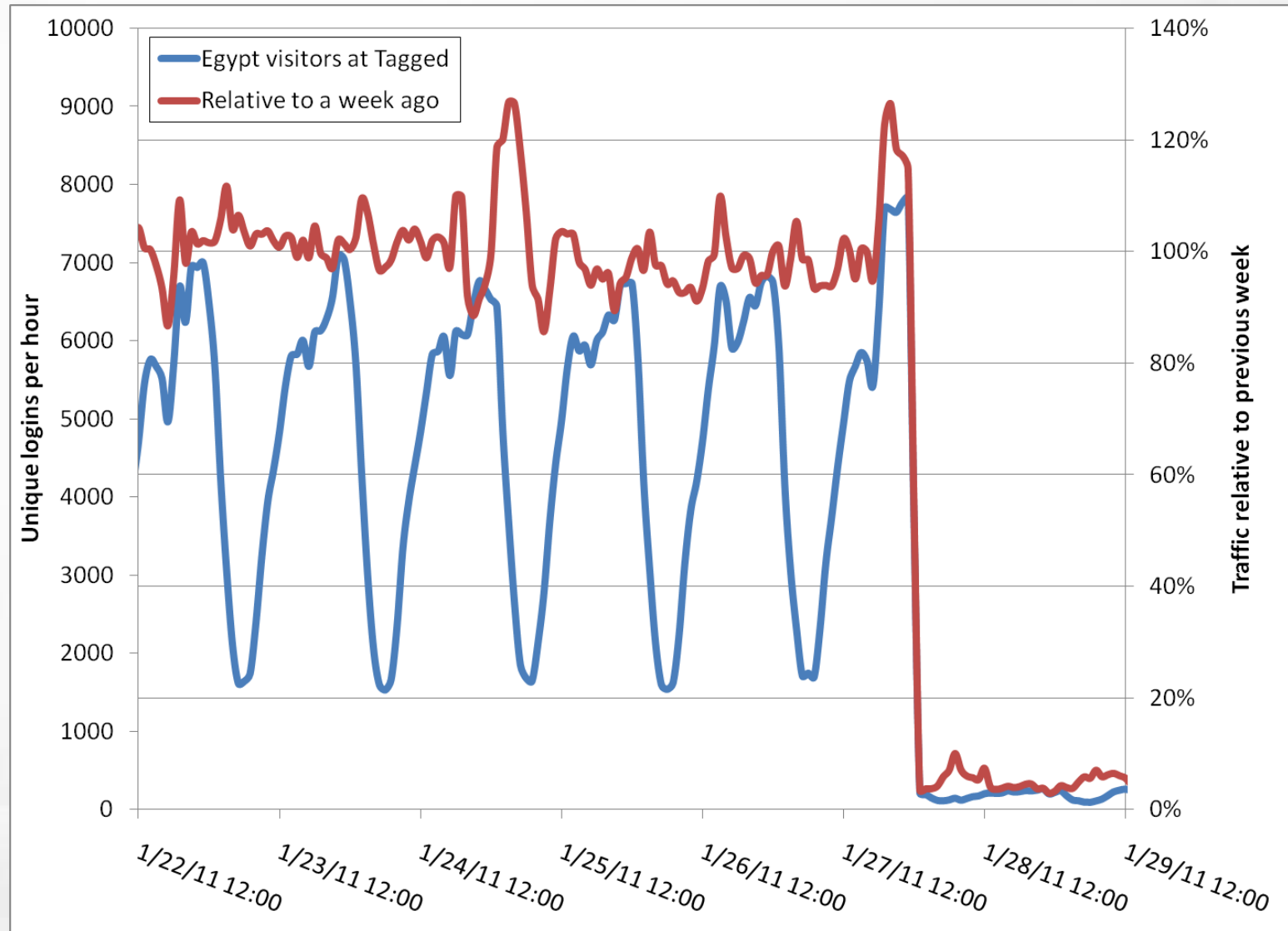
Best Practices

Case Study: Business Operations

Q: why are page views tracking 3% lower than usual today?

- Review dashboards, per-feature and per-country metrics
- Common operational problems
 - Bugs introduced by daily release
 - Slow pages?
 - E-mail problems?
 - Regional holidays
- Want to achieve automated alerting & problem isolation

Case Study: Protests in Egypt



Case Study: Protests in Egypt

<http://www.about-tagged.com/news/monitor?country=EG>

TAGGED

ABOUT

COMPANY

JOBS

NEWS

ADVERTISING

CONTACT

NEWS

IN THE NEWS

PRESS RELEASES

BLOG

MONITOR

Tagged Monitor

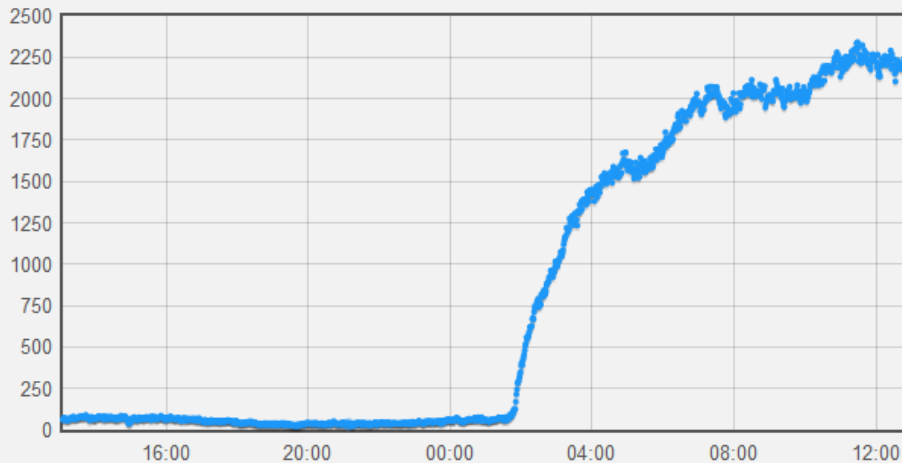
The Tagged International Traffic Monitor graph provides near real-time view of active users per minute by time of day and country.

View:

Country:

Updated: 2011-02-02 13:04:02.130

Active Users Per Minute vs Time of Day(PST)



Case Study: Mobile Usage

Q: how should Tagged direct its mobile web development?

- Parse logs
 - Identify User-Agents in HTTP headers that correspond to mobile platforms
 - Count number of users on mobile web browsers
 - Which features do mobile web users engage with most
 - Check messages
 - Browse profiles
 - Play meet me
 - Play pets
 - Now know what features to focus on in mobile web product
- Queries run in under one minute
- Quick iteration helps us develop rules for classifying headers
- Analysis done in one hour

Case Study: Tuning Mafia Game

- Find points of player confusion
- A/B test UI improvements
- Analyst must play game and understand its intricacies
- Log events with plenty of extra game state
- >3x revenue improvement

The screenshot displays the Mafia game interface. At the top, the player's stats are shown: Level 41 (64 xp left), Energy 155/155, Endurance 36/36, Health 213/213, and Cash (bank) \$26,129. The interface is divided into several sections:

- Jobs:** A sidebar on the left lists various job categories like Properties, Fight, Store, Character (11), and Rankings.
- The Big Heist:** A job listing with a reward of \$102,000, 204 XP, and 5 Skill Points. It requires 10 energy and has a 'Do It!' button.
- Finale:** A section containing 'Foil the Casino Heist' with a reward of \$1,020 and 38 XP, requiring 16 energy and various items (4x gun, 3x vest, 3x mask, 3x cash).
- Chapter 3:** Contains 'Bounce the Grifter' (reward: \$408, 18 XP, Loot) and 'Steal a Tiger from the Zoo' (reward: \$408, 22 XP).
- Chapter 2:** Contains 'Smuggle Illegal Fireworks across the Border' (reward: \$816, 32 XP).

Each job listing includes a 'Do It!' button and a progress bar indicating mastery level (e.g., 0%, 54%, 100%).

Case Study: Phishing from a Botnet

Q: phishing spam is going on now, how do we stop it?

- Forensic inquiry on phished accounts
 - Is there evidence of a brute-force password attack?
 - Have compromised users been active on Tagged recently?
 - Were credentials stolen from another web site?
 - Is spam highly distributed or coming from a few IP blocks?
 - Are there headers that identify the botnet?
 - What actions are common immediately after phishing takes place?
 - Were game players actively targeted?
- Bayesian classifier systems for spam identification
- Working on automated pattern extraction

Case Study: Matchmaking on 100 million accounts

- Meet Me is the dating feature on Tagged
- Connects people based on social patterns
- 10x increase in usage through A/B testing

Meet Me Play Yes List Matches

Are You Interested? No Yes


Gender
 Age to
 Country
 Location

 Reset Update

You might see people outside your filters if we think they may be good matches for you
 [Change Preferences](#)


How it works

- Click Yes on people you want to meet
- We will add them to your Yes List and show them your photo in Meet Me (and they'll know you clicked Yes)
- If they click Yes on you, it's a match
- If you're really interested in someone, send some gifts after clicking Yes!



Diane I
45, Female, San Francisco, CA
Interested in Friends, Networking
"Never let fear stop you from experiencing life to the fullest"
[View Profile](#)
[Previous User](#)

Give Thanks. **Free Shipping.**



START SAVING >

Crate&Barrel

Free Shipping on Select Orders of \$100 or more.
Offer ends Dec. 22 12PM (CT).

Analytics with Greenplum

Case Studies

 **Best Practices**

Best Practices in Social Media Analytics

- Application writes logs specifically for analytics
- Get plenty of space and log a lot
- Data can be queried in real time
- Use SQL for quick ad-hoc studies
- Automate reports with lots of graphs
- Compute statistical uncertainty in A/B tests

Data Computing Appliance Delivers

- Greenplum gobbles up all the data
- Analysts love the speed and SQL advantage
- DBAs find life is easy
- Business gets insightful and complete answers to questions, quickly
- Users enjoy the longest session times of any social network

