


Capítulo 1: Introducción al Big Data

Capítulo 2: El Big Data y la ciencia de datos

Capítulo 3: Exploración del Big Data



División de Alta Tecnología


1

Introducción al Big Data

Big Data

En Dios confiamos, el resto debe traer datos.
(W. Edwards Deming)

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.




Objetivos

Al finalizar el capítulo, el alumno logrará:

- Identificar el valor agregado del análisis de datos para las empresas.
- Conocer las tendencias del mercado en Big Data.

1 - 2

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Agenda

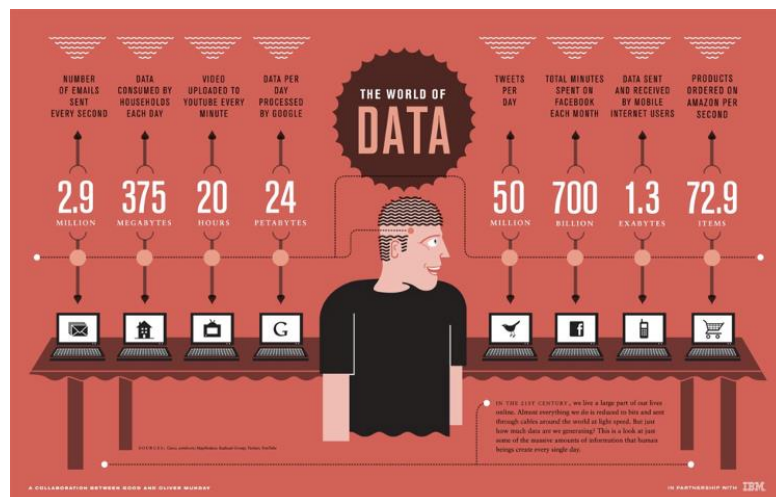
- Características del Big Data
- Las V's del Big Data
- El impacto del Big Data en los negocios
- Ejemplos de Big Data

1 - 3

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



El mundo de los datos

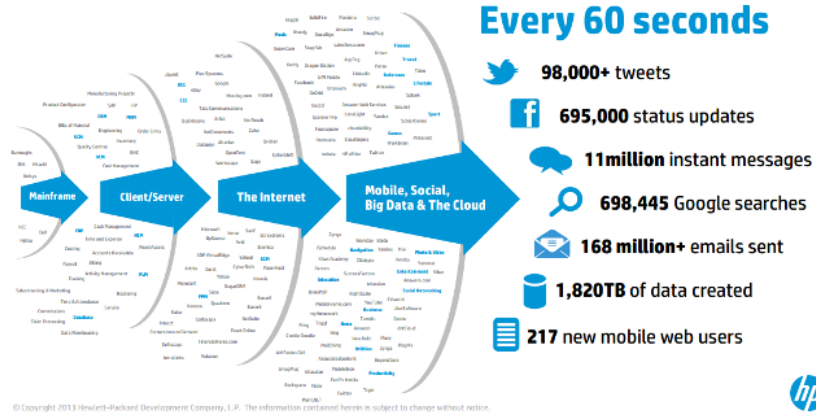


1 - 4

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



¿ Qué ocurre en Internet ?



1 - 5

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.

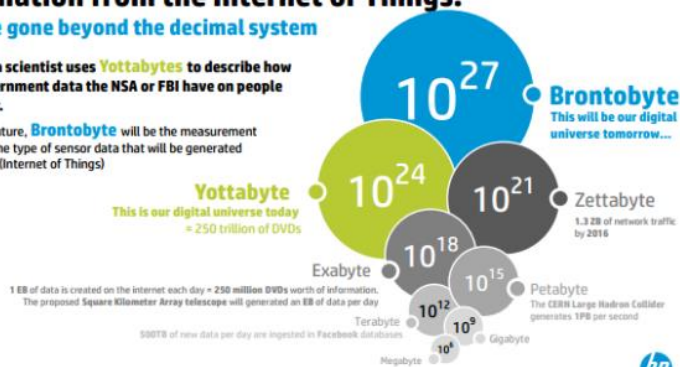
Medidas de almacenamiento

Information from the Internet of Things:

We have gone beyond the decimal system

Today data scientist uses **Yottabytes** to describe how much government data the NSA or FBI have on people altogether.

In the near future, **Brontobyte** will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things)



© Copyright 2013 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice.

1 - 6

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.

Definiciones de Big Data



Wikipedia hace referencia a una cantidad de datos tal que supera la capacidad del software convencional para ser capturados, administrados y procesados en un tiempo razonable.

IBM: toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales.



Google: serían demasiado caros de almacenar, gestionar y analizar utilizando sistemas de bases de datos tradicionales, son ineficientes debido a su inflexibilidad para almacenar datos no estructurados (imágenes, texto y video), acomodar datos de "alta velocidad" (en tiempo real) o escalar a soporte muy grande (petabytes).

Gartner: son activos de información de gran volumen, alta velocidad y / o gran variedad que demandan formas rentables e innovadoras de procesamiento de la información que permiten un mejor conocimiento, toma de decisiones y automatización de procesos.



1 - 7

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



¿Qué es Big Data?

Big Data es un conjunto de conocimientos, métodos y tecnologías orientados a facilitar la adquisición, gestión y uso de los datos generados en la actualidad.



1 - 8

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Características del Big Data

- **N = ALL**
 - Fin del muestreo
 - Todo el conjunto de datos es válido, no se descartan casos
 - Se puede eliminar el problema del sesgo
- La inexactitud de los datos ya no es un problema
 - El error de la muestra se minimiza, asumimos datos menos exactos.
 - Dicho de otra forma, cuando dependemos de una muestra, queremos que los datos sean exactos.
- Obtenemos el qué, pero no el porqué
 - Las técnicas del Big Data no explican la causalidad (correlaciones).

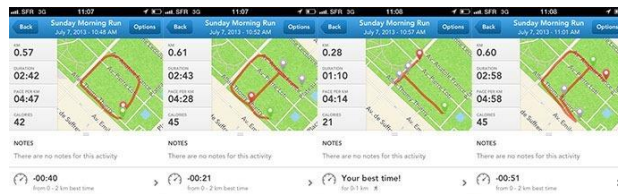
1 - 9

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



¿Qué es la datificación?

- Este concepto consiste en convertir cualquier acción o evento susceptible de ser medido en datos digitales.
- Ahora, toda acción, dato, preferencia, etc. está sujeta a ser medida y almacenada.
 - Localización de una persona.
 - Vibración del motor de un coche.
 - Recorrido de una camilla en el hospital.

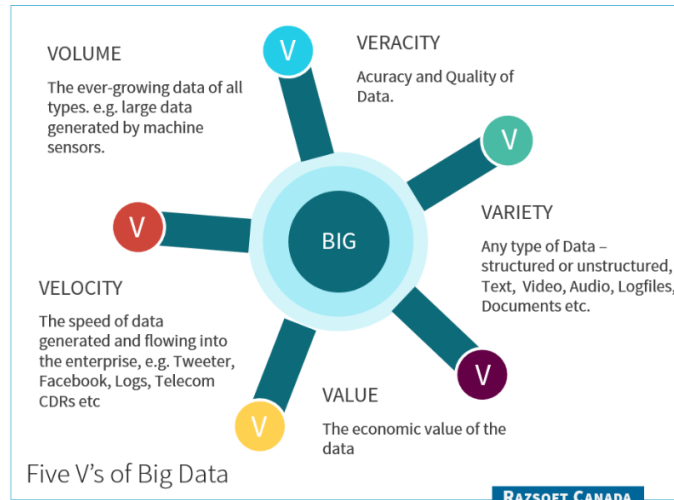


1 - 10

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Las V's del Big Data



1 - 11

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Volumen

- Aumento del Volumen de los datos disponibles para el análisis.
- Las empresas que han adoptado Big Data gestionan desde unos pocos Terabytes (103 Gb) hasta Petabytes (106Gb)
- Muchas de ellas ya superan los 10 Terabytes
 - En tres años lo habitual será ≥ 100 Terabytes
 - No es lo mismo trabajar con 1 Terabyte de texto con 1 Gigabyte imágenes médicas.



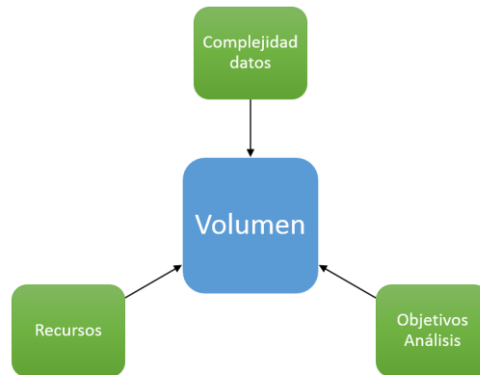
1 - 12

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Volumen

A la hora de valorar el volumen de los datos hemos de tener en cuenta otros factores, como los que se muestran en la imagen.



1 - 13

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Velocidad

- Incremento de la **velocidad** a la que se genera y se distribuye los datos en las fuentes.
 - Es una de las razones del incremento del volumen de datos
- Streaming: datos que se generan y distribuyen en tiempo real.
 - Generados por sensores, servidores web, redes sociales, etc.
- También hace referencia a la necesidad para extraer conocimiento de los datos en el momento oportuno.
 - Ejemplo: Datos de tipo financiero.



1 - 14

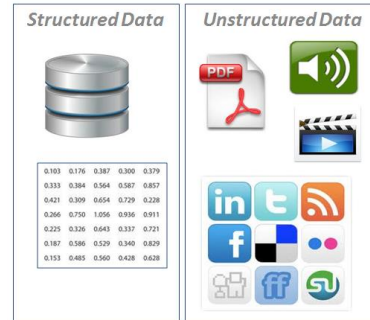
Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Variedad

La **variedad** se refiere al importante aumento en la heterogeneidad de las fuentes de datos debido a diversos factores como:

- Incremento en el número de fuentes disponibles.
- Fuentes con distintos nivel de estructura.
- Diversidad de formatos de distribución.



1 - 15

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Veracidad

- Aumento de la incertidumbre respecto a la veracidad o calidad de los datos disponibles

Incetidumbre datos => Incetidumbre conocimiento extraído.

- Es uno de los retos principales del nuevo contexto de datos.
 - El uso de datos incorrectos supone **grandes pérdidas**.



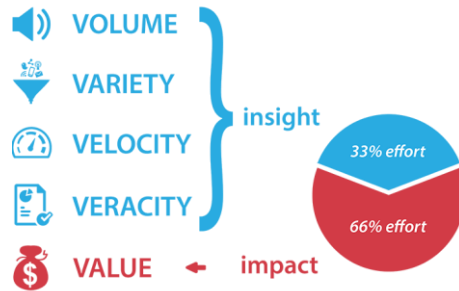
1 - 16

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Valor

- El valor es la medida de la utilidad de los datos seleccionados para nuestros objetivos finales.
- Determinar dicha utilidad a priori puede ser realmente complicado.



1 - 17

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



El impacto del Big Data en los negocios

El valor de los datos

- El uso intensivo de los datos ha pasado a ser el petróleo de muchas compañías.
- El nuevo enfoque es almacenar cualquier tipo de dato, por irrelevante que pueda parecer, para su posterior análisis.
 - Clics de ratón en la página web de mi negocio.
 - Vibración del motor del coche.
 - Movimiento del acelerómetro del Smartphone.
- Permite crear modelos para responder preguntas complejas, mostrar percepciones y brindar resultados únicos.

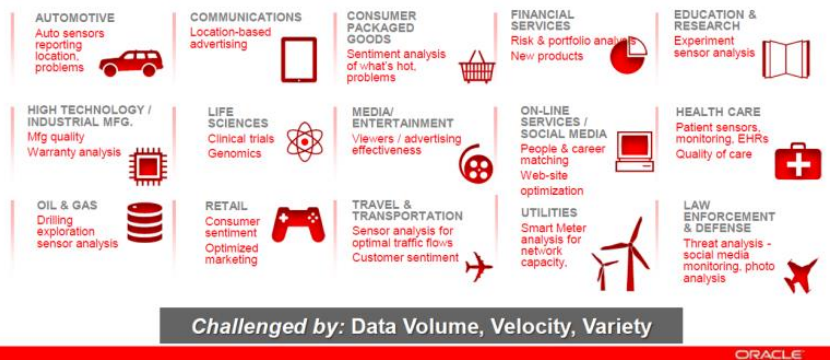
1 - 18

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



El impacto del Big Data en los negocios

El valor de los datos



1 - 19

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



El impacto del Big Data en los negocios



1 - 20

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



El impacto del Big Data en los negocios



1 - 21

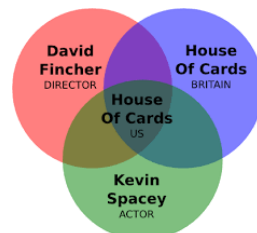
Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Ejemplo de Big Data - Netflix

Utiliza la información de sus suscriptores para predecir que contenidos tienen más probabilidades de triunfar.

- ¿Qué búsquedas realizan?
- ¿Qué dispositivos usan?
- ¿Cuál es su día preferido?
- ¿Cuánto tiempo emplean en el servicio y en cada uno de los contenidos?
- Si ven los capítulos enteros o parcialmente, incluso, ¿qué fragmentos vuelven a visionar?
- ¿En qué momento abandonan el visionado y si lo recuperan o abandonan?
- Las valoraciones de los consumidores.
- ¿Qué preferencias tienen en común con sus amigos o con la audiencia de su misma zona geográfica?
- La información de sus perfiles en redes sociales.



1 - 22

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Ejercicio N° 1.1: Beneficios de Big Data

Al finalizar, el alumno logrará:

- Identificar los diversos datos de su organización y sustentar los beneficios de utilizar Big Data.

1 - 23

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Tarea N° 1: Introduction to Data Science

Al finalizar la tarea, el alumno logrará:

- Aprender acerca de por qué la ciencia de datos, la inteligencia artificial (IA) y el aprendizaje automático están revolucionando la forma en que las personas hacen negocios e investigan en todo el mundo.

1 - 24

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Lecturas adicionales

Se sugiere revisar los siguientes enlaces para profundizar en los conceptos tratados en el presente capítulo:

- a) Términos y conceptos de big data
- b) Casos de éxito de soluciones de Big Data
- c) Tecnologías y herramientas para el almacenamiento y tratamiento de datos de Big Data.

1 - 25

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Resumen

En este capítulo hemos aprendido lo que es Big Data, las características que tiene y las 5Vs de Big Data.

Además, hemos hecho una breve revisión del impacto del Big Data en los negocios y que compañías actualmente vienen trabajando con esta tecnología.

1 - 26

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.

