

SESIÓN 3

FUNDAMENTOS DE CIENCIA DE DATOS

U UPAEP

Agenda

- Temas de interés del área de Ciencia de Datos.

(Alejandro Rondero García / Gustavo Adolfo Pérez Cruz)

- M1: Introducción a Big Data
 - El concepto de Big Data
 - El rol que el Big Data desempeña en las organizaciones
 - Cultura analítica en la organización: *Data Driven Business*
 - Modelos de negocio impulsados por los datos
 - Nuevas competencias y nuevas capacidades
 - Nuevos roles





M1: Introducción a Big Data

Características de Big Data (5V)

- ✓ Volumen
- ✓ Variedad
- ✓ Velocidad
- ✓ Veracidad
- ✓ Valor

Equipo 1

Equipo 2

Equipo 3

Equipo 4

Equipo 5

Características de Big Data (5V)

1

Volumen

El volumen de datos almacenados en los repositorios empresariales han crecido de los megabytes y gigabytes a los peta bytes.

2

Variedad

La variedad de los datos con los que se trabaja en las organizaciones, ya no es sólo la información estructurada que se almacena en los repositorios empresariales, sino que se ha extendido a la información no estructurada o la semiestructurada, como pueden ser: archivos de audio, video, XML, etc.

M1 – Introducción a Big Data

Características de Big Data (5V)

3

Velocidad: La velocidad con la que se capturan, procesan o mueven datos, interna o externamente, se ha incrementado significativamente. Los modelos basados en tecnologías de *Business Intelligence* normalmente tardan días en procesar la información, en comparación con los requisitos actuales de análisis “casi” en tiempo real, utilizando flujo de entrada de datos de alta velocidad.

4

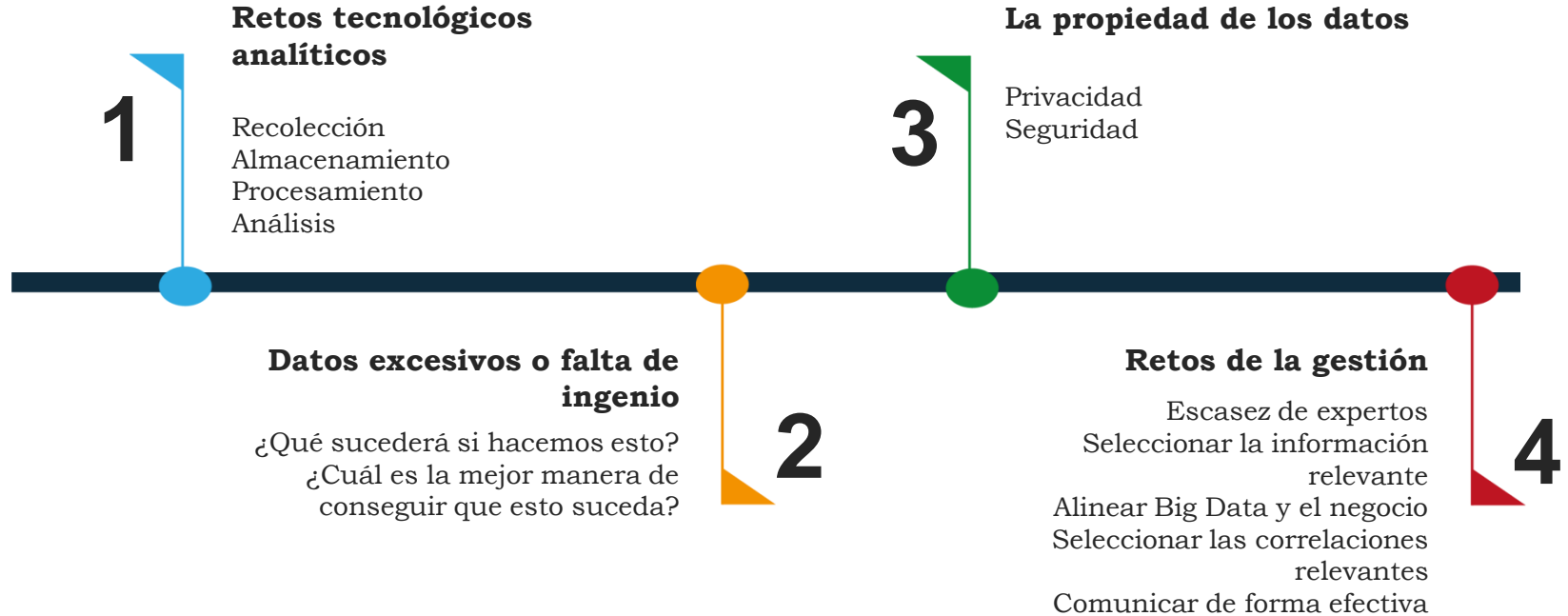
Veracidad: Los datos deben ser fieles a la realidad, no estar manipulados y ser fiables. Datos erróneos o mal interpretados pueden conducir a un análisis pobre de los mismos y a obtener conclusiones distorsionadas.

5

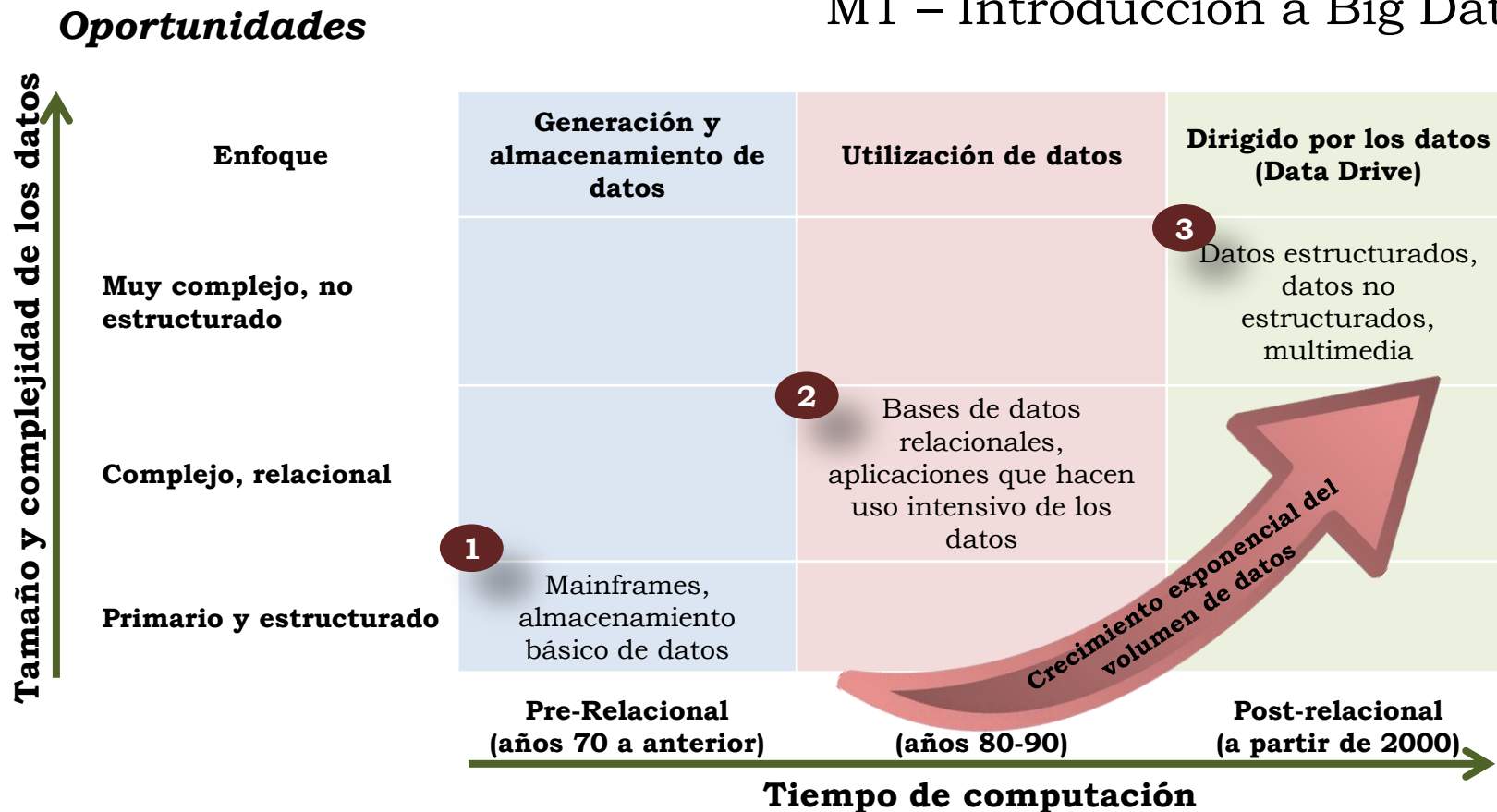
Valor: Esta propiedad hace referencia al hecho de que al disponer de una mayor cantidad de datos, estos se pueden cruzar y analizar para obtener un valor de negocio que los datos que se almacenaban tradicionalmente no eran capaces de revelar.

M1 – Introducción a Big Data

Retos a superar



M1 – Introducción a Big Data



M1 – Introducción a Big Data

Fuentes de información de Big Data



Datos generados por la organización

- Cada organización tiene prácticas de operación y modelos de negocio distintos.
- Cada una tiene una variedad de plataformas de generación de datos.
- Casi todos los eventos pueden almacenarse potencialmente.
- Almacenan los datos para su uso actual y futuro, así como para hacer un análisis del pasado.



Información obtenida de las personas

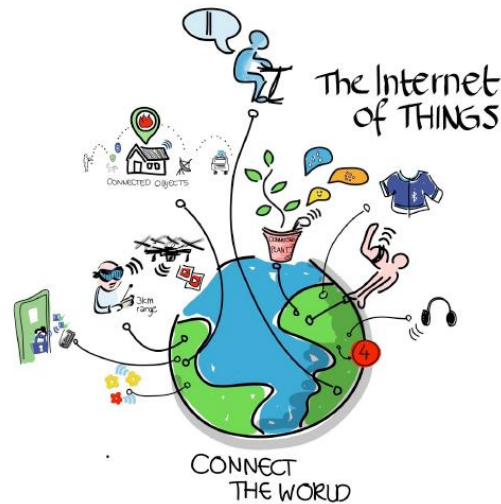
- El reto de este tipo de información es que es **NO ESTRUCTURADA**.

Compañía	Procesamiento diario de datos
eBay	100 Petabytes (PB= 10^{15})
Google	100 PB
Facebook	30+ PB
Twitter	100 Terabytes (=.1 PB)
Spotify	64 Terabytes (TB= 10^{12})

Entre el 80% - 90% son datos no estructurados.

¿Qué hace a un dispositivo inteligente?

1. Estar conectado con otros dispositivos y/o redes.
2. Recopilar y analizar datos de forma autónoma.
3. Proporcionar un contexto del medio ambiente.



M1 – Introducción a Big Data

Acercándonos a la realidad

“Today we can process Exabyte's of data at lightning speed, and these gives us the **potential to make bad decisions** far more quickly, efficiently and with greater impact than we did in the past.”

Susan Etlinger



Ted Talk: What do we do with all this big data?

https://www.ted.com/talks/susan_etlinger_what_do_we_do_with_all_this_big_data

M1 – Introducción a Big Data

¿Es siempre adecuado el uso de Big Data?

Cautelas

- ✓ Correlación no equivale a causalidad.
- ✓ Big Data es una herramienta excelente para la investigación pero no puede sustituir al investigador.
- ✓ No sustituye al trabajo humano.
- ✓ Las predicciones no son infalibles.
- ✓ Posibilidad de círculo vicioso.
- ✓ Acontecimientos poco frecuentes.

Limitaciones

- ✓ ¿Los perfiles de las redes sociales son engañosos?
- ✓ ¿La cartografía de las conversaciones sociales es estadísticamente significativa?
- ✓ ¿Es fiable el análisis automático de los sentimientos?
- ✓ Representatividad de las palabras clave y las búsquedas temáticas.
- ✓ Cambios continuos en las dinámicas de las redes sociales.
- ✓ Los datos no pueden sustituir una inversión significativa en horas de análisis.

Actividad 3 ***Data-Driven en la empresa/roles de ciencia de datos***

1. ¿Qué es una organización manejada por los datos (DDB, por sus siglas en inglés)?
2. Menciona tres beneficios de una organización impulsada por los datos.
3. Resume los puntos sobre cómo instaurar una cultura *data-driven* en la empresa.
4. ¿Cuáles son los tres pilares de una empresa impulsada por los datos?
5. Explica dos de los ejemplos reales (zoom-in), presentados en el documento de SAS.
6. Define los nuevos roles en el área de ciencia de datos: *Chief Data Officer*, *Data Scientist* y *Data Steward*.

M1 – Introducción a Big Data

Cultura analítica en la organización: Data Driven Business

- ✓ DDDM- Data Driven Decision Management: Toma de decisiones de gestión basada en los datos es un nuevo enfoque de la gestión empresarial que valora aquellas decisiones respaldadas con datos que pueden ser verificados.
- ✓ Un estudio de MIT reveló que aquellas organizaciones que basan la toma de decisiones en los datos, tenían índice de productividad superiores en un 4% a la media y beneficios superiores al 6%.

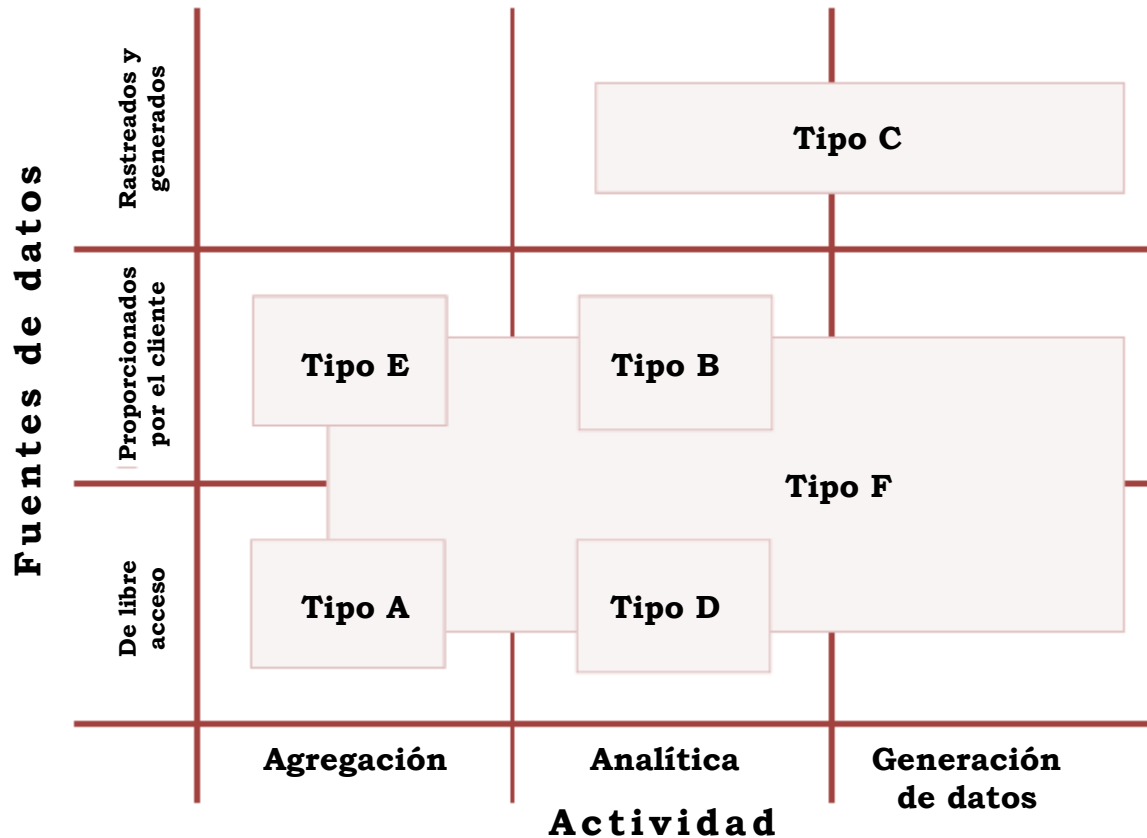
M1 – Introducción a Big Data

Comportamientos que facilitan la toma de decisiones basada en los datos

- ✓ Alguien es responsable de la calidad de los datos.
- ✓ En cualquier nuevo proyecto, la obtención de datos es una de las primeras prioridades.
- ✓ Cuando alguien da una opinión, la acompaña de números y datos.
- ✓ Los datos se comunican siempre, aunque muestren malos resultados.
- ✓ Todo el mundo puede acceder a los datos que tienen relación con su trabajo.
- ✓ Cada objetivo tiene un indicador y un valor a alcanzar.
- ✓ Los equipos reciben formación en análisis de datos.
- ✓ Los proyectos que implican recolección de datos no tienen problemas para recibir financiación.
- ✓ Los datos nunca se utilizan para señalar un culpable en caso de fallo.
- ✓ La realización de pruebas es algo natural.

M1 – Introducción a Big Data

Modelos de negocio impulsados por los datos



Tipo A: Recolección y agregación de datos de acceso libre.

Tipo B: Servicios de analítica de datos.

Tipo C: Generación y análisis de datos.

Tipo D: Descubrimiento de conocimiento sobre datos de acceso libre.

Tipo E: Servicios de agregación de datos.

Tipo F: Mezcla y análisis de datos de múltiples fuentes.

M1 – Introducción a Big Data

El dilema del directivo: Intuición vs. Datos

- ✓ Machine learning.
- ✓ La intuición como forma de tomar decisiones.
- ✓ La gestión basada en datos (Data Driven Management).
- ✓ Proceso de transición a la toma de decisiones basada en datos.
- ✓ Ser capaz de hacer las preguntas correctas.

M1 – Introducción a Big Data

Nuevas competencias y nuevas capacidades

- ✓ Las principales tareas del analista de datos:
 - Nueva mentalidad
 - Definir una estrategia
 - ¿Qué crear y qué comprar o alquilar?
 - Contar con un experto
 - Movilizar los recursos
 - Crear capacidades en la primera línea de gestión

Nuevos roles

- ✓ Chief Data Officer
Responsable de implantar la cultura
- ✓ Data Scientist
Analista de datos cualificado
- ✓ Data Steward
Gobierno de los datos

M2 – Bibliografía

- Schroeck, M., Analytics: el Uso de Big Data en el Mundo Real. Consultado el 8 de diciembre de 2020. IBM Institute for Business Value, 2012.
- BBVA Innovation Edge, Big Data: Es Hora de Generar Valor de Negocio con los Datos. Consultado el 27 de junio de 2020.
- Dasu T., Johnson T., Exploratory Data Mining and Data Cleaning, Wiley, 2003.
- Fayyad, U., Shapiro, G., From Data Mining to Knowledge Discovery in Databases, En: AI Magazine 17.3 (1996), págs. 37-54.
- <https://www.enmimaquinafunciona.com/pregunta/29364/como-conseguir-excel-para-interpretar-la-coma-como-delimitador-predeterminado-en-archivos-csv>
- Imágenes:
- <https://www.yorokobu.es/big-data-2/>
- <http://anband.spb.ru/undertale-mobile>
- <https://kawak.net/wp-content/uploads/2018/11/beneficios.png>