

# **Ensayo big-data-y-open-data-**

## **Tecnologías Avanzadas de la Web: Big data y Open Data.**

30 de enero de 2013

### **Introducción**

El presente ensayo tiene como objetivo realizar un breve análisis de la evolución de la web desde sus inicios hasta nuestros días, análisis necesario para abordar el tema de las Tecnologías Avanzadas de la Web haciendo especial énfasis en Big data y Open Data, tópicos que fueron abordados en el Seminario Isummit Loxa 2012, acentuando en como estas tecnologías pueden ser empleadas, los retos que presentan y algunos inconvenientes que se han detectado.

La Web a sufrido muchos cambios, desde la Web 1.0 que solo permitía que las personas sean observadoras de contenido, hacia la Web 2.0 que permite que los usuarios interactúen y colaboren entre sí en una comunidad virtual para generar conocimiento, fue un gran avance para la humanidad, pero luego se abrió paso la Web 3.0 que tiene como característica más importante la Web Semántica, incluso se habla de la Web 4.0 que apertura el camino hacia una red ubicua.

Junto con la evolución de la web, ingentes cantidades de datos se generan día a día en el planeta, es aquí donde nace el término “Big Data” que es visto como una oportunidad y reto para investigar en el volumen, velocidad y variedad de datos.

Por otra parte hay que destacar el momento histórico en el que nos encontramos, en el cual disponemos de tecnologías que nos permiten una mayor disponibilidad y acceso a datos científicos a nivel mundial, aumentando la colaboración e incrementando el ritmo y la profundidad del descubrimiento, aquí es donde surge el término “Open Data”, cuya filosofía fundamental es permitir que los datos estén disponibles para que todo el mundo tenga acceso a ellos.

Desarrollo Desde que Tim Berners-Lee desarrollo la web, esta ha ido sufriendo una constante evolución, así tenemos el inicio de la denominada Web 1.0 que era una de contenidos estáticos, es decir solo

Producto de esta constante evolución tenemos la Web 4.0 que trata de movernos hacia una Web Ubicua donde el objetivo será el de unir las inteligencias, para que tanto las personas como las cosas se comuniquen entre sí para la toma de decisiones (Quispe Ortega , 2011).



Imagen tomada del sifo consultora debian :  
<http://www.consultora-debian.net/inicio/noticias/52-de-la-web-10-a-la-web-40>

Por el inminente auge de la web y sus nuevas características, así como el desarrollo de la computación ubicua (Gonzales, 2003), y el mayor uso de internet y dispositivos móviles, los seres humanos estamos creando y almacenando información constantemente y cada vez más en cantidades astronómicas, es así que en el año 2012 alrededor de 2.5 quintillón ( $2.5 \times 10^{18}$ ) de bytes al día fueron creados, y se estima que este fenómeno crezca hasta el 60% anual. Tal es el crecimiento de la cantidad de datos que las medidas de procesamiento han ido desplazando su límite superior a lo largo del tiempo, llegando a contabilizarse en petabytes y zettabytes de datos.

WHAT'S A ZETTABYTE?

1 kilobyte	1,000	000 000 000 000 000 000
1 megabyte	1,000,000	000 000 000 000 000
1 gigabyte	1,000,000,000	000 000 000
1 terabyte	1,000,000,000,000	000 000
1 petabyte	1,000,000,000,000,000	000 000
1 exabyte	1,000,000,000,000,000,000	000
1 zettabyte	1,000,000,000,000,000,000,000	

Figura 2 Evolución en la medición de la cantidad de datos, diciembre 2012.

Imagen tomada del sitio paperblog:  
<http://es.paperblog.com/predicciones-para-los-proximos-100-anos-1611672/>

¿Pero de donde vienen todos esos datos?. Toda esta información se genera de: Social media, datos transaccionales, comportamientos del consumidor, multimedia, sensores y dispositivos embebidos, dispositivos de red, dispositivos móviles (iphones, tablets etc).

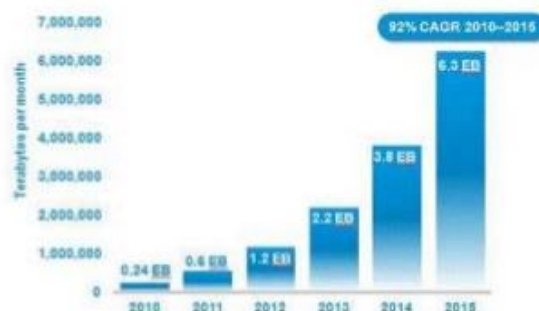


Figura 3 6,3 Exabytes de datos móviles mensuales previstos para 2015, Febrero 2011.

Imagen tomada del sitio cisco Latinoamérica:  
<http://www.blogciscolatam.com/2011/02/el-trafico-global-de-datos-moviles-se-multiplicara-por-26-entre-2010-y-2015/>

Estas grandes cantidades de datos que se generan, sobrepasan la capacidad del software tradicional para procesarlos y gestionarlos de una forma adecuada, aquí es donde surge “Big Data”, que es definido por algunos expertos como un gran volumen de datos que crecen de forma masiva (Mcclowry, 2012), tambien es definido como un sector de tecnologías de la información y la comunicación con respecto a los sistemas que manipulan grandes conjuntos de datos cuyo tamaño está más allá de la capacidad de la mayoría de los software utilizados para capturar, gestionar y procesar la información dentro de un lapso tolerable (Rattinger, 2013), en el seminario de Isummit Loxa 2012 surgió una definición interesante “Mas datos que los que una base de datos relacional puede manejar” (Pullas Aguilar).

Big data es visto como una oportunidad y que tendrá gran impacto en las áreas de negocio de las organizaciones (Jimenez, 2013), también es considerado como un reto para investigar en el volumen, la velocidad y la variedad de datos, con la finalidad de obtener acceso y procesamiento ágil tanto a datos estructurados y no estructurados desde diversos orígenes y obtener de ellos el máximo valor (Petthey & Goasduff, 2011), también big data está transformando la forma en que se conduce una investigación adquiriendo habilidades en el uso de Big Data para resolver problemas complejos relacionados con el descubrimiento científico, investigación ambiental y biomédica, educación, salud, seguridad nacional, y organizaciones en general.

Es increíble lo que se puede hacer con big data: se puede trabajar sobre los datos hacerlos hablar, ir detrás de “migajas de datos”, se puede mantener un rastro de las cosas, realizar descubrimientos, predicciones, “espíar clientes”, encontrar tendencias, analizar datos en tiempo real, realizar mediciones, lo que permite a las compañías tomar mejores decisiones incrementando la innovación, pero reduciendo costes. También se pueden realizar aplicaciones en forma de sistemas de recomendación y sistemas de negocio.

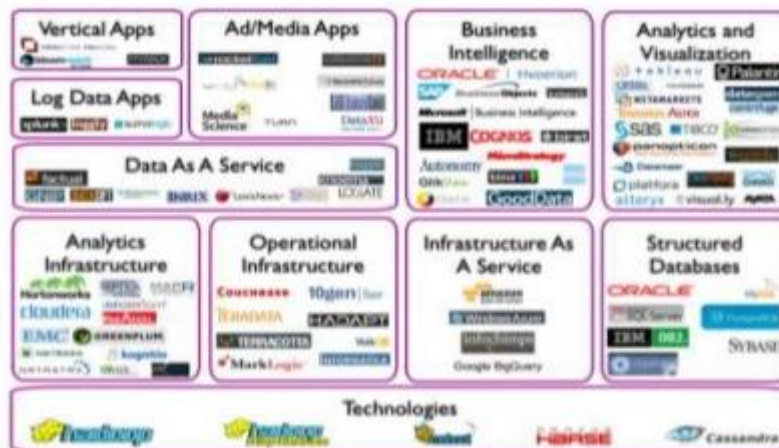


Figura 4. El Ambiente de big data, Junio 2012.

Imagen tomada del sitio forbes:

<http://www.forbes.com/sites/davefeinleib/2012/06/19/the-big-data-landscape>

El entorno big data a su vez se ha encontrado con algunos inconvenientes como que actualmente la información no está correctamente estructurada, procesada, agregada, ni filtrada, la información es repetitiva y generalmente desordenada y su volumen es enorme. Por ello los retos son grandes: el sistema de almacenamiento debe ser escalable, el procesamiento debe ser paralelo y masivo, se deben emplear cada vez mejores herramientas para la gestión de la información, debemos considerar los datos como el recurso más valioso y pensar que los datos son el futuro. De entre los proyectos que se pueden mencionar donde se ha llevado a cabo el uso de solución de Big Data se encuentran: Hadoop, Avro, Cassandra, Chukwa, Jaql las cuales son herramientas potentes para el análisis de big data. En el momento histórico en el que nos encontramos, en el cual disponemos de tecnologías que nos permiten una mayor disponibilidad y acceso a datos científicos a nivel mundial, aumentando la colaboración e incrementando el ritmo y la profundidad del descubrimiento, surge el término “Open Data”, cuya filosofía fundamental es permitir que los datos estén disponibles para que todo el mundo tenga acceso a ellos, tanto para su reutilización como para su redistribución, sin restricciones de copyright, patentes u otros mecanismos de control (org ciudad inteligente, 2012).



Figura 5. El movimiento open data, Abril 2012.

Imagen tomada del sitio La melle Numérique: <http://www.meleenumerique.com/contenu/bigopendata>

Para que un dato sea considerado abierto, tiene que ser accesible y reutilizable, sin exigir permisos específicos, aunque los tipos de reutilización pueden estar controlados mediante una licencia. Existen algunos parámetros que se deben cumplir para open data (Barros, 2012):

- Liderazgo y fuerte soporte de la administración
- Modelo de identificación y priorización de datasets
- Modelos y prácticas de licenciamiento abiertas y difundidas
- Uso de estándares internacionales de publicación de datos
- Portales y plataformas de publicación en la web
- Generación de interpretaciones y reportes estáticos
- Incentivos y generación de comunidades de intereses de uso
- Desarrollo de competencias en torno a los datos y su uso
- Mecanismos de retroalimentación
- Marco regulatorio para la promoción e incentivos de los datos abiertos.

Entre los principales desafíos que se presentan con open data están: Las fuentes primarias de datos deben estar sin ambigüedades, disponibles para a todo tipo de usuarios, deben ser procesables, licencias mas abiertas (Creative Commons), a todo esto las empresas, universidades, ong's deben aportar con esta iniciativa. Algunas iniciativas como OpenUniversity, OpenResearchOnline, OpenLearn y OpenGovernment, así como el uso de software libre, y el código

abierto, fortalecen los principios de open data. También se acuña el término de Open Linked Data que hace referencia al método con el que se pueden mostrar, intercambiar y conectar datos a través de URIs en la Web para acceder a los recursos, ofreciendo información usando y enlazando RDF's, para de esta forma enlazar los datos convirtiendo a la web semántica en una gran base de datos, permitiendo tanto a personas como computadoras explorar la información bien referenciada e interconectada. Entonces mientras que open data surge como una idea para publicar datos de forma libre y asequible a cualquier persona para usar y republicar sin restricciones de ningún tipo, en cambio linked data nos da el soporte para utilizar la web para conectar datos relacionados que no estaban enlazados previamente usando las mejoras que provee la web semántica.

### **Conclusiones:**

- El futuro de la web radica en la interconexión total, es decir personas conectándose con personas y aplicaciones Web de forma ubicua, tecnologías como la Inteligencia Artificial, Web semántica son fundamentales para lograr estos objetivos.
- Debido a la gran cantidad de medios por los cuales los seres humanos podemos generar información, esta es abundante y hace necesario el uso de big data y herramientas para poder realizar una mejor gestión de la misma.
- El uso apropiado de big data facilita la gestión de la información de las organizaciones y les ayuda a tomar mejores decisiones de forma ágil, decisiones correctivas e incluso realizar predicciones con respecto a los productos o servicios que ofrecen y a las necesidades y nivel de satisfacción de sus clientes. Big data aportar valor a las empresas y ayuda a acercarse a los clientes.
- Open data y Open linked data no son una moda pasajera, es una tecnología que permite un mejor manejo de la información, para hacerla disponible a todo el mundo sin restricciones de ningún tipo.



## **Bibliografía**

Barros, A. (19 de Septiembre de 2012). alejandrobarros. Obtenido de <http://www.alejandrobarros.com/open-data-por-donde-empezar>

Gonzales, D. C. (30 de Junio de 2003). Obtenido de [http://mobiquo.gsync.es/mobility inteligente](http://mobiquo.gsync.es/mobility%20inteligente), o. c. (2012). Obtenido de [http://2012.desarrollandoamerica.org/que-es-dal/que-son- datos-abiertos/](http://2012.desarrollandoamerica.org/que-es-dal/que-son-datos-abiertos/)

Jimenez,A.(2013, Enero29).  
Retrievedfrom[http://www.greenplum.com/blog/dive-in/disruptive- data-science-transforming-your-company-into-a-data-science-driven-enterprise](http://www.greenplum.com/blog/dive-in/disruptive-data-science-transforming-your-company-into-a-data-science-driven-enterprise)

Mcclowry, S. (2012). mike 2.0. Retrieved from [http://mike2.openmethodology.org/wiki/Big\\_Data\\_Definition](http://mike2.openmethodology.org/wiki/Big_Data_Definition) Pettey, C., &

Goasduff, L. (2011, Junio 27). Gartner. Retrieved from <http://www.gartner.com> Pullas Aguilar, P. (s.f.). oracle.

Quispe Ortega , L. M. (2011). consultora devian. Obtenido de [http://www.consultora- devian.net/inicio/noticias/52-de-la-web-10-a-la-web-40](http://www.consultora-devian.net/inicio/noticias/52-de-la-web-10-a-la-web-40)

Rattinger, A. (24 de Enero de 2013). merca20. Obtenido de <http://www.merca20.com/big-data/> Smart, J., Cascio, J., & Paffendorf, J. (2007). metaverseroadmap. Obtenido de <http://www.metaverseroadmap.org>

Cascio, J., & Paffendorf, J. (2007). metaverseroadmap. Obtenido de <http://www.metaverseroadmap.org>