

# How to Run a Top-10 Website, Publicly and Transparently

Kunal Mehta / User:Legoktm  
HOPE 2022

# WIKIPEDIA

The Free Encyclopedia

## English

6 458 000+ articles

## 日本語

1 314 000+ 記事

## Русский

1 798 000+ статей

## Deutsch

2 667 000+ Artikel

## Español

1 755 000+ artículos

## Français

2 400 000+ articles

## Italiano

1 742 000+ voci

## 中文

1 256 000+ 条目 / 條目



## Português

1 085 000+ artigos

## Polski

1 512 000+ hasel

EN ▾ 

 [Read Wikipedia in your language](#) ▾



# Wikipedia is

- An encyclopedia
- Edited by volunteers
- Available under a free Creative Commons license
- Trying to compile the sum of all human knowledge
- Makes the internet not suck



# We are only 5% there



The idea of compiling **all human knowledge** in a single work, although not in a single place,<sup>[3]</sup> is highly seductive. In this project, we attempt to study how many articles are needed to cover the sum of all human knowledge. As of July 2022, [English Wikipedia \(Q328\)](#) has 6,544,399 articles<sup>[4]</sup> and [Wikidata \(Q2013\)](#) includes 99,331,255 items.<sup>[5]</sup> On the one hand, this page still in expansion estimates that, according to its current [notability threshold](#), the **total notable articles figure for Wikipedia is over 119,511,420**. On the other hand, the potential number of items is much bigger for Wikidata, probably over 1,000,000,000 (1 billion), as its [notability threshold](#) is lower.

"This is a work that cannot be completed except by a society of men of letters and skilled workmen, each working separately on his own part, but all bound together solely by their zeal for the best interests of the human race and a feeling of mutual good will." -[Denis Diderot \(Q448\)](#), [Encyclopédie \(Q447\)](#) (1751-1766)

Many individuals devoted their lives to different efforts of knowledge compilation and preservation. Some inspiring cases are [Vivian Maier \(Q200890\)](#), a [nanny \(Q936969\)](#) that took 150,000 photographs during her lifetime primarily of people and architecture;<sup>[6]</sup> [Paul Mawhinney \(Q31818556\)](#), who archived a copy of every sold album growing a 3 million vinyl collection;<sup>[7]</sup> [Henry Spencer \(Q5358365\)](#), a computer scientist that preserved over 2 million [Usenet \(Q193162\)](#) messages onto magnetic tapes<sup>[8]</sup> or [Marion Stokes \(Q17612042\)](#), who recorded hundreds of thousands of hours of television news footage spanning 35 years.<sup>[9]</sup>

"Like all persons of the Library, I have traveled in my youth; I have wandered in search of a book, perhaps the [catalogue of catalogues...](#)" -[Jorge Luis Borges \(Q909\)](#), [The Library of Babel \(Q473\)](#) (1941)



Preserving *all human knowledge* for posterity. **It is all-important.**



Wikimedia servers, holding *terabyte (Q79741)* of text and



# Transparency is core to Wikipedia

Article Talk Read Edit View history

## *Webb's First Deep Field*

Coordinates: 07° 23′ N 19.5° W﻿ / ﻿73° 27′ W﻿ / 15.6° W

From Wikipedia, the free encyclopedia

***Webb's First Deep Field*** is the first [operational image](#) taken by the [James Webb Space Telescope](#). The [deep-field](#) photograph, covering a tiny area of sky visible from the [Southern Hemisphere](#), depicts [SMACS 0723](#), a [galaxy cluster](#) 4.6 billion [light-years](#) from Earth in the constellation of [Volans](#). Captured by the telescope's [Near-Infrared Camera](#) (NIRCam), the image was revealed to the public by [NASA](#) on 11 July 2022. Thousands of [galaxies](#) are visible in the image, which is the highest-resolution image of the [early universe](#) ever taken.

### Contents [hide]

- [Background](#)
- [Scientific results](#)
- [Explanation for the six rays emanating from bright sources of light](#)
- [Significance](#)
- [See also](#)
- [References](#)

## Background [edit]

The [James Webb Space Telescope](#) is a [space telescope](#) operated by [NASA](#) and designed primarily to conduct [infrared astronomy](#). Launched in December 2021, the spacecraft has been in a [halo orbit](#) around the second Sun–Earth [Lagrange point](#) (L<sub>2</sub>), about 1.5 million kilometres (900,000 mi) from Earth, since January 2022. At L<sub>2</sub>, the gravitational pulls of both the Sun and Earth keep the telescope's motion around the Sun synchronized with Earth's.<sup>[1]</sup>

*Webb's First Deep Field* was taken by the telescope's [Near-Infrared Camera](#) (NIRCam) and is a composite produced from images at different [wavelengths](#), totaling 12.5 hours.<sup>[2][3]</sup> The photo achieved depths at infrared wavelengths beyond the [Hubble Space Telescope](#)'s deepest fields, which took weeks.

### Webb's First Deep Field



*Webb's First Deep Field*

#### Observation data (Epoch)

**Constellation(s)** Galaxy cluster  
SMACS 0723

**Distance (co-moving)** 4.6 billion light years

See also: [Galaxy group](#), [Galaxy cluster](#),  
[List of galaxy groups and clusters](#)



# Transparency is core to Wikipedia



## Webb's First Deep Field

Coordinates: 07° 23′ 19.5″ S, −73° 27′ 15.6″ E﻿/﻿7° 23′ 19.5″ S, 73° 27′ 15.6″ E﻿/﻿-7.3918333° S, 73.4544444° E

From Wikipedia, the free encyclopedia

**Webb's First Deep Field** is the first [operational image](#) taken by the [James Webb Space Telescope](#). The [deep-field](#) photograph, covering a tiny area of sky visible from the [Southern Hemisphere](#), depicts [SMACS 0723](#), a [galaxy cluster](#) 4.6 billion [light-years](#) from Earth in the [constellation](#) of [Volans](#). Captured by the telescope's [Near-Infrared Camera](#) (NIRCam), the image was revealed to the public by [NASA](#) on 11 July 2022. Thousands of [galaxies](#) are visible in the image, which is the highest-resolution image of the [early universe](#) ever taken.

### Contents [hide]

- Background
- Scientific results
- Explanation for the six rays emanating from bright sources of light
- Significance
- See also
- References

## Background [ edit ]

The [James Webb Space Telescope](#) is a [space telescope](#) operated by [NASA](#) and designed primarily to conduct [infrared astronomy](#). Launched in December 2021, the spacecraft has been in a [halo orbit](#) around the second Sun–Earth [Lagrange point](#) (L<sub>2</sub>), about 1.5 million kilometres (900,000 mi) from Earth, since January 2022. At L<sub>2</sub>, the gravitational pulls of both the Sun and Earth keep the telescope's motion around the Sun synchronized with Earth's.<sup>[1]</sup>

*Webb's First Deep Field* was taken by the telescope's [Near-Infrared Camera](#) (NIRCam) and is a composite produced from images at different [wavelengths](#), totaling 12.5 hours.<sup>[2][3]</sup> The photo achieved depths at infrared wavelengths beyond the [Hubble Space Telescope's](#) deepest fields, which took weeks.

### Webb's First Deep Field



*Webb's First Deep Field*

#### Observation data (Epoch)

<b>Constellation(s)</b>	Galaxy cluster SMACS 0723
<b>Distance (co-moving)</b>	4.6 billion light years
<b>See also:</b>	Galaxy group, Galaxy cluster, List of galaxy groups and clusters



# Transparency is core to Wikipedia

## Webb's First Deep Field: Revision history

[View logs for this page](#) ([view filter log](#))

### Filter revisions

External tools: [Find addition/removal](#) (*Alternate*) • [Find edits by user](#) (*Alternate*) • [Page statistics](#) • [Pageviews](#) • [Fix dead links](#)

For any version listed below, click on its date to view it. For more help, see [Help:Page history](#) and [Help:Edit summary](#). (cur) = difference from current version, (prev) = difference from preceding version,

**m** = minor edit, **→** = section edit, **←** = automatic edit summary

(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

### Compare selected revisions

- [\(cur | prev\)](#) [16:28, 15 July 2022](#) [Xanzzibar](#) (talk | [contribs](#)) **m** . . (9,817 bytes) **(+4)** . . (→*Explanation for the 6 rays emanating from bright sources of light: MOS:NUMERAL*) ([undo](#))
- [\(cur | prev\)](#) [16:24, 15 July 2022](#) [Banana Republic](#) (talk | [contribs](#)) . . (9,813 bytes) **(+1,057)** . . (*Add an explanation for the very prominent feature of the 6 rays around light sources*) ([undo](#))
- [\(cur | prev\)](#) [15:37, 15 July 2022](#) [Unknownuncertain](#) (talk | [contribs](#)) **m** . . (8,756 bytes) **(−4)** . . (→*Background*) ([undo](#)) (*Tag: Visual edit*)
- [\(cur | prev\)](#) [01:42, 15 July 2022](#) [Davemck](#) (talk | [contribs](#)) **m** . . (8,760 bytes) **(−8)** . . (*renumber duplicate parms*) ([undo](#))
- [\(cur | prev\)](#) [01:26, 15 July 2022](#) [FederalBacon](#) (talk | [contribs](#)) . . (8,768 bytes) **(+965)** . . (*Reverted 1 edit by 138.88.198.177*) ([talk](#)) ([undo](#)) (*Tags: Twinkle, Undo*)
- [\(cur | prev\)](#) [01:23, 15 July 2022](#) [138.88.198.177](#) ([talk](#)) . . (7,803 bytes) **(−965)** . . ([undo](#)) (*Tag: Reverted*)
- [\(cur | prev\)](#) [00:46, 15 July 2022](#) [FederalBacon](#) (talk | [contribs](#)) **m** . . (8,768 bytes) **(+1)** . . (*Reverted edits by 166.48.126.45*) ([talk](#)) ([HG](#)) (3.4.10)) ([undo](#)) (*Tag: Rollback*)
- [\(cur | prev\)](#) [00:45, 15 July 2022](#) [166.48.126.45](#) ([talk](#)) . . (8,767 bytes) **(−1)** . . ([undo](#)) (*Tag: Reverted*)
- [\(cur | prev\)](#) [00:13, 15 July 2022](#) [Airborne84](#) (talk | [contribs](#)) . . (8,768 bytes) **(+992)** . . (*Added infobox galaxy cluster -- others will have to fill in the more scientific fields listed in the template*) ([undo](#))
- [\(cur | prev\)](#) [11:23, 14 July 2022](#) [Trappist the monk](#) (talk | [contribs](#)) **m** . . (7,776 bytes) **(−10)** . . (*cite repair;*) ([undo](#))
- [\(cur | prev\)](#) [09:13, 14 July 2022](#) [Yahya](#) (talk | [contribs](#)) . . (7,786 bytes) **(−23)** . . (*Undid revision 1098125910 by 2402:E280:2192:228:2495:EB27:FB26:5A05*) ([talk](#)), *unnecessary*) ([undo](#)) (*Tags: Undo, Mobile edit, Mobile web Advanced mobile edit*)
- [\(cur | prev\)](#) [08:57, 14 July 2022](#) [2402:e280:2192:228:2495:eb27:fb26:5a05](#) ([talk](#)) . . (7,809 bytes) **(+23)** . . (→*Background*) ([undo](#)) (*Tags: Reverted, Visual edit*)
- [\(cur | prev\)](#) [03:42, 14 July 2022](#) [Randy Kryn](#) (talk | [contribs](#)) . . (7,786 bytes) **(−25)** . . (→*Scientific results: removed unneeded descriptor portion*) ([undo](#))
- [\(cur | prev\)](#) [03:33, 14 July 2022](#) [MagicatthemovieS](#) (talk | [contribs](#)) . . (7,811 bytes) **(+24)** . . (→*References*) ([undo](#))
- [\(cur | prev\)](#) [23:47, 13 July 2022](#) [Somedifferentstuff](#) (talk | [contribs](#)) . . (7,787 bytes) **(+8)** . . (*Wikilinks*) ([undo](#))
- [\(cur | prev\)](#) [22:45, 13 July 2022](#) [MagicatthemovieS](#) (talk | [contribs](#)) . . (7,779 bytes) **(−24)** . . (→*References: I removed a redundant category*) ([undo](#))
- [\(cur | prev\)](#) [20:14, 13 July 2022](#) [97.118.73.28](#) ([talk](#)) . . (7,803 bytes) **(+1)** . . (*(minor) link brackets fix*) ([undo](#))
- [\(cur | prev\)](#) [20:03, 13 July 2022](#) [RScheiber](#) (talk | [contribs](#)) . . (7,802 bytes) **(+7)** . . (*WLS*) ([undo](#))
- [\(cur | prev\)](#) [19:33, 13 July 2022](#) [2603:6010:f02:fe53:e4c8:9972:9267:a7ab](#) ([talk](#)) . . (7,795 bytes) **(0)** . . (→*Significance*) ([undo](#)) (*Tag: Manual revert*)



# Transparency is core to Wikipedia

Revision as of 16:24, 15 July 2022 (edit)

Banana Republic (talk | contribs)

(Add an explanation for the very prominent feature of the 6 rays around light sources)

← Previous edit

Revision as of 16:28, 15 July 2022 (edit) (undo)

Xanzzibar (talk | contribs)

m (→Explanation for the 6 rays emanating from bright sources of light: **MOS:NUMERAL**)

Next edit →

Line 48:

```
The combined mass of the galaxy cluster acted as a [[gravitational lens]], magnifying much more distant galaxies behind it. Webb's NIRCam brought the distant galaxies into sharp focus, revealing tiny, faint structures that had never been seen before, including [[star cluster]]s and diffuse features.<ref name=":0" />
```

- == Explanation for the 6 rays emanating from bright sources of light ==

```
[[File:JWST-HST-primary-mirrors.svg|thumb|To first order, the Webb telescope's mirror rim is a hexagon, rather than the typical round rim]]
```

- The feature in the photo of six rays emanating from bright light sources is a result of [[diffraction]] from the mirror's edge, which is to first order a hexagon (per figure, the exact shape of the rim has additional minor cut-outs on top of the basic hexagonal shape). A telescope typically has a mirror/lens that is circular in shape. In such cases, the diffraction from the rim would create circular rings around the light source. The hexagonal rim of the Webb telescope's mirror gives rise to the 6 rays instead of rings of lights.<ref>{{cite web|title=Wondering About the 6 Rays Coming out of JWST's Test Image? Here's why They Happen|date=March 19, 2022|url=https://www.universetoday.com/155062/wondering-about-the-6-rays-coming-out-of-jwsts-test-image-heres-why-they-happen/|publisher=[[Universe Today]]|first=Matt|last=Williams}}</ref>

== Significance ==

Line 48:

```
The combined mass of the galaxy cluster acted as a [[gravitational lens]], magnifying much more distant galaxies behind it. Webb's NIRCam brought the distant galaxies into sharp focus, revealing tiny, faint structures that had never been seen before, including [[star cluster]]s and diffuse features.<ref name=":0" />
```

+ == Explanation for the six rays emanating from bright sources of light ==

```
[[File:JWST-HST-primary-mirrors.svg|thumb|To first order, the Webb telescope's mirror rim is a hexagon, rather than the typical round rim]]
```

+ The feature in the photo of six rays emanating from bright light sources is a result of [[diffraction]] from the mirror's edge, which is to first order a hexagon (per figure, the exact shape of the rim has additional minor cut-outs on top of the basic hexagonal shape). A telescope typically has a mirror/lens that is circular in shape. In such cases, the diffraction from the rim would create circular rings around the light source. The hexagonal rim of the Webb telescope's mirror gives rise to the six rays instead of rings of lights.<ref>{{cite web|title=Wondering About the 6 Rays Coming out of JWST's Test Image? Here's why They Happen|date=March 19, 2022|url=https://www.universetoday.com/155062/wondering-about-the-6-rays-coming-out-of-jwsts-test-image-heres-why-they-happen/|publisher=[[Universe Today]]|first=Matt|last=Williams}}</ref>

== Significance ==





# Transparency is core to Wikipedia



## Webb's First Deep Field

Coordinates: 07° 23′ 19.5″, −73° 27′ 15.6″﻿ / ﻿7.38861° 23′ 19.5″, −73.45439° 27′ 15.6″﻿ / 7.38861; -73.45439

From Wikipedia, the free encyclopedia

**Webb's First Deep Field** is the first [operational image](#) taken by the [James Webb Space Telescope](#). The [deep-field](#) photograph, covering a tiny area of sky visible from the [Southern Hemisphere](#), depicts [SMACS 0723](#), a [galaxy cluster](#) 4.6 billion [light-years](#) from Earth in the [constellation](#) of [Volans](#). Captured by the telescope's [Near-Infrared Camera](#) (NIRCam), the image was revealed to the public by [NASA](#) on 11 July 2022. Thousands of [galaxies](#) are visible in the image, which is the highest-resolution image of the [early universe](#) ever taken.

### Contents [hide]

- [Background](#)
- [Scientific results](#)
- [Explanation for the six rays emanating from bright sources of light](#)
- [Significance](#)
- [See also](#)
- [References](#)

## Background [ edit ]

The [James Webb Space Telescope](#) is a [space telescope](#) operated by [NASA](#) and designed primarily to conduct [infrared astronomy](#). Launched in December 2021, the spacecraft has been in a [halo orbit](#) around the second Sun–Earth [Lagrange point](#) (L<sub>2</sub>), about 1.5 million kilometres (900,000 mi) from Earth, since January 2022. At L<sub>2</sub>, the gravitational pulls of both the Sun and Earth keep the telescope's motion around the Sun synchronized with Earth's.<sup>[1]</sup>

*Webb's First Deep Field* was taken by the telescope's [Near-Infrared Camera](#) (NIRCam) and is a composite produced from images at different [wavelengths](#), totaling 12.5 hours.<sup>[2][3]</sup> The photo achieved depths at infrared wavelengths beyond the [Hubble Space Telescope's](#) deepest fields, which took weeks.

### Webb's First Deep Field



*Webb's First Deep Field*

#### Observation data (Epoch )

<b>Constellation(s)</b>	Galaxy cluster SMACS 0723
<b>Distance (co-moving)</b>	4.6 billion light years

See also: [Galaxy group](#), [Galaxy cluster](#), [List of galaxy groups and clusters](#)



# Transparency is core to Wikipedia

Where in the sky  [ [edit](#) ]

---

Just for reference, could we get the RA/Dec coordinates of this part of the sky? [Denelson83](#) 02:29, 12 July 2022 (UTC)  [ [reply](#) ]

User:[Denelson83](#) sorry it's not letting me reply to you directly in this talk page for some odd reason

anyway, you may know better than i do if this is the right information *at all*, but according to the NASA/IPAC Extragalactic Database (if i'm reading this right), the coordinates are 284.993575, -23.708009  [1]@ [Ayydoc](#) (talk) 02:48, 12 July 2022 (UTC)  [ [reply](#) ]

"As it appeared 4.6b years ago"  [ [edit](#) ]

---

Does this picture not depict the galaxy at different times, and different stages of evolution (if the scales involved are big enough?) due to the varying distances the light has traveled? Does anyone have accurate estimations of the ages? [102.219.79.151](#) (talk) 08:45, 12 July 2022 (UTC)  [ [reply](#) ]

Came here to make this (or a similar) point about this phrase: It is inaccurate and unworthy of us, as the image actually shows how that section of the sky appeared six months ago (or whenever exactly it was recorded). The light left its sources 4+ billion years ago (and *much* longer ago in some cases ~ the point of the whole exercise), but in now way can we say that is how it was nor how it appeared then. Happy days ~ [LindsayHello](#) 09:25, 12 July 2022 (UTC)  [ [reply](#) ]

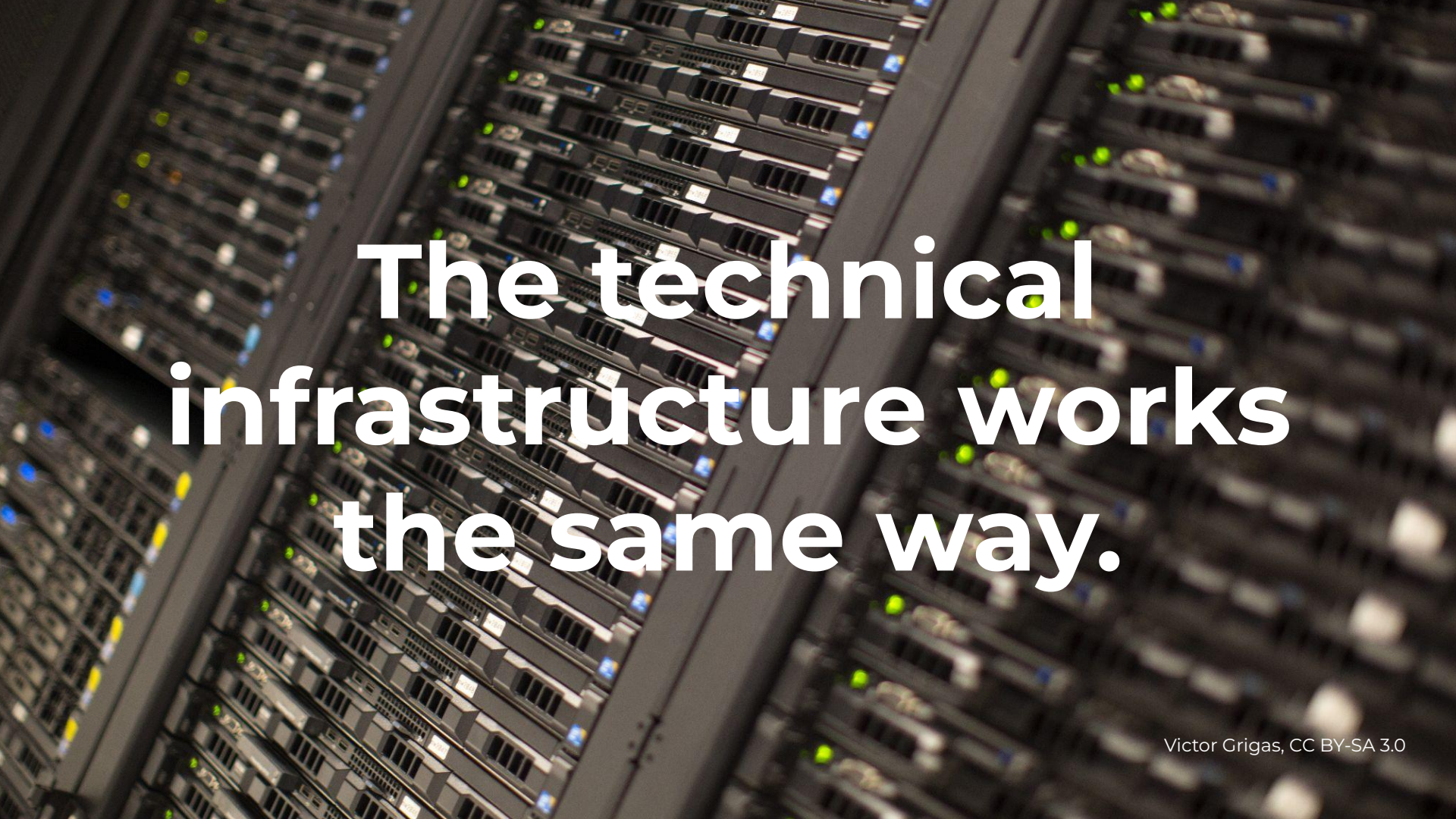
We're not really talking about how that section of the sky looked like, but how a "think" looked like, that is, it's actual appearance. However I agree there's a bit of ambiguity here as some objects might be younger/older depending on the distance, and not all objects are uniformly far away at the same distance. [BeŽet](#) (talk) 13:01, 12 July 2022 (UTC)  [ [reply](#) ]

Given that the light travels (generally) unimpeded through space -- what do you think would make it change in the intervening 4.6 billion years? That is the entire point. Think of it as a snapshot that is traveling through space -- the snapshot will not change. The snapshot eventually reaches you. Although it is correct to say that the snapshot reached you (reached the telescope) 6 months ago or whenever it was received -- it is ALSO TRUE to say that the snapshot is an unaltered glimpse of what was there 4.6 billion years ago. If the light had been impeded, blocked, changed, absorbed and re-emitted, distorted, deflected, etc. the image (snapshot) would be very, very different. So -- think of the primordial light as a non-changing snapshot (a Polaroid! LOL) traveling through space ... taking 4.6 billion years to get to you. [Chesspride216.144.161.51](#) (talk) 18:36, 12 July 2022 (UTC)  [ [reply](#) ]

I've already updated the article for [SMACS\\_J0723.3-7327](#) because of this. This galaxy cluster, which was the central target for this image is located at a redshift of  $z=0.390$ <sup>[1]</sup>. First of all this gives us a lookback time of ~4.35 Gyr (not 4.6 Gyr as it's stated on NASA's website (<https://www.nasa.gov/image-feature/goddard/2022/nasa-s-webb-delivers-deepest-infrared-image-of-universe-yet@>), but it could be easily an error made by the article writer). Also lots of articles over the internet cites this as "this cluster is located 4.6 billion light-years from Earth", which is even worse, since the  $z=0.390$  redshift gives us a proper distance of ~5.12 Glyr. (All these values were obtained using the cosmology determined by the Planck 2018 results<sup>[2]</sup>.) [Masterdesky](#) (talk) 13:38, 12 July 2022 (UTC)  [ [reply](#) ]

These kinds of small adjustments are needed, of course, but the general point remains. [Chesspride216.144.161.51](#) (talk) 18:38, 12 July 2022 (UTC)  [ [reply](#) ]





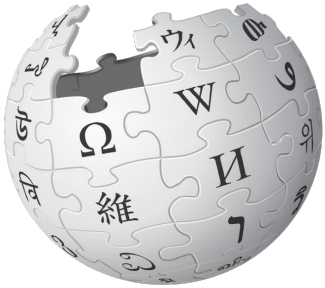
**The technical  
infrastructure works  
the same way.**

# Technical infrastructure is

- 7th most visited website
- Maintained by volunteers and staff
- Available under free licenses, mostly GNU GPL
- Developed openly in a collaborative manner



# Naming things is hard



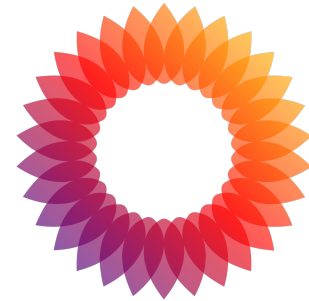
**WIKIPEDIA**

encyclopedia



**WIKIMEDIA**

a social movement  
also a non-profit  
(runs the servers)



**MediaWiki**

FOSS wiki software





[\[dismiss\]](#)

[See the current version of this page on Wikipedia.](#)

[Printable version](#) | [Disclaimers](#) | [Privacy policy](#) | [Latest revision](#)

**Welcome to Wikipedia**, a collaborative project to produce a complete encyclopedia from scratch. We started in January 2001 and already have **over 19,000 articles**. We want to make over 100,000, so let's get to work--*anyone* can edit any article--copyedit, expand an article, write a little, write a lot. See the [Wikipedia FAQ](#) for information on how to edit pages and other questions.

The content of Wikipedia is covered by the [GNU Free Documentation License](#), which means that it is free and will remain so forever. See [open content](#) and [free content](#) for background.

**Current Events and Breaking News**

Encyclopedia articles about topics behind the news.  
[2001 U.S. Attack on Afghanistan](#) - [More current events](#)

**Philosophy, Mathematics, and Natural Science**

[Astronomy and Astrophysics](#) - [Biology](#) - [Chemistry](#) - [Earth Sciences](#) - [Mathematics](#) - [Philosophy](#) - [Physics](#) - [Statistics](#)

**Social Sciences**

[Anthropology](#) - [Archaeology](#) - [Economics](#) - [Geography](#) - [History](#) - [History of Science and Technology](#) - [Language](#) - [Linguistics](#) - [Parapsychology](#) - [Political Science](#) - [Psychology](#) - [Sociology](#)

**Applied Arts and Sciences**

[Agriculture](#) - [Architecture](#) - [Business and Industry](#) - [Communication](#) - [Computer Science](#) - [Education](#) - [Engineering](#) - [Family and Consumer Science](#) - [Health Sciences](#) - [Law](#) - [Library and Information Science](#) - [Public Affairs](#) - [Technology](#) - [Transport](#)

**Culture**

[Classics](#) - [Cooking](#) - [Critical Theory](#) - [Dance](#) - [Entertainment](#) - [Film](#) - [Games](#) - [Hobbies](#) - [Literature](#) - [Music](#) - [Opera](#) - [Painting](#) - [Recreation](#) - [Religion](#) - [Sculpture](#) - [Sports](#) - [Theater](#) - [Tourism](#) - [Visual Arts and Design](#)

**Other Category Schemes**

[About Wikipedia category schemes](#) - [Complete list of encyclopedia topics](#) - [Library of Congress catalog scheme](#) - [Dewey Decimal System](#) - [Wikipedia arranged by topic](#) - [Historical timeline](#) - [Historical anniversaries](#) - [Reference tables](#) - [Biographical Listing](#) - [Current events](#)

**Wikipedias in other languages**

[About the non-English Wikipedias](#) - [Arabic \(Araby\)](#) - [Catalan \(Català\)](#) - [Chinese \(Hanyu\)](#) - [Danish \(Dansk\)](#) - [Dutch \(Nederlands\)](#) - [German \(Deutsch\)](#) - [Esperanto](#) - [French \(Français\)](#) - [Hebrew \(Ivrit\)](#) - [Hungarian \(Magyar\)](#) - [Italian \(Italiano\)](#) - [Japanese \(Nihongo\)](#) - [Polish \(Polska\)](#) - [Portuguese \(Português\)](#) - [Russian \(Russkiy\)](#) - [Simplified English](#) - [Spanish \(Castellano\)](#) - [Swedish \(Svenska\)](#)

**Wikipedia**

**About the project:** [Announcements](#) - [How to edit a page](#) - [FAQ](#) - [Policy](#) - [Feature requests](#) - [Bug reports](#) - [Mailing list](#) - [Wikipedia and Nupedia](#) - [Wikipedians](#) - [Wikipedia Chat](#)  
**Article pointers:** [Article news](#) - [New topics](#) - [Brilliant prose](#) - [Receive articles daily by e-mail](#) - [Go to a random page](#) - [Index of all pages](#)

**Content requests:** [Requested articles](#) - [Help desk](#)

**Useful resources for us:** [Wikipedia utilities](#) - [Tarballs of Wikipedia data](#) - [Public domain resources](#) - [Commentary](#), which is being transferred to [meta.wikipedia.com](#) - [SandBox](#)

**Selected press coverage:** *The New York Times* -- [December 9, 2001](#) *The New York Times* -- [September 20, 2001](#) - MIT's *Technology Review* (and [a letter](#))

**Links:** [Friends of Wikipedia](#)

Today is Thursday, [December 20, 2001](#), servertime (U.S. Pacific Time)

/talk

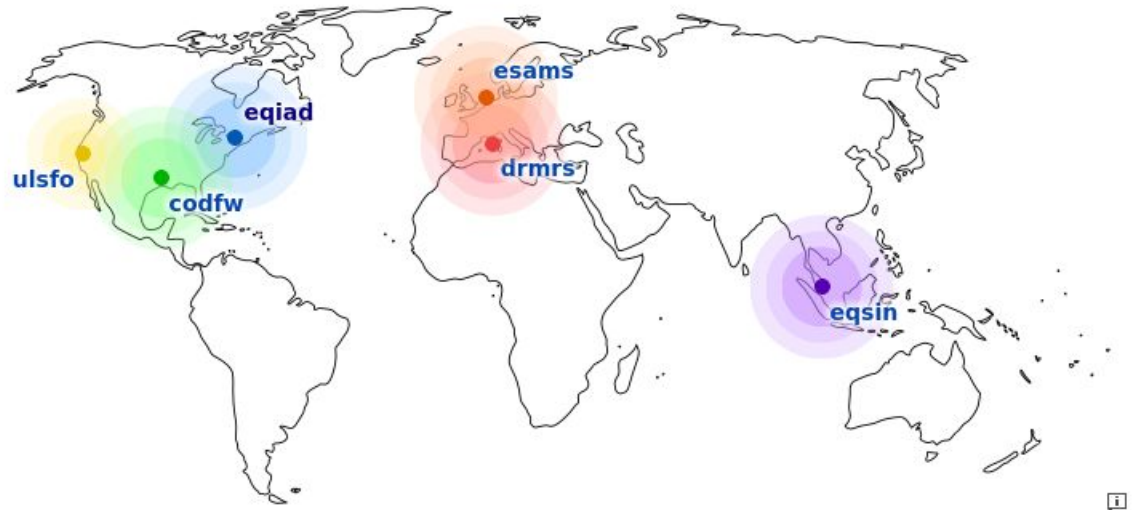
Note: Unless you have the administrator password, you cannot currently edit this page. While this is unfortunate, it has turned out to be necessary to prevent vandalism of this page, which has occurred on several occasions.

# Brief history

- 2001: Starts as a side project of Bomis, a different approach from Nupedia. Servers were in San Diego, CA, but volunteers got access.
- 2003: Wikimedia Foundation split off as a non-profit
- 2004: Servers move to Tampa, FL. First backup taken because of hurricane threats.

# Datcenters

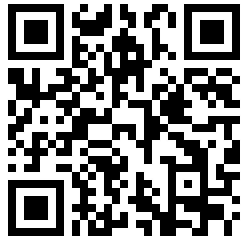
Map of Wikimedia Foundation **data centers**.



- 2 “core” datacenters that run MediaWiki and services
- 4 caching PoPs

Present

DC ↕	Location	Usage	Since ↕
<a href="#">Eqiad</a>	Ashburn, Virginia, United States (North America)	Core services	2010
<a href="#">Codfw</a>	Carrollton, Texas, United States (North America)	Core services	2014
<a href="#">Esams</a>	Amsterdam, Netherlands (Europe)	Edge caching	2009 <sup>[1]</sup>
<a href="#">Ulsfo</a>	San Francisco, California, United States (North America)	Edge caching	2014 <sup>[2]</sup>
<a href="#">Eqsin</a>	Singapore (Asia)	Edge caching	2017 <sup>[3]</sup>
<a href="#">Drmrs</a>	Marseille, France (Europe)	Edge caching	2022 <sup>[4]</sup>



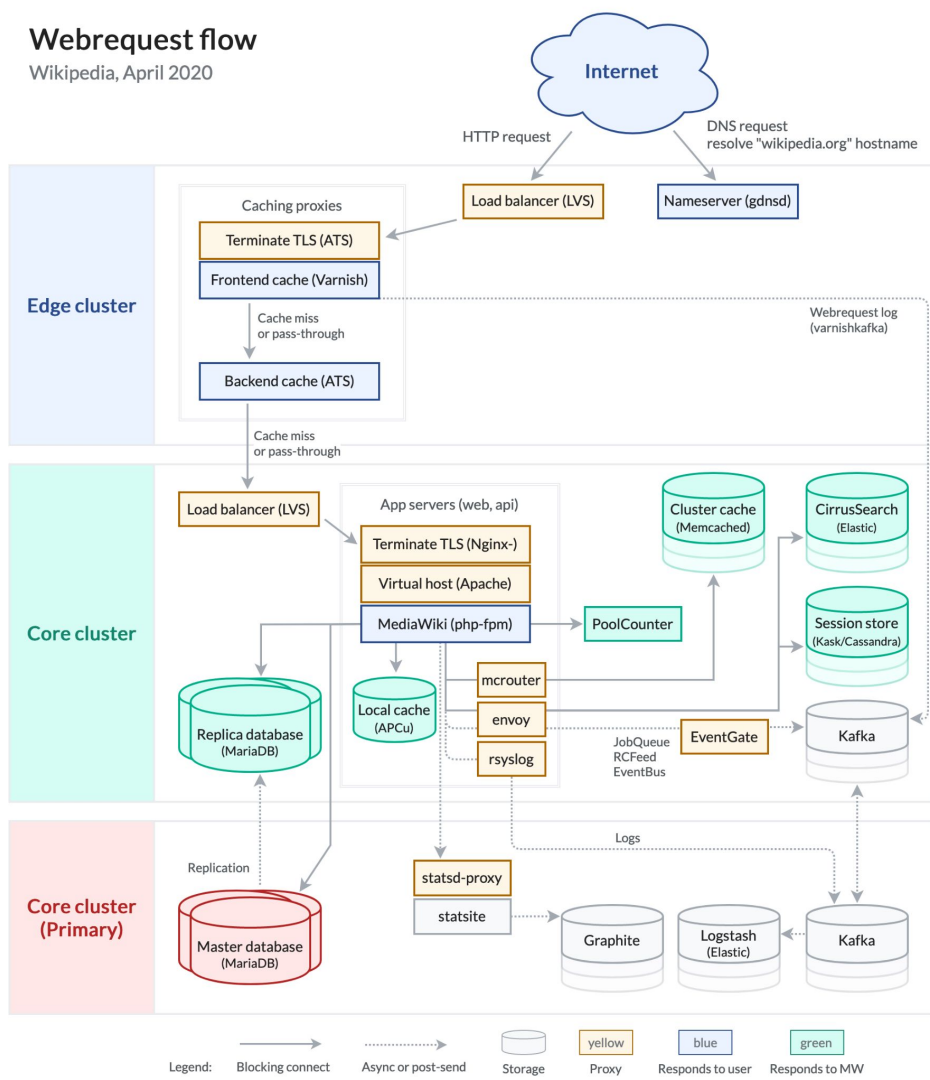
# How it works

- Cache:
  - Apache Traffic Server
  - Varnish
- Appservers:
  - Apache httpd
  - PHP-FPM
- Storage:
  - MariaDB
  - Cassandra
  - Kafka
  - OpenStack Swift
  - ElasticOpenSearch



## Webrequest flow

Wikipedia, April 2020





# Today's landscape

- Code: [gerrit.wikimedia.org](https://gerrit.wikimedia.org) (mirrored on GitHub)
- Metrics/stats: [grafana.wikimedia.org](https://grafana.wikimedia.org)
- Bugs: [phabricator.wikimedia.org](https://phabricator.wikimedia.org)
- Documentation: [wikitech.wikimedia.org](https://wikitech.wikimedia.org) and [www.mediawiki.org](https://www.mediawiki.org) and [doc.wikimedia.org](https://doc.wikimedia.org)



# Gerrit, Git and code

- Gerrit is a code review platform for Git
- Last 90 days: 12,964 patches, 367 different authors
- All patches to MediaWiki code must be approved by someone with merge aka “+2” rights. Code is deployed in weekly “deployment trains”.
- Servers are declaratively maintained using Puppet. Code is deployed immediately.



# Who submits MediaWiki patches?

Over the past 30 days:

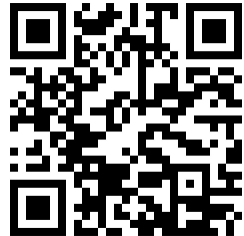
1. 145 patches by Sam Reed (staff)
2. 36 patches by Daniel Kinzler (staff)
3. 33 patches by Umherirrender (volunteer)
4. 32 patches by Zabe (volunteer)
5. 28 patches by Matěj Suchánek (volunteer)
6. 27 patches by Timo Tijhof (staff)
7. 25 patches by DannyS712 (volunteer)
8. 25 patches by MarkAHershberger (volunteer)
9. 18 patches by Derick Alangi (staff)
10. 18 patches by Ladsgroup (staff)



# Who approves MediaWiki patches?

Over the past 30 days:

1. 47 patches by DannyS712 (volunteer)
2. 44 patches by Krinkle (staff)
3. 23 patches by Zabe (volunteer)
4. 23 patches by Tim Starling (staff)
5. 22 patches by Umherirrender (volunteer)
6. 21 patches by Jforrester (staff)



# Puppet

- Declaratively states what should be installed or running on a server (similar to Ansible, etc.)
- Effectively is root access, so limited to mostly Site Reliability Engineers (SRE) and a few volunteers.
- Private repository contains passwords and secret keys
- Others can test puppet patches in virtual machines



# Puppet

[gerrit.wikimedia.org](https://gerrit.wikimedia.org/) / [operations](#) / [puppet](#) / [HEAD](#) / [.](#) / [modules](#) / [profile](#) / [manifests](#) / [mediawiki](#) / [webserver.pp](#)


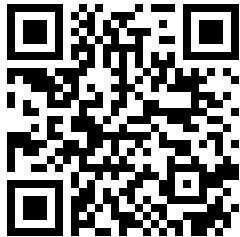
blob: 89d2589bf4b6c569fbc137e81f68f70514d5ad0a [\[file\]](#) [\[log\]](#) [\[blame\]](#)

```
1 class profile::mediawiki::webserver(  
2     Boolean $has_lvs = lookup('has_lvs'),  
3     Boolean $has_tls = lookup('profile::mediawiki::webserver::has_tls'),  
4     Boolean $stream_to_logstash = lookup('profile::mediawiki::webserver::stream_to_logstash', {'default_value' => false}),  
5     Optional[Stdlib::Port::User] $fcgi_port = lookup('profile::php_fpm::fcgi_port', {'default_value' => undef}),  
6     String $fcgi_pool = lookup('profile::mediawiki::fcgi_pool', {'default_value' => 'www'}),  
7     Array[Wmflib::Php_version] $php_versions = lookup('profile::mediawiki::php::php_versions', {'default_value' => ['7.2']}),  
8     Mediawiki::Vhost_feature_flags $vhost_feature_flags = lookup('profile::mediawiki::vhost_feature_flags', {'default_value' => {})  
9     # Sites shared between different installations  
10    Array[Mediawiki::SiteCollection] $common_sites = lookup('mediawiki::common_sites'),  
11    # Installation/site dependent sites  
12    Array[Mediawiki::SiteCollection] $sites = lookup('mediawiki::sites'),  
13    Boolean $install_fonts = lookup('profile::mediawiki::webserver::install_fonts', {'default_value' => false}),  
14 ) {  
15     include ::profile::mediawiki::httpd  
16     $versioned_port = php::fpm::versioned_port($fcgi_port, $php_versions)  
17     $fcgi_proxies = $php_versions.map |$idx, $version| {  
18         $fcgi_pool_name = $idx? {  
19             0 => $fcgi_pool,  
20             default => "${fcgi_pool}-${version}"  
21         }  
22         $default = ($idx == 0)  
23         $retval = [$version, mediawiki::fcgi_endpoint($versioned_port[$version], $fcgi_pool_name, $default)]  
24     }  
25  
26     # Declare the proxies explicitly with retry=0  
27     httpd::conf { 'fcgi_proxies':  
28         ensure => present,  
29         content => template('mediawiki/apache/fcgi_proxies.conf.erb')  
30     }  
31  
32     # we may not need fonts anymore! (T294378)  
33     $font_ensure = $install_fonts.bool2str('installed', 'absent')  
34     class { '::mediawiki::packages::fonts':  
35         ensure => $font_ensure  
36     }  
37 }
```



# Cloud Services

- Wikimedia Cloud Services provides computing resources (OpenStack) for volunteers and staff
- Can apply same puppet roles for testing and debugging without private user data
- “Beta cluster” replica of wikis to catch integration issues



Main page  
Current events  
Contents  
Random article  
About Wikipedia  
Contact page  
Donate to Wikipedia  
Wikipedia store

Interaction

Help  
Community portal  
Recent changes  
Upload file

Tools

What links here  
Related changes  
Upload file  
Special pages  
Permanent link  
Page information  
Get shortened URL  
Wikidata item

Print/export

Download as PDF  
Download as TXT  
Printable version

In other projects

Wikimedia Commons

★ Meta-Wiki  
Wikibooks  
Wikidata  
Wikinews  
Wikiquote  
Wikisource

Not logged in | Talk | Contributions | Create account | Log in

Main Page | Discussion | Read | View source | View history | Search Wikipedia

Test banner two of three (CentralNotice) [X]

This is a message from MediaWiki:Sitentice. [hide]

Hello, and welcome to the WMF integration environment known as Beta Cluster. The software running this wiki is automatically updated whenever there is a change to the software (currently running 1.39.0-alpha (b07027f) ). It is used for testing purposes by Wikimedia developers. This wiki is configured to be like en.wikipedia.org. For more information, see [Beta cluster Meta-Wiki](#) and [wikitech:Beta Cluster](#).

**From today's featured test articles**

For language testing try [Polar bear](#).


**Did you know ...**

Our release team is rad.

**In the news**

- [Deployment calendar](#)
- [open unbreak news](#)

**Today's featured picture**



Goats are very cute.

**Other areas of Wikipedia**

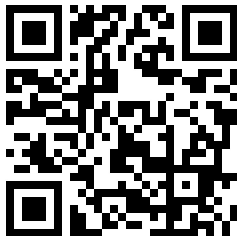
- **Community portal** - Bulletin board, projects, resources and activities covering a wide range of Wikipedia areas.
- **Village pump** - Ask questions and discuss technical issues and policies about Beta Wikipedia.

**Wikipedia's sister projects**

Wikipedia is hosted by the [Wikimedia Foundation](#), a non-profit organization that also hosts a range of other projects:

# DB replicas

- Redacted replicas of DBs are available to Cloud Services
- Quarry is a web tool to run SQL queries (like phpMyAdmin)
- Even traditionally “non-technical” users learn SQL via copy-paste-modify



Quarry Home New Query Recent Queries Discuss Database tables Login

This query has been published by Pppery.

enwiki

SQL [Toggle Highlighting](#)

```
select concat("Module:",page_title),page_len from page where page_namespace=828
and not /**/ exists (select 1 from templatelinks where tl_namespace=page_namespace and tl_title=page_title
and tl_from namespace != 2 and tl_from_namespace != 10)
and page_title not like "%/sandbox"
and page_title not like "User:%"
and page_title not like "Sandbox/%"
and page_title not like "Unicode_data/names/%"
and page_content_model!="wikitext"
order by page_id desc;
```

All SQL code is licensed under [CC0 License](#).

Query status: **complete**

Executed in 13.41 seconds as of Sat, 26 Feb 2022 18:38:51 UTC.

Resultset (321 rows) [Download data](#)

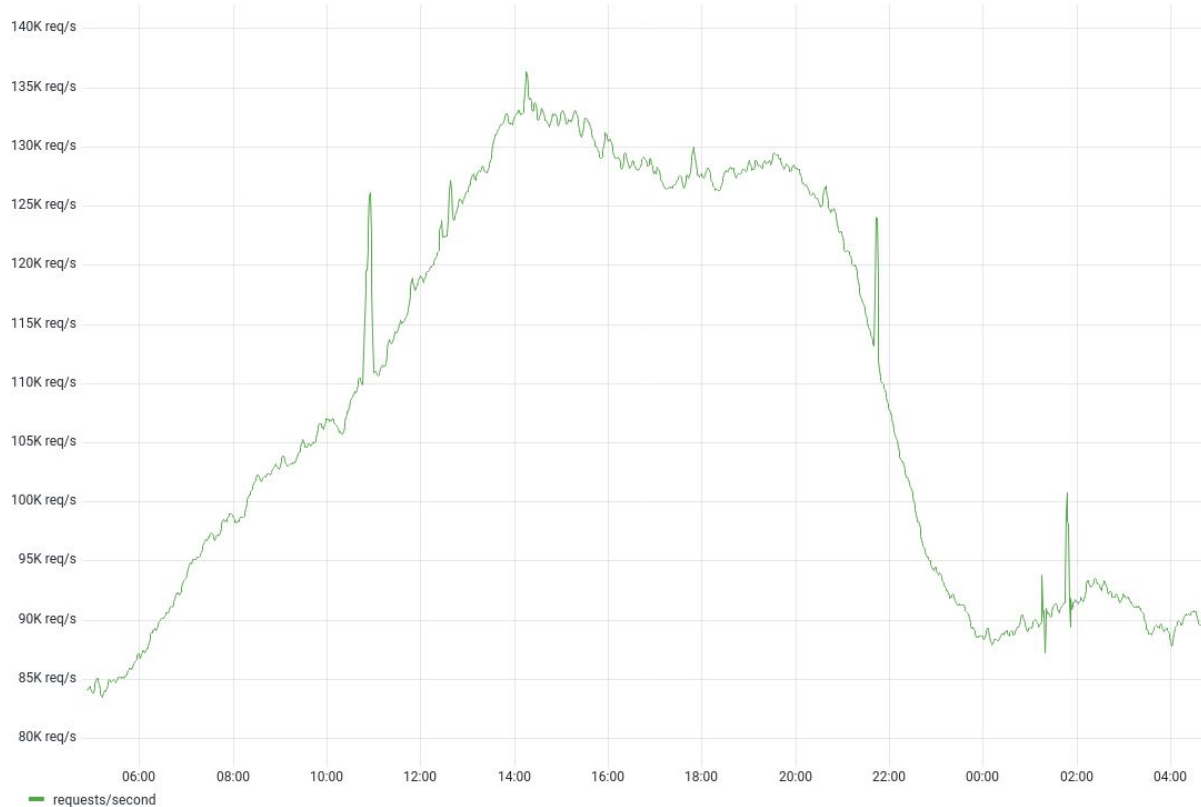
Show  entries Search:

concat("Module:",page_title)	page_len
Module:Ranking_movements/styles.css	155
Module:Ranking_movements	5338
Module:ACmplNotes	808
Module:Deletion_sorting	1396
Module:Testrand	297
Module:Talk_quote_inline	891
Module:IATA_and_ICAO_code/data/sandbox3	840273
Module:Article_list/testcases	0
Module:Navbar/div	16068

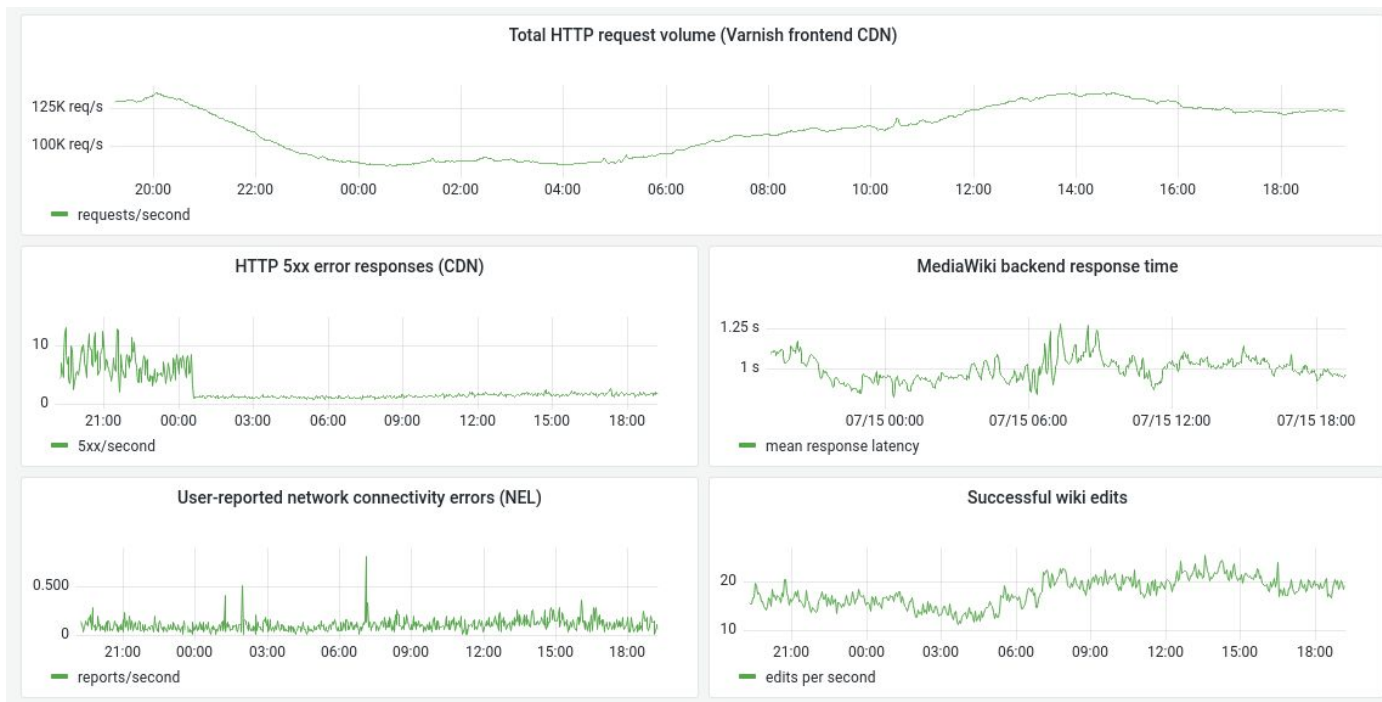


# Open statistics: 130k+ req/s

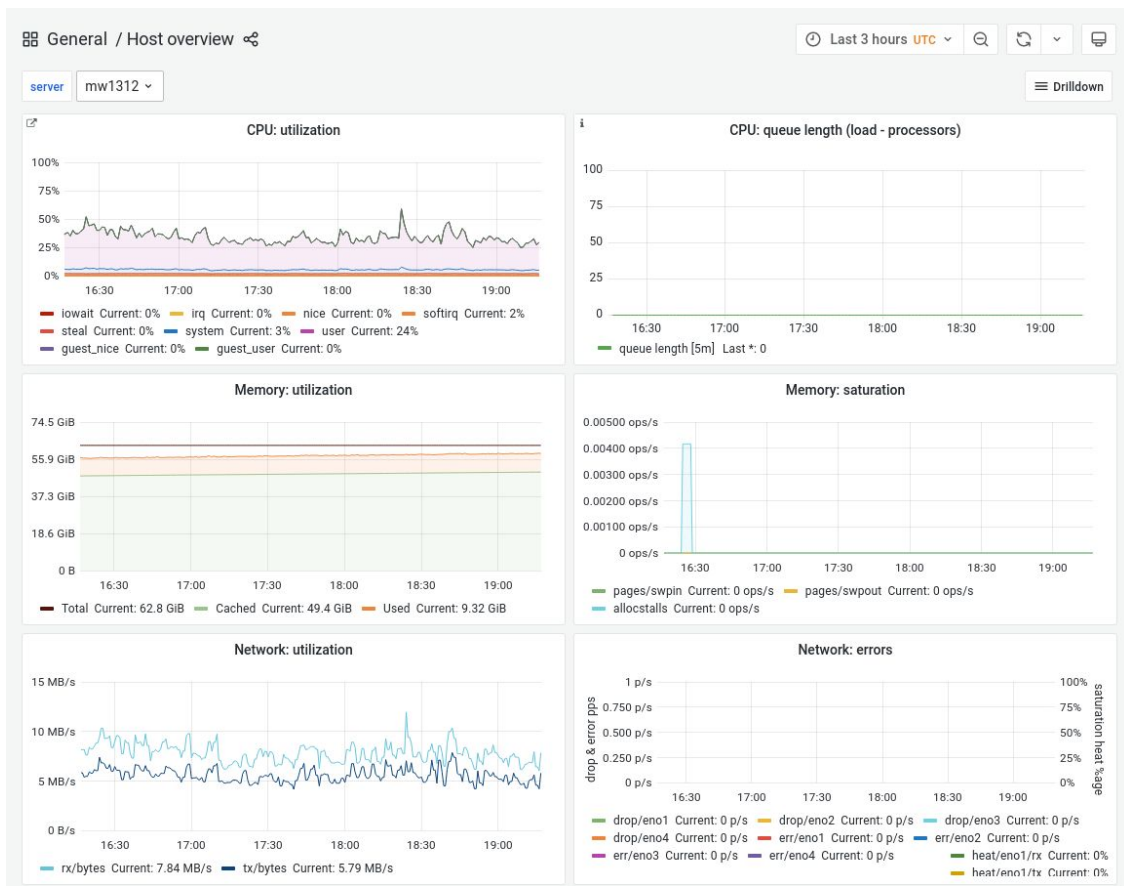
Total HTTP request volume (Varnish frontend CDN)



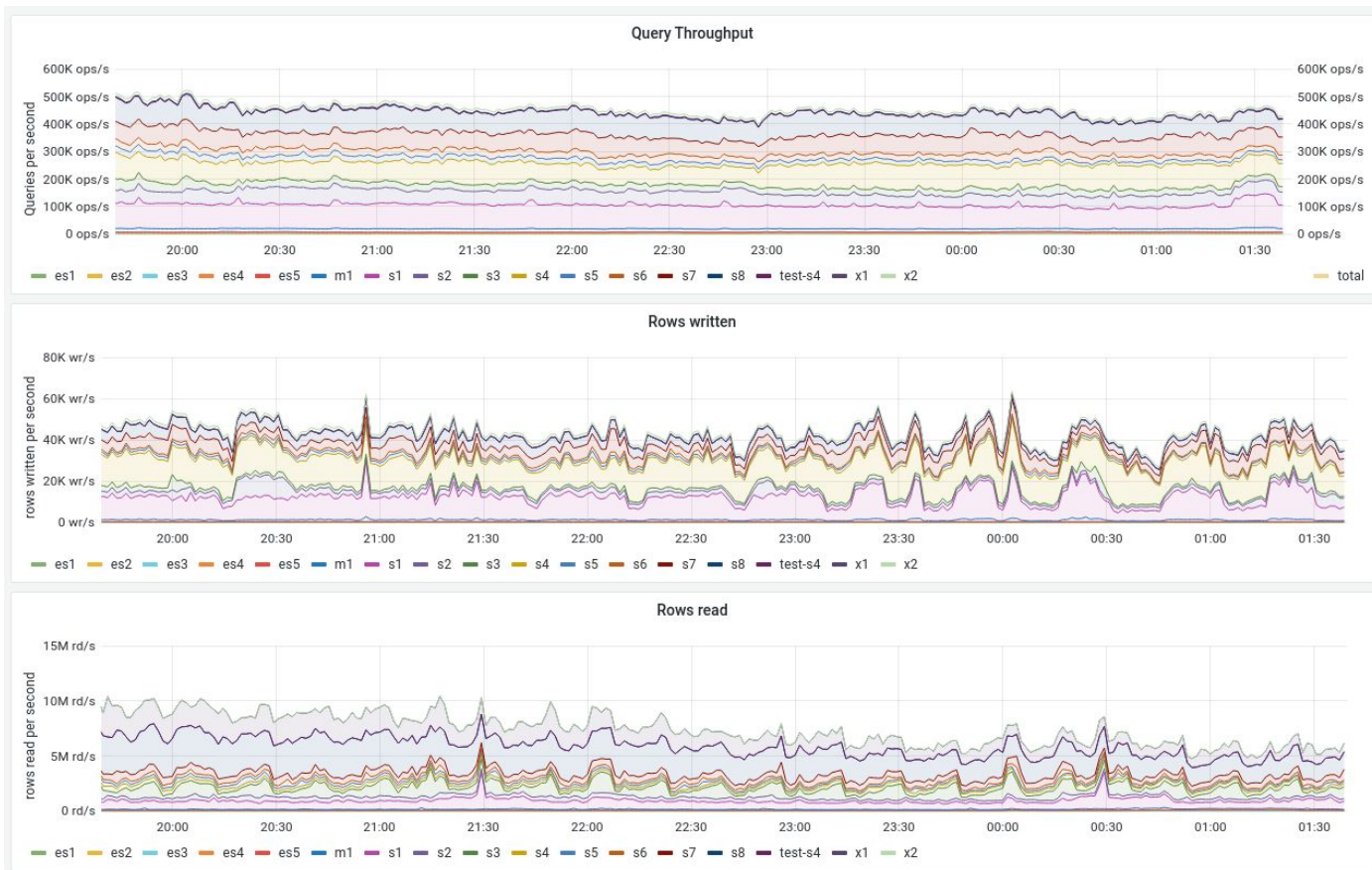
# Open statistics



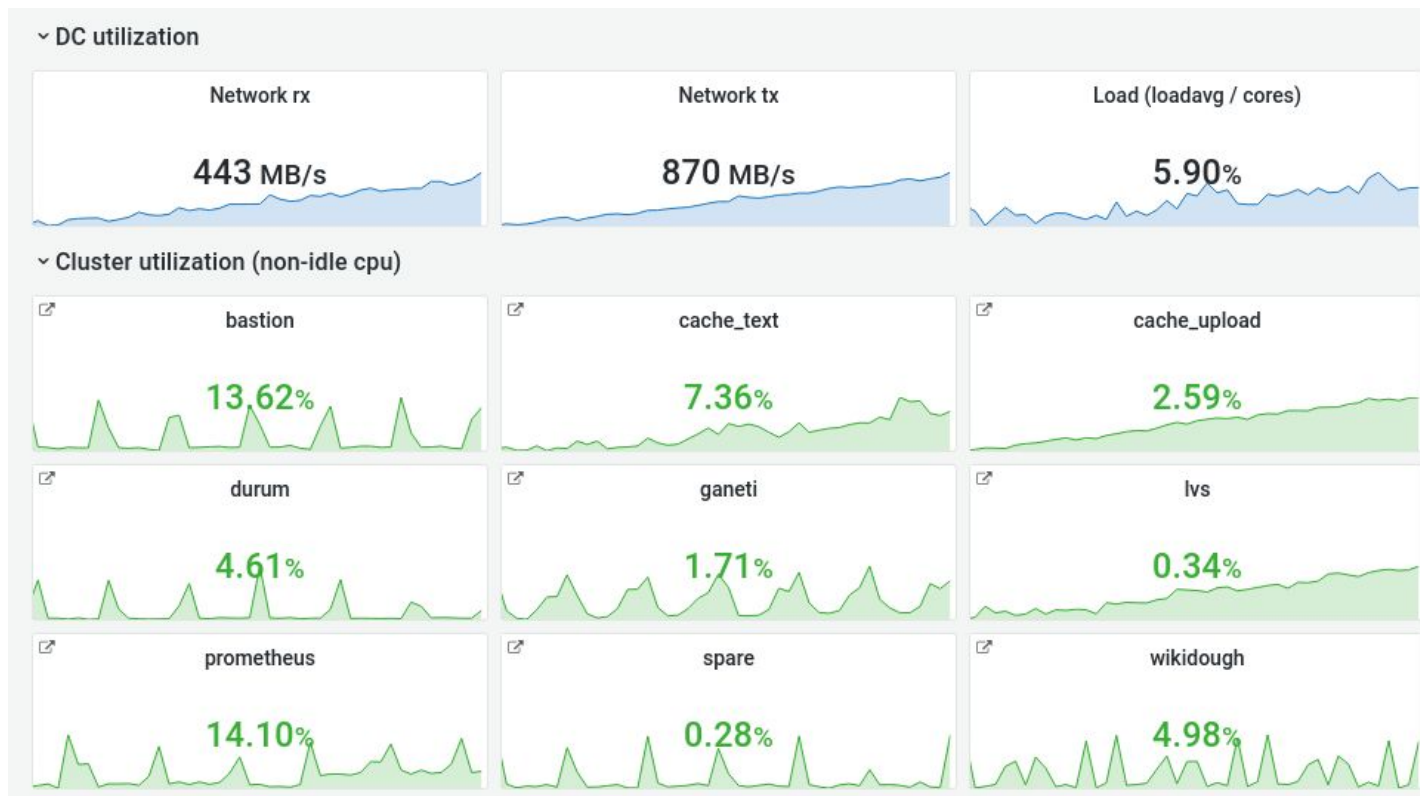
# Even for individual servers



# Or for databases



# Or entire datacenters



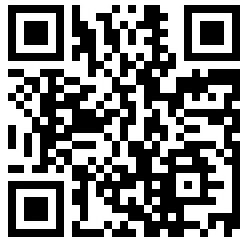
# Public communications

- Mailing lists: wikitech-l on lists.wikimedia.org
- Phabricator, Gerrit comments
- IRC channels on Libera Chat: #wikimedia-operations, #wikimedia-sre, #wikimedia-tech, etc. (Biggest channels are bridged to Matrix)



# Phabricator transparency

“The problem: mw times out on a database transaction e.g. on writing stashed files to swift. Turns out that files are being written from eqiad to codfw at about ~2Mbps, thus a ~600MB file can hit the 300s mw timeout...”



## ✔ Jobrunner timeouts on cross-DC file uploads because of HTTP/2

✔ Closed, Resolved    Public

### Description

This is a followup investigation of [T274589: No atomic section is open \[got LocalFile:lockingTransaction\]](#) to be more focused on the specific "slow PUTs" problem.

The problem: mw times out on a database transaction e.g. on writing stashed files to swift. Turns out that files are being written from eqiad to codfw at about ~2Mbps, thus a ~600MB file can hit the 300s mw timeout (more context at [↓ https://phabricator.wikimedia.org/T274589#6850912](https://phabricator.wikimedia.org/T274589#6850912)).

One example of this problem is the following upload from mw1305 (a jobrunner), which takes 6s in eqiad and 300s in codfw:

```
Feb 23 10:52:40 ms-fe1007 proxy-server: 10.64.16.105 10.64.32.220 23/Feb/2021:10/52/40 PUT /v1/AUTH_mw/wikipedia-commons-local-public.9d/9/9d/TheLostWorld1925.webm HTTP/1.0 201 - wikimedia/multi-http-client%20v1.0 AUTH tk6fb0dd3ce... 624059195 - 3a6f10dc474ed8113e1498a5751bb075 tx4a56556fd5e44e01bab13-006034de71 - 7.8468 - - 1614077553.124433041 1614077560.971237898 0
```

```
Feb 23 10:58:13 ms-fe2007 proxy-server: 10.64.16.105 10.192.32.155 23/Feb/2021:10/58/13 PUT /v1/AUTH_mw/wikipedia-commons-local-public.9d/9/9d/TheLostWorld1925.webm HTTP/1.0 201 - wikimedia/multi-http-client%20v1.0 AUTH tk5c39f2dc9... 624059195 - 3a6f10dc474ed8113e1498a5751bb075 tx3a392c4173db430aa9e05-006034de82 - 323.8699 - - 1614077570.121001959 1614077893.990805019 0
```

I've talked on centrallog hosts the average transfer time in codfw for commons PUTs larger than 300MB, to get a better idea of the time frames involved (note that the day boundaries are not exact, log files are rotated at ~6 UTC)

```
zcat ms-fe2*/swift.log-2021$(date).gz | awk '$16 > 300000000 && /PUT /v1/AUTH_mw/wikipedia-commons-local-public/ {print $21 }'
```

```
jan 01 7.72208
jan 02 8.493843333333333
jan 03 7.75212173913044
...
jan 23 7.75591935483871
jan 24 6.15061379310345
jan 25 9.28589795918367
jan 26 99.3627666666667
jan 27 23.3935756097561
jan 28 9.05734482758621
jan 29 28.5930095238095
jan 30 149.271148
jan 31 70.0038066666667
feb 01 15.1718
feb 02 156.918878571429
Feb 03 74.7627030461520
```

During the Feb 11th time window we were, among other things, in the process of decom'ing swift codfw hosts ([T272837](#)) and thus pushing more weight to existing hosts. The codfw swift cluster is now in steady state, in the sense that there are no planned com/decom.

The timeframe also coincided with Buster upgrades for mw hosts, and indeed the reported slow jobrunners are all the Buster ones. See also

[↓ https://phabricator.wikimedia.org/T275752#6864889](https://phabricator.wikimedia.org/T275752#6864889) for further context.

✎ Edit Task

↓ Edit Related Tasks... ▶

🔍 Edit Related Objects... ▶

🔔 Subscribe

🔕 Mute Notifications

🔒 Protect as security issue

🏆 Award Token

📌 Flag For Later

Assigned To

👤 Legoktm

Authored By

👤 fgiunchedi  
Feb 25 2021, 11:20 AM

Tags

🏷️ SRE (Backlog)

🏷️ serviceops (Incoming 🚩)

🏷️ Sustainability (Tag)

🏷️ Patch-For-Review

📌 MW-1.38-notes (1.38.0-wmf.6; 2021-10-26)

Referenced Files

📄 F34715054: 2021-10-29 12-55.png  
Oct 29 2021, 9:59 AM

Subscribers

👤 Aklanter

# Phabricator transparency

“I think I narrowed it down. If I upload using plain CLI curl, it finishes instantaneously:...”

“Now when I use a stripped down version of SwiftFileBackend/MultiHttpClient, I get:”



L Legoktm added a comment. Oct 29 2021, 5:12 AM

I think I narrowed it down. If I upload using plain CLI curl, it finishes instantaneously:

```
P17633 (An Untitled Masterwork)
1 legoktm@mw1305:~$ time curl -v -X PUT --data-binary @legoktm_test_T275752.djvu -H 'x-content-dimensions:' -H 'content-type: image/vnd.djvu' -H 'x-auth-token: [redacted]' -H 'etag: 069fc83e88bc112f309f2548c386853c' -H 'x-object-meta-shalbase36: 56e0liu04c2u9avcrvlz7m5k4nh6gxf' --http1.0 https://ms-fe.svc.codfw.wmnet/v1/AUTH_mw/wikipedia-test-local-temp/legoktm_test_T275752_2.djvu
2 * Expire in 0 ms for 6 (transfer 0x5650c5041fb0)
3 * Expire in 1 ms for 1 (transfer 0x5650c5041fb0)
4 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
5 * Expire in 1 ms for 1 (transfer 0x5650c5041fb0)
6 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
7 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
8 * Expire in 1 ms for 1 (transfer 0x5650c5041fb0)
9 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
10 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
11 * Expire in 1 ms for 1 (transfer 0x5650c5041fb0)
12 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
13 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
14 * Expire in 1 ms for 1 (transfer 0x5650c5041fb0)
15 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
16 * Expire in 0 ms for 1 (transfer 0x5650c5041fb0)
17 * Expire in 1 ms for 1 (transfer 0x5650c5041fb0)
```

Now when I use a stripped down version of SwiftFileBackend/MultiHttpClient, I get:

```
P17632 (An Untitled Masterwork)
1 legoktm@mw1305:~$ time php test.php legoktm_test_T275752.djvu
2 array(2) {
3   [0]=>
4     string(8) "test.php"
5   [1]=>
6     string(25) "legoktm_test_T275752.djvu"
7 }
8 Uploading to https://ms-fe.svc.codfw.wmnet/v1/AUTH_mw/wikipedia-test-local-temp/legoktm_test_T275752.djvu
9 array(7) {
```



# Phabricator transparency

**L** Legoktm added a comment. Oct 29 2021, 6:14 AM

If I use PHP's stream wrappers, literally:

```
$opts = [ 'http' => [
    'method' => 'PUT',
    'header' => $realHeaders,
    'content' => $contents,
] ];

$ctx = stream_context_create($opts);
$fh = fopen( $url, "r", false, $ctx );
$result = stream_get_meta_data( $fh );
```

Where `$contents` is the file read into memory, then the same exact upload to codfw takes ~1s.

**L** Legoktm added a comment. Oct 29 2021, 6:16 AM

Also note that during the buster upgrade we did move from curl 7.52.1 to 7.64.0, so it could also be a regression from that. I haven't looked at the libcurl changelog yet.

**L** Legoktm added a comment. Oct 29 2021, 6:40 AM

I tried a few different variations on the PHP curl script, none of which made a difference:

- Using `curl_setopt( $ch, CURLOPT_POSTFIELDS, $contents );` instead of `CURLOPT_INFILE/CURLOPT_READFUNCTION`
- Writing the file to `php://memory` which is supposed to be a in-memory stream

The `stream_context_create` solution only works for files that fit under the memory limit, which is currently ~600MB.



# Phabricator transparency

“So @Legoktm hypothesis seems overall correct - but it's not the syscalls that cause the slowness, it's the 820 round trips between datacenters because we're sending such small chunks.”



Joe added a comment.

Oct 29 2021, 9:09 AM

So, what I think we know at this point is:

- The problem is completely within php / curl; curl from the command line or pretty much any other client behaves correctly.
- PHP's curl usage will write/read responses in 65k chunks, so there seems to be a back-forth of request/response partial data across the wire
- for the test file, I see 820 such req/resp partial stages (the fread in lego's paste)
- If we multiply 820 by 31.5 ms (the rt between mw1305 and ms-fe.svc.codfw.wmnet) we get 25.8 seconds... which is more or less the overhead we're seeing

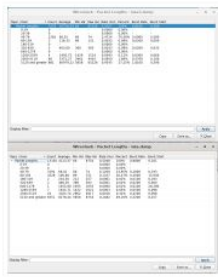
Now I would just need to get a full dump of the tcp communications between client and server to confirm what exactly is happening - I suppose it's waiting for "100-continue" messages from the server.

So @Legoktm hypothesis seems overall correct - but it's not the syscalls that cause the slowness, it's the 820 round trips between datacenters because we're sending such small chunks.



akosiaris added a comment.

Oct 29 2021, 9:59 AM



Adding a statistical analysis of packet lengths in wireshark from 2 captures. Upper one, with 3323 packets in total is standard curl call, lower one, with 13326 packets is test.php.

# Phabricator transparency

“But then, as Lego's output above shows, cmdline curl is using HTTP/1.0 and libcurl from PHP is using HTTP/2...Can we force cmdline curl to HTTP/2, or PHP libcurl to HTTP/1.1, to test that?”



Xover added a comment.

Oct 29 2021, 10:12 AM

I don't think curl sends `Expect: 100-continue` for chunked transfers to begin with, and I don't think chunks need to be ack'ed before sending the next in HTTP/1.1.

But then, as Lego's output above shows, cmdline curl is using HTTP/1.0 and libcurl from PHP is using HTTP/2. And in HTTP/2 chunked transfer encoding isn't supported, replaced with some new streaming and multiplexing stuff that I've never looked into but which sounds suspiciously like reimplementing TCP on top of HTTP. In particular, I wouldn't be at all surprised to find that HTTP/2 has what amounts to a TCP ACK but without SACK.

Can we force cmdline curl to HTTP/2, or PHP libcurl to HTTP/1.1, to test that?

In any case, with ~15ms latency any synchronous ack is going to kill performance with small chunks (http level "packets"), and upping the size of the chunks massively the best bet for improving throughput.



akosiaris added a comment.

Oct 29 2021, 10:24 AM



*Can we force cmdline curl to HTTP/2, or PHP libcurl to HTTP/1.1, to test that?*

Right on target! Thanks for noticing that, that's the issue. HTTP2 shouldn't really be used in internal scenarios, it was never meant for that.

setting `curl_setopt( $ch, CURLOPT_HTTP_VERSION, CURL_HTTP_VERSION_1_1)`; in the code fixes it and conversely passing `--http2` to the curl call reproduces it.



Joe added a comment.

Oct 29 2021, 10:27 AM

Yes, thanks for noticing @Xover ; a posteriori, it's pretty obvious what is going on and it's interesting how much more inefficient using http2 is in this case (not surprising, once you think about it, but still unexpected, at least for me).

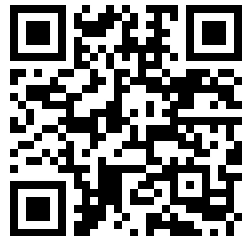
# Phabricator transparency

- Verbosely pasting, publicly, allowed a volunteer with no server access to participate, figuring out the issue.
- It is likely that SREs (staff) would've figured out the real cause eventually, but it would've taken more time.
- Further reading:  
[https://wikitech.wikimedia.org/wiki/Incidents/2021-11-04\\_large\\_file\\_upload\\_timeouts](https://wikitech.wikimedia.org/wiki/Incidents/2021-11-04_large_file_upload_timeouts)



# IRC channels

<a href="#">#wikimedia-mobileconnect</a>	en	All things related to mobile devices within the Wikimedia projects <a href="#">Publicly logged</a> .
<a href="#">#wikimedia-multimediaconnect</a>	en	Wikimedia Foundation <a href="#">Multimedia</a> team. <a href="#">Publicly logged</a>
<a href="#">#wikimedia-operationsconnect</a>	en	Operation of the <a href="#">Wikimedia servers</a> . Used by <a href="#">sysadmins</a> to discuss and coordinate server administration. Please do <b>not</b> report technical problems here, but in <a href="#">#wikimedia-techconnect</a> instead. <a href="#">Publicly logged</a> .
<a href="#">#wikimedia-observabilityconnect</a>	en	Monitoring, alerting, logging and all things relating to the observability of <a href="#">Wikimedia servers</a> , covering Prometheus, Grafana, ELK, Icinga, VictorOps, etc.
<a href="#">#wikimedia-perfconnect</a>	en	Wikimedia Foundation <a href="#">Wikimedia Performance Team</a>
<a href="#">#wikimedia-perf-botsconnect</a>	en	Wikimedia Foundation <a href="#">Wikimedia Performance Team Bot Channel</a>
<a href="#">#wikimedia-qaconnect</a>	en	Formerly Wikimedia Foundation <a href="#">Quality Assurance</a> team. <a href="#">Publicly logged</a> . Used to redirect to <a href="#">#wikimedia-relengconnect</a> .
<a href="#">#wikimedia-qteconnect</a>	en	Wikimedia Foundation <a href="#">Quality and Test Engineering Team</a> channel.
<a href="#">#wikimedia-quibbleconnect</a>	en, fr, de	Quibble a test runner for MediaWiki - <a href="https://doc.wikimedia.org/quibble/">https://doc.wikimedia.org/quibble/</a>
<a href="#">#wikimedia-readingconnect</a>	en	Wikimedia Foundation <a href="#">Reading</a> team
<a href="#">#wikimedia-relengconnect</a>	en	Wikimedia Foundation <a href="#">Release Engineering Team</a> . <a href="#">Publicly logged</a> .
<a href="#">#wikimedia-researchconnect</a>	en	Wikimedia research channel used by the <a href="#">Wikimedia Research and Data</a> team. Ancillary channel: <a href="#">#wikimedia-labs2connect</a> (R:L2).
<a href="#">#wikimedia-rustconnect</a>	en	<a href="#">Wikimedia Rust developers user group</a>
<a href="#">#wikimedia-searchconnect</a>	en	Wikimedia Foundation <a href="#">Search Platform Team</a> .
<a href="#">#wikimedia-serviceopsconnect</a>	en	<a href="#">Service Operations</a> . <a href="#">Publicly logged</a> .
<a href="#">#wikimedia-servicesconnect</a>	en	Wikimedia Foundation <a href="#">Services team</a>
<a href="#">#wikimedia-teampracticesconnect</a>	en	<a href="#">Team Practices Group</a> channel for anyone to discuss engineering/product development practices, processes, philosophy, theory, etc, as it pertains to Wikimedia and beyond. Also used for <a href="#">Team Practices Group</a> team coordination/conversation. <a href="#">Publicly logged</a> .
<a href="#">#wikimedia-tool-devsconnect</a>	en	<a href="#">Wikimedia Tool Developers User Group</a>
<a href="#">#wikimedia-toolforge-standards-committeecconnect</a>	en	[private] <a href="#">Toolforge standards committee</a>
<a href="#">#wikimedia-trafficconnect</a>	en	Traffic Operations on <a href="#">Wikimedia servers</a> , covering Varnish, Apache, networking, DNS, domains, HTTPS, etc.
<a href="#">#wikimedia-dev-africacconnect</a>	en	<a href="#">Africa Wikimedia Developers</a> project to on-board new/old developers into the Wikimedia movement from Africa.



# IRC channels

- #wikimedia-operations: Main coordination channel
- #wikimedia-sre: Discussion amongst SREs
- #wikimedia-tech: End-user support and help
- Private security channel for dealing with DDoS attacks or active security issues



# IRC transparency

15:07:17 <taavi> jeena: jnuche: hey, can we rollback group1 due to T313432? commons uploading interfaces (Special:Upload and Special:UploadWizard) are completely broken at least for me

15:07:18 <+stashbot> T313432: Error: Call to a member function getConfig() on null - <https://phabricator.wikimedia.org/T313432>

15:07:47 <jeena> okay, I'll roll back

15:08:05 <taavi> I'm looking if I can find any obvious causes for that

15:13:06 <taavi> found the cause, let's see if I can fix it easily

15:13:26 <+logmsgbot> !log jhuneidi@deploy1002 rebuilt and synchronized wikiversions files: Revert "group[0|1] wikis to [VERSION]"



# Server Admin Log

- Log for actions that aren't reflected elsewhere like Git
- Deployment tools automatically send logs
- !log in IRC ends up on a wiki page, Mastodon, Twitter
- Wiki archives go back to June 2004





# IRC transparency

23:18:08 <+icinga-wm> PROBLEM - BGP status on cr2-eqiad is CRITICAL: BGP CRITICAL - AS64601/IPv6: Connect - kubernetes-eqiad, AS64601/IPv6: Connect - kubernetes-eqiad, AS64601/IPv4: Connect - kubernetes-eqiad, AS64601/IPv4: Connect - kubernetes-eqiad, AS64605/IPv4: Connect - Anycast, AS64605/IPv6: Active - Anycast

[https://wikitech.wikimedia.org/wiki/Network\\_monitoring%23BGP\\_status](https://wikitech.wikimedia.org/wiki/Network_monitoring%23BGP_status)

23:18:16 <+icinga-wm> PROBLEM - Host db1146 #page is DOWN: PING CRITICAL - Packet loss = 100%

23:18:18 <+icinga-wm> PROBLEM - Host wtp1039 is DOWN: PING CRITICAL - Packet loss = 100%

23:18:33 <+icinga-wm> PROBLEM - Host pc1013 #page is DOWN: PING CRITICAL - Packet loss = 100%

23:19:04 <+icinga-wm> PROBLEM - Host mwdebug1001 is DOWN: PING CRITICAL - Packet loss = 100%

23:19:06 <+icinga-wm> PROBLEM - Host matomo1002 is DOWN: PING CRITICAL - Packet loss = 100%

23:19:06 <+icinga-wm> PROBLEM - Host logstash1025 is DOWN: PING CRITICAL - Packet loss = 100%

23:19:10 <+icinga-wm> PROBLEM - Host ml-etcd1002 is DOWN: PING CRITICAL - Packet loss = 100%

23:19:10 <+icinga-wm> PROBLEM - Host kubetcd1004 is DOWN: PING CRITICAL - Packet loss = 100%

23:19:15 <@rzl> wuh oh



# Private information

- Wikipedia has very little private information compared to most websites! (Excluding donor data, which is stored separately)
- Volunteers need to sign an NDA to get access to debug logs, web request logs, slow SQL queries, security tickets, etc.
- Bar has been lowered over time, previously needed C-level signoff, now just any WMF employee can vouch.
- Currently exactly 100 users in the “nda” group.



# Server access



CC-BY 2.0: Tim Avatar Bartel

- Volunteers need to sign an NDA and go through deployment training, must be approved by WMF Release Engineering team
- About 7 volunteers with MediaWiki deployment access
- Only 2 volunteers with root, both former staff



# Transparency is hard

- It is a constant fight to keep things transparent.
- Open defaults really help with keeping things public.
  - But people will still trend toward closed platforms (e.g. Slack).
- Some things must be private, like legal advice.
- People can be intimidated by having to do everything publicly! It often runs counter to our value of privacy.
- Every slip up, mistake, silly comment will be public and archived forever. Do you really want that?



# What can you do?

- Other public interest websites should try to do the same.
- Start gradually! Documentation is the easiest place to start
  - Wikis are the best (of course), other people will document things for you
  - It's OK to lose control
- Publish your server configuration, just like people publish their dotfiles
- Track issues publicly, people might start to pitch in
  - Have an “easy” or “good first task” category
- Keep up the fight for transparency!



# What can you do?

- Wikimedia is always looking for more contributors!
- Lurk in IRC/Matrix, maybe something will catch your attention
- How to become a MediaWiki hacker (first QR code)
- Wikimedia Developer Portal (second QR code)



CC-BY-SA 3.0: Supercarwaar

# Thank you

- Slides: <https://w.wiki/5VRS>
- Email: [legoktm@debian.org](mailto:legoktm@debian.org)
- IRC: legoktm on Libera Chat
- Matrix: [@legoktm:matrix.org](https://matrix.org/@legoktm:matrix.org)



# ~~Security through obscurity~~

## Installed software

Product	Version
MediaWiki	1.39.0-wmf.21 (0950074) 23:03, 18 July 2022
PHP	7.2.34-18+0~20210223.60+debian10~1.gbp21322+wmf5 (fpm-fcgi)
MariaDB	10.4.25-MariaDB-log
ICU	63.1
Pygments	2.10.0
LilyPond	2.22.0
Elasticsearch	6.8.23
LuaSandbox	3.0.3
Lua	5.1.5

## Entry point URLs

Entry point	URL
Article path	<a href="#">/wiki/\$1</a>
Script path	<a href="#">/w</a>
index.php	<a href="#">/w/index.php</a>
api.php	<a href="#">/w/api.php</a>
rest.php	<a href="#">/w/rest.php</a>

## Installed skins

Skin	Version	License	Description	Authors
<b>Cologne Blue</b>	- (3450f1d) 06:17, 18 July 2022	GPL-2.0-or-later	A lightweight skin with minimal formatting	Lee Daniel Crocker and others

