

Digitalización aplicada a los sectores productivos

María Gracia López Olivencia



GRADO SUPERIOR

Paraninfo
ciclos formativos

UNIDAD 5 Evaluación de datos

Contenidos

- 5.1. Datos *versus* información
- 5.2. Ciclo de vida de los datos
- 5.3. *Big data*. Análisis de los datos
- 5.4. Almacenamiento de datos en la nube
- 5.5. Etapas de la ingeniería de datos
- 5.6. Aplicación a las empresas de la ciencia de datos
- 5.7. Herramientas para analizar los datos

Resultados de aprendizaje (RA) y criterios de evaluación (CE)

5

Evalúa la importancia de los datos, así como su protección en una economía digital globalizada, definiendo sistemas de seguridad y ciberseguridad tanto a nivel de equipo/sistema, como globales.

CE

- a Se ha establecido la diferencia entre «dato» e «información».
- b Se ha descrito el ciclo de vida del dato.
- c Se ha identificado la relación entre *big data*, análisis de datos, *machine/deep learning* e inteligencia artificial.
- d Se han descrito las características que definen *big data*.
- e Se han descrito las etapas típicas de la ciencia de datos y su relación en el proceso.

- f Se han descrito los procedimientos de almacenaje de datos en la nube.
- g Se ha descrito la importancia del *cloud computing*.
- h Se han identificado los principales objetivos de la ciencia de datos en las diferentes empresas.
- i Se ha valorado la importancia de la seguridad y su regulación en relación con los datos.

5.1. Datos *versus* información

Los datos guían la toma de decisiones y se usan para obtener más información de los clientes, ganar más consumidores y optimizar procesos empresariales.

El ecosistema de datos es un grupo de elementos que interactúan entre sí para producir, analizar y compartir datos. Son gestionados por hardware, software y analistas.



Figura 5.1. El analista de datos es una persona con formación en bases de datos y lenguajes de programación, con conocimientos matemáticos y estadísticos y con una gran capacidad para trabajar con grandes volúmenes de datos.

Estructura de datos	Nivel de acceso (seguridad)	Tipo de información (estadístico)
Datos simples. Es un dato propio e indivisible; por ejemplo, el año de nacimiento.	Datos restringidos. Tienen la categoría de mayor control de seguridad. Si estos datos se borran o se modifican pueden dañar significativamente a la compañía o al particular. Por ejemplo, los datos gubernamentales.	Cuantitativa o numérica. Es un dato que se puede medir y contar; por ejemplo, la temperatura.
Datos semicomuestos. Este dato se representa mediante una estructura parcial o variable, ya que puede ser modificada según el contexto. Por ejemplo, una página web.	Datos privados. Tienen la categoría media de control de seguridad en cuanto a su divulgación, modificación o borrado. Por ejemplo, el nombre de los empleados.	Cualitativa nominal. Es un dato no numérico que no sigue un orden. Por ejemplo, los tipos de podcasts, informática, ciencia, psicología, cocina.
Datos compuestos. Es la unión de dos o más datos simples o compuestos con una estructura fija. Por ejemplo, una radiografía bucal.	Datos públicos. Tienen una nula o muy pequeña categoría de control de seguridad en cuanto a su divulgación, modificación o borrado. Por ejemplo, la publicación en redes sociales, prensa y otros medios de comunicación.	Cualitativa ordinal. Es un dato no numérico susceptible de ser ordenado. Por ejemplo, las calificaciones de los alumnos: insuficiente, suficiente, bien, notable, sobresaliente, matrícula de honor.

Tabla 5.1. Los tipos de datos se almacenan en bases de datos para posteriormente ser analizados e interpretados

5.1.1. Diferencias clave entre datos e información

Datos	Información
Colección de hechos.	Sitúa esos datos en contexto.
Los puntos de datos son individuales y, en ocasiones, no están relacionados.	La información traza esos datos para proporcionar una visión general de cómo encajan todos.
Los datos no dependen de la información.	La información depende de los datos.
Los datos no son suficientes para la toma de decisiones.	Se pueden tomar decisiones basadas en la información.
Los datos, por sí solos, no tienen sentido.	Cuando los datos se analizan e interpretan, se convierten en información significativa.
Los datos están en bruto y desorganizados.	La información está organizada.
Los datos suelen presentarse en forma de gráficos, números, cifras o estadísticas.	La información generalmente se presenta a través de palabras, lenguaje, pensamientos e ideas.

Tabla 5.2. Diferencias entre datos e información

5.2. Ciclo de vida de los datos

DLM o *Data Lifecycle Management* se refiere al ciclo de vida de los datos porque estos se crean, se desarrollan y evolucionan.

Figura 5.2.
Ejecución del ciclo
de vida de los
datos.

1. Planificación

- Tipos de datos
- Arquitectura

2. Captura

- Datos estructurados
- Datos no estructurados

3. Utilización

- Herramientas de protección
- Limpieza

4. Análisis

- Resolución de problemas
- Toma de decisiones
- Respaldo objetivo empresarial

5. Archivado

- Duración determinada

6. Borrado

- Diferentes unidades de almacenamiento
- Borrado seguro a bajo nivel
- Trituración de documentos físicos

5.2.1. Gestión del ciclo de vida de los datos

Las grandes cantidades de datos que generan los dispositivos hace que la revisión de esos datos sea esencial para optimizar su utilidad y reducir la posibilidad de que se produzcan errores.

Eliminar o almacenar datos al final de su vida útil garantiza que no consuman más recursos de los necesarios.

5.3. *Big data*. Análisis de los datos

Big data se refiere al enorme conjunto de datos que son difíciles de procesos por bases de datos tradicionales.

Estos datos se obtienen de sensores, actuadores y de internet.

5.3.1. Características del *big data*

- Volumen:** para gestionar los datos se recurre a nubes o proveedores de servicios.
- Velocidad:** los datos envejecen rápido, por eso se necesitan herramientas que permitan tomar decisiones a tiempo.
- Variedad:** se trata de los distintos tipos de datos, cómo procesarlos y extraerlos para obtener información.
- Veracidad:** los datos solo son útiles si están limpios, sin errores, son fiables y coherentes.
- Valor:** los macrodatos se consolidan como la nueva ventaja competitiva si se pueden convertir en información útil y de calidad.

Uso de herramientas de gestión de *big data* para optimizar las 5V

Herramientas sólidas de gestión del *big data* brindan a las empresas almacenes de datos masivos, su limpieza y un rápido procesamiento en tiempo real.

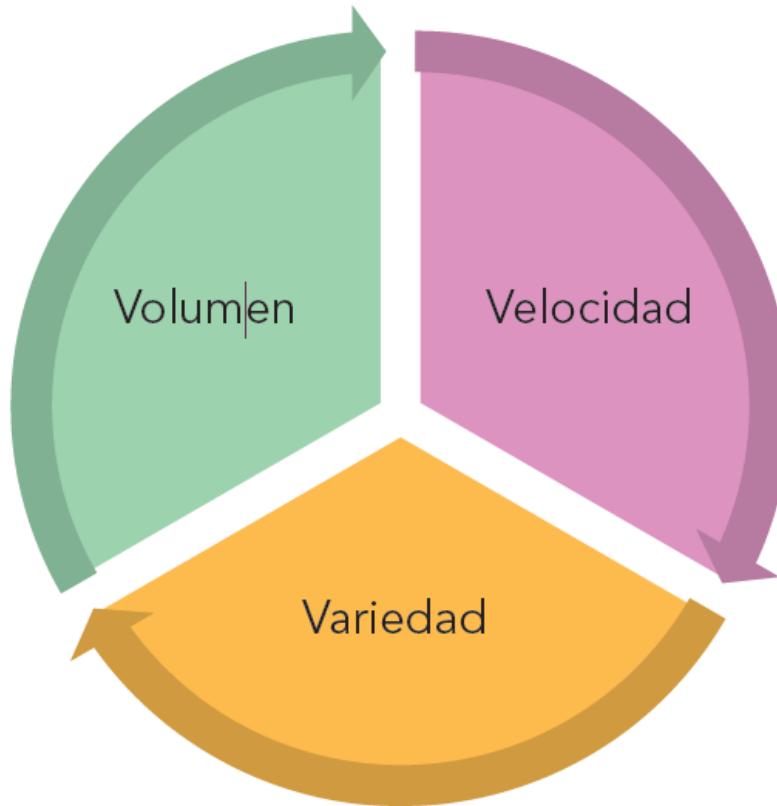


Figura 5.3. Las 3 V del *big data*: volumen en terabytes, registros y tablas; variedad de datos estructurados y no estructurados; velocidad del procesamiento por lotes y procesamiento en tiempo real.

5.3.2. Relación entre *big data*, análisis de datos, *machine/deep learning* e inteligencia artificial

La ciencia de los datos vuelve útiles los datos.

Disciplinas de la ciencia de los datos

Estadística: para tomar decisiones importantes con incertidumbre.

Machine learning e IA: para automatizar muchas decisiones con incertidumbre.

Deep learning: procesa grandes volúmenes de datos sin estructura para comenzar con su aprendizaje en profundidad.

El **análisis de datos** se encarga de recopilar, transformar y organizar los datos para obtener conclusiones y también permite hacer predicciones e impulsar una toma de decisiones informada.

5.3.3. Tipos de análisis de datos

Análisis descriptivo: se usa para conocer el estado actual de una compañía. Se analiza el histórico de datos para justificar las decisiones.

Análisis predictivo: se basa en el histórico de datos y los datos del presente para predecir el futuro. Se usa para saber si un producto va a funcionar o no.

Análisis prescriptivo: sugiere decisiones o acciones específicas para alcanzar un resultado. Se basa en análisis anteriores y el uso de técnicas avanzadas que pueden sugerir y hacer recomendaciones.



Figura 5.4. YouTube fue adquirida por Google por 1650 millones de dólares en el año 2006. Todo un éxito para una empresa que todavía no había celebrado su primer aniversario.

5.4. Almacenamiento de datos en la nube

La nube es una localización virtual que permite mantener y acceder a los datos en línea.

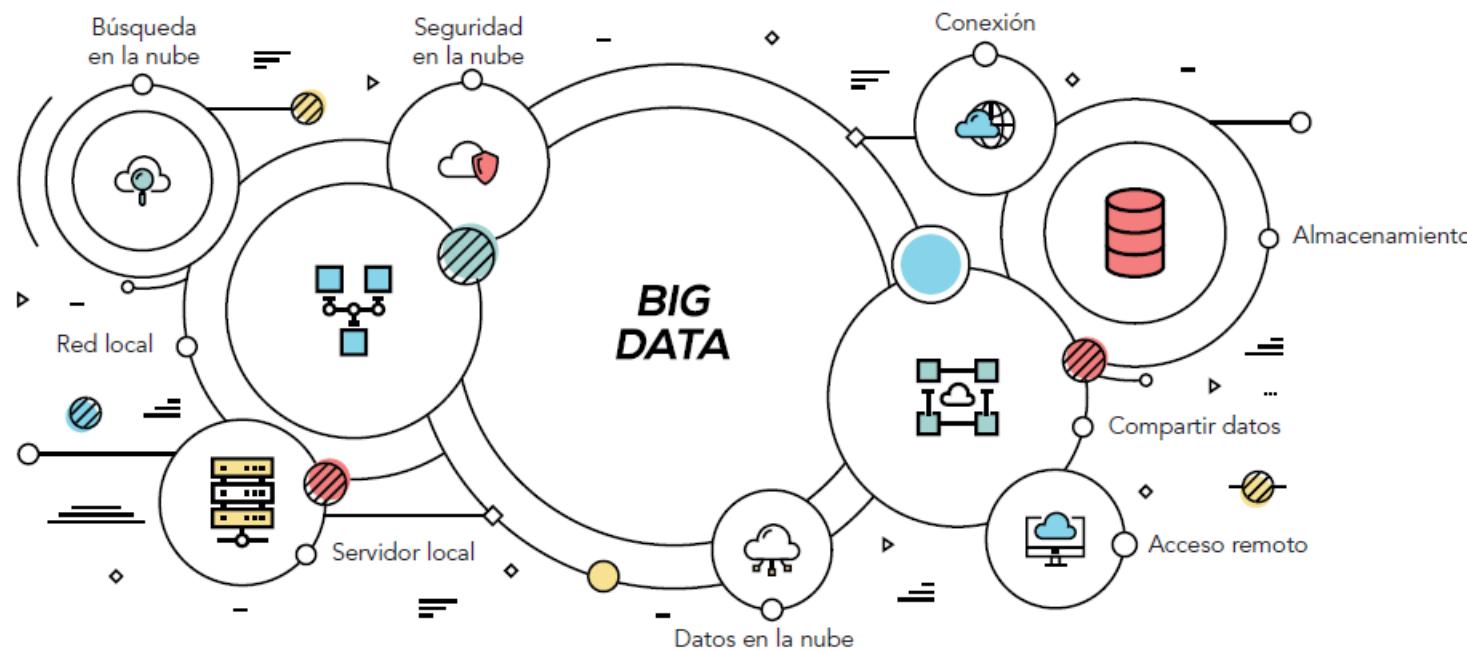
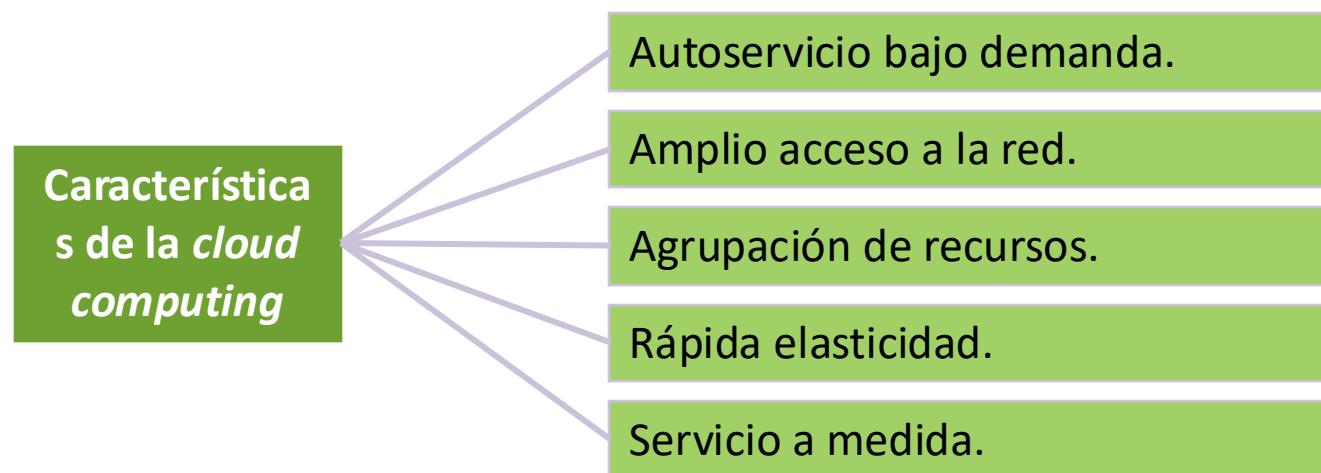


Figura 5.5. Ecosistema formado por *big data* y la *cloud computing*.

5.4.1. Importancia de la *cloud computing*

La computación en la nube permite un acceso de red cómodo y bajo demanda a un conjunto de recursos informáticos que no necesitan mucho esfuerzo de administración o interacción con el proveedor de servicios.



5.5. Etapas de la ingeniería de datos

Las etapas u objetivos de la ingeniería de datos se desarrollan para encontrar la respuesta a una pregunta de negocio.

La pregunta surge porque se ha identificado un problema, un incidente o una posible mejoría en el proceso empresarial.

No se deben confundir las etapas del proceso de análisis de datos con las etapas del ciclo de vida de los datos.

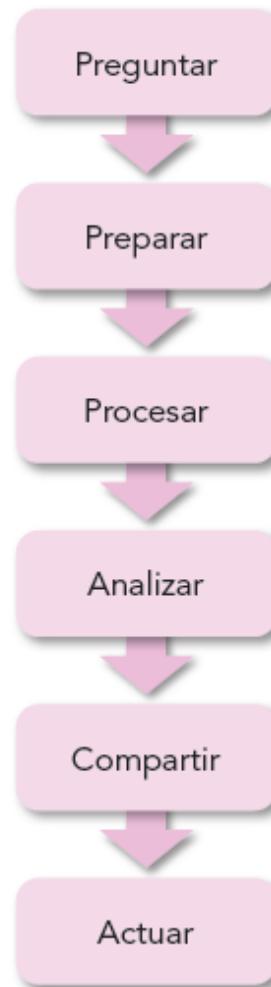


Figura 5.6. Etapas de la ingeniería de datos.

5.5.1. Escenarios en los que se puede utilizar la minería de datos

- Industria financiera:** Permite la detección del fraude con tarjetas de crédito.
- Fábricas:** Puede detectar mal funcionamiento de forma oportunidad, la necesidad de mantenimiento de los equipos y predecir el uso de energía en la fabricación.
- Educación:** Los datos pueden brindar apoyo al aprendizaje de cada estudiante de forma personalizada y desarrollar un plan de estudios basados en los datos recopilados.

5.6. Aplicación a las empresas de la ciencia de datos

5.6.1. Obtención de información sobre los clientes

Los datos acerca de la clientela se presentan de varias formas y se capturan por varias vías y cada dato contiene una clave muy valiosa.

Su análisis permite a la empresa reconocer las necesidades de sus clientes y comprender sus comportamientos, preferencias y motivaciones.

El verdadero obstáculo es desenredar esa compleja red de información que se obtiene de los clientes.



Figura 5.7. Muchas preguntas pueden llevar a un acierto.

5.7. Herramientas para analizar los datos

Hojas de cálculo: se pueden realizar operaciones matemáticas, aplicar fórmulas simples o complejas y funciones.

- Microsoft Excel
- Google Sheets

Lenguajes de consultas: las consultas son preguntas que se lanzan a una base de datos formada por tablas que contiene atributos. Las tablas se relacionan entre ellas para que se puedan producir las consultas. También se pueden insertar y eliminar datos.

Herramientas de visualización: sirven para representar la información.

- Tableau
- Looker Studio

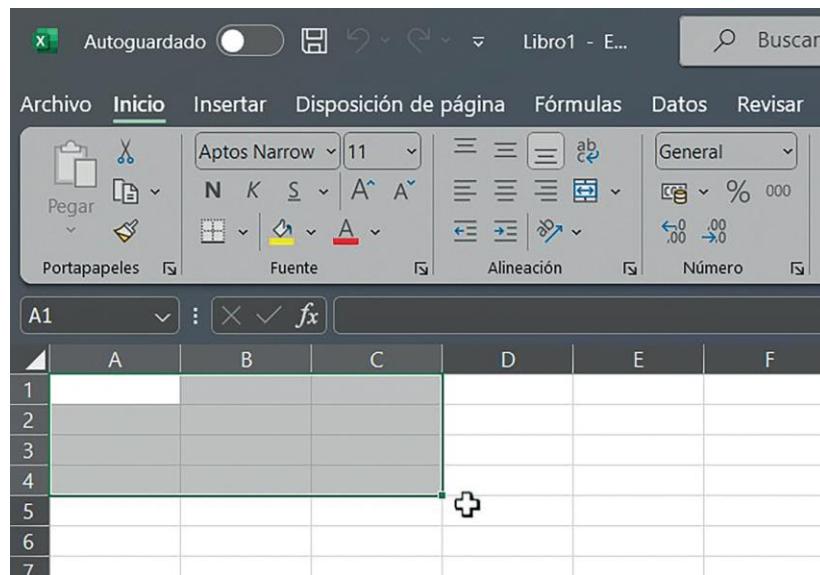


Figura 5.8. Menú principal de Microsoft Excel.



Figura 5.9. Google Sheets cada vez se emplea más en los dispositivos móviles.

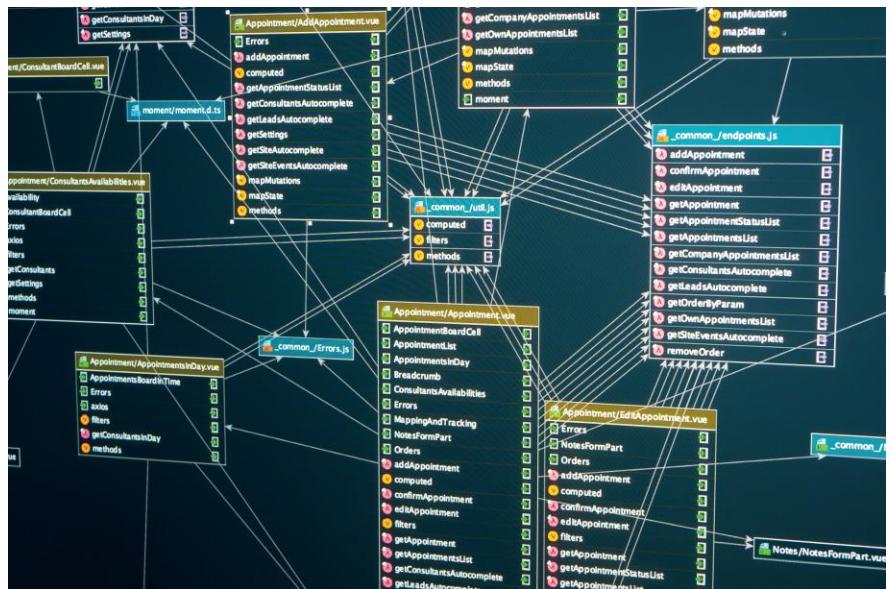


Figura 5.10. Relaciones entre entidades de una base de datos MySQL Server.

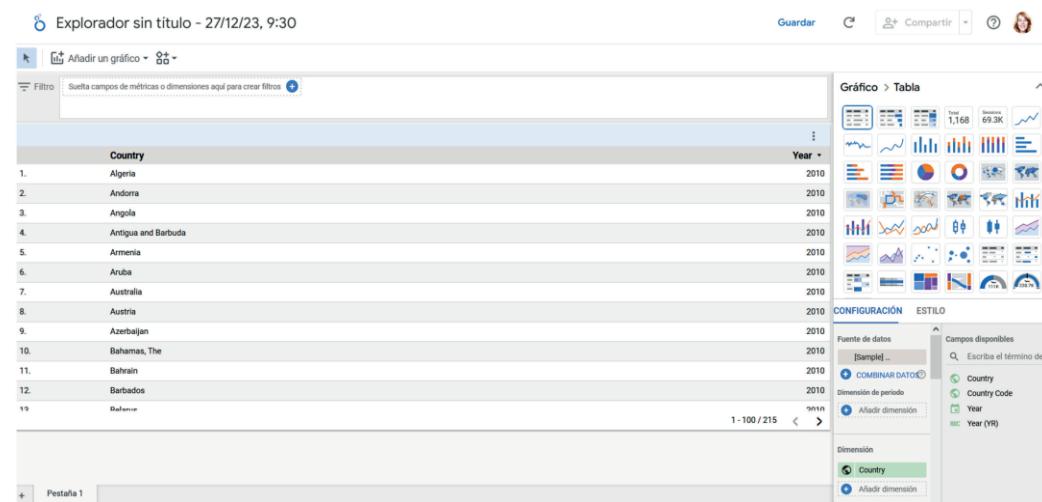


Figura 5.11.

Preparación de un informe utilizando el menú de Looker Studio.