

# Wikimedia case study

According to its own Wikipedia entry: the site attracted “at least 684 million visitors yearly by 2008” with more than 75,000 active contributors working on ten million articles in more than 250 languages

## Background

Most people with a passing acquaintance with a browser or Google search know Wikipedia, the web-based encyclopedia spanning topics from the ridiculous to the sublime. Want Britney Spear’s bio? It’s there. Want a quick briefing in quantum mechanics? It too, is there.

Type the name of a subject into any online search engine and the first hit will typically be its Wikipedia entry. The site’s streamlined interface and quick paragraphs make it easy on the eyes and easy to navigate

## Business Challenge

Behind the scenes, the challenge “is getting a major web site on the air—or web—and running it without too much money and not too many resources in terms of people,” said Brion Vibber, CTO for the Wikimedia Foundation, the non-profit organization behind the online encyclopedia.

Wikipedia is famously dependent on thousands of contributors who update, edit or add entries as needed. That means many points of entry. According to its own Wikipedia entry: the site attracted “at least 684 million visitors yearly by 2008” with more than 75,000 active contributors working on ten million articles in more than 250 languages.

## Ubuntu Solution

Wikipedia has been Linux-focused from the beginning. The effort started seven years ago running various varieties of Red Hat, mostly because that’s what the other hosting servers were using at the time, according to Vibber. The glow faded over the years after three or four versions of Fedora running on a couple of architectures “and we couldn’t figure out what was going on in terms of [Red Hat] packaging,” he said.

So, when Wikipedia’s infrastructure gurus wanted to standardize on one platform, they started looking around. Vibber said, “We looked at the possibilities including [Red Hat] Fedora, but Fedora moves a little too fast and we were not too happy about some of the configuration management features.” In addition, Red Hat Enterprise was not quite totally free, while some of the clones were free but were also harder to manage.





In fact, the whole idea that Canonical backstops the Ubuntu distribution with well organized updates and patches was a big draw for Wikimedia. For a workload this large, the predictability and stability of updates and patches is critical.

Meanwhile, several Wikimedia administrators favored Debian flavors of Linux— especially Canonical Ubuntu. Canonical’s organized and well timed updates helped as well. As did the level of support the company offered.

In fact, the whole idea that Canonical backstops the Ubuntu distribution with well organized updates and patches was a big draw for Wikimedia. For a workload this large, the predictability and stability of updates and patches is critical.

Wikimedia started the transition in 2006 with the Ubuntu 6.04 release. The organization now runs Ubuntu 8.04 and “will stick with that as long as possible depending on the server,” said Amsterdam-based Wikipedia network administrator, Mark Bergsma.

While some of the older hardware still runs Fedora, all of the more recently added servers run Ubuntu—as will additional servers coming online. The new machines run custom versions of Ubuntu, including Wikipedia’s own custom packages for applications and software configuration. Wikipedia is a poster child for the Linux-Apache-MySQL-PHP or LAMP stack. It uses Squid proxy servers and the Subversion open-source code repository for version tracking and a Bugzilla open-access bug tracking system.

Wikipedia runs at least 350 servers, mostly Dell 1U and 2U boxes mostly at three data centers—in Tampa, Fla., Amsterdam and South Korea, said Bergsma.

The emphasis is on running industry-standard software and hardware across the board. There is nothing proprietary in the data center platform “except the switches and routers and of course all the BIOS in the servers,” Bergsma said. There is nothing in the stack from Microsoft although technically speaking some of the Cisco routers run proprietary software. He said, “We consider them to be appliances.”

Contributors also use IRC chat and mail to communicate.

The staff includes three or four coders, two in Wikimedia’s San Francisco headquarters, along with a handful system administrators and technicians.

And Ubuntu is making its presence felt on Wikipedia’s desktops as well. Several staff members, including the executive director, now run Ubuntu on the desktop. There is one remaining Windows PC to run QuickBooks. Wikimedia runs Ubuntu front to back.

The Wikimedia team lauds Canonical’s support and backing. They’re particularly happy with its bug fix and security patch methodology. They also like the year and a half of security support that comes as part and parcel of use of the distribution.

The open-source centric, techie-heavy organization’s emphasis is on self-maintenance and fixes. “We have a very small but very talented group of engineers all with different specialties,” said Vibber. “If we have a problem



usually the person with the most knowledge of that area will fix it.” There is great pride taken in submitting fixes to the open source community.

Wikimedia is also considering a service subscription to address software problems that might occur outside of Wikimedia’s core areas of expertise, Bergsma said.

## Result

Wikipedia’s reliance on open source LAMP stack mirrors its community focused existence. The organizations’ contributor/editor model mimics open source methodology. Ubuntu scales up to meet spikes in traffic and facilitates the contributor process: Wikipedia’s use of Ubuntu is a significant endorsement of the distribution as an enterprise foundation.

“Ours is an open source infrastructure that millions of people use and thousands of people work on , so it has to be reliable, scalable and secure. Ubuntu fits the bill,” Vibber said.

The site peaks at about 50,000 requests per second. Again, citing Wikipedia:Statistics, there are more than 2.5 million articles in English and tens of thousands of edits by hundreds of thousands of visitors.

