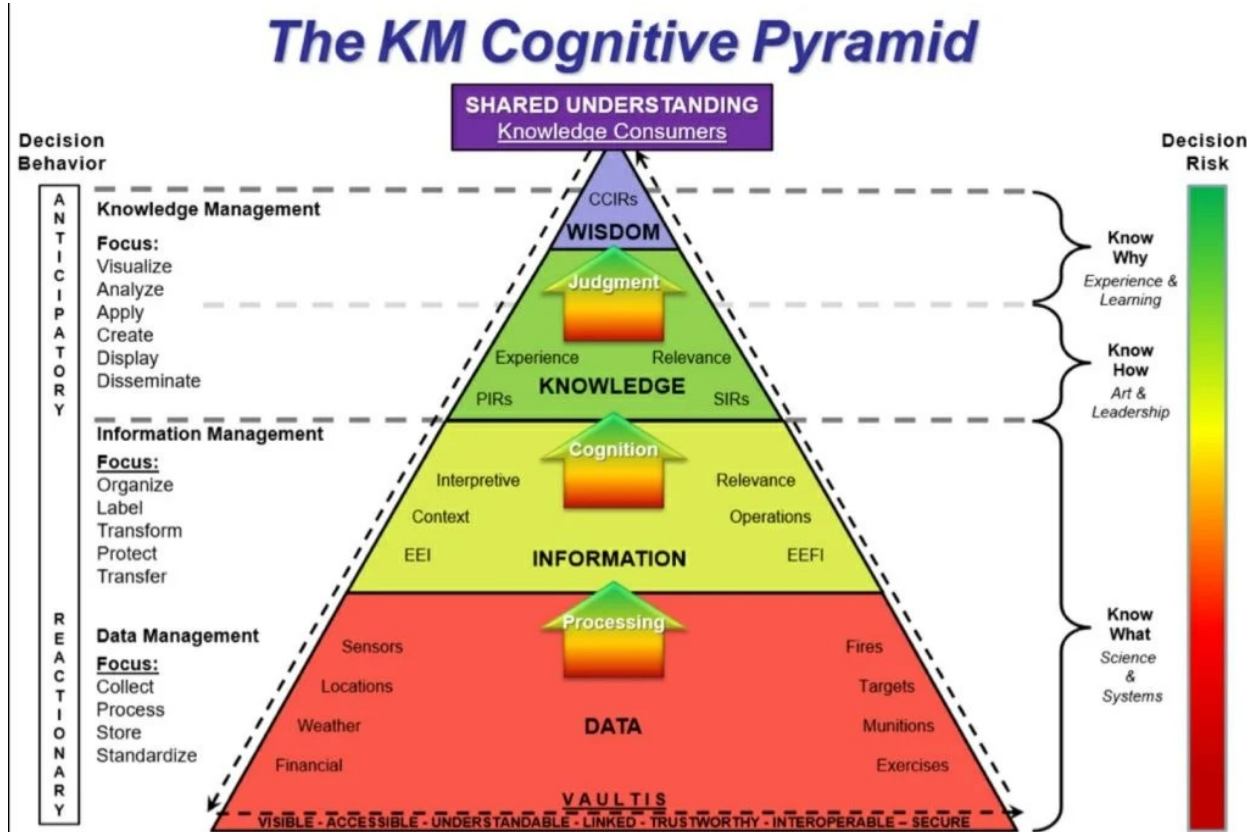


BIG DATA:

Del Dato a la Información

Datos, Información y Conocimiento



<https://keepcoding.io/blog/que-es-y-como funciona-dikw/>

Datos, Información y Conocimiento

¿Cómo funciona la pirámide **DIKW**?

La **pirámide DIKW** representa un proceso ascendente, donde los datos se transforman en información, luego en conocimiento y finalmente en sabiduría. Cada nivel construye sobre el anterior, y la sabiduría es el resultado de la integración de todos los niveles.

Datos, Información y Conocimiento

Ejemplo:

Imagina que tienes una base de datos con información sobre las ventas de tu empresa.

Datos: Cada registro de venta es un dato individual (fecha, producto, cantidad, precio).

Información: Al analizar los datos, puedes generar informes que muestren las ventas totales por mes, los productos más vendidos o los clientes más recurrentes.

Datos, Información y Conocimiento

Conocimiento: A partir de estos informes, puedes identificar tendencias en las ventas, predecir la demanda futura y ajustar tus estrategias de marketing.

Sabiduría: Utilizando este conocimiento, puedes desarrollar nuevos productos, mejorar la experiencia del cliente y tomar decisiones estratégicas para el crecimiento de tu negocio.

Datos, Información y Conocimiento

Desde el punto de vista de las tecnologías de la información a medida que ascendemos en la pirámide nos enfrentamos con conceptos menos programables y susceptibles de ser manipulados mediante algoritmos.

Datos, Información y Conocimiento

Desde el punto de vista de las tecnologías de la información a medida que ascendemos en la pirámide nos enfrentamos con conceptos menos programables y susceptibles de ser manipulados mediante algoritmos.

Vamos a ver los roles involucrados en cada uno de los procesos en que los datos se convierten en información y luego en conocimiento.

BIG DATA:

Caracterización del Dato

Caracterización del Dato

- Dato en cuanto al tipo.
- Dato en cuanto al formato.
- Dato en cuanto al generador.
- Dato en cuanto al tamaño.
- Dato en cuanto a su rol.
- Dato en cuanto a su latencia.
- Dato en cuanto a su sensibilidad.

Dato en cuanto al TIPO

- **Tipos simples:** Representan un único valor. Estos se dividen en lógicos, caracteres y numéricos, cada uno de ellos representado por un número de bits.

Dato en cuanto al TIPO

- **Tipos complejos:** Son el resultado de la combinación de los tipos simples y aparecen tipos compuestos, que representan un conjunto de valores a modo de estructura.

Con tipos compuestos podemos representar la imagen de la matrícula de un coche, en forma de matriz de bits, etc.

Dato en cuanto a su FORMATO

- **Datos estructurados:** Una colección de datos está estructurada cuando presenta un modelo o esquema organizativo. Una idea que se nos puede venir a la cabeza como ejemplo de datos estructurados es la de una base de datos SQL, donde las colecciones se hacen presentes mediante tablas y las relaciones como referencias.

Por ejemplo: un cliente en una tabla puede estar almacenado como tipo compuesto en una tabla, constituyendo un registro y estará compuesto por un conjunto de datos simples.

Dato en cuanto a su FORMATO

- **Datos semiestructurados:** Estos datos no son estructurados, pero que presentan cierta organización.

Por ejemplo: **MIME** (Multipurpose Internet Mail Extension), **XML**(eXtensible Markup Language) y **JSON** (JavaScript Object Notation).

Dato en cuanto a su FORMATO

- **Datos no estructurados:** Los datos no estructurados se caracterizan por su formato variable y la dificultad para almacenarlos en bases de datos relacionales tradicionales.

Dato en cuanto a su FORMATO

Ejemplos de datos NO ESTRUCTURADOS.

- Mensajes de Blog
- Imágenes de redes sociales.
- Datos geoespaciales.
- Lecturas de sensores de una ciudad.
- Etc...

Dato en cuanto a su GENERADOR

- **Datos generados por personas:** Estos son los que más rápidamente están creciendo por el volumen de interacciones en las redes sociales y el comercio electrónico.

Dato en cuanto a su GENERADOR

Ejemplos de datos generados por personas:

- Información sobre nuestro día a día online.
- Evaluaciones online de productos.
- Blog y vídeos subidos por usuarios.
- Publicaciones en redes sociales.
- Texto en la red social X.
- Información recopilada de la apps móviles...

Dato en cuanto a su GENERADOR

- **Datos generados por máquinas:** En esencia, los datos generados por máquinas se refieren a la información creada, capturada y transmitida por dispositivos, sensores y sistemas automatizados, sin intervención humana directa.

Dato en cuanto a su GENERADOR

Ejemplos de datos generados por máquinas:

- Lectura de sensores.
- Registros de transacciones comerciales online.
- Registros de accesos web.
- Datos generados por dispositivos IoT.
- Etc...

Dato en cuanto a su TAMAÑO

Ejemplos:

- La **biblioteca del ICMAT** (Instituto de Ciencias Matemáticas), con 15.000 volúmenes, se asocia a 75 GB.
- La **Bolsa de Nueva York**, antes de la llegada de los sistemas de trading algorítmico, generaba **1 TB** de información diaria.

Dato en cuanto a su TAMAÑO

Ejemplos:

- Se estima que **Google** procesa **20 PB** de información al día. Esto significa que un solo petabyte representa la cantidad de datos que Google procesa en una hora.
- El tráfico anual de internet en **2015** se estimó en **2021 EB.**

Dato en cuanto a su TAMAÑO

Unidades de almacenamiento:

- 1 kilobyte (kB) = 1.024 bytes (B)
- 1 megabyte (MB) = 1.024 kB = **1.024²** = 1.048.576 B
- 1 gigabyte (GB) = 1.024 MB = **1.024³** = 1.073.741.824 B
- 1 terabyte (TB) = 1.024 GB = **1.024⁴** = 1.099.511.627.776 B
- 1 petabyte (PB) = 1.024 TB = **1.024⁵** = 1.125.899.906.842.624 B
- 1 exabyte (EB) = 1.024 PB = **1.024⁶** = 1.152.921.504.606.846.976 B
- 1 zettabyte (ZB) = 1.024 EB = **1.024⁷** = 1.180.591.620.717.411.303.424 B
- 1 yottabyte (YB) = 1.024 ZB = **1.024⁸** = 1.208.925.819.614.629.174.706.176 B

Dato en cuanto a su TAMAÑO

El Impacto del Tamaño en la Tecnología:

- Sistemas de almacenamiento.
- Capacidad de procesamiento.
- Bases de datos NoSQL.

Dato en cuanto a su TAMAÑO

El Impacto del Tamaño en el ANÁLISIS de DATOS:

- Complejidad del análisis.
- Necesidad de escalabilidad.
- Importancia de la visualización.

Dato en cuanto a su ROL

- **Datos maestros:** Los datos maestros se refieren a un conjunto de información **consistente** y **unificada** sobre entidades **clave** de una organización. Estos datos son **esenciales** para las operaciones comerciales y se comparten entre diferentes áreas y sistemas.

Dato en cuanto a su ROL

Ejemplos de datos maestros:

- Clientes
- Productos
- Proveedores
- Empleados
- Ubicaciones

Dato en cuanto a su ROL

Características de los datos maestros:

- Consistencia.
- Uniformidad.
- Centralización.
- Actualización.

Dato en cuanto a su ROL

Beneficios de los datos maestros:

- Mejora de la calidad de los datos.
- Facilitación de la integración de datos.
- Obtención de una visión unificada del negocio.
- Toma de decisiones más informada.

Dato en cuanto a su ROL

- **Datos operacionales:** En el contexto empresarial, los datos operacionales se refieren a la información generada por las operaciones diarias del negocio.

Dato en cuanto a su ROL

Ejemplos de datos operacionales:

- Datos de transacciones de clientes.
- Datos de inventario.
- Datos de navegación web.

Dato en cuanto a su ROL

- **Datos externos:** Se pueden entender como **aquellos datos generados y recopilados fuera del propio negocio**, pero que tienen relación él, siendo susceptibles de influir y aportar valor.

Dato en cuanto a su ROL

Estos datos pueden provenir de una amplia variedad de fuentes, incluyendo:

- Redes sociales.
- Blogs.
- Sitios webs.
- Dispositivos móviles.
- Información financiera.

Dato en cuanto a su ROL

- **Datos analíticos:** Aquí ya se habla de información en lugar de datos. El dato analítico se genera a partir de los datos operacionales dentro del contexto de los datos maestros, denominados **dimensiones**, y relatado a lo largo de una perspectiva temporal.

Dato en cuanto a su LATENCIA

La latencia, en el contexto de los datos, se refiere **al tiempo que transcurre entre la solicitud de un dato y su disponibilidad para su uso**. Es un factor crítico en muchas aplicaciones, especialmente en aquellas que requieren respuestas en tiempo real o procesamiento de alta frecuencia.

Dato en cuanto a su LATENCIA

- **Datos en tiempo real:** Los datos en tiempo real son aquellos que se procesan y analizan a medida que se generan, permitiendo tomar decisiones y acciones inmediatas. El **procesamiento** en tiempo real es una característica inherente a los **datos operacionales**.

Dato en cuanto a su LATENCIA

Ejemplos de datos en tiempo real:

- Monitorización de seguridad en redes.
- Comercio electrónico.
- Redes sociales.
- Sistemas de transporte.
- ETC...

Dato en cuanto a su LATENCIA

Beneficios de los datos en tiempo real:

- Toma de decisiones más rápida e informada.
- Mejor experiencia del cliente.
- Mayor eficiencia operativa.
- Innovación.

Dato en cuanto a su LATENCIA

- **Datos en lotes:** Los datos son almacenados en un lote (**batch**) cuando son recibidos, permaneciendo así durante un periodo de tiempo o hasta que alcancen un volumen determinado. Por ejemplo, una empresa puede procesar las transacciones de ventas del día al final del día, o los datos de sensores cada hora.

Dato en cuanto a su LATENCIA

Ventajas de procesar datos en lotes:

- Eficiencia para grandes volúmenes de datos.
- Simplificación de la infraestructura.

Dato en cuanto a su LATENCIA

Desventajas de procesar datos en lotes:

- Mayor latencia.
- Menor flexibilidad.

Dato en cuanto a su SENSIBILIDAD

Son los datos clasificados en términos de **privacidad e intimidad**. Esto marca quien accede a estos datos y durante cuanto tiempo, estableciéndose también las condiciones en que deben ser almacenados.

Dato en cuanto a su SENSIBILIDAD

Son los datos clasificados en términos de **privacidad e intimidad**. Esto marca quien accede a estos datos y durante cuanto tiempo, estableciéndose también las condiciones en que deben ser almacenados.

Dato en cuanto a su SENSIBILIDAD

Las empresas suelen categorizar estos datos en cuanto a su sensibilidad y acceso, pero además estas tienen en cuenta tanto datos recogidos de las interacciones con los clientes y proveedores, como aquellos generados por el propio negocio.

Dato en cuanto a su SENSIBILIDAD

Varían de un país a otro en función de las regulaciones que se apliquen:

- **GDPR.**

- **SOC2.**

- **HIPAA.**

- **PCI SSC.**

Dato en cuanto a su SENSIBILIDAD

De forma genérica, podemos hablar de tres clases de de datos en términos del riesgo que supone su manipulación indebida:

- Riesgo alto.
- Riesgo medio.
- Riesgo bajo.

Dato en cuanto a su SENSIBILIDAD

Riesgo bajo	Riesgo medio	Riesgo alto
<ul style="list-style-type: none">• Ofertas de empleo.• Nota de prensa.• Material de marketing aprobado para uso público	<ul style="list-style-type: none">• Identificadores internos de empleados• Datos de desarrollo, investigación y patentes no publicadas.• Contenidos con propiedad intelectual licenciada de tercero o restringida por contrato.• Datos de recursos humanos relacionados con los empleados.	<ul style="list-style-type: none">• Identificadores personales (números de la Seguridad Social, DNI, pasaporte)• Contraseñas y claves de acceso a sistemas, aplicaciones, etc.• Información identificable sobre salud o pólizas.• Números de tarjetas de crédito o débito.