

Software libero per un'enciclopedia libera

Giuseppe Lavagetto (Wikimedia Foundation)



**Emilia Romagna
Linux User Group**



25/10/2014

ErLUG - Linux Day 2014 Bologna



Wikipedia: qualche numero (approssimativo)

- Quinto sito al mondo per traffico
- 505 milioni di utenti al mese
- 80 mila redattori attivi
- 20 miliardi di pagine viste al mese
- 30 milioni di articoli
- ... non solo wikipedia!

Problema #1

- Gli utenti non sono il nostro prodotto
- Wikipedia vive delle donazioni dei suoi utenti
- WMF ha circa 200 dipendenti, Google 40 mila, Facebook 10 mila...

Come si rende veloce e affidabile un sito di tali dimensioni con risorse così limitate?

Free software!

- Wikipedia condivide i principi e le pratiche del software libero
- Tutti i contenuti hanno una licenza libera
- Tutto il software è rilasciato sotto licenze libere
- Il contributo dei volontari è fondamentale anche per il funzionamento dei siti
- Si usa il lavoro di migliaia di ingegneri

Qualche altro numero

- ~1000 server
- ~ 16000 core
- 4 Datacenters (3 in USA, 1 in Europa)
- 190000 richieste/s al picco
- 40000 edit/ora
- 30 milioni di immagini originali, 320 M thumbnails
- 800 TB di storage

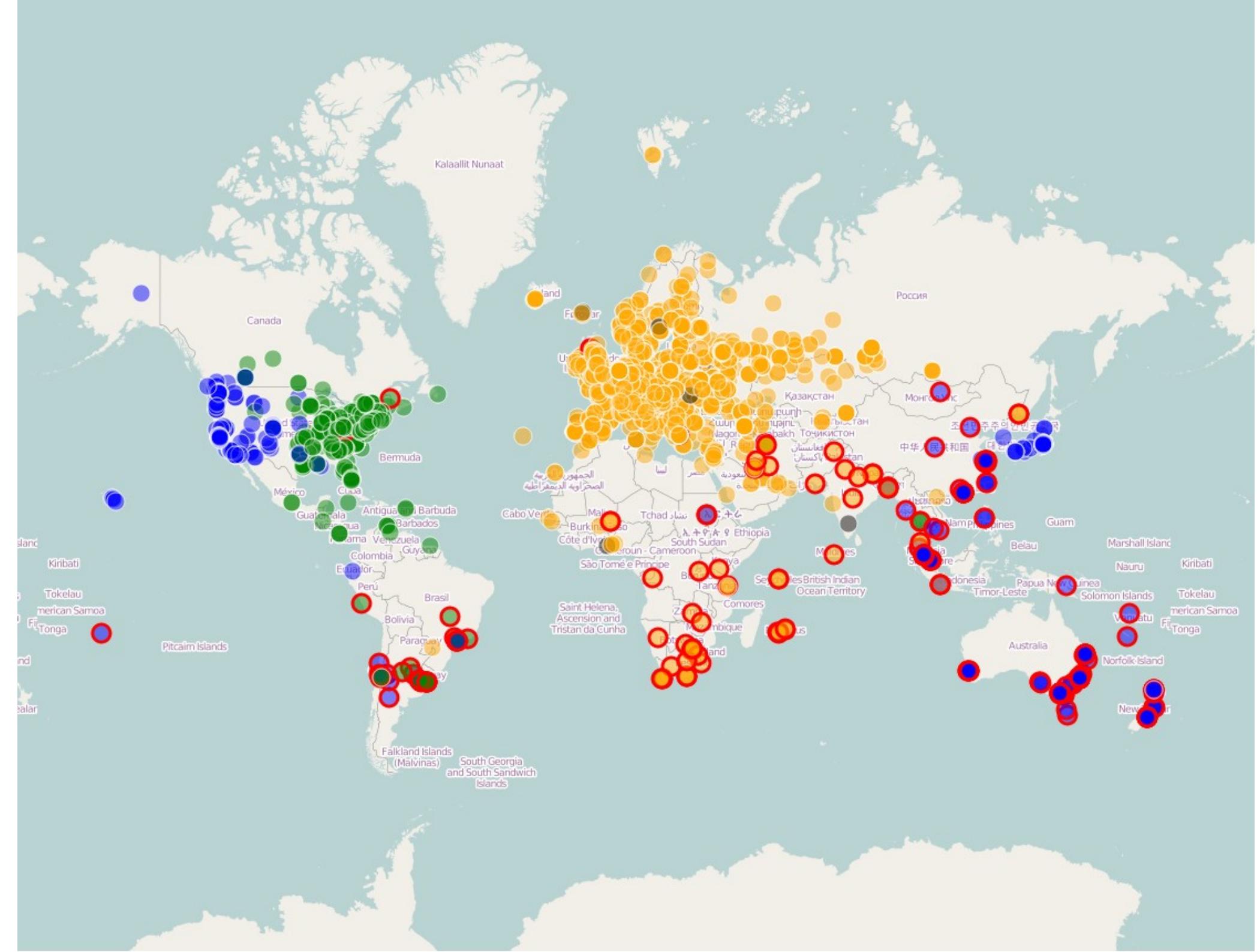
I Datacenter

- Un Content Delivery Network permette di raggiungere facilmente gli utenti in tutto il globo
- La maggior parte dei grandi siti web usa una CDN commerciale
- Oltre ai problemi di privacy, \$\$\$

...quindi abbiamo creato la nostra CDN!

Come funziona

- DNS geolocalizzato dirige gli utenti al datacenter piu' vicino (se cercate "it.wikipedia.org" in diversi luoghi nel mondo il risultato cambia)
- Velocita' misurata con l'aiuto del progetto ATLAS del RIPE



Caching

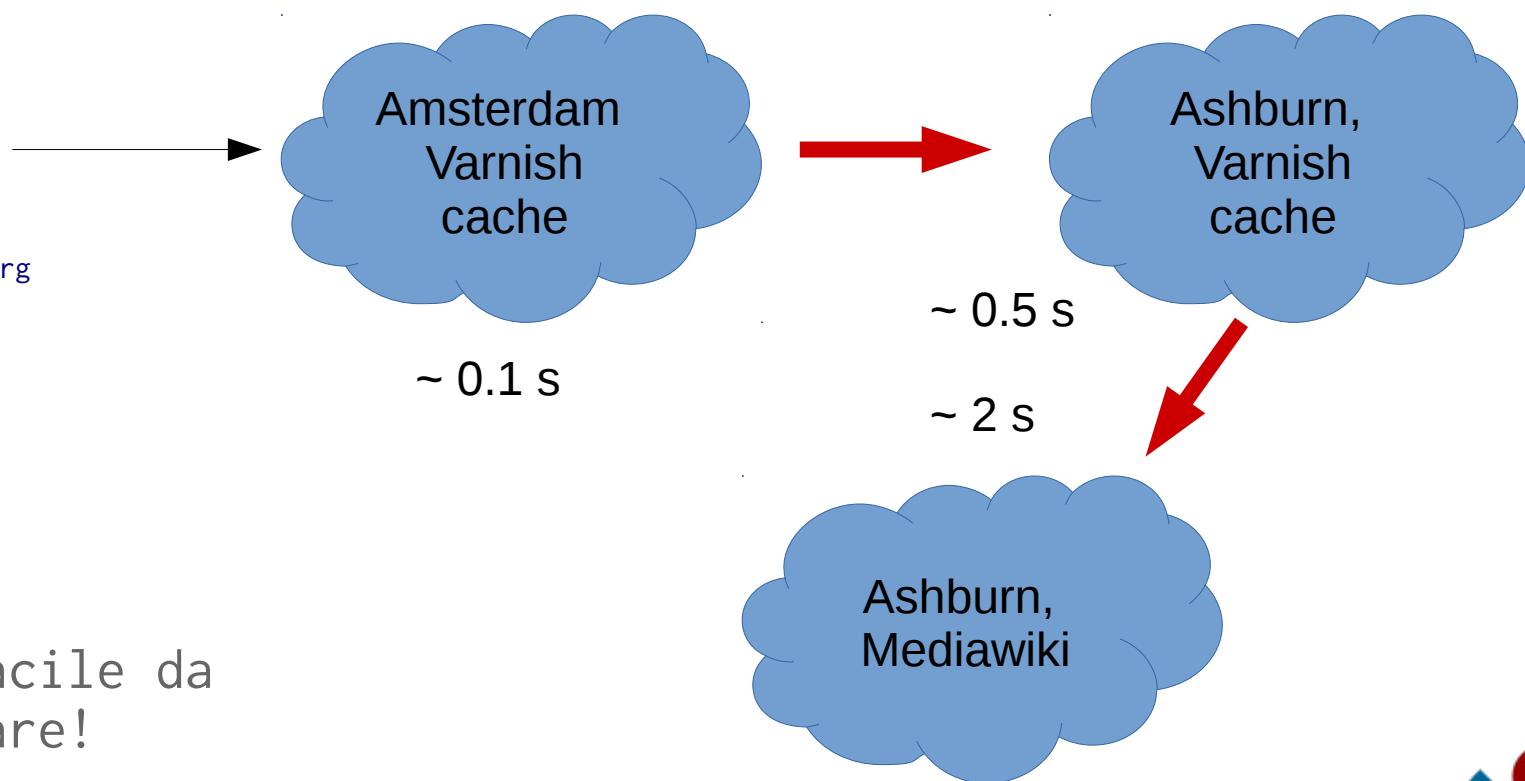
- Letture >> Scritture
- Ogni datacenter ospita un cluster di cache
- Esperienza utente migliore
- Meno risorse computazionali
- Varnish (<https://www.varnish-cache.org/>)

Caching/2

La storia di una richiesta



[https://it.wikipedia.org
/wiki/Mus_musculus](https://it.wikipedia.org/wiki/Mus_musculus)



... ed è facile da
controllare!

Mediawiki

- Il software per wiki più diffuso, FLOSS fin dal concepimento (2002)
- Wikipedia, Wikia, Wikileaks... e il wiki di molti LUG
- La differenza è che oggi mediawiki per far funzionare wikipedia è installato su 250 servers, a cui si aggiungono 15 server di Cache locale (memcached) e alcune decine di database – quasi tutti su MariaDB 10.

Open source infrastructure

Ogni aspetto del funzionamento dei nostri siti è trasparente e documentato in rete:

- Mailing Lists
- Canali IRC
- [Wikitech](#)
- Server Admin Log (<https://bit.ly/wikisal>)
- “Infrastructure as Code” – Puppet
- Il contributo di chiunque è il benvenuto.

Open monitoring

Diversi strumenti:

- Ganglia (
<https://ganglia.wikimedia.org/latest>)
- Gdash (<https://gdash.wikimedia.org/>)
- Status (<http://status.wikimedia.org>)
- ... altri tool liberamente consultabili in arrivo.

Wikimedia Labs

- Public cloud basata su openstack
- Si puo' ottenere l'uso per chiunque voglia sperimentare/abbia progetti rilevanti
- Ogni VM puo' essere configurata usando le "ricette" che vengono dalla produzione

Come contribuire

- Leggere la documentazione su wikitech
- Chiedete un account per wikimedia labs!
- <https://git.wikimedia.org>
- entrare in IRC e contattateci...
- ...oppure donate :)

Grazie dell'attenzione

Questa presentazione è rilasciata sotto licenza Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License