

U.T.N. – Facultad Regional Buenos Aires

Ingeniería en sistemas de Información

Análisis de Sistemas – 2013

TURNO:Noche CURSO: K2053

DOCENTE A CARGO: Ramiro Garbarini

AYUDANTES DE CÁTEDRA: Laura Recchini

TRABAJO PRÁCTICO

BIG DATA

GRUPO N° 8

Cervera Ivana

Barbeito Manuel

Filotti Emilio

Mariño Camila

Parra Matías

Perez De Ambrogio Mariana

Redelico Ariel

FECHA DE VENCIMIENTO: 10/09/2013

FECHA DE PRESENTACIÓN: 10/09/2013

FECHA DE DEVOLUCIÓN: \_\_/\_\_/\_\_

CALIFICACIÓN: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ FIRMA PROFESOR: \_\_\_\_\_\_\_\_\_

TP BIG DATA

**ÍNDICE**

**Introducción ………………………………………………………………………………. 2**

**¿Qué es Big Data? .................................................................................................. 3**

**Características ……………………………………………………………………………. 3**

**Fuentes y Tipos …………………………………………………………………….…….. 4**

**Beneficios ...…..…………………………………………………………………………… 5**

**Conclusión ………………………………………………………………………………… 6**

**Bibliografía ………………………………………………………………………………… 7**

*Análisis de Sistemas* Grupo N°: 8 Hoja 1 de 7

TP BIG DATA

**Introducción**

Debido al gran avance que existe día a día en las tecnologías de información, las organizaciones se han tenido que enfrentar a nuevos desafíos que les permitan analizar, descubrir y entender más allá de lo que sus herramientas tradicionales reportan sobre su información, al mismo tiempo que durante los últimos años el gran crecimiento de las aplicaciones disponibles en Internet han sido parte importante en las decisiones de negocio de las empresas. Este trabajo práctico tiene como propósito introducir al lector en el concepto de Big Data y describir algunas características de los componentes principales que constituyen una solución de este tipo.

*Análisis de Sistemas* Grupo N°: 8 Hoja 2 de 7

TP BIG DATA

**¿Qué es Big Data?**

En términos generales podríamos referirnos como a la tendencia en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos que tomaría demasiado tiempo y sería muy costoso cargarlos a un base de datos relacional para su análisis.

De tal manera que, el concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales. Sin embargo, Big Data no se refiere a alguna cantidad en específico, ya que es usualmente utilizado cuando se habla en términos de petabytes y exabytes de datos. Los tamaños del "big data" se hallan constantemente en aumento.

IBM define la estrategia de Big Data como el desafío de las cuatro ‘V’: Volumen, Variedad, Velocidad y Veracidad y es utilizada para describir, organizar y sistematizar enormes cantidades de datos (estructurados, no estructurados) que pueden ser almacenados, convertidos a información y buscados de forma inteligente. Todo con el objetivo de producir una visión de negocios más enriquecida.

**Características:**

En un informe de investigación del 2001, el analista Doug Laney del [META Group](http://es.wikipedia.org/w/index.php?title=META_Group&action=edit&redlink=1) (ahora [Gartner](http://es.wikipedia.org/wiki/Gartner_%28empresa%29)) definió el crecimiento constante de datos en tres dimensiones: Volumen, velocidad y variedad. Actualmente gran parte de la industria, sigue utilizando este modelo de las "3vs" para describir los grandes volúmenes de datos al cual se le ha añadidio una “4V” Veracidad.

* **Volumen:** la cantidad de datos. Siendo quizá la característica que se asocia con mayor frecuencia a big data, el volumen hace referencia a las cantidades masivas de datos que las organizaciones intentan aprovechar para mejorar la toma de decisiones en toda la empresa. Los volúmenes de datos continúan aumentado a un ritmo sin precedentes.
* **Velocidad:** En los procesos en los que el tiempo cuenta como descubrir fraudes, los grandes volúmenes de datos deben utilizarse como una corriente en su empresa para maximizar el valor.
* **Variedad:** diferentes tipos y fuentes de datos. La variedad tiene que ver con gestionar la complejidad de múltiples tipos de datos, incluidos los datos estructurados, semiestructurados y no estructurados. Las organizaciones necesitan integrar y analizar datos de un complejo abanico de fuentes de información tanto tradicional como no tradicional procedentes tanto de dentro como de fuera de la empresa.

*Análisis de Sistemas* Grupo N°: 8 Hoja 3 de 7

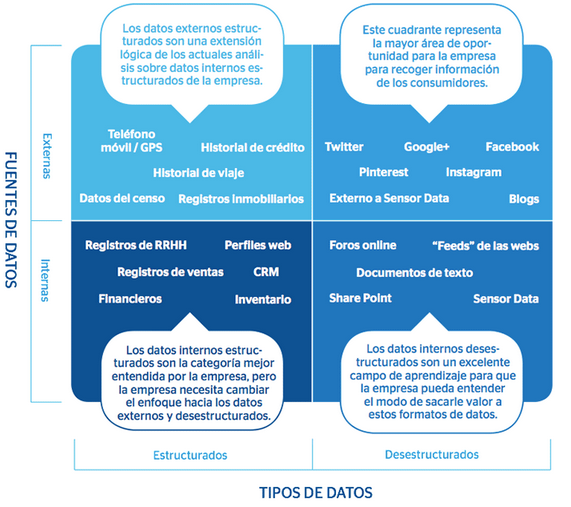
TP BIG DATA

* ***Veracidad:*** la incertidumbre de los datos. La veracidad hace referencia al nivel de fiabilidad asociado a ciertos tipos de datos. Esforzarse por conseguir unos datos de alta calidad es un requisito importante y un reto fundamental de big data, pero incluso los mejores métodos de limpieza de datos no pueden eliminar la imprevisibilidad inherente de algunos datos, como el tiempo, la economía o las futuras decisiones de compra de un cliente. La necesidad de reconocer y planificar la incertidumbre es una dimensión de big data que surge a medida que los directivos intentan comprender mejor el mundo incierto que les rodea.

**Fuentes y tipos**

Los datos **estructurados** son los datos tradicionalmente presentes en los sistemas corporativos (bases de datos, archivos jerárquicos y secuenciales, etc.), y los datos **no estructurados** se relacionan principalmente, con el contenido digital más reciente, y se ponen a disposición previamente en un formato no digital, tales como archivos de imagen, audio, texto, entre otros.

Las fuentes de datos **internas** son aquellas que provienen del interior de la organización, y las **externas** del exterior de la misma.



*Análisis de Sistemas* Grupo N°: 8 Hoja 4 de 7

TP BIG DATA

**Beneficios que encuentra una organización cuando implanta Big Data en su operación:**

* Análisis de la información en tiempo real para la toma de decisiones o incluso detección y predicción de posibles acontecimientos.

**Por ejemplo, el sector financiero puede analizar transacciones en tiempo real para detectar posibles operaciones fraudulentas y poder tomar acciones rápidas, como la de detener dicha operación. El sector salud puede detectar patrones al analizar todos los datos de ciertos pacientes y prever riesgos –es el caso de una unidad de bebés prematuros que analiza datos generados por sensores para preveer posibles acontecimientos en la salud del bebé, como una falla renal en las siguientes 24 horas.**

**Entidades como la policía también analizan en tiempo real diferentes fuentes de información como fotos, videos y audios para detectar zonas donde se puede presentar eventuales acontecimientos de violencia y reaccionar más rápidamente. O, finalmente, toda empresa puede usar el análisis de sentimiento con respecto a un nuevo producto o a su marca para conocer qué piensa el consumidor y tomar decisiones más acertadas con respecto a su negocio.**

* **Mejora de la relación con el cliente.**

Big Data proporciona nuevas y mejores herramientas para satisfacer a los clientes, ya que se le puede ofrecer una atención personalizada basada en un conocimiento mayor y mejor de él. Otras funcionalidades son: mejorar las medidas antifraude y reducir la portabilidad hacia la competencia.

**¿Qué se puede hacer entonces con grandes volúmenes de datos?**

-Entender mejor a sus clientes.

-Anticiparse a los cambios en el comportamiento del mercado.

-Predecir fallas de los equipos.

-Optimizar los procesos.

-Reducir los riesgos.

Sólo como un ejemplo, algunos casos útiles donde el uso de la tecnología de predicción es avanzada:

-Identificación de los productos más rentables.

-Segmentación de clientes.

-Identificación de la siguiente mejor oferta.

-Detección del fraude.

-Retención de clientes o pérdida de clientes.

*Análisis de Sistemas* Grupo N°: 8 Hoja 5 de 7

TP BIG DATA

**Big Data y el campo de investigación**

Los científicos e investigadores han analizado datos desde ya hace mucho tiempo, lo que ahora representa el gran reto es la escala en la que estos son generados.

Esta explosión de "grandes datos" está transformando la manera 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, entre otros.

De entre los proyectos que se pueden mencionar donde se ha llevado a cabo el uso de una solución de Big Data se encuentran:

* El Language, Interaction and Computation Laboratory (CLIC) en conjunto con la Universidad de Trento en Italia, son un grupo de investigadores cuyo interés es el estudio de la comunicación verbal y no verbal tanto con métodos computacionales como cognitivos.
* [Lineberger Comprehensive Cancer Center - Bioinformatics Group](https://lbg.unc.edu/) utiliza Hadoop y HBase para analizar datos producidos por los investigadores de The Cancer Genome Atlas(TCGA) para soportar las investigaciones relacionadas con el cáncer.
* El [PSG College of Technology, India](http://www.psgtech.edu/), analiza múltiples secuencias de proteínas para determinar los enlaces evolutivos y predecir estructuras moleculares. La naturaleza del algoritmo y el paralelismo computacional de Hadoop mejora la velocidad y exactitud de estas secuencias.
* La Universidad Distrital Francisco Jose de Caldas utiliza Hadoop para apoyar su proyecto de investigación relacionado con el sistema de inteligencia territorial de la ciudad de Bogotá.
* La Universidad de Maryland es una de las seis universidades que colaboran en la iniciativa académica de cómputo en la nube de IBM/Google. Sus investigaciones incluyen proyectos en la lingüistica computacional (machine translation), modelado del lenguaje, bioinformática, análisis de correo electrónico y procesamiento de imágenes

**Empresas que brindan el servicio de Big Data.**

SAP: La herramienta emblema de SAP para el análisis de Big Data es [SAP HANA](http://www.businessoftware.net/sap-hana/).  HANA es la aplicación in-memory SAP que consiste en un hardware y software ya enlatado que permite una velocidad de procesamiento nunca visto hasta la actualidad. Permite obtener tiempos de respuesta altísimos al momento de realizar una consulta

Microsoft: SQL Server 2012 es la herramienta que propone Microsoft para contener grandes volúmenes de datos.

Oracle: ofrece una gama de productos para la organización de grandes volúmenes de datos, como:

* [Oracle Big Data Appliance](http://www.oracle.com/lad/products/database/big-data-appliance/overview/)
* [Oracle Data Integrator](http://www.oracle.com/us/products/middleware/data-integration/enterprise-edition/overview/index.html?ssSourceSiteId=ocomlad)
* [Oracle Big Data Connectors](http://www.oracle.com/us/products/database/big-data-connectors/overview/index.html?ssSourceSiteId=ocomlad)

*Análisis de Sistemas* Grupo N°: 8 Hoja 6 de 8

TP BIG DATA

**Conclusión**

Big Data les sirve a todas las empresas y organizaciones interesadas en mejorar de forma inteligente el sistema de almacenamiento, búsqueda y uso de información, con el objetivo de mejorar la toma de decisiones de su negocio. De igual forma, les sirve a los usuarios que desean mejorar sus sistemas de almacenamiento.

Big data permite obtener una imagen más completa de las preferencias y demandas de los clientes; a través de esta profunda comprensión empresas de todo tipo encuentran nuevas formas de interactuar con sus clientes actuales y futuros.

*Análisis de Sistemas* Grupo N°: 8 Hoja 7 de 8

TP BIG DATA

**Bibliografía**

<http://www-03.ibm.com/software/products/es/es/category/SWP10>

<http://www.lanacion.com.ar/1490304-el-nuevo-oro-se-llama-big-data>

<http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html?cmp=BS&ct=SocialMedia&cr=twitter>

<http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>

<http://en.wikipedia.org/wiki/Big_data>

<http://empresas.it/2013/05/big-data-preguntas-y-respuestas-de-un-experto/>

<http://alt1040.com/2013/08/big-data>

<https://www.centrodeinnovacionbbva.com/magazines/innovation-edge/publications/21-big-data/posts/153-big-data-en-que-punto-estamos>

<http://www.techroi.com.pe/techroi/thechroi/13/82/la-hora-del-big-data>

*Análisis de Sistemas* Grupo N°: 8 Hoja 8 de 8