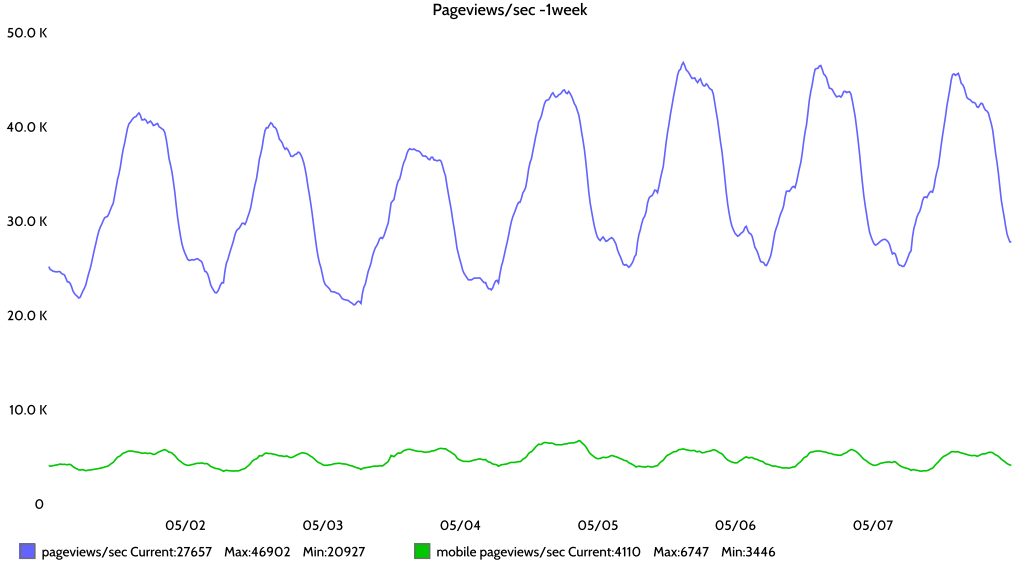
The Wikimedia infrastructure

Faidon Liambotis
faidon@wikimedia.org


WIKIPEDIA
The Free Encyclopedia

Design principles

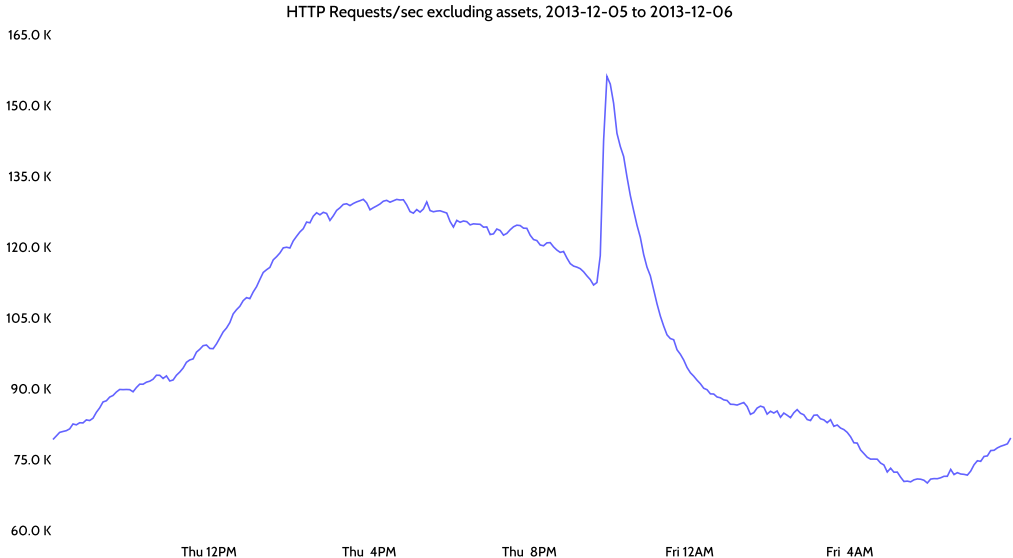


WIKIMEDIA FOUNDATION

Operating at scale

- Large, popular website
  - Wikipedia: 5^{th} largest website globally (comScore)
  - $\approx$ 500 million uniques, $\approx$ 20 billion pageviews per month
  - $\approx$ 190.000 HTTP req/s at peak
- Dynamic, collaborative
  - $\approx$ 80.000 active editors (active = 5+ edits per month)
  - $\approx$ 40.000 edits/hour
- Massive growth during the early years
- …but relatively constant traffic nowadays

WIKIMEDIA FOUNDATION



WIKIMEDIA FOUNDATION

Operating at scale (cont'd)

- Global in nature
  - No such thing as a 4am maintenance window
- Site needs to always be:
  - **Up.**
  - **Fast!**
- But also delivered **continuously**, using **agile** software practices

WIKIMEDIA FOUNDATION

Open-source, freedom, community & transparency

- Deeply rooted in the free culture and free software movements
- Infrastructure is being built *exclusively* with open-source components
- Design and build *in the open*, together with volunteers
- “Right to fork”
  - Anyone should be able to fork/clone
  - No secret sauce

WIKIMEDIA FOUNDATION

Limited resources

- Nonprofit, charitable organization
- Entirely funded by small donors
- No ads or VC money
- Small number of employees (not counting volunteers)
  - 2007: < 10
  - 2010: 61
  - 2014: 207 (65 SWE + 17 field ops/netops/SAs/DevOps)

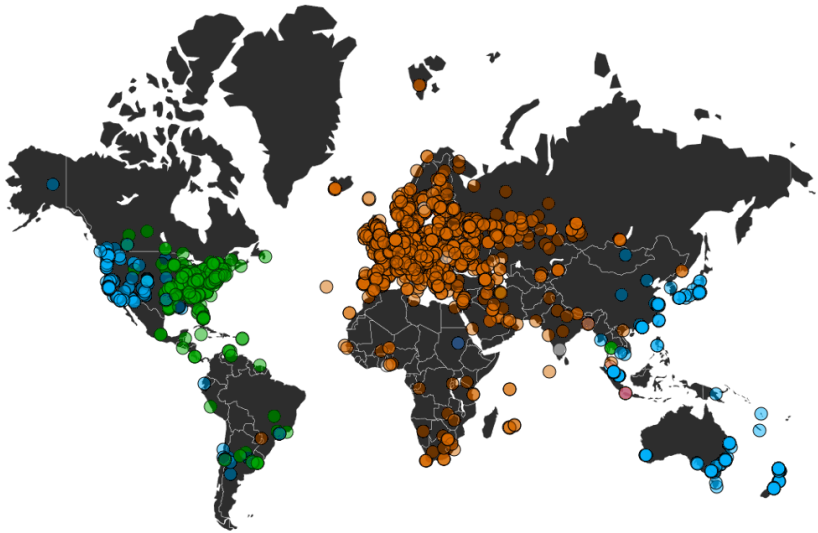
WIKIMEDIA FOUNDATION

Components

Physical topology

- Not using any third-party CDN or cloud provider
  - Usually involves secret sauces
  - Autonomy, privacy, risk of censorship
- Medium-sized infrastructure, $\approx$ 1.000 servers
- Two “primary” datacenters: Ashburn, VA (2011) & Dallas, TX (2014)
- Caching PoPs for CDN purposes
  - Amsterdam (EMEA) & San Francisco (NA West Coast, Oceania, Asia)

WIKIMEDIA FOUNDATION



WIKIMEDIA FOUNDATION

Network architecture

- Own user-facing & backhauling IP network
- AS **14907**, AS 43821
- 10G waves or MPLS redundant links between PoPs for backhauling
- Multiple 10Gs with tier-1/2s for transit on each location
- Present in multiple IXPs; peering settlement-free with everyone
- Proprietary network hardware for switches/routers :(

WIKIMEDIA FOUNDATION

System architecture

- Mostly one server vendor so far
- 1U/2U servers; no blades
- All physical; no virtualization (yet)
- Running exclusively **Ubuntu** Linux LTS (10.04, 12.04, 14.04)
- **Puppet** for configuration management
- **Salt** for remote execution/orchestration
- Automation, automation, automation

WIKIMEDIA FOUNDATION

Production architecture

Load balancing: layer 1

- Mapping users to PoPs: **GeoDNS**
  - Different responses per region to e.g. en.wikipedia.org
  - Europe resolves to Amsterdam; Oceania/East Asia to San Francisco
  - State/city load-balancing for US & Canada
- Using **gdnsd** since last year (switched from PowerDNS)
  - Highly-scalable, performant, stable, featureful
  - Uses MaxMind's GeoIP databases
- Serving ≈ 9.000 DNS req/s at peak

WIKIMEDIA FOUNDATION

Load balancing: layer 2/3

- Linux IPVS (LVS) for load-balancing
- LVS-DR, no need for big pipes
- Cheap scalability
  - No chokepoints
  - Commodity hardware (low-spec ordinary servers)
  - No expensive load-balancers or licenses
- Availability
  - Pybal: in-house monitoring daemon in Python
  - Health monitoring, pools/depools realservers
  - BGP with routers for IPVS availability failover

WIKIMEDIA FOUNDATION

Load balancing (& caching): layer 7

- nginx for (optional) SSL termination
- Multiple tiers of daisy-chained Varnish (≈ 80 in total)
  - High performance, generally very stable
  - Powerful but efficient custom DSL (VCL)
  - Based off the 3.0-plus branch, stack of custom patches on top
- Varnish for traffic routing
  - Consistent hashing per URL (custom director)
- Varnish for backend caching
  - Persistent on-disk caching
  - Backed with arrays of SSDs
  - Not as stable or supported anymore :(

WIKIMEDIA FOUNDATION

Main appserver stack

- LAMP stack on steroids
- Apache/PHP + a few custom PHP C extensions
- MediaWiki
  - Continuously evolving
- memcached
  - aggressive backend caching
  - twemproxy for connection pipelining & fault tolerance
- Redis
  - Job queue, etc.

WIKIMEDIA FOUNDATION

Main appserver stack (cont'd)

- MariaDB
  - Split into fairly static 7 shards, project/language-based
  - Beefy masters, multiple read-only slaves per shard
  - 1 master, 5-10 slaves each, < 100 servers in total
- ElasticSearch
  - (in progress)
  - Replacing old custom-built search on top of Lucene
  - Awesomeness.

WIKIMEDIA FOUNDATION

Internal services

- (Slow) move to SOA
- Multiple, smaller RESTful services
  - New wikitext⇔HTML parser (*Parsoid*)
  - HTML/RDF to PDF rendering
  - LaTeX/Math processor (*Mathoid*)
- Mostly in Node.js (so far)
- More to come!

WIKIMEDIA FOUNDATION

Media storage infrastructure

- Storing mainly images, but also audio & video
- Original uploads & arbitrarily-sized thumbnails
- ≈30 million originals, ≈ 320 million thumbnails
- ≈ 800 TB of raw storage
- Entirely based on OpenStack **Swift**
  - Horizontally scalable, region-aware, well-defined API, middlewares

WIKIMEDIA FOUNDATION

Production-like infrastructure

Wikimedia Labs

- Infrastructure for staff & **volunteers**
- OpenStack private cloud
- VMs running on the production puppet tree (sans passwords)
- Development, experimenting, QA, staging
- Public, participatory, collaborative
- https://wikitech.wikimedia.org/

WIKIMEDIA FOUNDATION

Thank you!

@faidonl