facebook

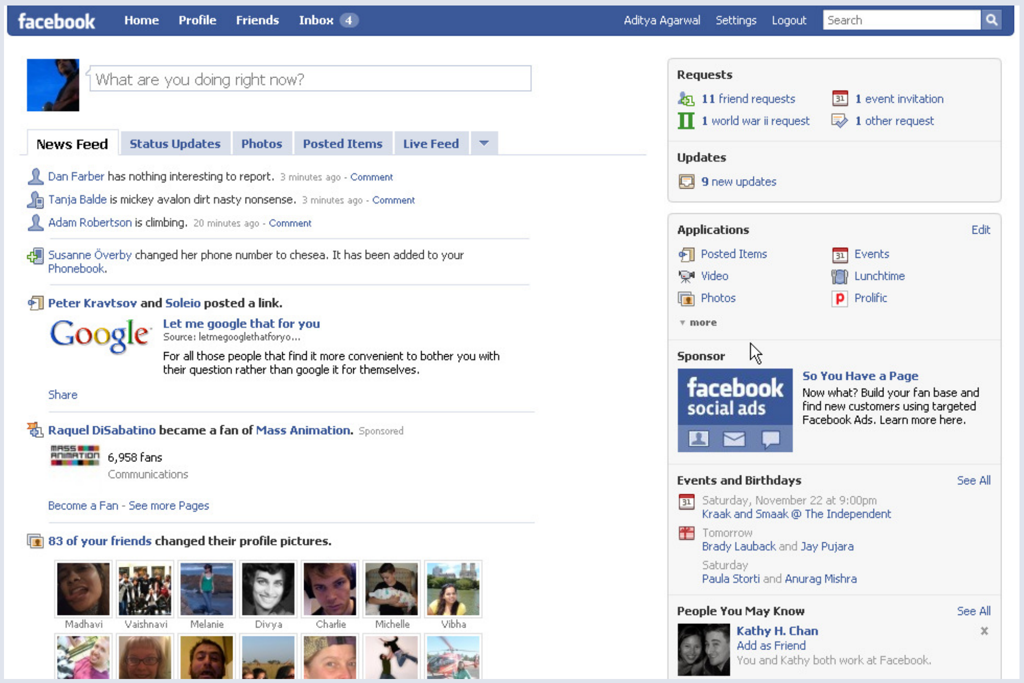
facebook

Facebook Architecture

Aditya Agarwal
Director of Engineering
11/22/2008

Agenda

facebook At a Glance



facebook Home Profile Friends Inbox 4 Aditya Agarwal Settings Logout Search

What are you doing right now?

News Feed | Status Updates | Photos | Posted Items | Live Feed

Dan Farber has nothing interesting to report. 3 minutes ago · Comment

Tanja Balde is mickey avalon dirt nasty nonsense. 3 minutes ago · Comment

Adam Robertson is climbing. 20 minutes ago · Comment

Susanne Överby changed her phone number to chesea. It has been added to your Phonebook.

Peter Kravtsov and Soleio posted a link.

Let me google that for you

Source: letmegooglethatforyo...

For all those people that find it more convenient to bother you with their question rather than google it for themselves.

Share

Raquel DiSabatino became a fan of Mass Animation. Sponsored

6,958 fans

Communications

Become a Fan · See more Pages

83 of your friends changed their profile pictures.

Madhavi, Vaishnavi, Melanie, Divya, Charlie, Michelle, Vibha

Requests

11 friend requests · 1 event invitation

1 world war ii request · 1 other request

Updates

9 new updates

Applications Edit

Posted Items · Events

Video · Lunchtime

Photos · Prolific

more

Sponsor

So You Have a Page

Now what? Build your fan base and find new customers using targeted Facebook Ads. Learn more here.

Events and Birthdays See All

Saturday, November 22 at 9:00pm

Kraak and Smaak @ The Independent

Tomorrow

Brady Laubeck and Jay Pujara

Saturday

Paula Storti and Anurag Mishra

People You May Know See All

Kathy H. Chan

Add as Friend

You and Kathy both work at Facebook.

The Social Graph

120M+ active users

50B+ PVs per month

10B+ Photos

1B+ connections

50K+ Platform Apps

400K+ App Developers

General Design Principles

- Use open source where possible
  - Explore making optimizations where needed
- Unix Philosophy
  - Keep individual components simple yet performant
  - Combine as necessary
  - Concentrate on clean interface points
- Build everything for scale
- Try to minimize failure points
- Simplicity, Simplicity, Simplicity!

Architecture Overview

LAMP + **Services**

PHP

Memcache

MySQL

php!

AdServer
Search
Network Selector
News Feed
Blogfeeds
CSSParser
Mobile
ShareScraper

!php

Thrift
Scribe
ODS
Tools

PHP

- Good web programming language
  - Extensive library support for web development
  - Active developer community

- Good for rapid iteration
  - Dynamically typed, interpreted scripting language

PHP: What we Learnt

- Tough to scale for large code bases
  - Weak typing
  - Limited opportunities for static analysis, code optimizations

- Not necessarily optimized for large website use case
  - E.g. No dynamic reloading of files on web server

- Linearly increasing cost per included file

- Extension framework is difficult to use

PHP: Customizations

- Op-code optimization
- APC improvements
  - Lazy loading
  - Cache priming
  - More efficient locking semantics for variable cache data
- Custom extensions
  - Memcache client extension
  - Serialization format
  - Logging, Stats collection, Monitoring
  - Asynchronous event-handling mechanism

MySQL

- Fast, reliable

- Used primarily as ‹key,value› store
  - Data randomly distributed amongst large set of logical instances
  - Most data access based on global id

- Large number of logical instances spread out across physical nodes
  - Load balancing at physical node level

- No read replication

MySQL: What We Learnt (ing)

- Logical migration of data is *very* difficult
- Create a large number of logical dbs, load balance them over varying number of physical nodes
- No joins in production
  - Logically difficult (because data is distributed randomly)
- Easier to scale CPU on web tier

MySQL: What we Learnt (ing)

- Most data access is for recent data
    - Optimize table layout for recency
    - Archive older data

- Don’t ever store non-static data in a central db
    - CDB makes it easier to perform certain aggregated queries
    - Will not scale

- Use services or memcache for global queries
    - E.g.: What are the most popular groups in my network

MySQL: Customizations

- No extensive native MySQL modifications

- Custom partitioning scheme
  - Global id assigned to all data

- Custom archiving scheme
  - Based on frequency and recency of data on a per-user basis

- Extended Query Engine for cross-data center replication, cache consistency

MySQL: Customizations

- Graph based data-access libraries
  - Loosely typed objects (nodes) with limited datatypes (int, varchar, text)
  - Replicated connections (edges)
  - Analogous to distributed foreign keys

- Some data collocated
  - Example: User profile data and all of user's connections

- Most data distributed randomly

memcached

Memcache

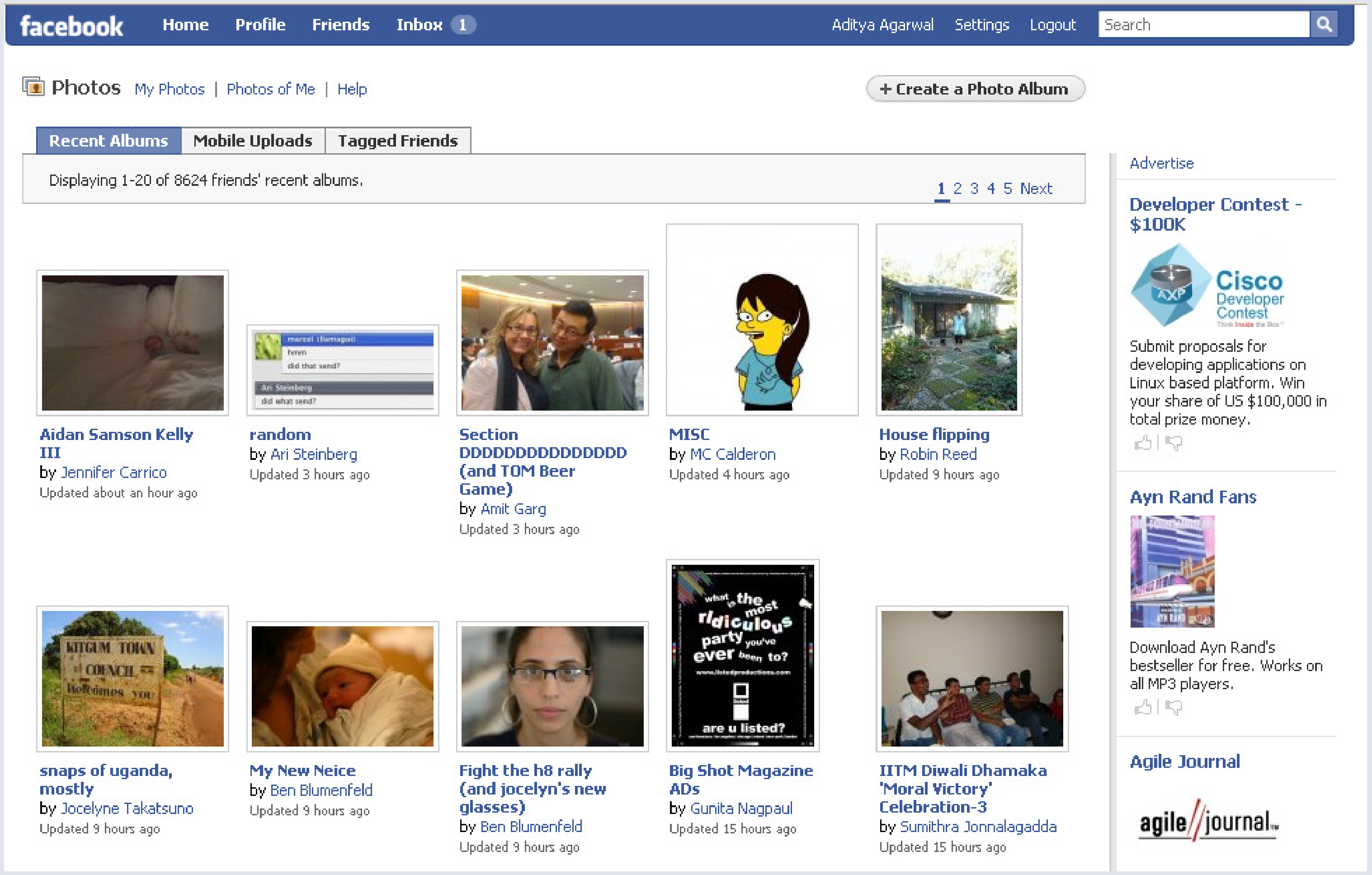
- High-Performance, distributed in-memory hash table
- Used to alleviate database load
- Primary form of caching
- Over 25TB of in-memory cache
- Average latency < 200 micro-seconds
- Cache serialized PHP data structures
- Lots and lots of multi-gets to retrieve data spanning across graph edges

memcached

Memache: Customizations

- Memache over UDP
  - Reduce memory overhead of thousands of TCP connection buffers
  - Application-level flow control (optimization for multi-gets)
- On demand aggregation of per-thread stats
  - Reduces global lock contention
- Multiple Kernel changes to optimize for Memcache usage
  - Distributing network interrupt handling over multiple cores
  - Opportunistic polling of network interface

Let’s put this into action



facebook Home Profile Friends Inbox 1 Aditya Agarwal Settings Logout Search

Photos My Photos | Photos of Me | Help + Create a Photo Album

Recent Albums | Mobile Uploads | Tagged Friends

Displaying 1-20 of 8624 friends' recent albums. 1 2 3 4 5 Next

Aidan Samson Kelly III
by Jennifer Carrico
Updated about an hour ago

random
by Ari Steinberg
Updated 3 hours ago

Section DDDDDDDDDDDDDDD (and TOM Beer Game)
by Amit Garg
Updated 3 hours ago

MISC
by MC Calderon
Updated 4 hours ago

House flipping
by Robin Reed
Updated 9 hours ago

snaps of uganda, mostly
by Jocelyne Takatsuno
Updated 9 hours ago

My New Neice
by Ben Blumenfeld
Updated 9 hours ago

Fight the h8 rally (and jocelyn's new glasses)
by Ben Blumenfeld
Updated 9 hours ago

Big Shot Magazine ADs
by Gunita Nagpaul
Updated 15 hours ago

IITM Diwali Dhamaka 'Moral Victory' Celebration-3
by Sumithra Jonnalagadda
Updated 15 hours ago

Advertise

Developer Contest - $100K

Submit proposals for developing applications on Linux based platform. Win your share of US $100,000 in total prize money.

Ayn Rand Fans

Download Ayn Rand's bestseller for free. Works on all MP3 players.

Agile Journal

Under the Covers

- Get my profile data
  - Fetch from cache, potentially go to my DB (based on user-id)
- Get friend connections
  - Cache, if not DB (based on user-id)
- In parallel, fetch last 10 photo album ids for each of my friends
  - Multi-get; individual cache misses fetches data from db (based on photo-album id)
- Fetch data for most recent photo albums in parallel
- Execute page-specific rendering logic in PHP
- Return data, make user happy

LAMP is not Perfect

LAMP is not Perfect

- PHP+MySQL+Memcache works for a large class of problems but not for everything
  - PHP is stateless
  - PHP not the fastest executing language
  - All data is remote
- Reasons why services are written
  - Store code closer to data
  - Compiled environment is more efficient
  - Certain functionality only present in other languages

Services Philosophy

- Create a service iff required
  - Real overhead for deployment, maintenance, separate code-base
  - Another failure point
- Create a common framework and toolset that will allow for easier creation of services
  - Thrift
  - Scribe
  - ODS, Alerting service, Monitoring service
- Use the right language, library and tool for the task

Thrift

C++

JAVA™

php

python

High-Level Goal: Enable transparent interaction between these.

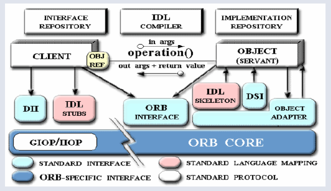
...and some others too.

Thrift

- Lightweight software framework for cross-language development
- Provide IDL, statically generate code
- Supported bindings: C++, PHP, Python, Java, Ruby, Erlang, Perl, Haskell etc.
- Transports: Simple Interface to I/O
  - Tsocket, TFileTransport, TMemoryBuffer
- Protocols: Serialization Format
  - TBinaryProtocol, TJSONProtocol
- Servers
  - Non-Blocking, Async, Single Threaded, Multi-threaded

Hasn’t this been done before? (yes.)

- SOAP
    - XML, XML, and more XML
- CORBA
    - Bloated? Remote bindings?
- COM
    - Face-Win32ClientSoftware.dll-Book
- Pillar
    - Slick! But no versioning/abstraction.
- Protocol Buffers



Thrift: Why?

- It’s quick. Really quick.

- Less time wasted by individual developers
  - No duplicated networking and protocol code
  - Less time dealing with boilerplate stuff
  - Write your client and server in about 5 minutes

- Division of labor
  - Work on high-performance servers separate from applications

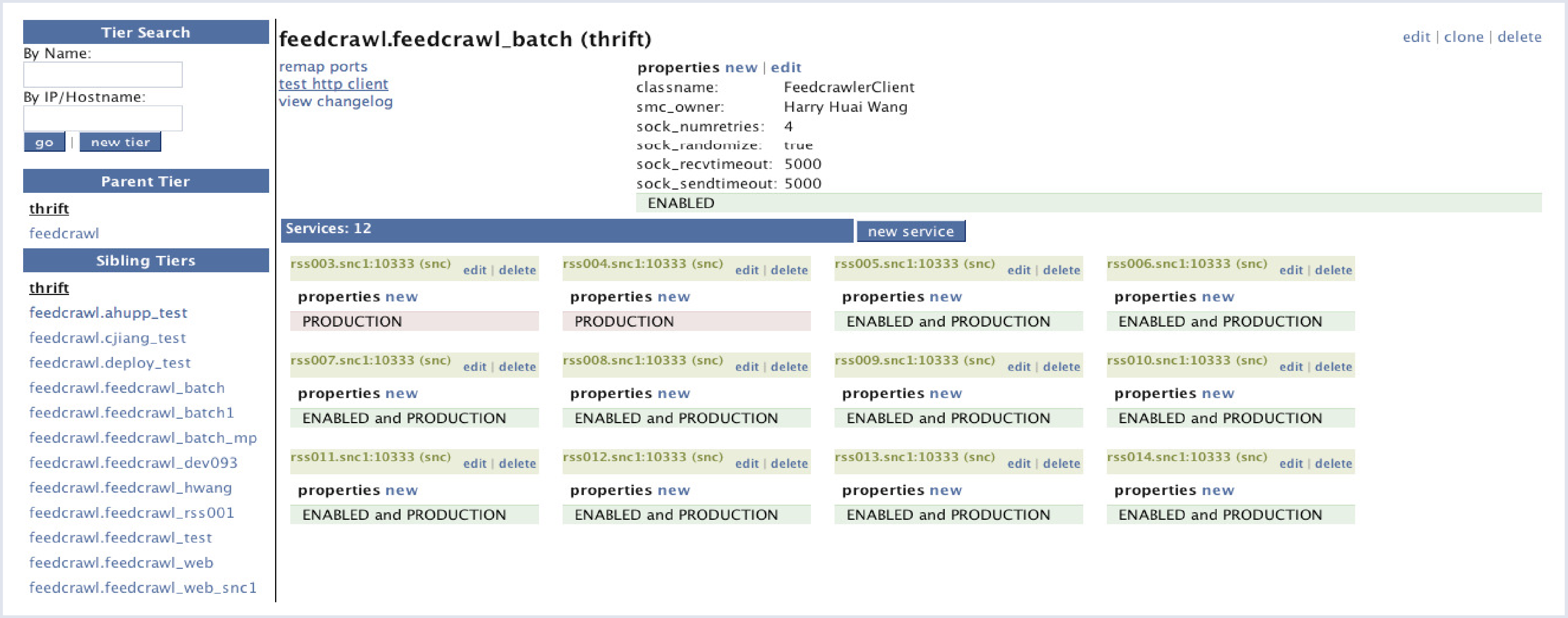
- Common toolkit
  - Fosters code reuse and shared tools

Scribe

- Scalable distributed logging framework
- Useful for logging a wide array of data
  - Search Redologs
  - Powers news feed publishing
  - A/B testing data
- Weak Reliability
  - More reliable than traditional logging but not suitable for database transactions.
- Simple data model
- Built on top of Thrift

Other Tools

- SMC (Service Management Console)
  - Centralized configuration
  - Used to determine logical service -> physical node mapping



Tier Search
By Name:
By IP/Hostname:
go | new tier

Parent Tier
thrift
feedcrawl

Sibling Tiers
thrift
feedcrawl.ahupp_test
feedcrawl.cjiang_test
feedcrawl.deploy_test
feedcrawl.feedcrawl_batch
feedcrawl.feedcrawl_batch1
feedcrawl.feedcrawl_batch_mp
feedcrawl.feedcrawl_dev093
feedcrawl.feedcrawl_hwang
feedcrawl.feedcrawl_rss001
feedcrawl.feedcrawl_test
feedcrawl.feedcrawl_web
feedcrawl.feedcrawl_web_snc1

feedcrawl.feedcrawl_batch (thrift)

edit | clone | delete

remap ports
test http client
view changelog

properties new | edit

classname:	FeedcrawlerClient
smc_owner:	Harry Huai Wang
sock_numretries:	4
sock_randomize:	true
sock_recvtimeout:	5000
sock_sendtimeout:	5000

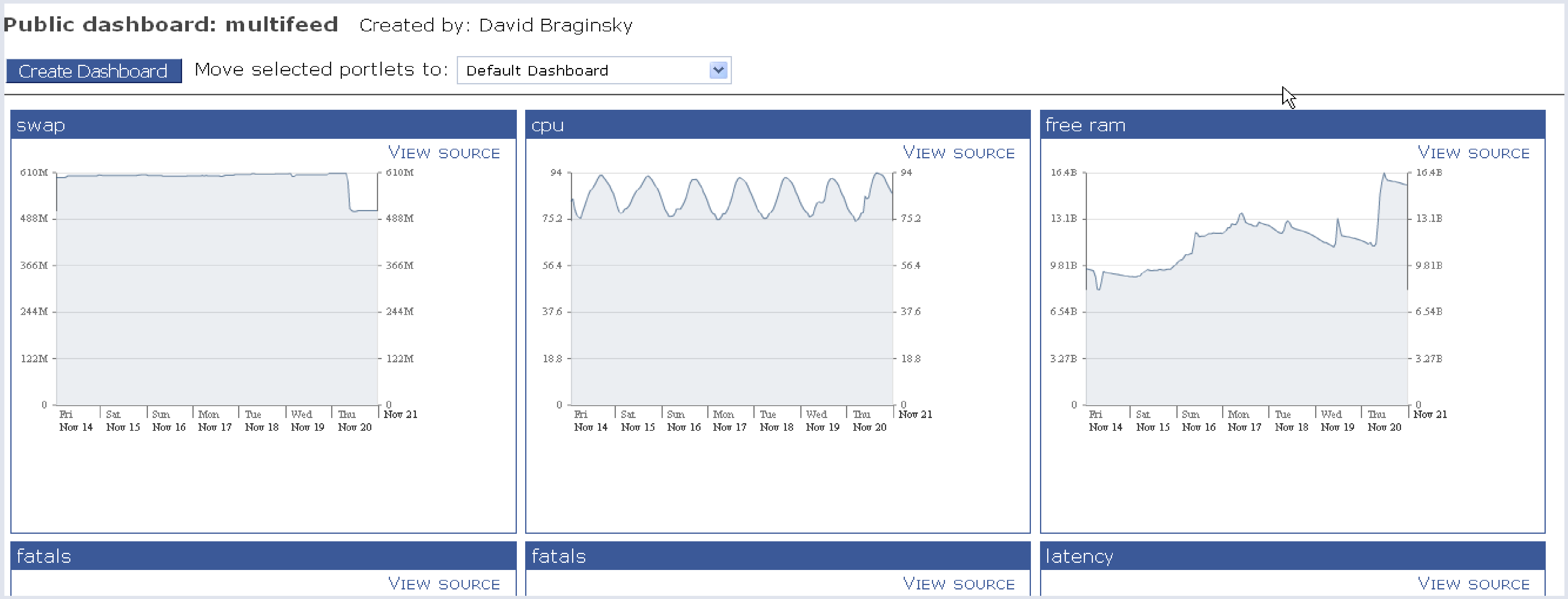
ENABLED

Services: 12 — new service

rss003.snc1:10333 (snc) edit | delete	rss004.snc1:10333 (snc) edit | delete	rss005.snc1:10333 (snc) edit | delete	rss006.snc1:10333 (snc) edit | delete
properties new	properties new	properties new	properties new
PRODUCTION	PRODUCTION	ENABLED and PRODUCTION	ENABLED and PRODUCTION
rss007.snc1:10333 (snc) edit | delete	rss008.snc1:10333 (snc) edit | delete	rss009.snc1:10333 (snc) edit | delete	rss010.snc1:10333 (snc) edit | delete
properties new	properties new	properties new	properties new
ENABLED and PRODUCTION	ENABLED and PRODUCTION	ENABLED and PRODUCTION	ENABLED and PRODUCTION
rss011.snc1:10333 (snc) edit | delete	rss012.snc1:10333 (snc) edit | delete	rss013.snc1:10333 (snc) edit | delete	rss014.snc1:10333 (snc) edit | delete
properties new	properties new	properties new	properties new
ENABLED and PRODUCTION	ENABLED and PRODUCTION	ENABLED and PRODUCTION	ENABLED and PRODUCTION

Other Tools

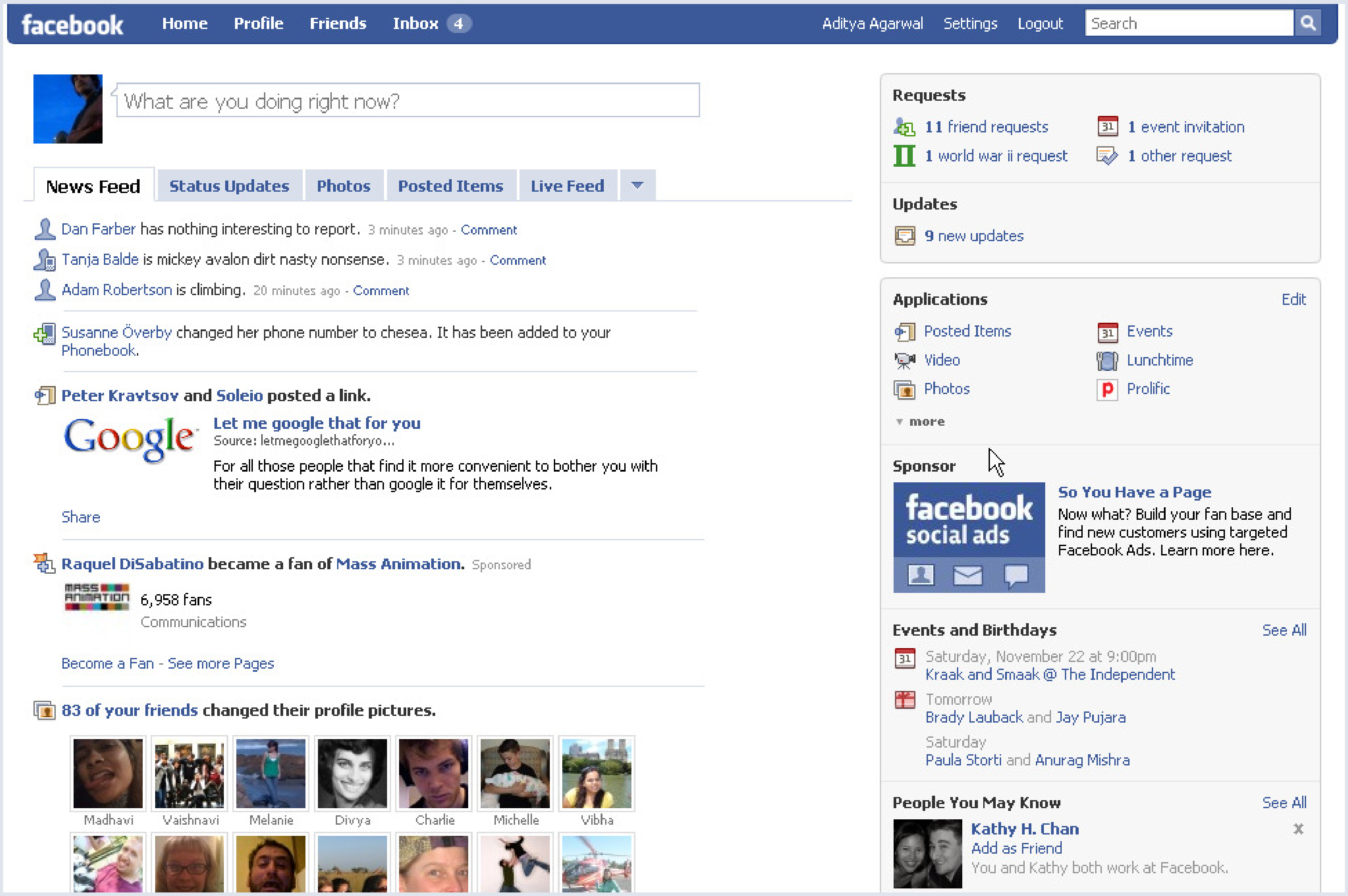
- ODS
  - Used to log and view historical trends for any stats published by service
  - Useful for service monitoring, alerting


Public dashboard: multifeed
Created by: David Braginsky
Create Dashboard
Move selected portlets to:
Default Dashboard
swap
cpu
free ram
fatals
fatals
latency
View source

Open Source

- Thrift
  - http://developers.facebook.com/thrift/
- Scribe
  - http://developers.facebook.com/scribe/
- PHPEmbed
  - http://developers.facebook.com/phpembed/
- More good stuff
  - http://developers.facebook.com/opensource.php

NewsFeed – The Goodz



facebook Home Profile Friends Inbox 4 Aditya Agarwal Settings Logout Search

What are you doing right now?

News Feed | Status Updates | Photos | Posted Items | Live Feed

Dan Farber has nothing interesting to report. 3 minutes ago - Comment

Tanja Balde is mickey avalon dirt nasty nonsense. 3 minutes ago - Comment

Adam Robertson is climbing. 20 minutes ago - Comment

Susanne Överby changed her phone number to chesea. It has been added to your Phonebook.

Peter Kravtsov and Soleio posted a link.

Let me google that for you
Source: letmegooglethatforyo...

For all those people that find it more convenient to bother you with their question rather than google it for themselves.

Share

Raquel DiSabatino became a fan of Mass Animation. Sponsored

6,958 fans
Communications

Become a Fan - See more Pages

83 of your friends changed their profile pictures.

Madhavi, Vaishnavi, Melanie, Divya, Charlie, Michelle, Vibha

Requests
11 friend requests
1 event invitation
1 world war ii request
1 other request

Updates
9 new updates

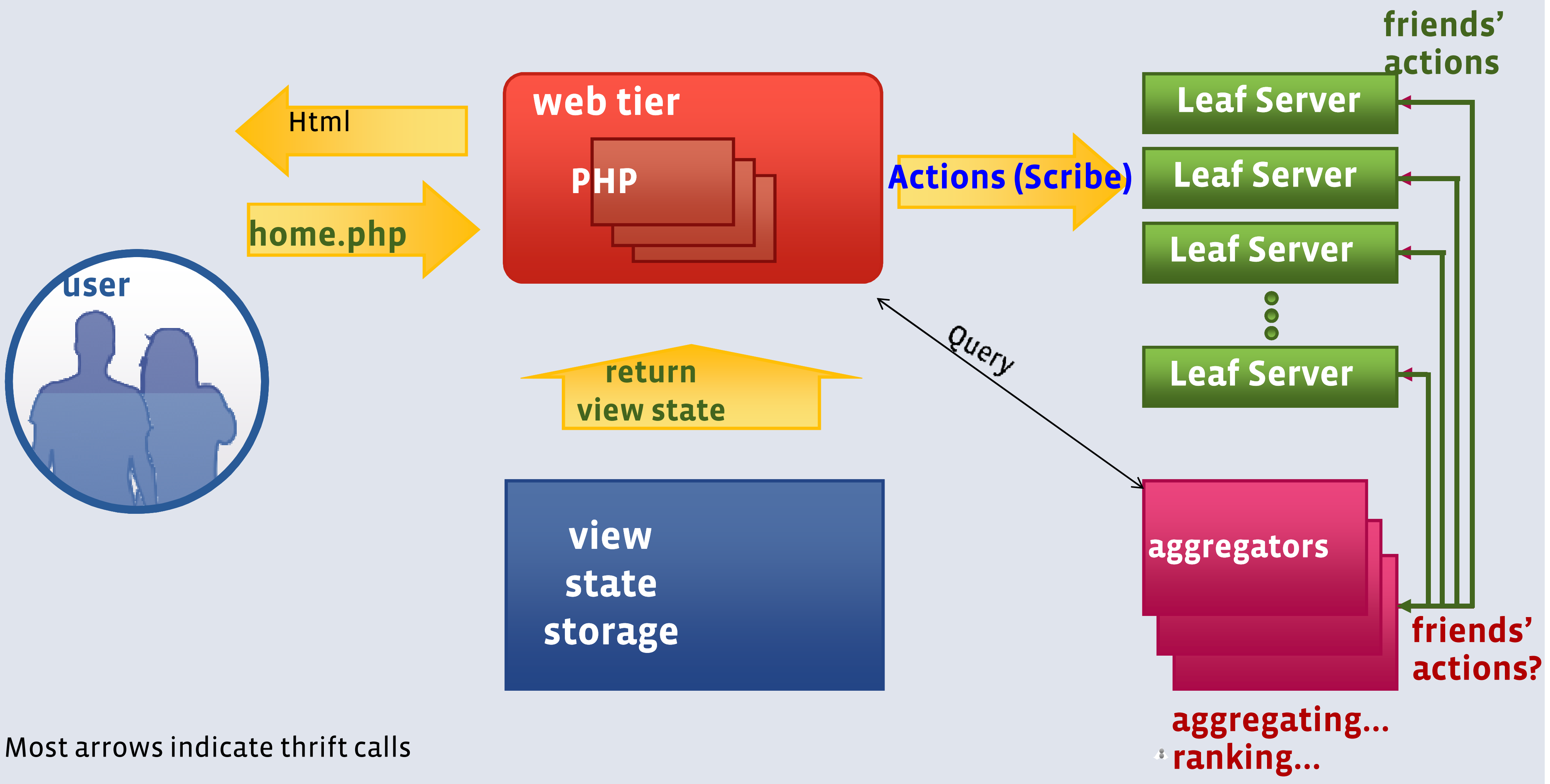
Applications Edit
Posted Items
Events
Video
Lunchtime
Photos
Prolific
more

Sponsor
facebook social ads
So You Have a Page
Now what? Build your fan base and find new customers using targeted Facebook Ads. Learn more here.

Events and Birthdays See All
Saturday, November 22 at 9:00pm
Kraak and Smaak @ The Independent
Tomorrow
Brady Lauback and Jay Pujara
Saturday
Paula Storti and Anurag Mishra

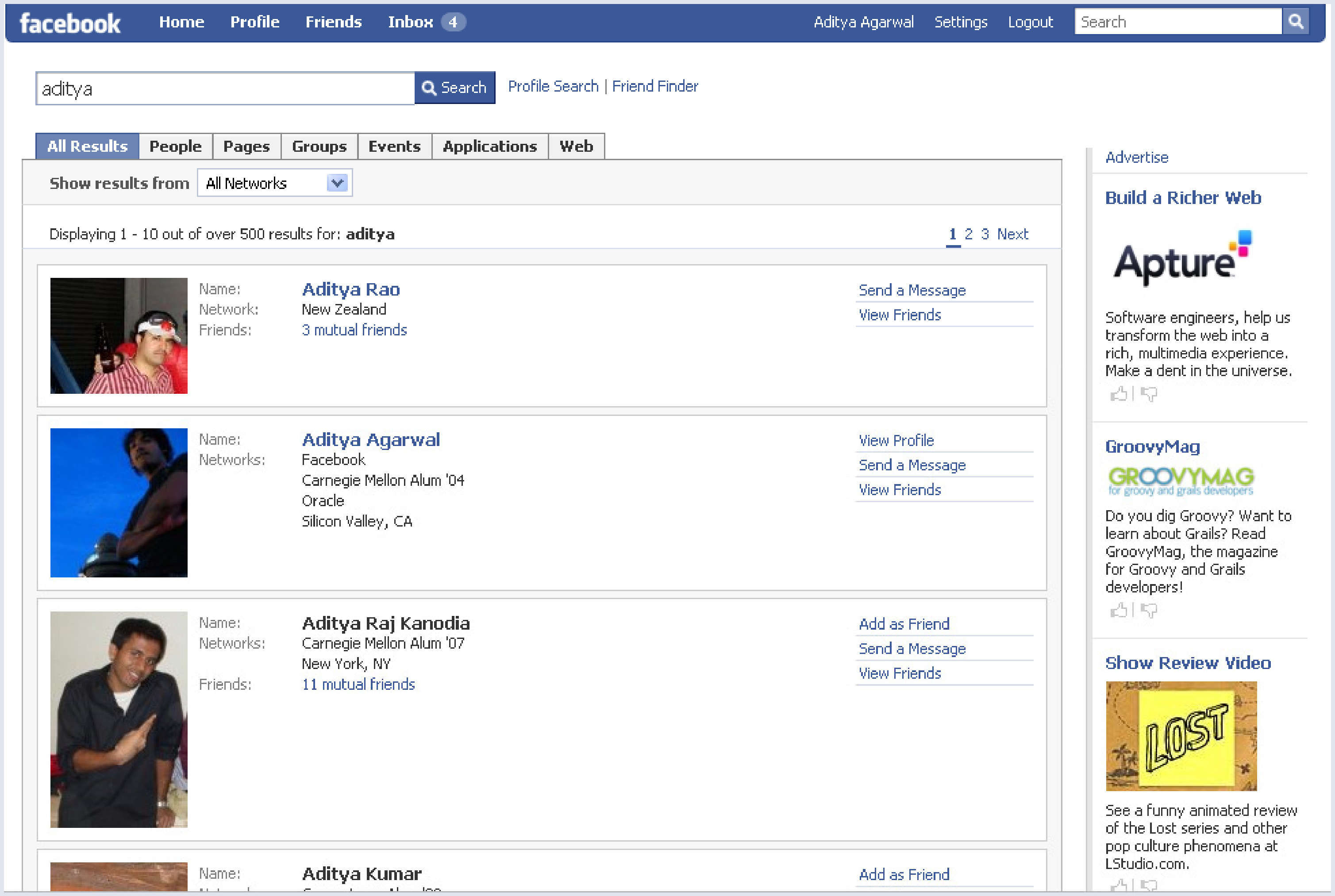
People You May Know See All
Kathy H. Chan
Add as Friend
You and Kathy both work at Facebook.

NewsFeed – The Work



- Most arrows indicate thrift calls

Search – The Goodz



Displaying 1 - 10 out of over 500 results for: **aditya**

1 2 3 Next

Name: Aditya Rao
Network: New Zealand
Friends: 3 mutual friends
Send a Message
View Friends

Name: Aditya Agarwal
Networks: Facebook
Carnegie Mellon Alum '04
Oracle
Silicon Valley, CA
View Profile
Send a Message
View Friends

Name: Aditya Raj Kanodia
Networks: Carnegie Mellon Alum '07
New York, NY
Friends: 11 mutual friends
Add as Friend
Send a Message
View Friends

Name: Aditya Kumar
Add as Friend

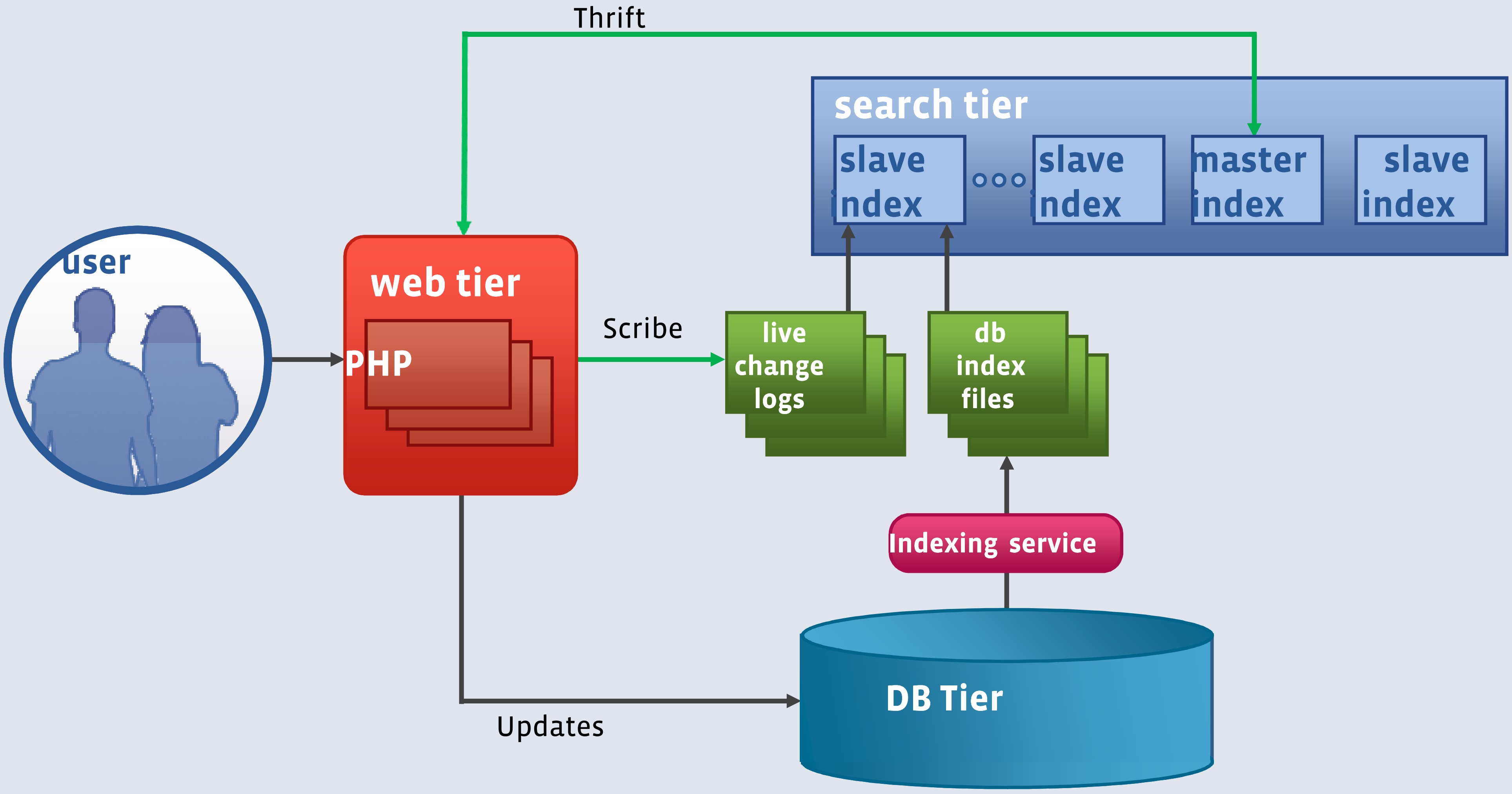
Advertise

Build a Richer Web
Apture
Software engineers, help us transform the web into a rich, multimedia experience. Make a dent in the universe.

GroovyMag
GROOVYMAG for groovy and grails developers
Do you dig Groovy? Want to learn about Grails? Read GroovyMag, the magazine for Groovy and Grails developers!

Show Review Video
LOST
See a funny animated review of the Lost series and other pop culture phenomena at LStudio.com.

Search – The Work



Questions?

More info at www.facebook.com/eblog

Aditya Agarwal
aditya@facebook.com