



IFCT128PO: Big data

Unidad 2: La importancia del dato



ÍNDICE

INTRODUCCIÓN.....	3
OBJETIVOS / CAPACIDADES.....	4
1. CONTEXTUALIZACIÓN PRÁCTICA DE LA PRODUCTIVIDAD DEL DATO.....	5
2. TIPOLOGÍA DE LOS DATOS.....	7
3. TRATAMIENTO DEL DATO.....	9
3.1. Estructura arquitectónica en big data.....	9
RESUMEN.....	11
MAPA CONCEPTUAL.....	12
RECURSOS PARA AMPLIAR.....	13
BIBLIOGRAFÍA.....	14
GLOSARIO.....	15

INTRODUCCIÓN

A lo largo de esta unidad realizaremos una **contextualización de la relevancia que ha ido adquiriendo el dato** en nuestra sociedad.



Si bien es cierto que hace unos años el **concepto de dato** se relacionaba únicamente con grandes empresa y el uso de importantes tecnologías, con el devenir de los años, en este siglo XXI hemos ido asistiendo de una **amplificación cualitativa y cuantitativa** de la imagen del dato, extrapolándose su uso a multitud de situaciones en nuestra vida cotidiana.

Para ello, y con el aumento de datos que se generan, es necesario establecer cierto **orden en ellos** y saber identificar cuáles son datos **estructurados, no estructurados** o **híbridos** para poder tratarlos de manera más coherente y eficaz para obtener una mayor rentabilidad y transformarlos en información útil.



OBJETIVOS / CAPACIDADES

En esta unidad de aprendizaje, las capacidades que más se va a trabajar son:

- ✓ Tomar consciencia de la utilidad de la gestión de los datos en un entorno social, económico y empresarial.
- ✓ Saber diferenciar los diferentes tipos de datos con los que podemos trabajar y las fuentes desde las que podemos extraerlos.
- ✓ Conocer cómo deben tratarse estos datos y cómo debemos realizar el proceso de ejecución de ese tratamiento.



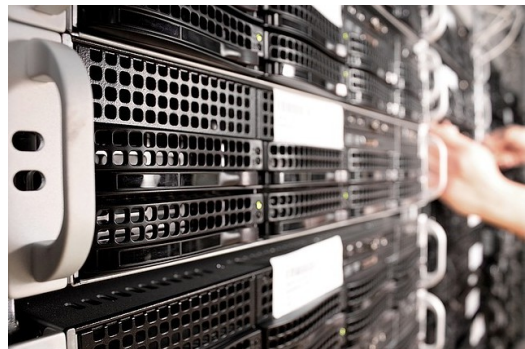
1. CONTEXTUALIZACIÓN PRÁCTICA DE LA PRODUCTIVIDAD DEL DATO



La historia moderna está llena de acontecimientos en los que los datos, aunque sean en cantidades ínfimas, nos han dado la oportunidad de mirar los acontecimientos desde una perspectiva más acertada o clarificadora.

Desde el papiro o los volúmenes ilustrados, los datos siempre han sido una de las claves para **acercarnos a un mayor conocimiento**, con el avance que eso supone tanto a nivel económico como humano.

Algunos de los rasgos del desarrollo de los datos:



→ Crecimiento vertiginoso:

En el siglo XXI, la conciencia sobre la importancia y la necesidad de los datos ha aumentado a un ritmo vertiginoso. El volumen masivo de datos, unido a un muy asequible coste de almacenamiento, ha facilitado todo esta acogida del big data, y los científicos de datos están encontrando nuevas herramientas vanguardistas que proporcionan un conocimiento más cualitativo gracias a estas grandes cantidades de datos.

Por ejemplo:



Si hacemos una comparación histórica, el censo de los E.E.U.U del año 1880 tuvo una duración de tabulación de ocho años y atendiendo a los recursos de la época, los expertos aseguran que se hubiera tardado 10 años en realizar el de 1890. Sin embargo, como ya comentábamos, el avance de los medios a lo largo de los siglos ha conseguido que tan solo en 2017 el volumen de datos generados al día fuera de 2.500 millones de gigas.

→ Extensión del uso de datos:

Conforme estas nuevas tecnologías adquieren virtudes más transformadoras, sus efectos son más plenos y las oportunidades cada vez más crecientes.

Tanto es así, que su uso ya no solo está destinado únicamente a grandes fines empresariales. Los datos predictivos ya se utilizan para ayudarnos a cosas tan cotidianas como saber de antemano si debemos llevar paraguas al trabajo o, por ejemplo, en el caso de los datos de tráfico, para sincronizar los semáforos.

→ La incidencia de la productividad:

Sin duda, la innovación de los datos está demostradamente ligada al aumento de productividad. Según un estudio realizado por BSA, “las empresas que ya están realizando una toma de decisiones guiada por los datos informan un aumento de entre el 5 y 6 por ciento en la productividad”.

Los economistas aseguran que utilizando los datos de manera efectiva podemos alcanzar ganancias, aunque sean en un pequeño margen, un número significativo de industrias podrían conseguir ser al menos un 1% más eficientes. Tan solo este porcentaje conseguiría que se sumara aproximadamente 15 billones de dólares al PIB global de aquí al año 2030.



Además, en los próximos cuatro años, la optimización en el uso de los datos podría generar más de 1 billón y medio de dividendos de datos en el planeta.

Por tanto, podemos afirmar que los datos alimentan de manera efectiva negocios más competitivos, con mayor capacidad de respuesta, más productivos y más rentables.



En **Recursos para ampliar** podrás encontrar información sobre los beneficios del Big Data en la empresa.

2. TIPOLOGÍA DE LOS DATOS

Como hemos comentado en la unidad anterior, uno de elementos que nos ayudarán a definir si nuestro proyecto debe realizarse a través del análisis big data es la **variedad de los datos**.



Los datos se rigen por una clasificación muy básica, por la cual debemos atenernos a si los datos de los que disponemos y creemos conveniente usar en nuestro proyecto son estructurados o no estructurados. No obstante, también existen los datos híbridos que se mueven entre las dos categorías anteriores.

A modo didáctico, vamos a realizar un análisis para conocer cuál es la diferencia entre ambos y cuáles son las fuentes de las que pueden obtenerse.

→ Estructurados:

Poseen un formato predefinido por lo que los campos que contiene ocupan un sitio estático, fijo. Comúnmente se almacenan en tablas.



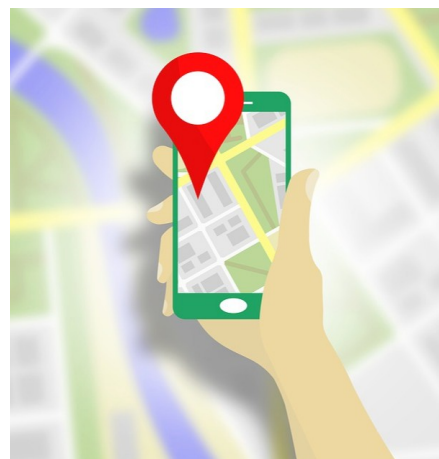
- ✓ **Generados por la empresa:** Tablas, ficheros XML...
- ✓ **Provocados:** Creados indirectamente en acciones previas. Por ejemplo, acciones de marketing antes del lanzamiento de un producto o valoraciones en foros.
- ✓ **Dirigidos por transacciones:** Datos generados tras una acción cerrada correctamente. Por ejemplo, tickets de compra, recibos...
- ✓ **Compilados:** Datos provenientes de empresas o servicios públicos, pero siempre a nivel grupal. Por ejemplo, viviendas públicas o el propio censo electoral
- ✓ **Experimentales:** Los datos que se generan en el proceso de un análisis.

→ No estructurados:

No poseen un formato determinado. Esto lleva a que sea imposible su organización en bases de datos. Pueden venir, por ejemplo, de PDF o correo electrónico, entre otras opciones.

✓ **Capturados:** Se generan a través de los movimientos de los usuarios: GPS, pulseras que recogen información biométrica...

✓ **Provenientes de usuarios:** Datos que genera el usuario voluntariamente a través de acciones y que quedan inevitablemente registrados: los vídeos que ve en Internet, las búsquedas que hace...



→ Semiestructurados o híbridos:

Mezclan información desestructurada pero que recogen algunos marcadores para diferenciar los diferentes datos.

No hay una manera estándar de trabajarla. Suelen gestionarse en XML o JSON.

Ejemplo:



En base a todo lo expuesto, podemos decir que los datos nos muestran un entorno concreto pero con una amplitud diferente relacionando todo tipo de información. Por poner un ejemplo, de Twitter podemos extraer tanto datos como sexo, edad... hasta datos referentes a los post o los seguidores de un usuario.



*En **Recursos para ampliar** podrás encontrar más información sobre los datos no estructurados y la diferencia entre estos y los estructurados.*



ACTIVIDAD

Actividad de aprendizaje 1

Para comenzar un análisis exhaustivo de los datos con los que contamos para nuestro proyecto, deberemos tener en cuenta antes de empezar si esos datos poseen un formato predefinido o no. En base a esto, realiza una búsqueda de ejemplos prácticos de cada tipología de datos (con imágenes incluidas) y debate con tus compañeros cómo influiría la mezcla de datos en nuestro análisis sin tener en cuenta esta clasificación. ¿Tendría coherencia de cara a la extracción de conocimiento?

3. TRATAMIENTO DEL DATO

Como hemos visto, la bajada de costes del almacenamiento de datos ha influido notablemente en la aparición de **nuevas tecnologías dedicadas al procesamiento de grandes volúmenes de datos**.

Así, con el **Internet de las Cosas (IoT)** comienzan a surgir nuevos datos gracias a las conexiones de nuevos dispositivos y objetos que antes eran sencillos y que ahora llevan incorporados chips y sensores para recoger datos.



Todo esto contribuye, lógicamente, a realizar de manera más sencilla y precisa análisis prescriptivos y productivos.

3.1. Estructura arquitectónica en big data



La arquitectura más productiva es aquella que se organiza en capas permitiéndonos pulir, en cierto modo, la complejidad de ese gran volumen de datos y su tratamiento.

En esta metodología por capas hace que en cada una de ellas se lleve a cabo una **función exacta** logran que los datos se canalicen en función de lo que el sistema vaya necesitando a través de flujos más ordenados de información.

No obstante, para que este proceso se realice de manera segura, esta arquitectura debe contar al menos con las capas que se especifican a continuación:



- **Capa 1. Ingestión de datos:** en este primer paso se clasifican los datos de diversas fuentes y se dividen en función de la prioridad.
- **Capa 2. Recolección de datos:** se pone especial atención en esta capa, a la transacción de datos desde la primera, la capa 1, al resto del flujo y canalización de los datos.
- **Capa 3. Procesamiento de datos:** es el primer punto donde comienza realmente la analítica.
- **Capa 4. Almacenamiento de datos:** en esta etapa deberemos tener en cuenta el crecimiento progresivo de los datos para no errar en la solución de almacenamiento escogida.
- **Capa 5. Consulta de datos:** comienza aquí la analítica sólida ya que es en este punto donde se examina el valor de los datos y su utilidad.
- **Capa 6. Visualización:** en esta última capa, el dato ya es conocimiento y, por tanto, los usuarios ya pueden tener su valor exacto. Esto facilitará al usuario la toma de decisiones y la innovación en la resolución de sus objetivos.

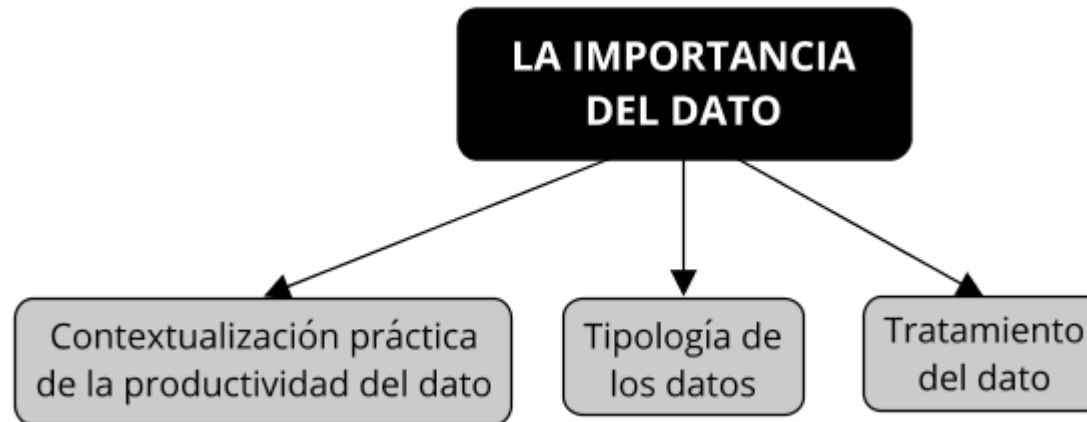
RESUMEN

Para extraer unas conclusiones claras de esta unidad referente a la **importancia de los datos** podemos decir que la relevancia de este concepto **ha ido creciendo con el paso de los años**, la toma de conciencia de la utilidad de un análisis de calidad de los mismos y la aparición de nuevas tecnologías y sistemas que nos facilitan esta recopilación y analítica.

Por otro lado, una de las claves fundamentales en este punto es **tomar conciencia de que la generación masiva de datos nos obliga a realizar una clasificación** previa al análisis, sin la cual nos será muy complicado obtener los resultados esperados. Tras esta clasificación es cuándo podremos proceder al tratamiento de los mismos para extraer conocimiento y transformarlos en información para una acertada y productiva toma de decisiones.

MAPA CONCEPTUAL

Para concluir la unidad la podemos resumir los **apartados e ideas principales** por medio de este mapa conceptual:



RECURSOS PARA AMPLIAR



TEXTOS ELECTRÓNICOS

- Beneficios del Big Data en la empresa: <https://aggity.com/beneficios-big-data-empresa/> [Consulta agosto 2021].
- Diferencia entre datos estructurados y no estructurados: <https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/insights-hub/articles/diferencia-entre-datos-estructurados-y-no-estructurados.html> [Consulta agosto 2021].



PÁGINAS WEB

- Datos no estructurados: <https://www.youtube.com/watch?v=b-RqDLoBVoY> [Consulta agosto 2021].





TEXTOS ELECTRÓNICOS

- ¿Por qué son tan importantes los datos?: http://data.bsa.org/wp-content/uploads/2015/10/BSADataStudy_es.pdf [Consulta agosto 2021].
- ¿Qué es el almacenamiento de datos?: <https://www.hpe.com/es/es/what-is/data-storage.html> [Consulta agosto 2021].
- ¿Qué es el procesamiento de [datos](https://www.nextu.com/blog/que-es-el-procesamiento-de-datos/)? <https://www.nextu.com/blog/que-es-el-procesamiento-de-datos/> [Consulta agosto 2021].
- Nuevos cursos, tecnológicas, novedades del sector y toda la actualidad Ibertech: <https://www.ibertech.org/big-data-y-analytics-el-tratamiento-del-dato/> [Consulta agosto 2021].
- Qué tipos de datos hay en Big Data: <https://agenciab12.com/noticia/tipos-de-datos-big-data> [Consulta agosto 2021].
- Tratamiento de datos de big data en el internet de las cosas: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/tratamiento-de-datos-de-big-data-en-en-el-internet-de-las-cosas> [Consulta agosto 2021].



GLOSARIO

- **Almacenamiento de datos:** El almacenamiento de datos se refiere a la utilización de medios de grabación para conservar los datos utilizando ordenador y otros dispositivos. Algunas de las maneras más comunes de almacenamiento de datos son: el almacenamiento en archivo, almacenamiento en bloque y el de objetos.
- **Datos estucturados:** Datos que poseen un formato predefinido por lo que los campos que contiene ocupan un sitio estático, fijo. Comúnmente se almacenan en tablas.
- **Datos híbridos:** Datos que mezclan información desestructurada pero que recogen algunos marcadores para diferenciar los diferentes datos.
- **Datos no estructurados:** Datos que no poseen un formato determinado. Esto lleva a que sea imposible su organización en bases de datos. Pueden venir, por ejemplo, de PDF o correo electrónico, entre otras opciones.
- **Ingestión de datos:** Proceso en el cual se clasifican los datos de diversas fuentes y se dividen en función de la prioridad.
- **Procesamiento de datos:** Acumulación y manipulación de elementos de datos para generar información relevante que facilita la toma de decisiones.

