

Wikimedia Operations Overview



A high level overview of Wikimedia Operations

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Top Websites *(estimates pulled xkcd-style)*

Company	Users	Revenue (USD)	Employees	Server Count
Google	920 million	23 billion	24,400	1,000,000+
Facebook	750 million	300 million	2,000+	60,000+
Microsoft	740 million	69 billion	90,412	220,000+
Yahoo	600 million	6 billion	13,900	50,000+
Wikipedia	411 million	20 million	73	370 (+400)

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Operations Team

- Seven full-time and one part-time personnel now work in Operations.
- All ops personnel are on-call for any major outages.
- Located across multiple time zones.



comic from xkcd.com

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Operations Tasks & Communication

- Maintaining site uptime.
- Configuration & deployment of software services.
- Hardware allocation & support of servers/network.
- Most communication is via IRC.
 - #wikimedia-tech & #wikimedia-operations

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Datacenters

- We currently have five deployments in total.
 - Tampa, FL, USA: pmtpa & sdtpa
 - Amsterdam/Haarlem, Netherlands: knams & esams
 - Ashburn, VA, USA: eqiad

Presented by: Rob Halsell

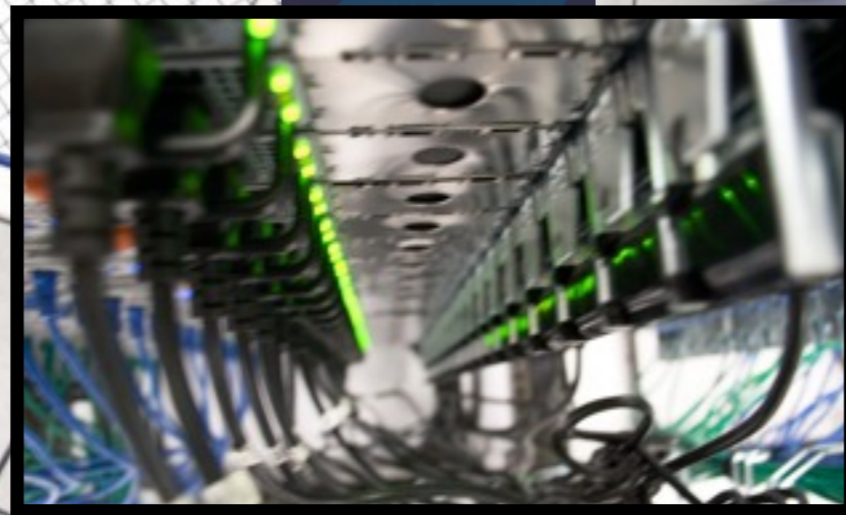
Wikimania 2011 - Haifa, Israel
August 4-7, 2011

pmtpa & sdtpa

- Located within the same building in Tampa, FL, USA on differing floors.
- Shares network, routing, and transit.
- Differing floors due to deployment schedule and mitigating factors.
- No peering.
- Approximately 300 servers.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011



Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

knams & esams

- knams facility is used for peering, also initially housed our servers.
- esams houses our caching servers, as well as some other services.
- private link between sites, ~15km apart.
- All transit for these servers are via peering agreements
- Approximately 110 servers.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011



eqiad

- Started deployment in February 2011.
- Redundancy & Reliability improvements.
 - Dual power(supplies) in all racks/servers.
 - Redundant network topography.
- Peering opportunities & cheap(er) transit.
- Already routing the majority of our traffic.
- Approximately 350 servers.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011



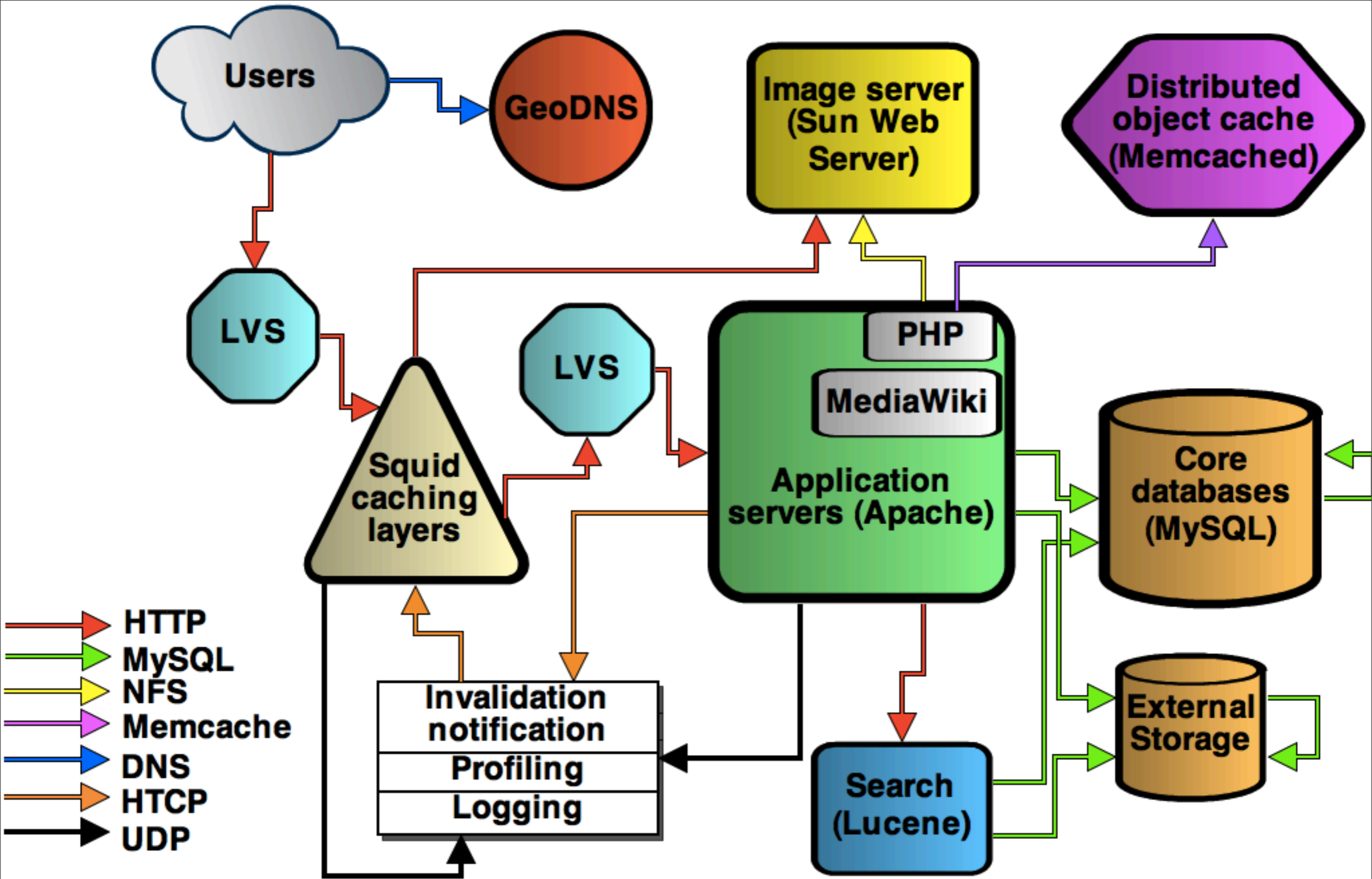
Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011



Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011



Presented by: Rob Halsell

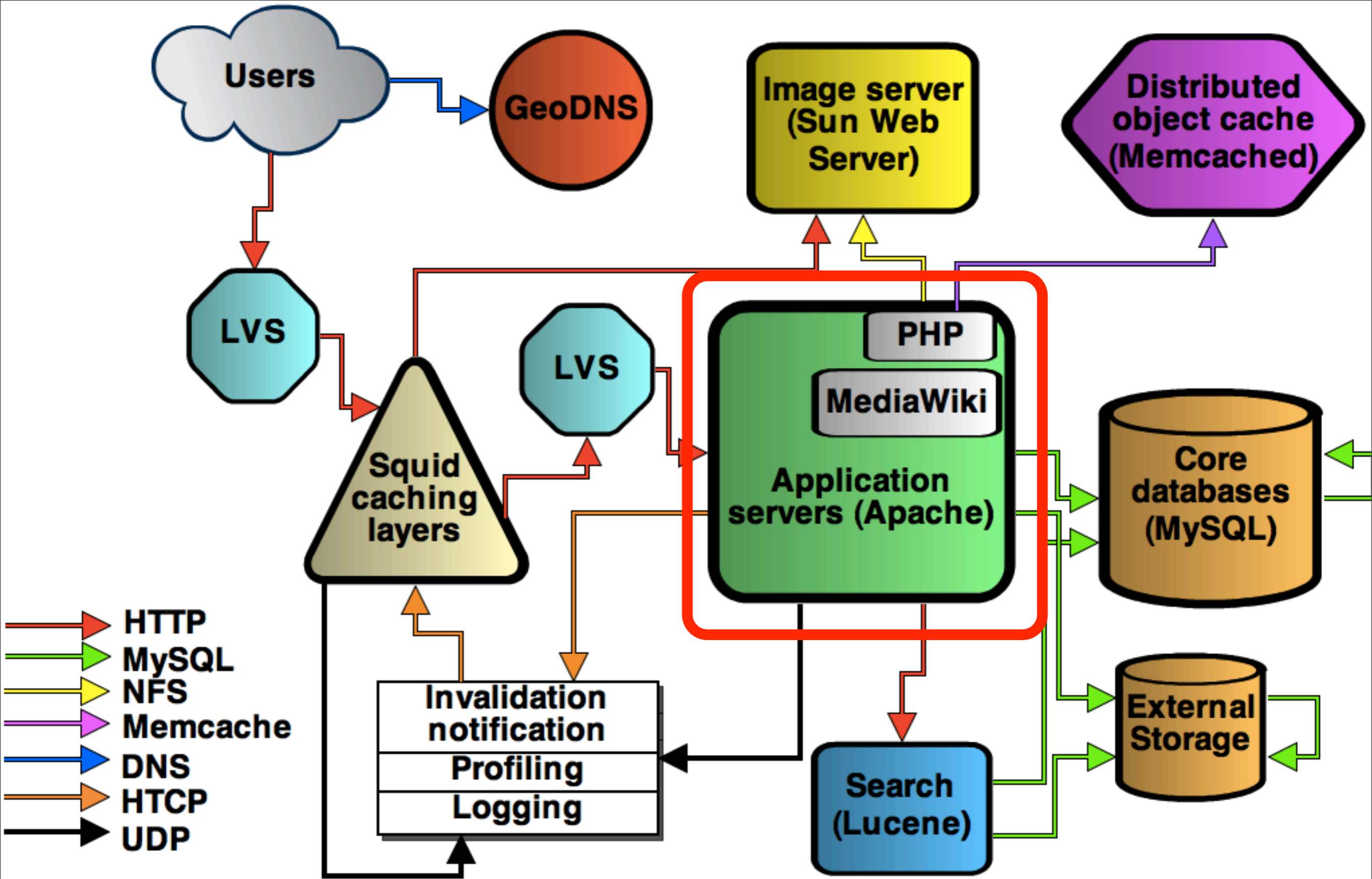
Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Wikimedia Cluster

- Caching is done for anonymous users via squid/varnish.
- Apache, coupled with Mediawiki, does our rendering.
- MySQL is used for our database storage, with memcached as an intermediary.
- External storage of old data is kept in dedicated mysql instances we refer to as external storage.
- All databases and apache servers are in Tampa or Ashburn.
- Our esams deployment is a caching center, no databases or apaches.
 - Toolserver and some other services also run out of here.
- eqiad is in initial deployment, no services hosted from there except network routing.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011



Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Apaches & Mediawiki

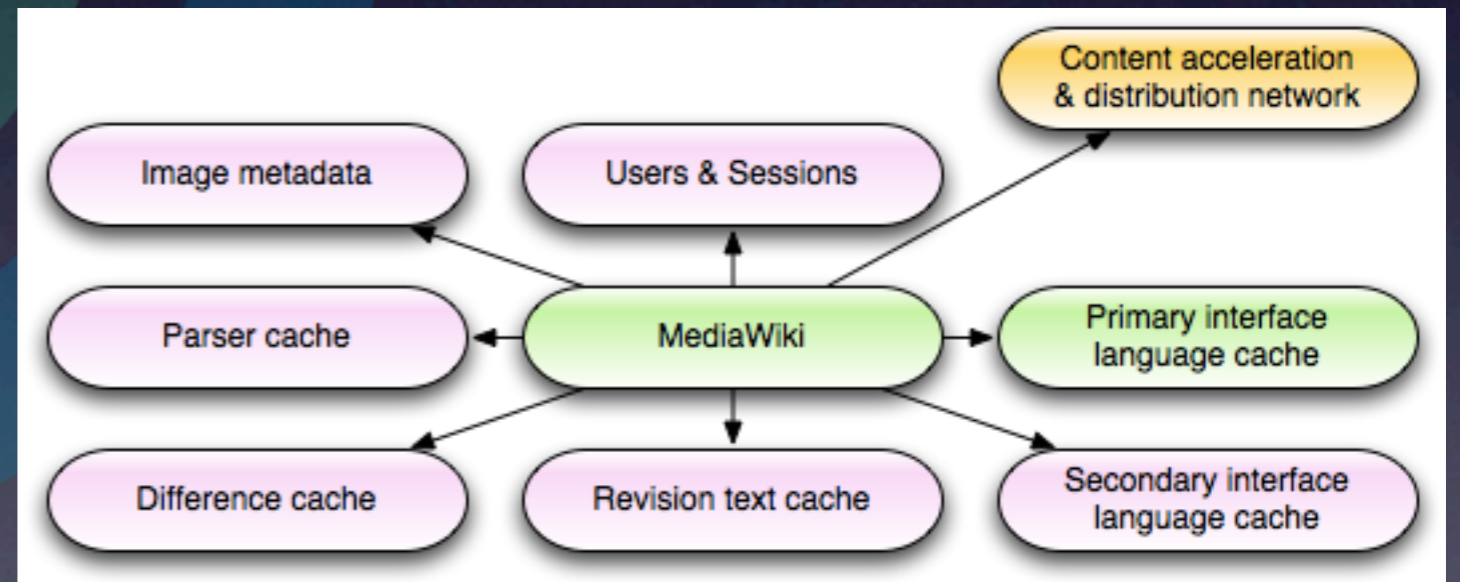
- Apache does primary web serving/rendering.
- Primary and API clusters.
- Logged in users hit these directly.
 - secure (in its old form) uses this directly.
- Mediawiki does the heavy lifting.
 - MW is developed primarily for this.
 - Open-source PHP software
 - Optimized by:
 - Not doing stupid things.
 - caching expensive operations
 - profiling and close focus in code.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

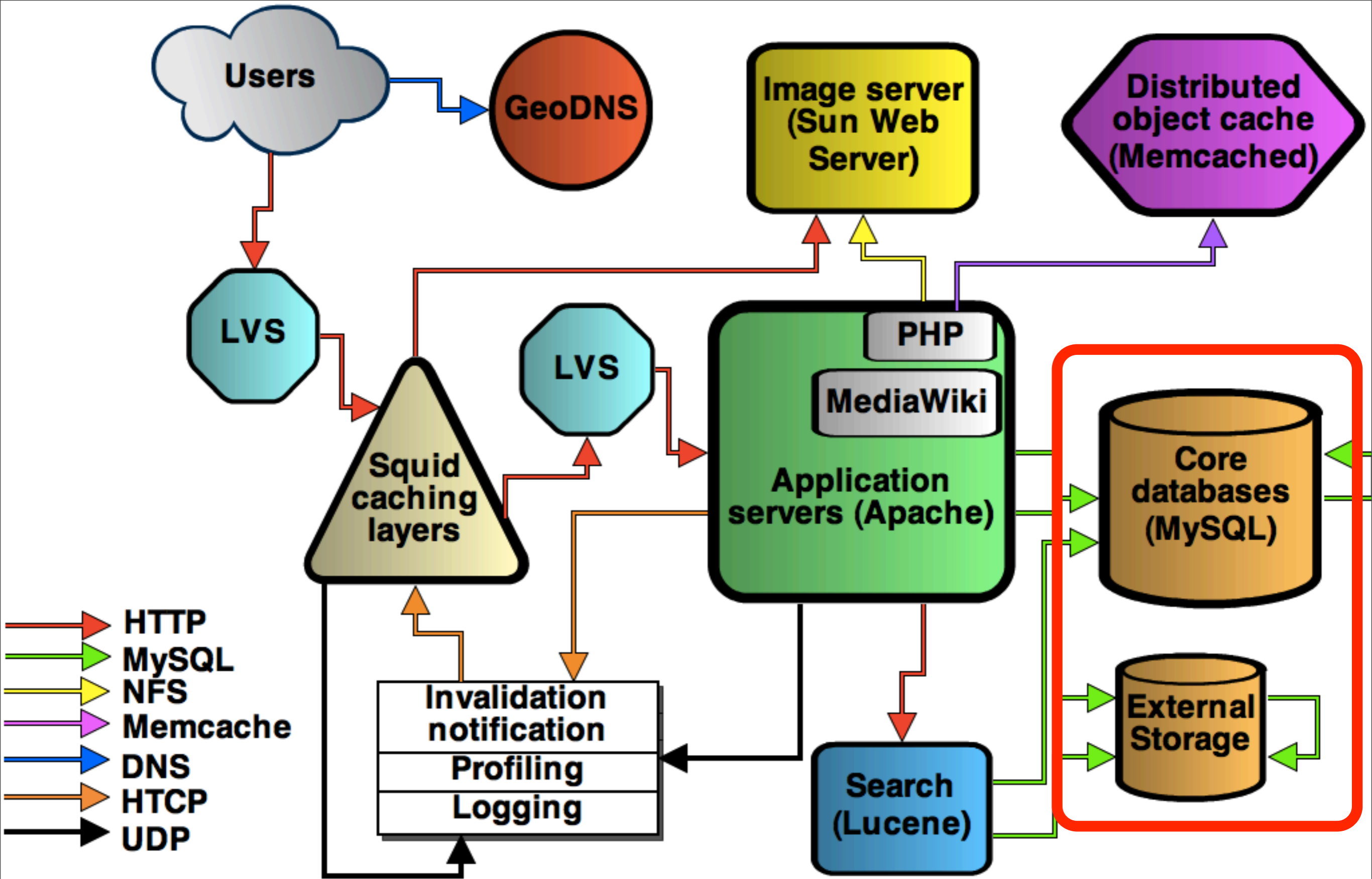
Mediawiki caching

- Practically everything within Mediawiki can be cached.
- Most of this resides in Memcached, distributed object cache.



Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

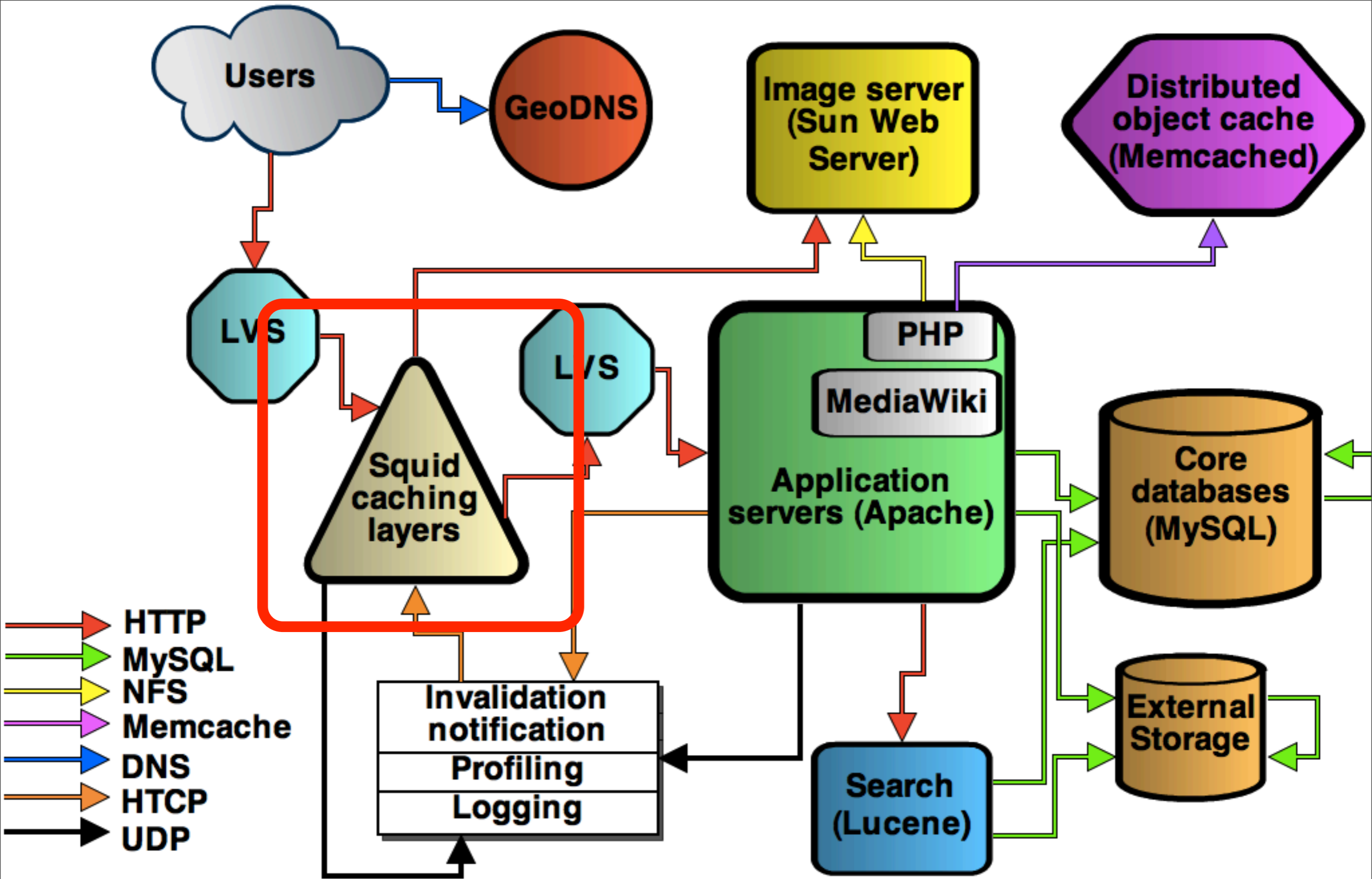


Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

MySQL Databases

- Databases are split into 7 distinct groups.
- Each group has one dedicated master and two dedicated replication hosts.
- Larger projects have less databases on the cluster shard; smaller wikis have more databases on each cluster.
- Each project/language is its own database.
- Old revision data is kept in external storage.
 - sdtpa/pmtpa share these roles on some apache servers.
 - eqiad has dedicated external storage servers.



Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Caching Layer

- Caching uses reverse HTTP proxy
- All non-logged in users hit squid.
- Squid clusters are located in every datacenter cluster
- sdtpa/pmtpa is primary, esams squids hit Tampa squids.
- eqiad not yet online.
- We have three groups, text, upload, and bits.
 - Squid runs text and upload, varnish runs bits.
- Hit rates: ~85% text, ~98% media.

Presented by: Rob Halsell

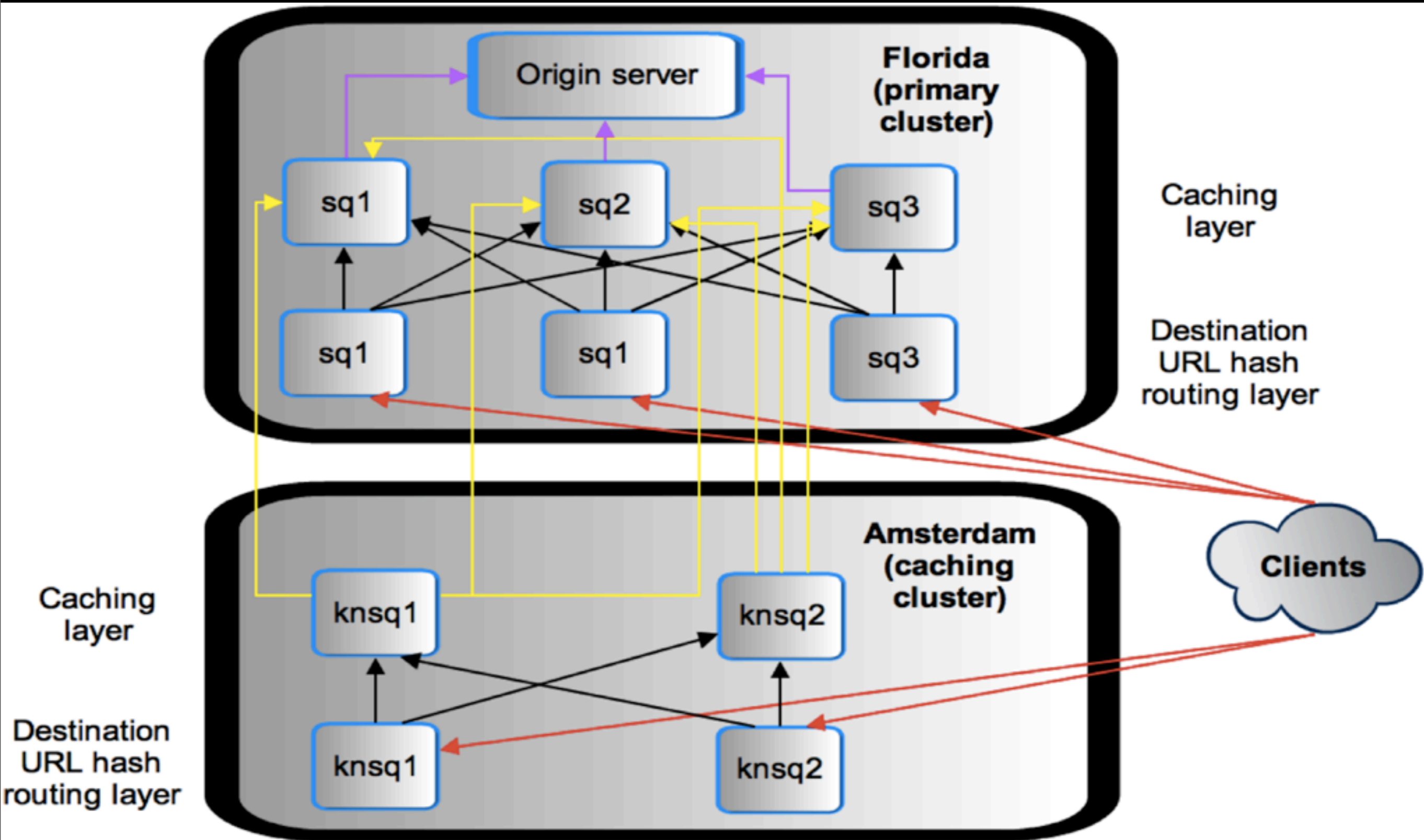
Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Squid Caching Invalidation

- Wiki pages are changed and invalidated at a very unpredictable rate.
- Users always need to see the latest revision.
- Cannot depend on invalidating based on time.
- Squid purge is implemented via mulitcast UDP based HTCP protocol.
- Squid handles text and upload.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011



Presented by: Rob Halsell

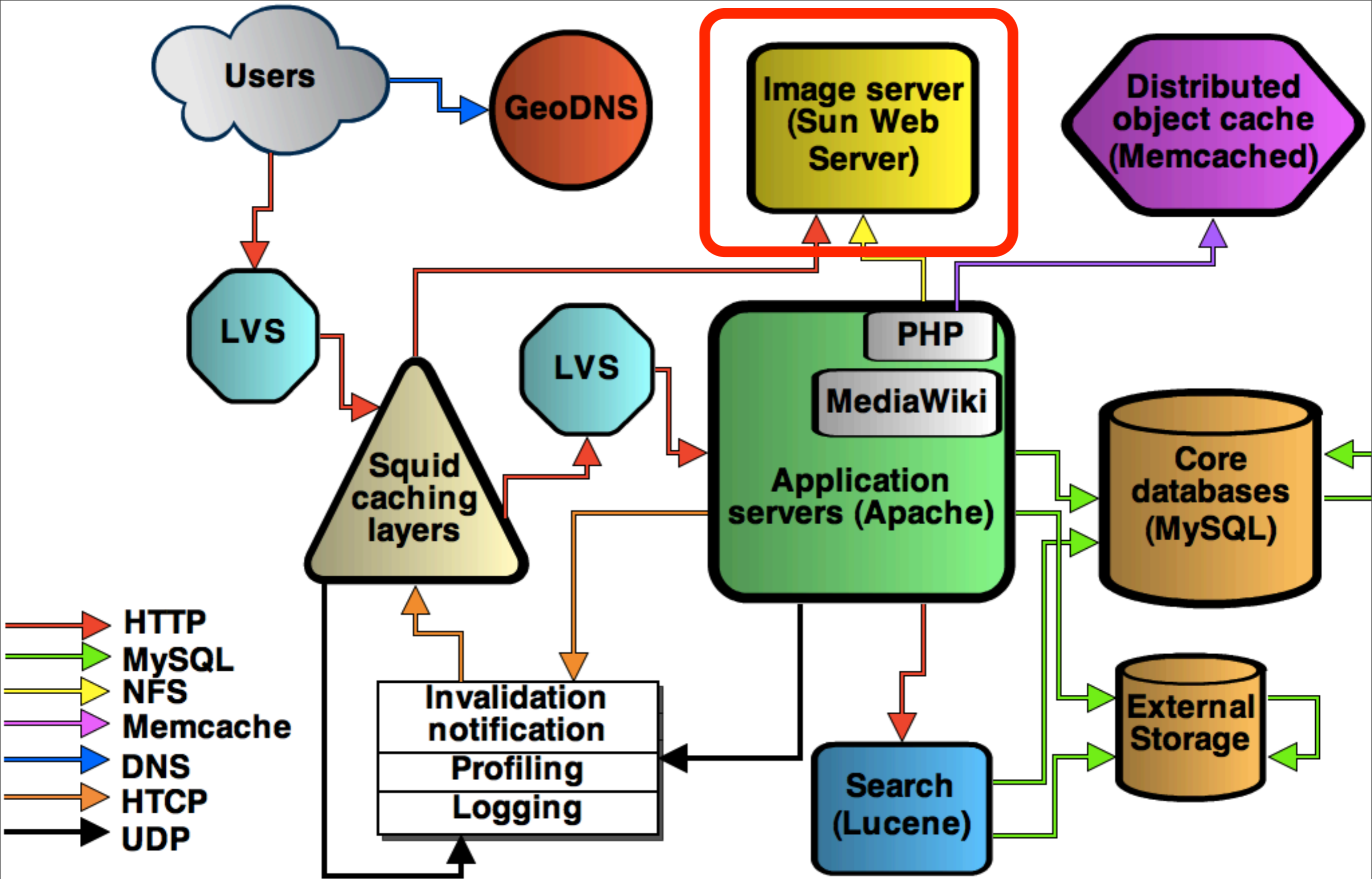
Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Varnish Caching Invalidation

- Currently Varnish just runs bits.
- Serves static content such as javascript and css files.
- Currently in development to replace squid for all caching.
- More efficient than squid, better able to use resources.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011



Presented by: Rob Halsell

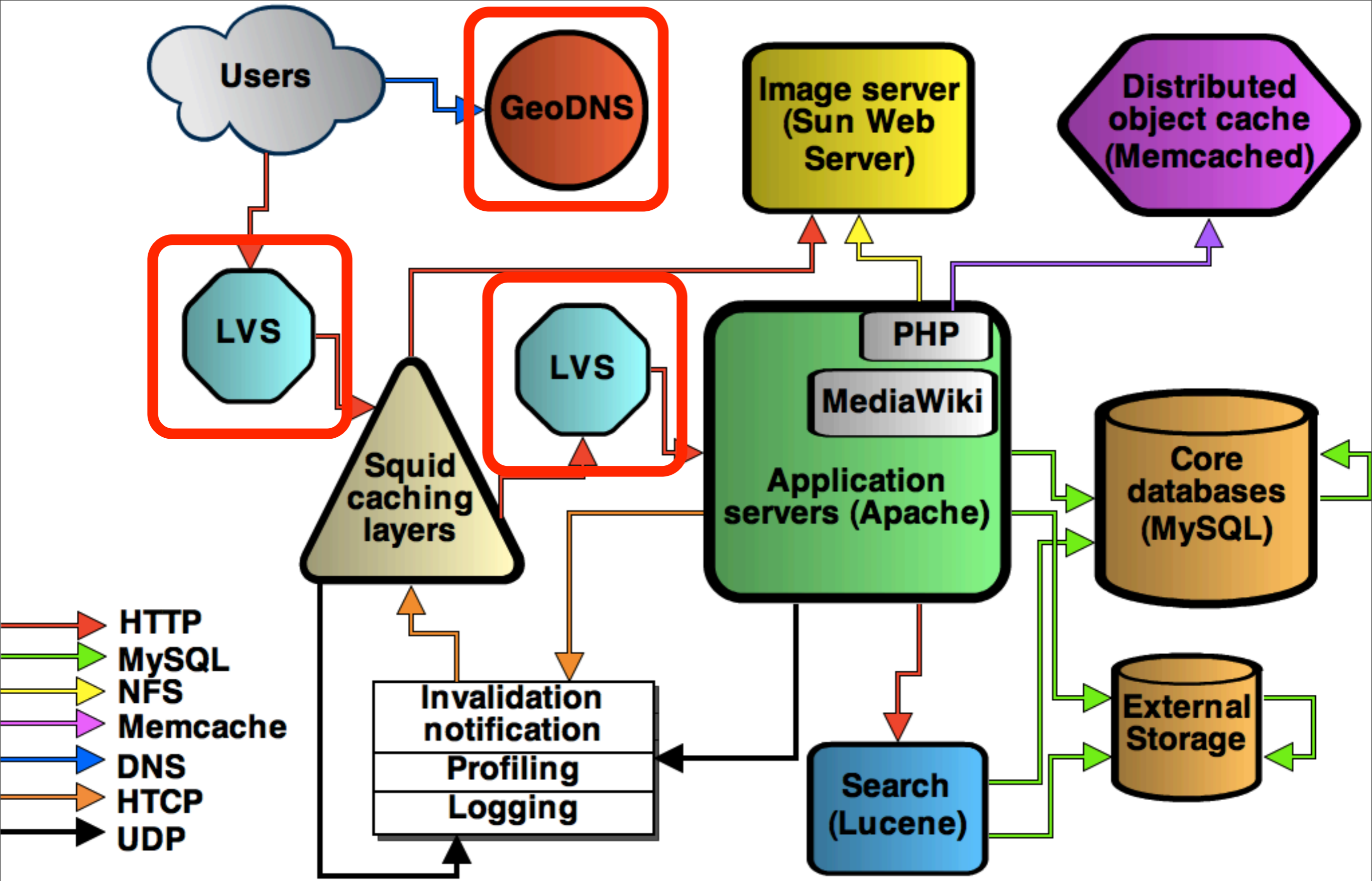
Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Media Storage

- Primary storage is currently used on 45XX series Sun servers.
- Not open source, not scalable.
- Seeking a new solution, currently a couple options in development.
- Thumbnails are served on same hardware, but different actual servers running Ubuntu.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

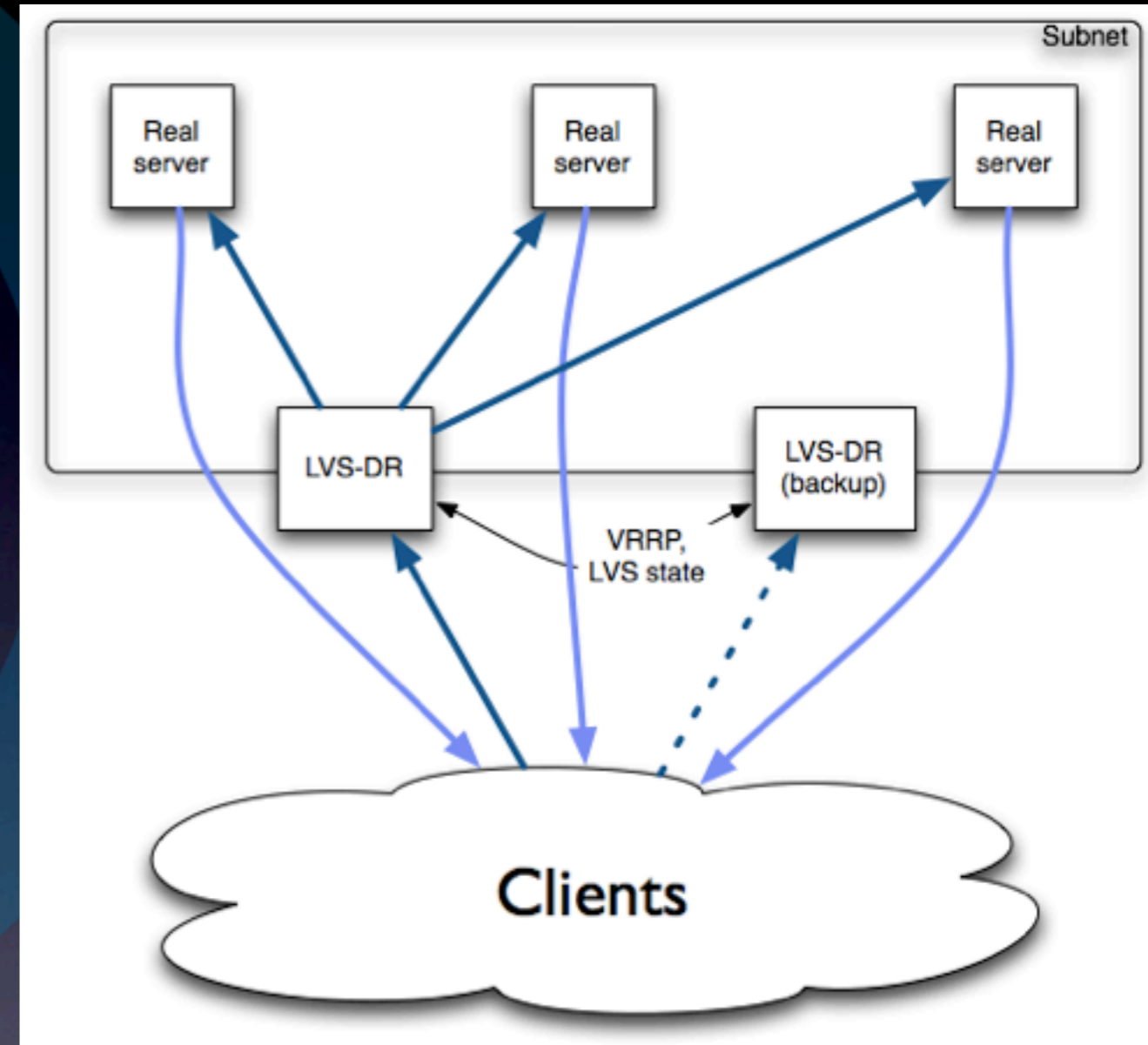


Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

LVS Load Balancing

- Linux Virtual Server
- Direct Routing mode
- All real servers share the same IP address
- Load balancer shares load across all real servers
- Return traffic is routed directly from the caching servers, not via the LVS server.



Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Geographic Load Balancing

- DNS resolves to the datacenter closest to the user.
- Not always physically closer.
- Map IP address of resolver to the country code.
- PowerDNS is used with Geobackend.

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Puppet

- <http://www.puppetlabs.com/>
- Uses ruby
- Handles all configuration file management & package installations for servers.
- All services must be documented and deployed via this tool.

```
CLASS APACHES::PACKAGES {
    # WIKIMEDIA-TASK-APPSERVER MOVED TO MEDIAWIKI.PP

    PACKAGE { [ "LIBAPACHE2-MOD-PHP5", "PHP-PEAR", "PHP5-CLI", "PHP5-
COMMON", "PHP5-CURL", "PHP5-MYSQL", "PHP5-XMLRPC" ]:
        ENSURE => LATEST;
    }
    IF ( $LSBDISTCODENAME == "HARDY" ) {
        PACKAGE { [ "PHP5-WIKIDIFF2", "PHP5-WMERRORS" ]:
            ENSURE => LATEST;
        }
    }
    IF ( $LSBDISTCODENAME == "LUCID" ) {
        PACKAGE { [ "PHP5", "PHP-WIKIDIFF2", "PHP5-WMERRORS" ]:
            ENSURE => LATEST;
        }
    }
}

NODE "HOOPER.WIKIMEDIA.ORG" {
    INCLUDE BASE,
        GANGLIA,
        NTP::CLIENT,
        EXIM::SIMPLE-MAIL-SENDER,
        ADMINS::ROOTS,
        SVN::CLIENT,
        MISC::ETHERPAD,
        MISC::BLOG-WIKIMEDIA,
        CERTIFICATES::STAR_WIKIMEDIA_ORG
}
```

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Wikitech

- <http://wikitech.wikimedia.org>
- Mediawiki Installation that we use to document setup, maintenance, recovery, etc...
- Exists outside the cluster, so if the cluster goes down, we still have notes on how to fix it.



Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Nagios

- <http://nagios.wikimedia.org>
- We use both active and passive checks as well as paging via direct attached SMS device.
- Checks are added to services via puppet and updated to our nagios server.
- Monitor all normal system specifications, including hardware faults and software response.

The screenshot displays the Nagios web interface for Wikimedia. The top navigation bar includes the Nagios logo and the URL <http://nagios.wikimedia.org/>. The main content area is divided into several sections:

- Current Network Status:** Shows the last update time (Thu Jul 28 16:53:56 UTC 2011) and provides links for history, notifications, and host status details.
- Summary Tables:** Two tables showing the overall system health. The first table shows 14 Up, 0 Down, 0 Unreachable, and 0 Pending. The second table shows 2 Ok, 0 Warning, 0 Unknown, 165 Critical, and 4 Pending.
- Display Filters:** A section for filtering the service status details, currently set to show all problems.
- Service Status Details For All Hosts:** A table listing various services across different hosts. The table has columns for Host, Service, Status, Last Check, Duration, Attempt, and Status Information.

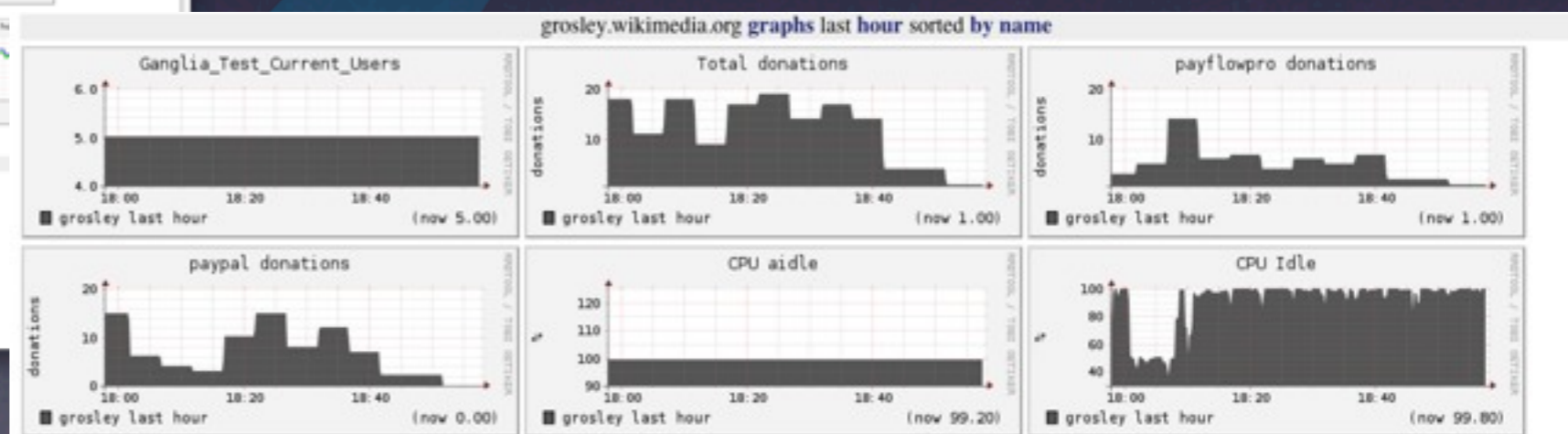
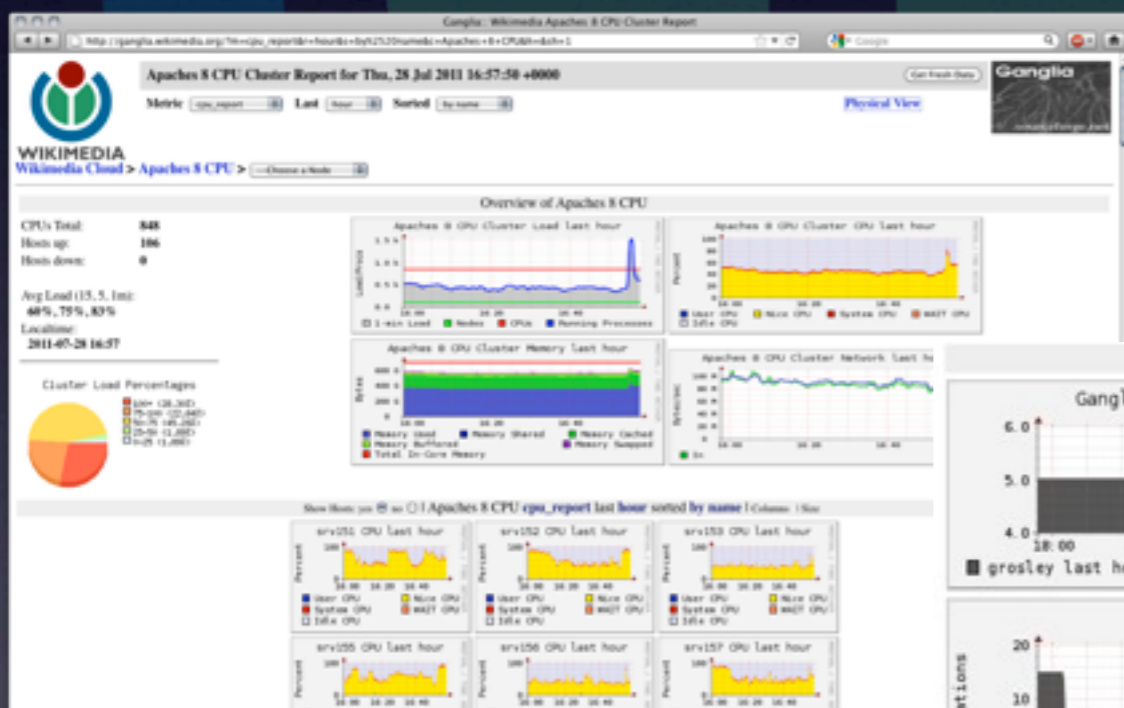
Host	Service	Status	Last Check	Duration	Attempt	Status Information
cr1-sfpoa	Router interfaces	CRITICAL	07-28-2011 16:53:09	8d 22h 50m 38s	3/3	CRITICAL: host '208.80.152.196': interfaces up: 58, down: 1, dormant: 0, excluded: 0, unused: 0 em1: down ->
db10	Disk space	WARNING	07-27-2011 20:12:35	1d 3h 59m 31s	3/3	DISK WARNING - free space: is 15460 MB (6% inode=99%);
db10	MySQL disk space	WARNING	07-28-2011 16:53:35	1d 3h 59m 56s	3/3	DISK WARNING - free space: is 15770 MB (6% inode=99%);
db1025	DPKG	CRITICAL	07-27-2011 20:12:53	6d 23h 25m 12s	3/3	DPKG CRITICAL: dpkg reports broken packages
db12	NTP	CRITICAL	07-27-2011 20:12:34	6d 20h 47m 23s	7/15	NTP CRITICAL: Offset unknown
db14	RAID	CRITICAL	07-27-2011 20:12:17	6d 2h 2m 51s	3/3	CRITICAL: logical devices: 1 defunct
db22	RAID	CRITICAL	07-28-2011 16:53:11	189d 15h 4m 7s	3/3	CRITICAL: logical devices: 1 defunct
db2	Disk space	CRITICAL	07-27-2011 20:12:11	34d 21h 16m 16s	3/3	DISK CRITICAL - free space: is 3394 MB (1% inode=99%);
db2	MySQL disk space	CRITICAL	07-27-2011 20:12:25	34d 21h 16m 0s	3/3	DISK CRITICAL - free space: is 3394 MB (1% inode=99%);
db20001	Storag	CRITICAL	07-27-2011 20:12:55	221d 2h 48m 36s	3/3	(null)
db2	RAID	CRITICAL	07-27-2011 20:12:27	100d 22h 33m 19s	3/3	CRITICAL: Active: 44, Working: 45, Failed: 1, Spare: 1
db2	Disk space	CRITICAL	07-28-2011 16:53:27	106d 17h 42m 1s	3/3	DISK CRITICAL - free space: is 3624 MB (0% inode=99%);
db2	check_all_mmcachefs	CRITICAL	07-27-2011 20:12:42	4d 16h 27m 50s	3/3	(Return code of 255 is out of bounds)
db2	DPKG	CRITICAL	07-28-2011 16:53:23	41d 18h 21m 30s	3/3	Connection refused by host
db2	Disk space	CRITICAL	07-28-2011 16:53:34	41d 18h 21m 17s	3/3	Connection refused by host
db2	RAID	CRITICAL	07-27-2011 20:12:54	41d 18h 21m 55s	3/3	Connection refused by host

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Ganglia

- <http://ganglia.wikimedia.org>
- Distributed monitoring software, graphing, and reviewing tool for servers.
- Shows us how hard the servers are working.

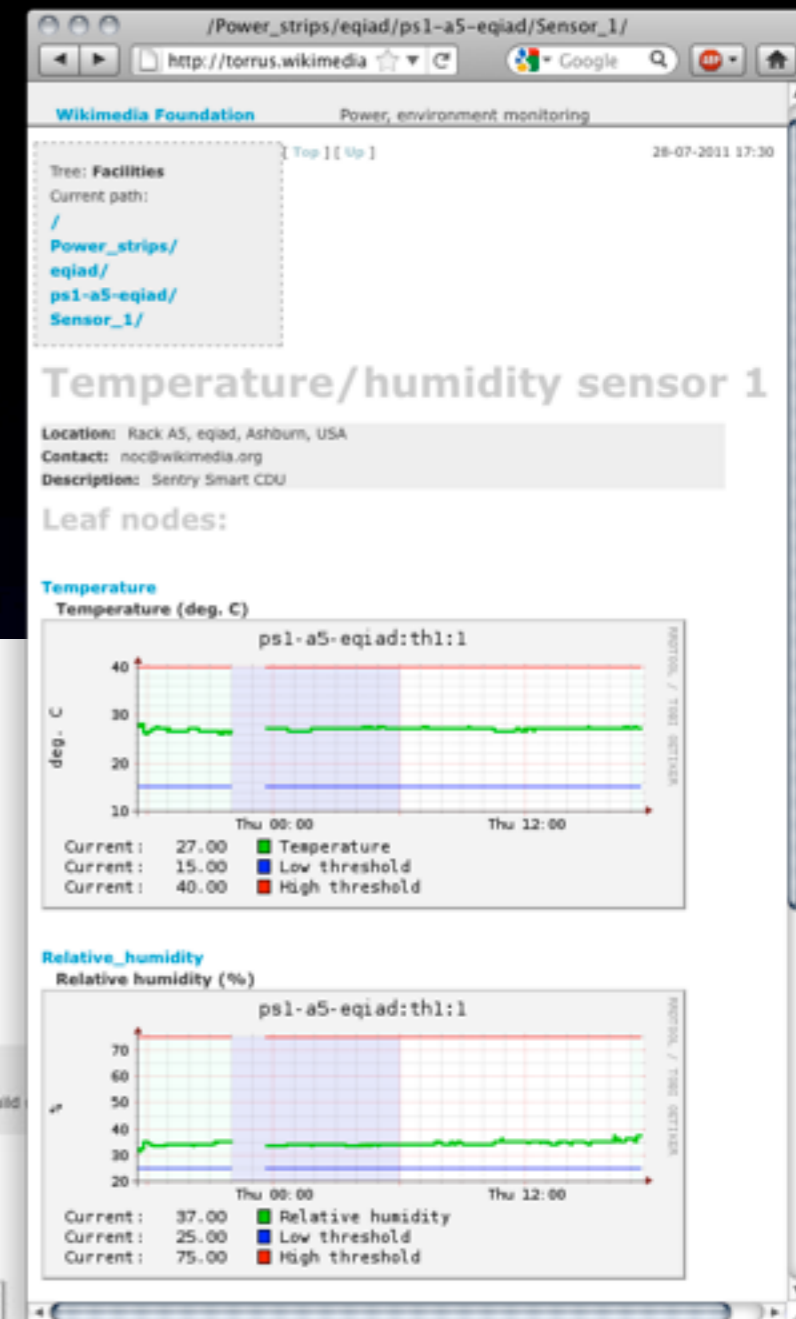
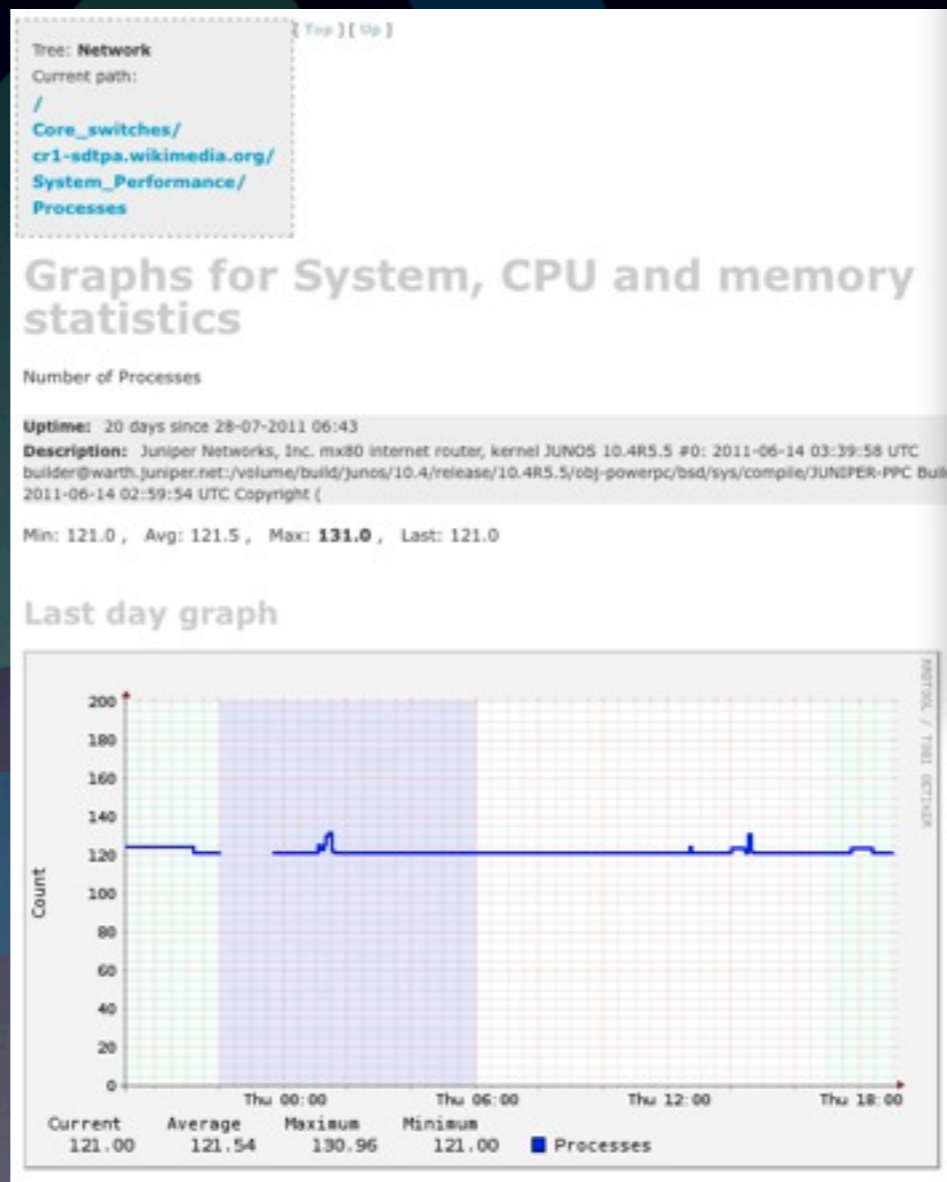


Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Torrus

- <http://torrus.wikimedia.org>
- SNMP monitoring and graphing software.
- We capture and display network usage, power and temperatures in our datacenters, squid performance, and other items.

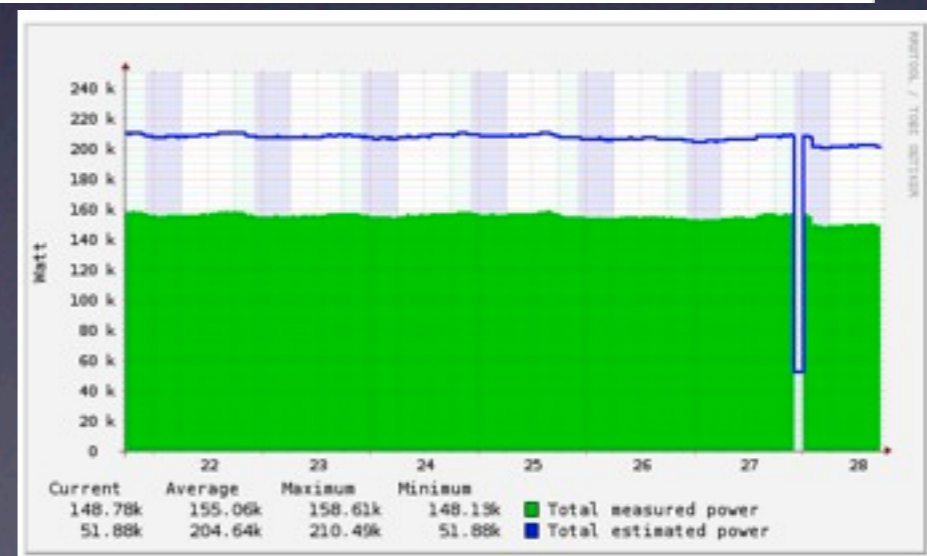
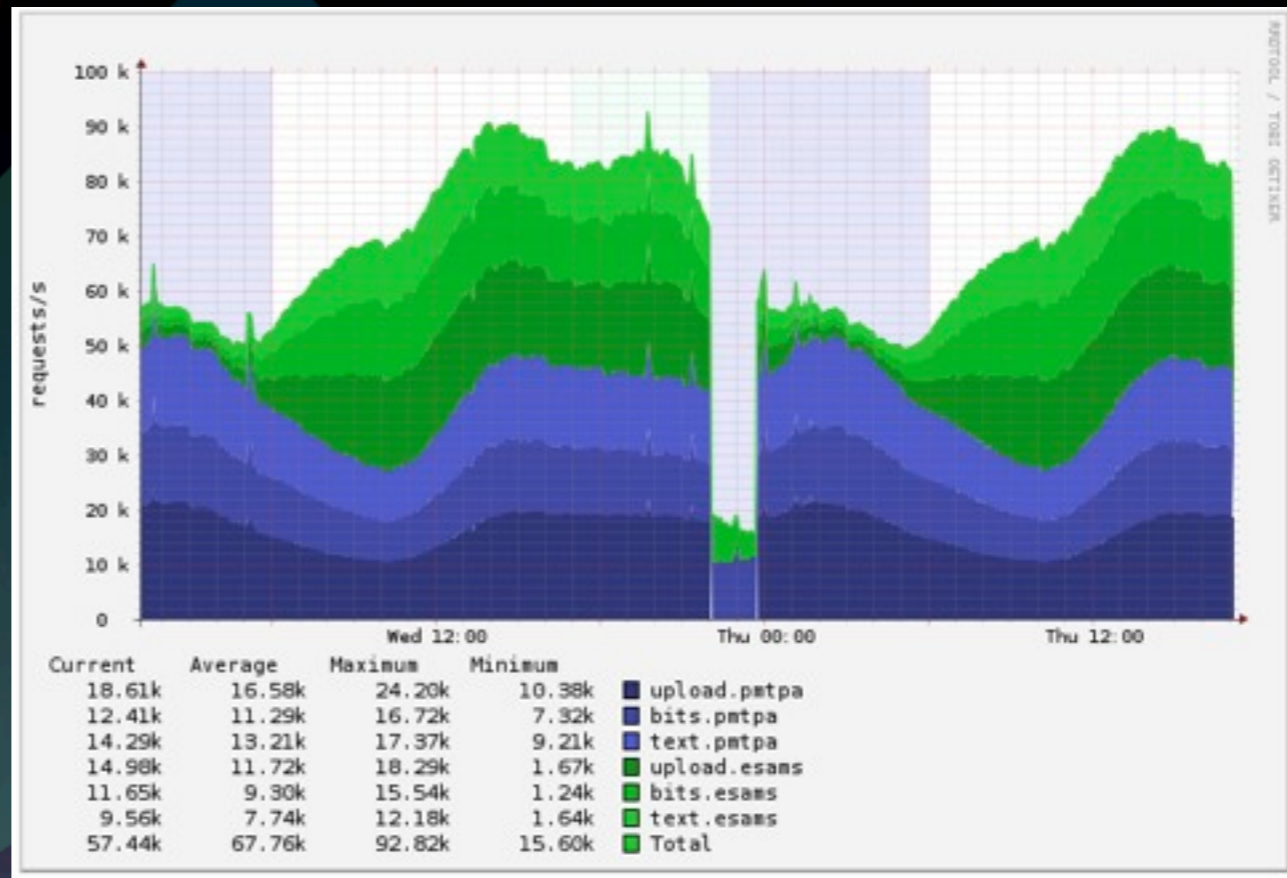


Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Statistics

- <http://stats.wikimedia.org/>
- Top request rate of approx. 141k requests per second.
- Approx: 760 servers.
- At minimum 160 kw power used. We only measure eqiad & sdtpa.



Presented by: Rob Halsell

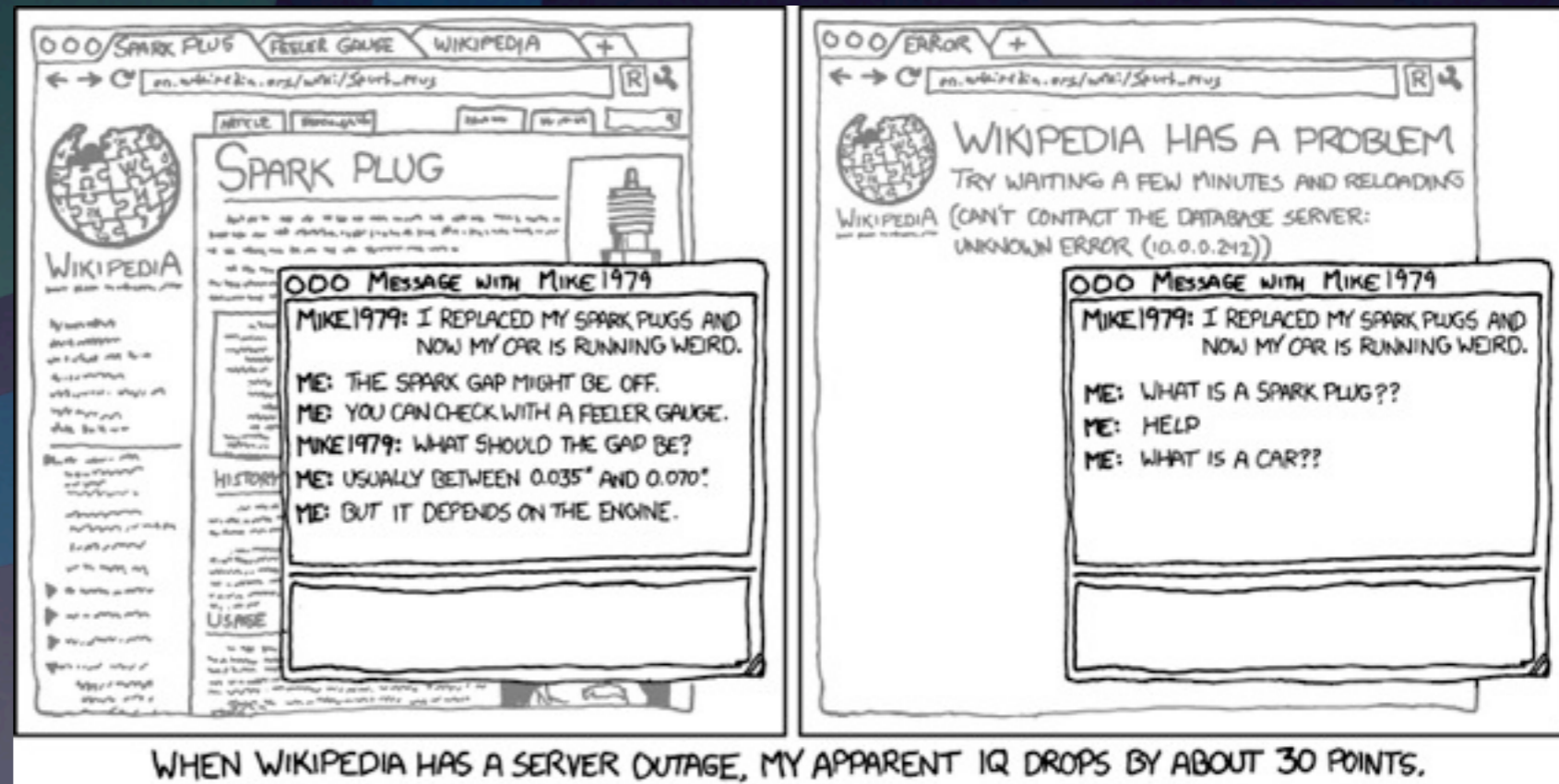
Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Wikipedia is down

(but my paper is due)



- <http://status.wikimedia.org>
- #wikimedia-tech
- <http://nagios.wikimedia.org>



Presented by: Rob Halsell

comic from xkcd.com

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Service / Website	Performance and Availability Status	Current Performance	Uptime Last 24h
✓ API	Service is operating normally	876 ms	100.0%
✓ DNS	Service is operating normally	338 ms	100.0%
✓ Dumps download	Service is operating normally	390 ms	100.0%
✓ GeoIP lookup	Service is operating normally	365 ms	100.0%
✓ https services (unsupported)	Service is operating normally	1271 ms	100.0%
✓ Images & media	Service is operating normally	124 ms	100.0%
✓ IRC RecentChanges	Service is operating normally	152 ms	100.0%
✓ Mail (SMTP)	Service is operating normally	211 ms	100.0%
✓ Mobile site	Service is operating normally	472 ms	99.9%
✓ payments.wikimedia.org	Service is operating normally	597 ms	100.0%
✓ Static assets (CSS/JS)	Service is operating normally	558 ms	100.0%
✓ Subversion (SVN)	Service is operating normally	368 ms	100.0%
✓ Ubuntu mirror	Service is operating normally	2125 ms	99.9%
✓ Wiki commons (s4)	Service is operating normally	2065 ms	100.0%
✓ Wiki commons (s4) - UNCACHED	Service is operating normally	1703 ms	100.0%
✓ Wiki platform [[w:de:Main Page]] (s5)	Service is operating normally	522 ms	100.0%
✓ Wiki platform [[w:de:Main Page]] (s5) - UNCACHED	Service is operating normally	349 ms	100.0%
✓ Wiki platform [[w:dsb:Main Page]] (s3)	Service is operating normally	356 ms	100.0%
✓ Wiki platform [[w:dsb:Main Page]] (s3) - UNCACHED	Service is operating normally	389 ms	100.0%
✓ Wiki platform [[w:en:Main Page]] (s1)	Service is operating normally	376 ms	100.0%

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

Open-source software used on cluster

mediawiki apache squid varnish
pybal memcached mysql php
ubuntu powerdns lighttpd exim
nfs wordpress limesurvey etherpad
nagios torrus requesttracker bugzilla
racktables observium drupal civicrm
otrs svn mailman openstack
nginx and others...

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011

In closing

- Nearly all operations are handled via open source.
- Always looking for more ops folks - future ops work will allow for volunteers.
- contact info: rhalsell@wikimedia.org, IRC: RobH

Presented by: Rob Halsell

Wikimania 2011 - Haifa, Israel
August 4-7, 2011