

**La arquitectura de software libre de wikipedia**  
**Por Manuel Dávila Sguerra**  
[mdavila@uniminuto.edu](mailto:mdavila@uniminuto.edu)

Una curiosidad que siempre tenemos quienes trabajamos con software libre se refiere a conocer qué arquitecturas tecnológicas utilizan las soluciones de gran impacto. En esta ocasión me referiré a Wikipedia la gran enciclopedia libre que ha transformado en mundo del conocimiento basado en un informe oficial de ellos del cual me tomé el trabajo de hacer este resumen.

Un informe oficial de Wikipedia en Abril del 2007 cuenta que fue fundada en Junio del 2003, solo tiene 9 empleados permanentes y muchos voluntarios con un presupuesto de dos millones de dólares que llegan a través de donaciones.

Si bien lo más conocido es la enciclopedia libre, en realidad tiene 8 proyectos más de los cuales hice hacer una recopilación para los que quieran investigar sobre ellos. Estos son:

Wikipedia, <http://en.wikipedia.org/> la enciclopedia libre  
Wiktionary <http://www.wiktionary.org/> un diccionario que maneja 150 idiomas  
Wikinews <http://www.wikinews.org/> un contenedor de noticias  
Wikibooks <http://es.wikibooks.org/wiki/Portada> un contenedor de libros libres  
Wikiquote [http://en.wikiquote.org/wiki/Main\\_Page](http://en.wikiquote.org/wiki/Main_Page) compendio de citas sobre gente y trabajos prominentes  
Wikisource [http://en.wikisource.org/wiki/Main\\_Page](http://en.wikisource.org/wiki/Main_Page) recopilación de textos originales  
Wikiversity <http://www.wikiversity.org/> materiales de aprendizaje (e-learning)  
Wikispecies [http://species.wikimedia.org/wiki/Main\\_Page](http://species.wikimedia.org/wiki/Main_Page) contenidos sobre especies  
Wikimedia [http://en.wikipedia.org/wiki/Wikimedia\\_Foundation](http://en.wikipedia.org/wiki/Wikimedia_Foundation) fundación para recolectar fondos

Wikipedia maneja en la actualidad 8 millones de artículos soportando cientos de idiomas, 110 millones de revisiones y según sus propias estadísticas dobla su tamaño cada 6 meses en versiones, tráfico y servidores.

En los momentos pico sus servidores web reciben 30.000 (treinta mil) solicitudes con un cúmulo de 2 giga bits de tráfico en datos, 350 (tres cientos cincuenta) servidores con 16 Giga Bytes de memoria principal en 3 centros de datos ubicados en Tampa, Amsterdam, Seul manejados por 6 personas.

Usa arquitectura LAMP [http://en.wikipedia.org/wiki/LAMP\\_%28software\\_bundle%29](http://en.wikipedia.org/wiki/LAMP_%28software_bundle%29) que es el acrónimo que indica la unión del sistema operacional Linux, el servidor web Apache, la base de datos Mysql y el lenguaje, php.

Los servidores se enlazan en 3 "cluster" ubicados en 3 continentes. El primario en Tampa Florida, el secundario para el manejo de la memoria cache en Amsterdam Holanday y en Seul Corea del sur los cuales necesitan hacer balance de carga geográfico basado en un "resolver" que es un programa que ayuda a resolver los IP's contra los nombre de los dominios y que forma parte del servicio del protocolo tcp/ip llamado el DNS.

El manejo del cache lo hace con el servidor de proxy llamado Squid sobre Linux en dos grupos, uno para el manejo de texto y otro para el manejo de imágenes con 55 servidores de cache que usan discos de 40GB con 4 por servidor y memoria principal de 8 gigas.

Usa servidores virtuales de Linux LVS para balanceo de carga  
[http://en.wikipedia.org/wiki/Linux\\_Virtual\\_Server](http://en.wikipedia.org/wiki/Linux_Virtual_Server)

Como plataforma para el manejo de los wikis <http://en.wikipedia.org/wiki/Wiki> usa Mediawiki <http://en.wikipedia.org/wiki/Mediawiki> que es un software escrito en php e instalado en 125 servidores con múltiples cpu's de a 8 por caja.

El almacenamiento usa 1.3 tera bytes, 4 millones de archivos usa un sistema de archivos nfs  
[http://en.wikipedia.org/wiki/Network\\_file\\_system](http://en.wikipedia.org/wiki/Network_file_system).

Naturalmente son muy cuidadosos y sacan backup cada 2 semanas.

# Arquitectura de la plataforma de wikipedia

