

Life Story of a Pageview

Madhumitha Viswanathan

@madhuvishy

Software Engineer, Analytics @ Wikimedia Foundation

Wikimedia Pageviews API

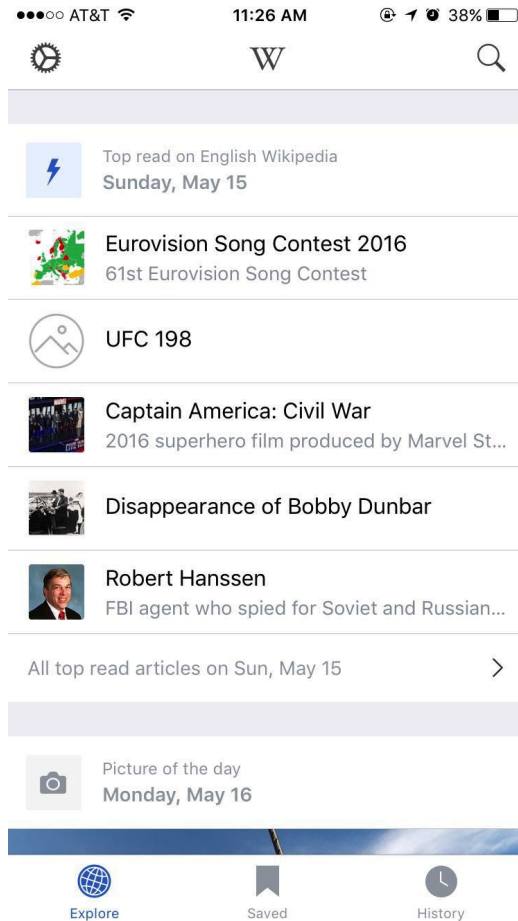
← → ↻ https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/de.wikipedia/all-access/user/Johann_Wolfgang_von_Goethe/daily/...

```
{
  "items": [
    {
      "project": "de.wikipedia",
      "article": "Johann_Wolfgang_von_Goethe",
      "granularity": "daily",
      "timestamp": "2016051300",
      "access": "all-access",
      "agent": "user",
      "views": 2299
    }
  ]
}
```

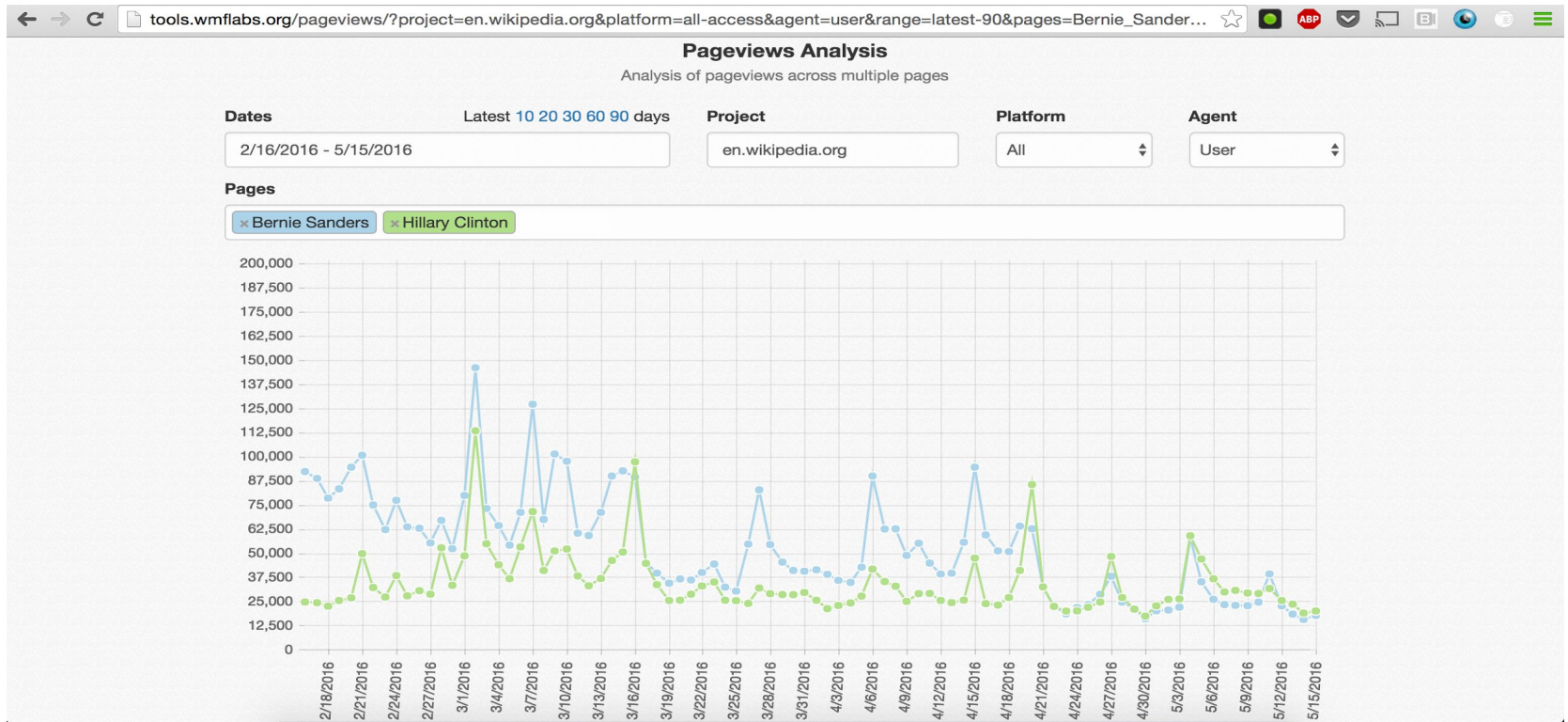
Announcement: <http://blog.wikimedia.org/2015/12/14/pageview-data-easily-accessible/>

Documentation: <https://wikitech.wikimedia.org/wiki/Analytics/PageviewAPI>

Used in Wikipedia iOS APP



Pageview Stats Tool



Screenshot from: <http://tools.wmflabs.org/pageviews>

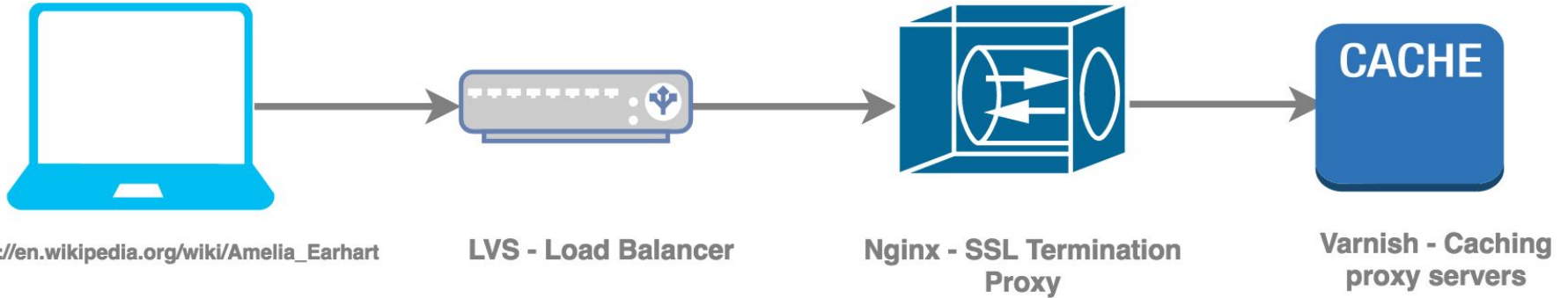
The BIG Picture

- ~ 820 Wikimedia Projects
- ~ 150K requests/second
- ~ 13B requests/day
- ~ 1.6T request data/day

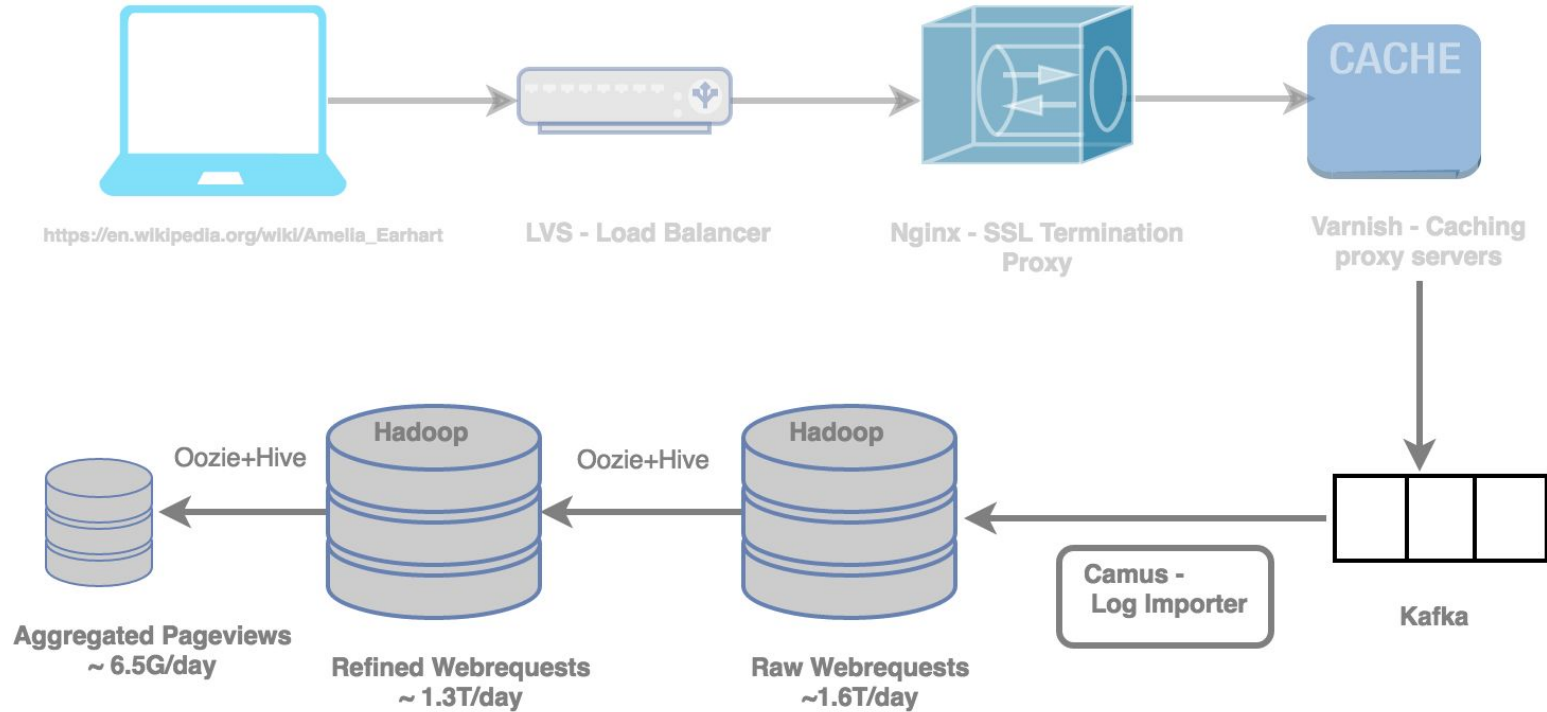
It starts with you



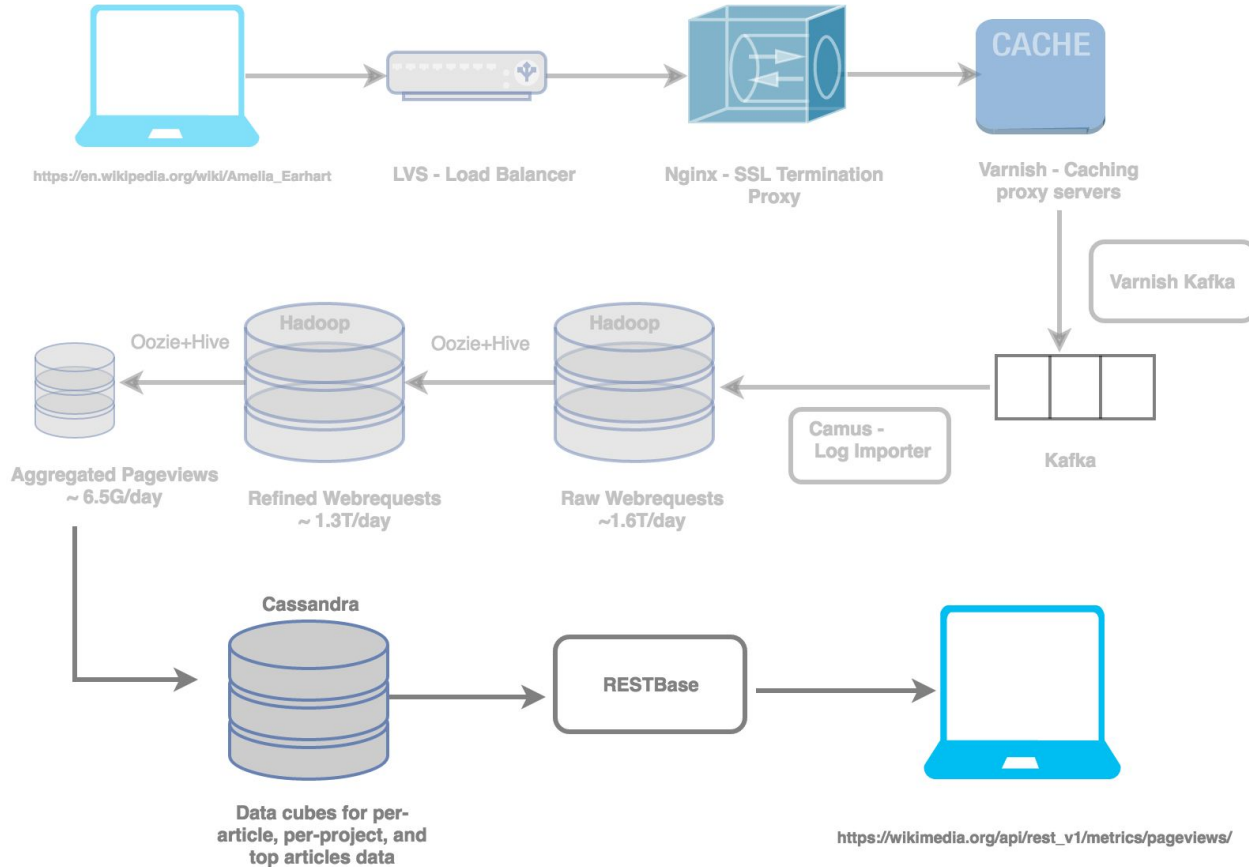
Travels through the web to our servers



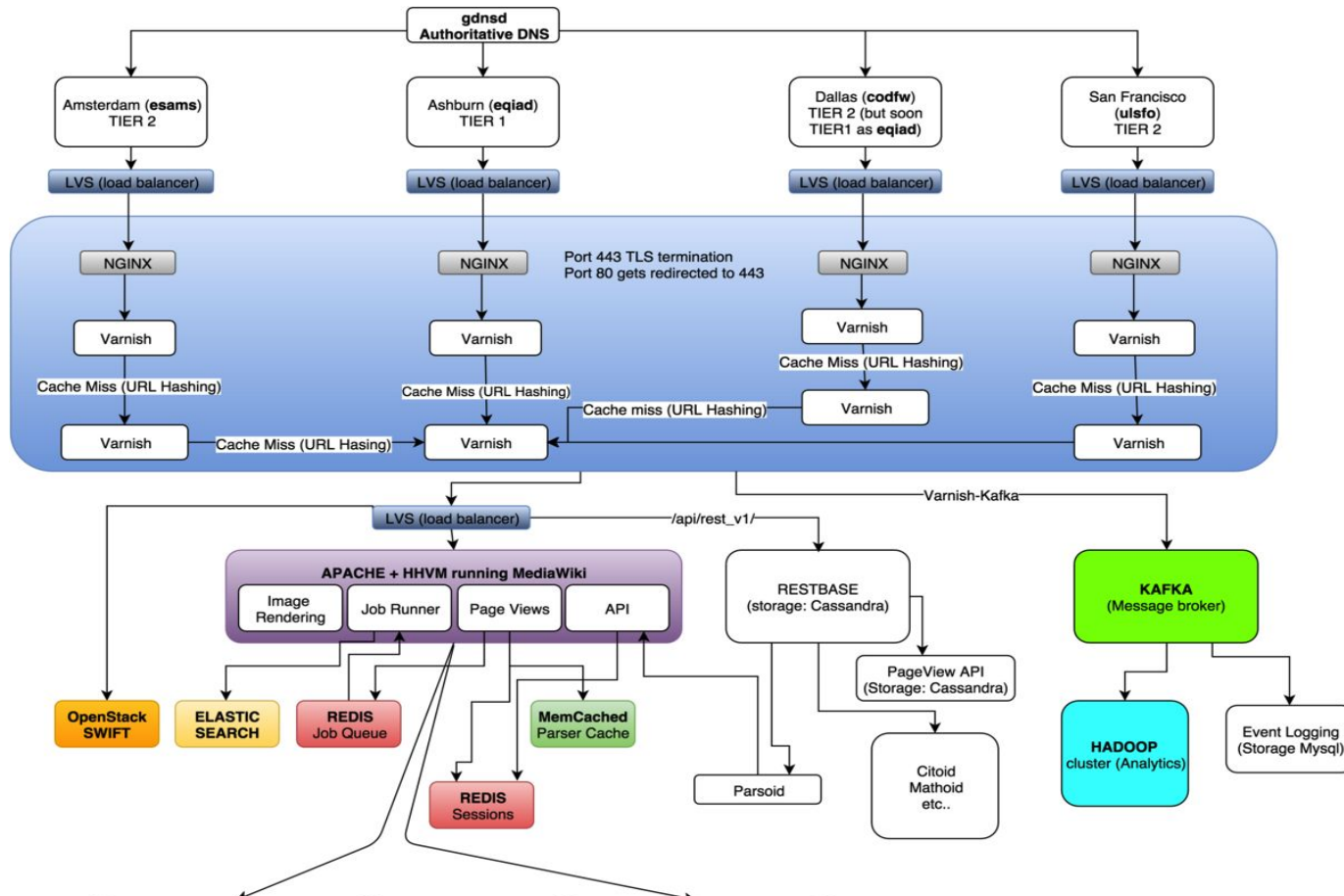
Ends up in Hadoop Land



And back to you



The full overview



Use the APIs!

https://wikimedia.org/api/rest_v1/?doc

<https://wikitech.wikimedia.org/wiki/Analytics/PageviewAPI>

https://wikitech.wikimedia.org/wiki/Analytics/Unique_Devices

Image Credits

Infrastructure Overview - By User:Elukey [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons

Questions?