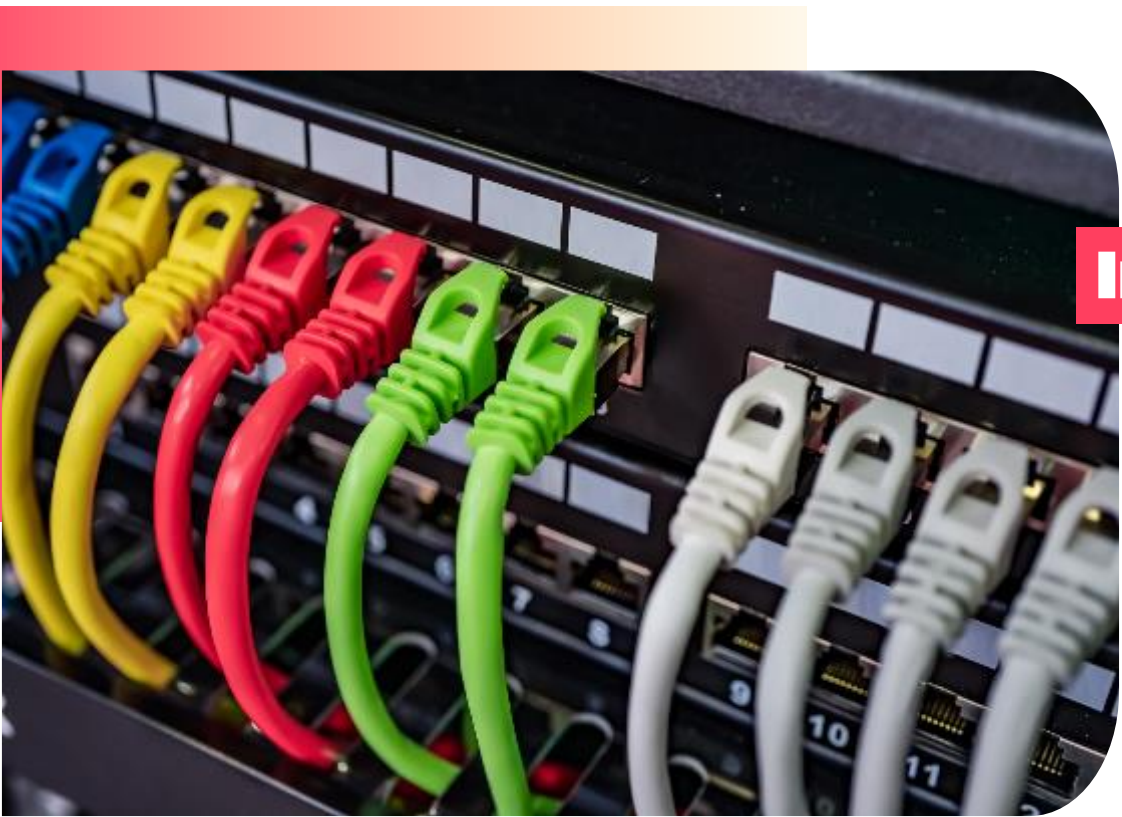


BIG DATA



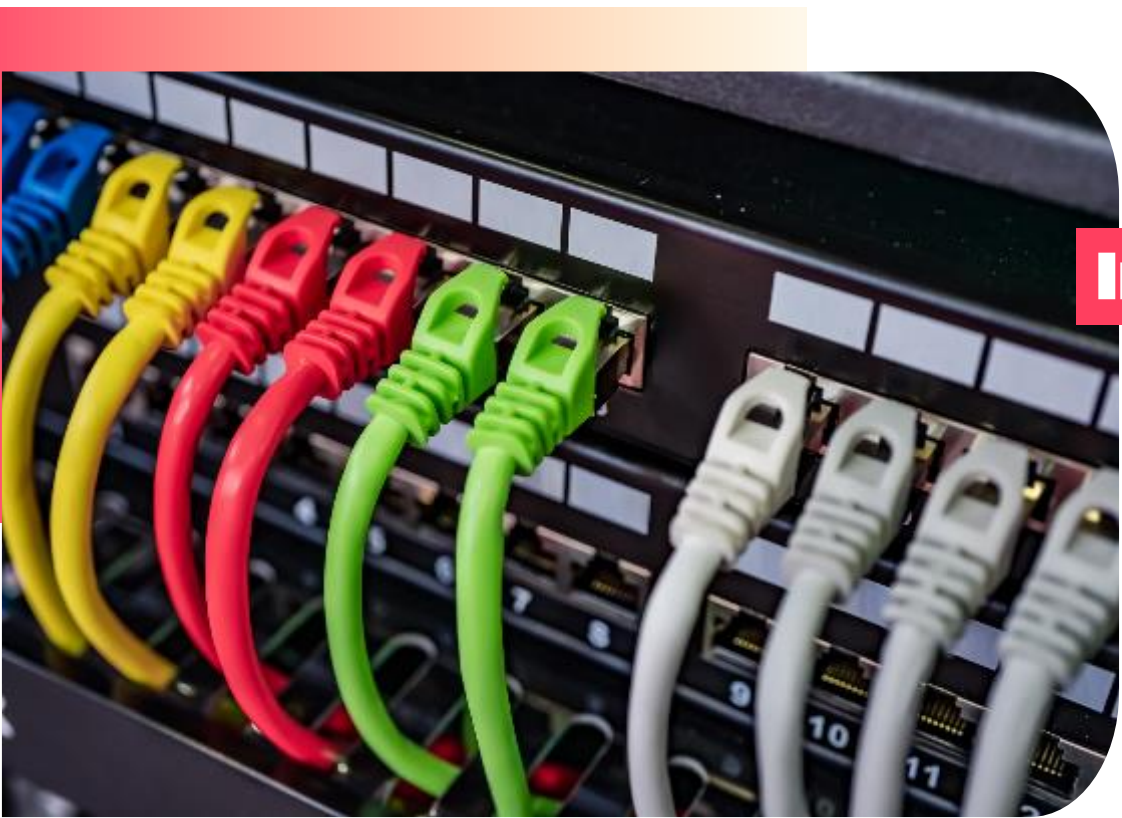


1. Antecedentes, definiciones y bases para un correcto entendimiento



Introducción

Ya todos conocemos la **vorágine evolutiva** de las tecnologías de la información y comunicaciones, esto implica una mayor interacción de las mismas, acceso y generación de datos, factores clave para el inicio del Big Data. Todo esto conlleva a una tendencia de **análisis de grandes volúmenes de datos**, de diferentes orígenes y formas procesados a gran velocidad.



Introducción

Por otra parte, los **sistemas de costos** de las empresas dan la posibilidad de tener conocimiento de la situación a través de procesos que facilitan una mayor eficiencia en el uso de recursos obteniendo como resultado un consumo efectivo.

Introducción

La operación de los modelos de costos basados en actividades permite a las compañías realizar analítica descriptiva para conocer el estado actual de la empresa; sin embargo, hacer analítica predictiva para conocer el estado futuro, sugiere considerar información histórica, lo cual puede generar considerables volúmenes de datos tratables fácilmente a través de herramientas Big Data. Por su parte, la analítica prescriptiva para sacar mayor provecho a situaciones futuras, hace necesario considerar información de terceros, de redes sociales o de streaming de audio y video, para cuyo tratamiento se hacen indispensables las herramientas disponibles en entornos Big Data. Este artículo muestra el potencial de integrar disciplinas como el costeo basado en actividades y la analítica de datos en el marco de Big Data, ilustrando su implementación en el caso de aplicación del análisis de datos de una universidad.

Antecedentes

01

La información en la era digital

02

Business Intelligence en los negocios

03

Evolución del análisis de datos: Data Mining



Invención de la computadora

La primera computadora fue creada durante el siglo XIX, esta, recibió el nombre de máquina analítica. Dicha máquina surgió con el objetivo de elaborar tablas matemáticas. Algunos años después el gobierno Inglés las utilizaría para realizar cuentas. A fines de la década de los cuarenta del siglo XX se creó en la Universidad de Harvard la computadora llamada Mark I. Los intentos seguirían a lo largo de Estados Unidos con el apoyo del gobierno.

Cuando la cantidad de diseños comenzó a aumentar comenzaron a ser clasificadas por sus características en generaciones:

Invención de la computadora

Primera generación:

A lo largo de los años cincuenta, se caracterizaban por que sus programaciones eran mediante un lenguaje de máquinas y eran diseñadas con tubos de vacío. Además ocupaban habitaciones enteras por su gran tamaño y resultaban demasiado caras. Al igual que en la segunda generación la información que ingresaba a la maquina era mediante tarjetas perforadas.

Segunda generación:

Esta generación abarcó la década de los sesenta. Durante esta generación, las computadoras podían procesar más datos y eran un poco más pequeñas. Fue durante estos años que comenzaron a surgir las comunicaciones entre máquinas. A su vez surgieron las primeras computadoras personales, de todos modos eran muy complejas de utilizarlas que era necesario conocer la programación.

Invención de la computadora

● Tercera generación:

Fue durante la tercera generación cuando la producción comenzó a ser en serie. Las computadoras podían ser manejadas por medio de los sistemas operativos. Muchas de las técnicas utilizadas en esos años se convirtieron en estándares que hasta el día de hoy son utilizados.

● Cuarta generación:

La cuarta generación es reconocida por la aparición de los microchips. Esto significó un hito en las ciencias de la computación. Los circuitos no solo comenzaron a ser mucho más pequeños y veloces, si no que a su vez resultaban mucho más económicos. Es por ello que la producción aumentó de manera notable y muchas más personas tuvieron acceso a las mismas. Esta generación es también conocida como la Revolución de la Computación.

La información en la era digital

Internet

La invención del telégrafo, el teléfono, la radio y el ordenador sentó las bases para esta integración de funcionalidades sin precedentes. Internet es a la vez una herramienta de emisión mundial, un mecanismo para diseminar información y un medio para la colaboración y la interacción entre personas y sus ordenadores, sin tener en cuenta su ubicación geográfica. Internet representa uno de los ejemplos más exitosos de los beneficios de una inversión y un compromiso continuos en el campo de la investigación y el desarrollo de la infraestructura de la información.

Primeras bases de datos

En 1969 se crea el concepto de base de datos siendo los años setenta cuando se gestaron las grandes bases de datos y aplicaciones para empresas, gestando un nuevo espacio para el desarrollo del Business Intelligence.

Sin embargo, a pesar de que estas bases de datos permitían acceder a la información que albergaban, su proceso era lento y complicado, siendo su eficacia baja y de mala calidad.

Recogida y Análisis de datos o reporting

En la década de los ochenta continúa la evolución gracias al reporting y el desarrollo del concepto de Data Warehouse. A pesar de la existencia de potentes mecanismos de bases de datos no se disponían de aplicaciones que pudiese manejar fluidamente la utilización de la información.

En el año 1985 Microsoft saca a la luz el programa **Excell**, cuyo uso fue hegemónico.

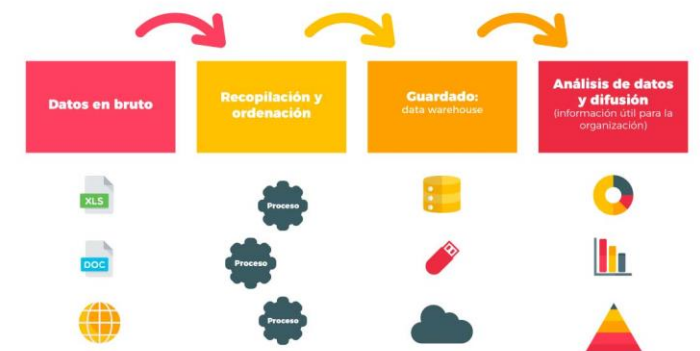
Data warehouse

En su comienzo situado en el año 1998, los data warehouses se pensaron para utilizar **datos repetitivos estructurados** que eran filtrados antes de entrar en el data warehouse.

Aquellos primeros datos relacionales estructurados no podían ser mezclados y emparejados para temas analíticos con datos textuales no estructurados.

Business intelligence 2.0

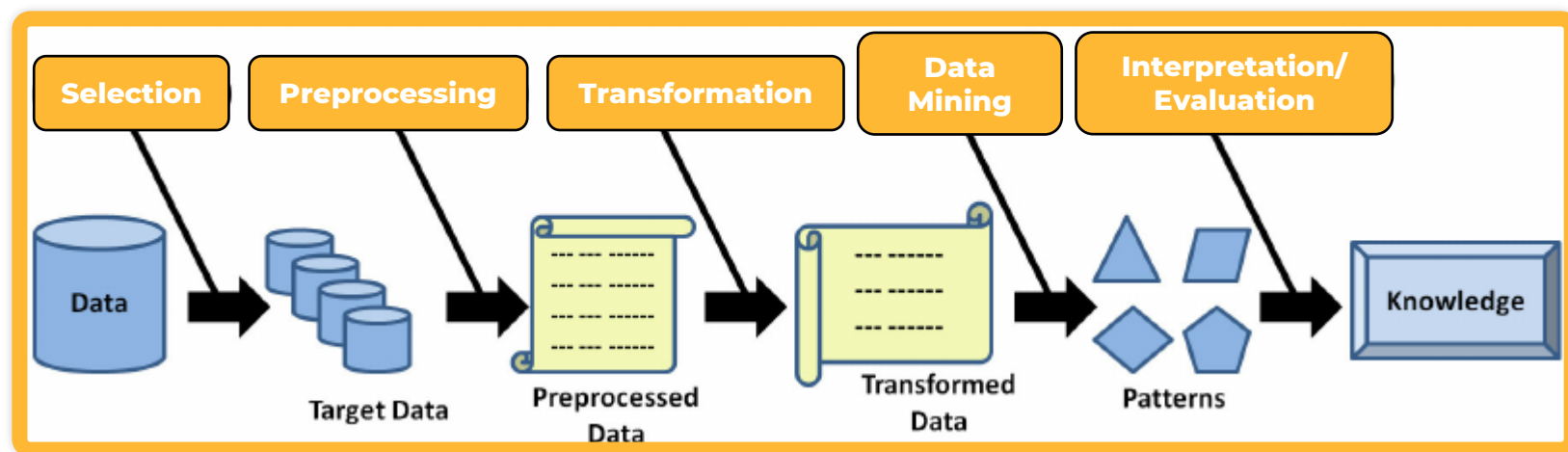
Hasta ahora los datos no importaban porque eran estructurados. Con el nuevo siglo se comienzan a desarrollar tecnologías de **análisis para datos no estructurados**, nace así el business intelligence 2.0 creándose de este modo el concepto de big data.



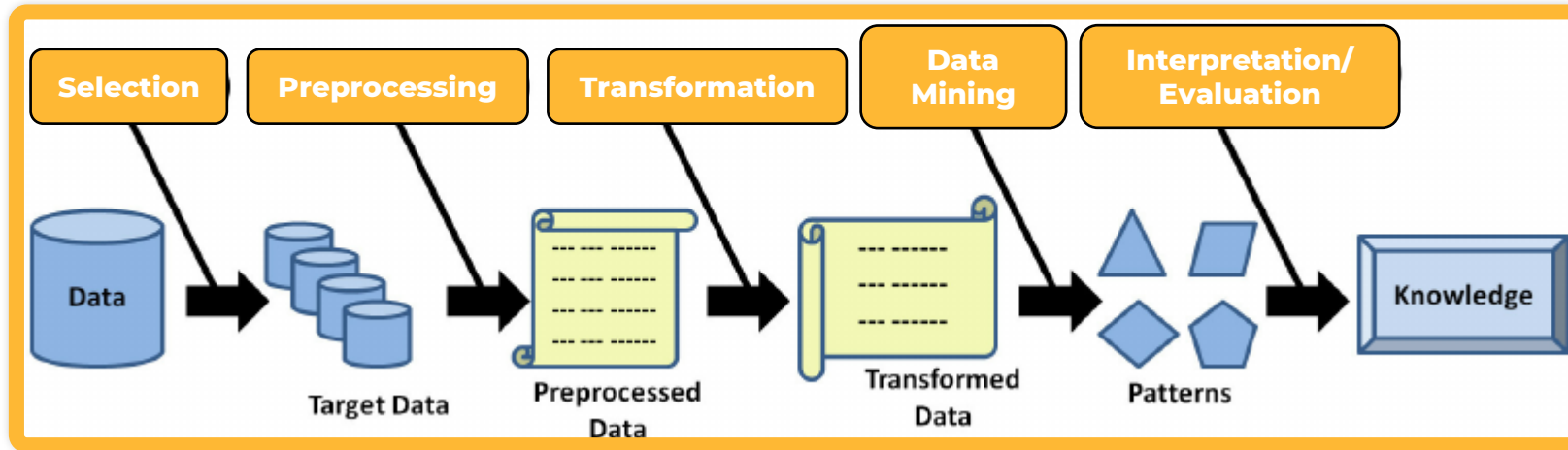
Business intelligence 2.0

KDD o Knowledge Discovery in database

Se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados.



KDD o Knowledge Discovery in database



01

Selección de
datos

02

Análisis de propiedades
de los datos

03

Transformación de
los datos

04

Técnicas de
Data Mining

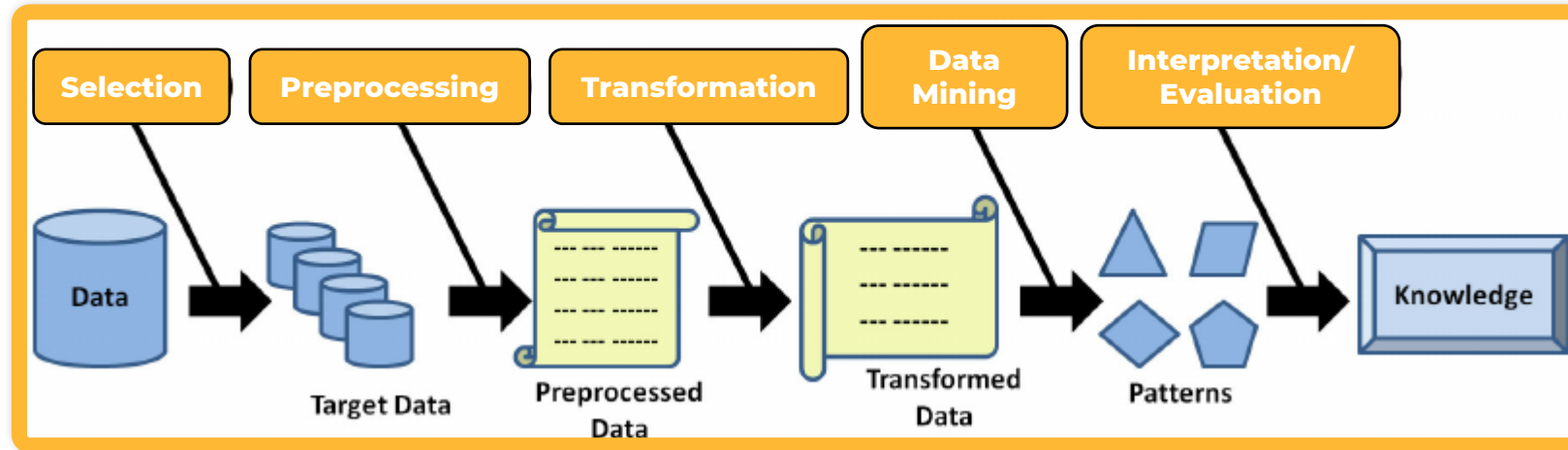
05

Evaluación de
los datos

06

Conocimiento

KDD o Knowledge Discovery in database



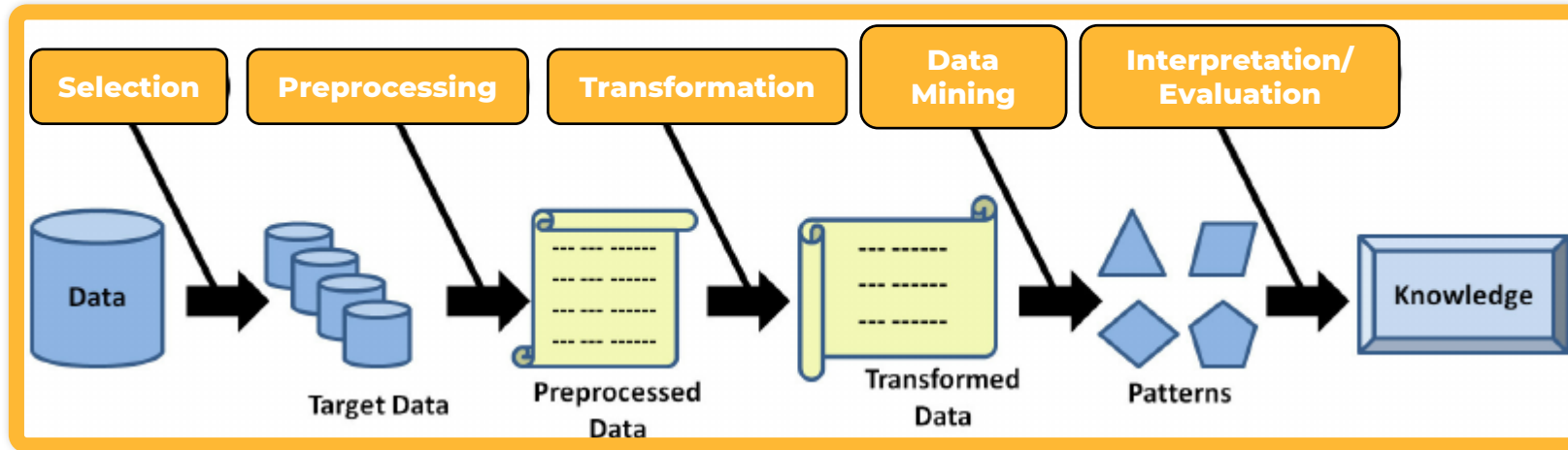
01

Selection (Selección de datos):

Escogemos tanto las variables que se quieren conocer o predecir como las variables que se utilizarán para el cálculo

Selección de
datos

KDD o Knowledge Discovery in database



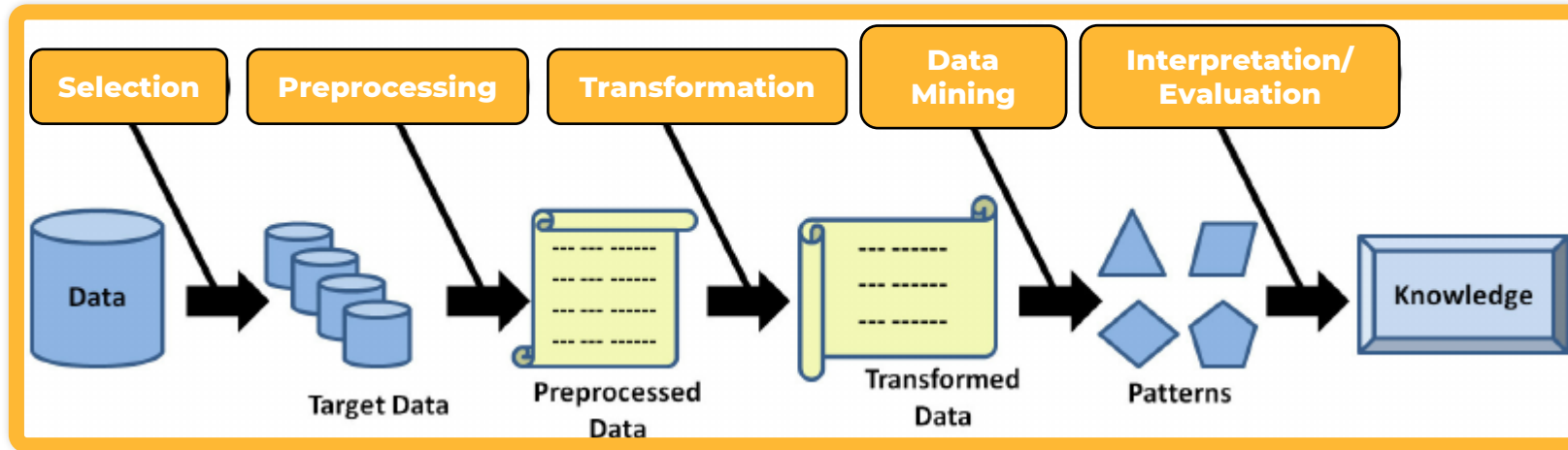
02

Preprocessing (Análisis de propiedades de los datos):

Analizaremos los datos descritos en diagramas e histogramas para tener una visión sobre posibles anomalías al respecto.

Análisis de propiedades
de los datos

KDD o Knowledge Discovery in database

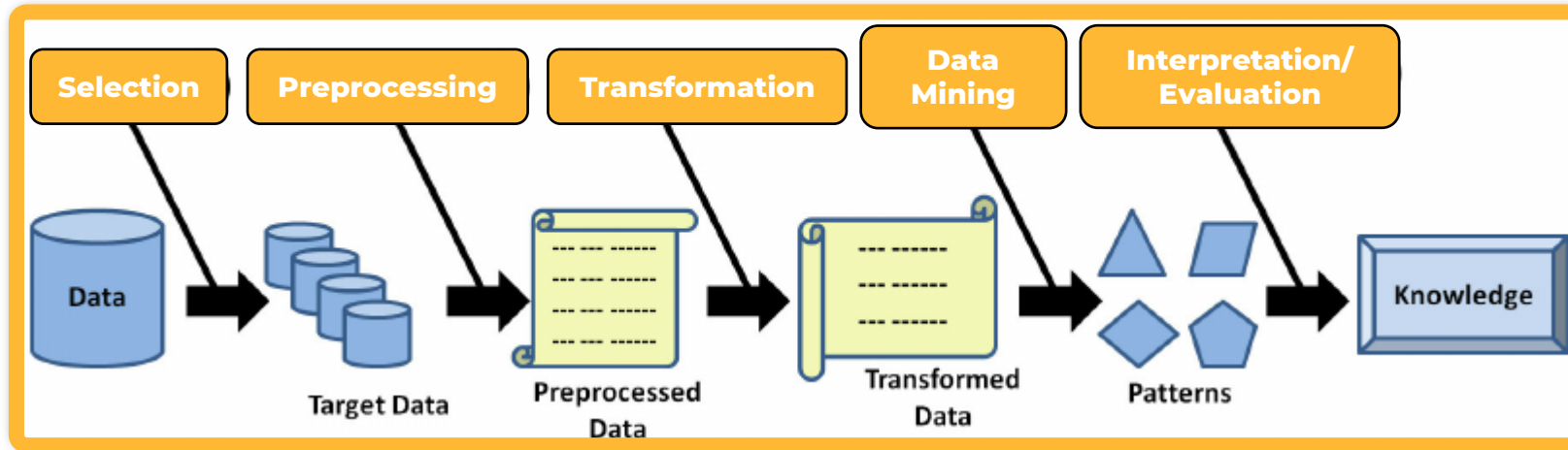


Transformation (Transformación de los datos):

Prepararemos los datos para el proceso de Data Mining.

Transformación de
los datos

KDD o Knowledge Discovery in database

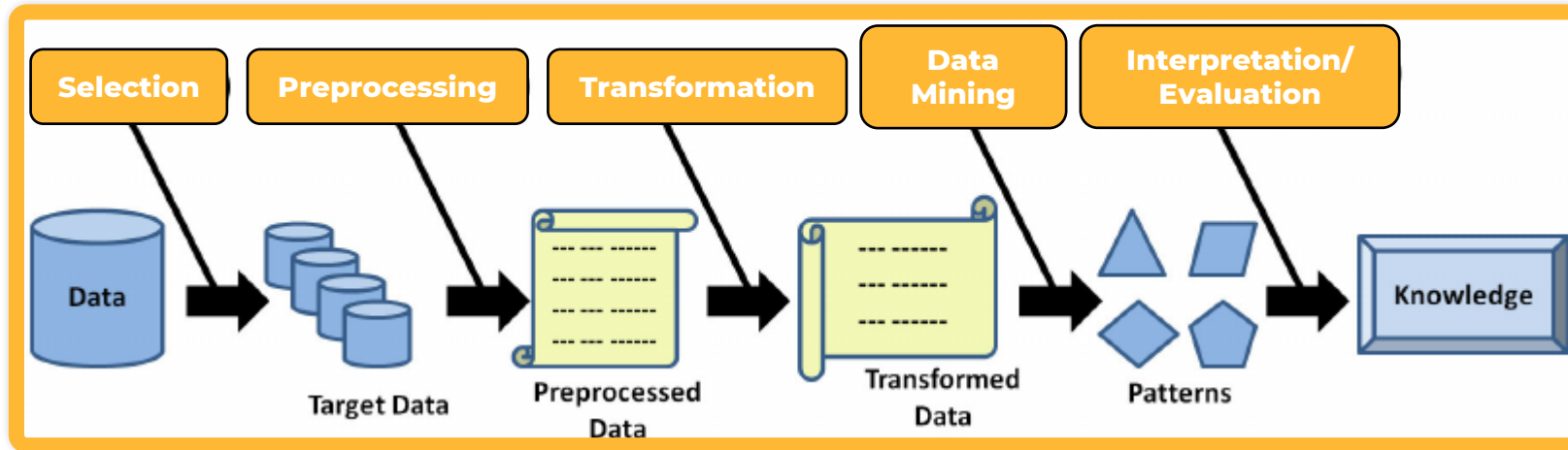


04

Data Mining (Técnica de Data Mining):

Una vez preparados los datos comenzaremos la construcción de un modelo.

KDD o Knowledge Discovery in database



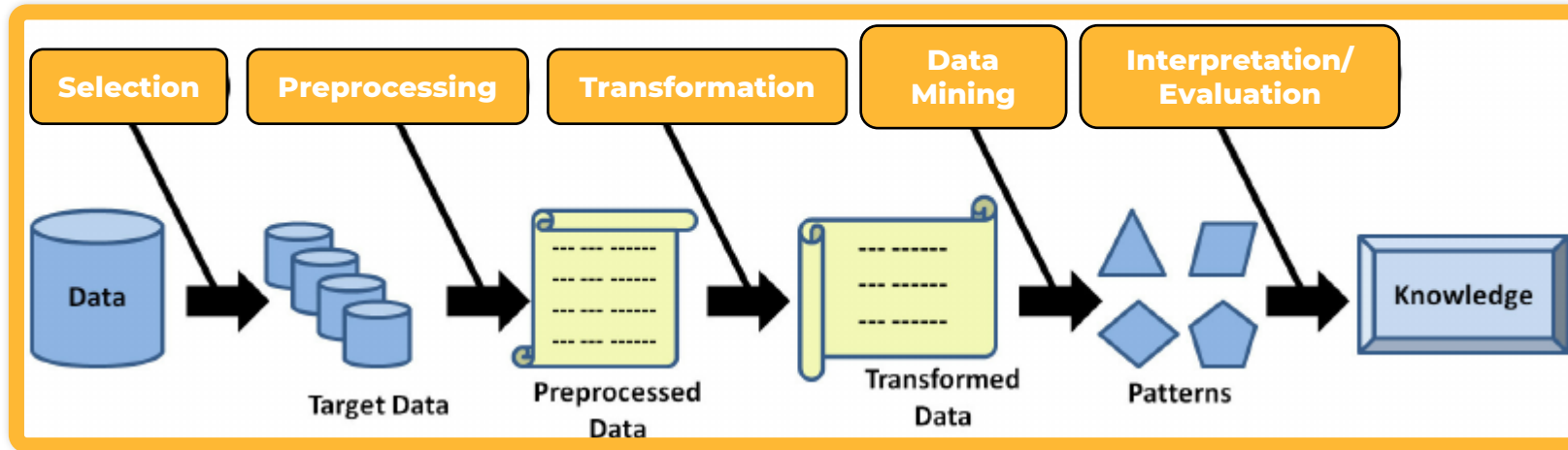
05

Interpretation/Evaluation (Evaluación de los datos):

Una vez obtenido los resultados, los estudiaremos para comprobar si han sido satisfactorios y presentan utilidad, en caso de negativa tendremos que repetir de nuevo el procesado.

Evaluación de
los datos

KDD o Knowledge Discovery in database




06

Knowledge (Conocimiento):

Obtendremos respuestas para cada variable objeto de estudio. Generando respuestas o preguntas que impulsen el conocimiento.

Conocimiento



011000010111001
100100010101001
010011010100100

BIG DATA

Definiciones

Después de haber visto someramente el contexto en el que surge el Big Data podemos definir qué implica esta tecnología revolucionaria atendiendo a sus **diferentes connotaciones**, dependiendo del campo donde se utilice.





01

Definición original

La mayoría de las definiciones se refieren a Big Data como un conjunto colosal de datos caracterizados por las famosas 3V del analista Doug Lane: volumen, velocidad y variedad que superan la capacidad del software habitual para ser capturados, gestionados y procesados. Desde que Lane propuso estos conceptos, hace más de una década, se han ido añadiendo nuevas V, como la validez, la veracidad, el valor y la visibilidad.

**Definiciones Big
Data**

Big Data como tecnología

Como tecnología de la información y de la comunicación, Big Data es un ecosistema que no deja de crecer. En particular, los software de código abierto como Hadoop y NoSQL han supuesto un hito tecnológico en almacenamiento y gestión de datos que se engloba en el término general de Big Data, en ocasiones de forma errónea al incluirse todo tipo de tecnología o herramienta de reciente aparición.



**Definiciones Big
Data**

Big Data como señales

Entender Big Data como una gran oportunidad de negocio que hay que coger al vuelo, con una gestión empresarial que realiza las transacciones adelantándose a los acontecimientos. Extraer valor de los datos significa utilizarlos como señales para anticiparse a lo que va a pasar y sacar partido de esa información. De igual modo, esa información puede utilizarse más allá de la empresa, constituyendo un instrumento clave en otros muchos ámbitos, desde la sanidad (prevenir una epidemia) hasta la administración pública, fomentando la sostenibilidad o, por ejemplo, la planificación urbana.



**Definiciones Big
Data**

Big Data como revolución del conocimiento

Si la información es poder, entonces Big Data se entiende como una tecnología que permite revelar aquellos datos que estaban ocultos. Si antes los investigadores se servían de los datos para apoyar o rechazar teorías, ahora recurren a ellos para inspirarlas.



**Definiciones Big
Data**

Big Data como cultura

La parte tecnológica del Big Data se desarrolla en un determinado contexto social, una dimensión cultural que afecta a aspectos muy distintos de nuestro alrededor, afectando a lo más cotidiano, a los estilos de vida y al pensamiento mismo. Si por un lado se ve como un movimiento positivo, por otro se teme una posible vulneración ética de las libertades fundamentales.



**Definiciones Big
Data**

Big Data como negocio

Big data es una palabra que ha prosperado en el mundo de los negocios, pudiendo describir cualquier oportunidad de negocio relacionada con los datos, estén o no relacionadas con las nuevas tecnologías.



**Definiciones Big
Data**

Big Data como metáfora

Rick Smolan dice que los grandes datos “ayudan al planeta a que tenga un sistema nervioso, en el que los seres humanos sólo somos otro tipo de sensor”. Una metáfora que alberga un profundo significado y podría acabar siendo premonitoria o, en todo caso, claramente alusiva al Internet de las cosas (IoT). El 90% de todos los datos que se han creado, fue creado en los últimos dos años. A partir de ahora, la cantidad de datos en el mundo se duplicará cada dos años.



**Definiciones Big
Data**

Bases del Big Data

Hasta ahora hemos explicado como el big data proviene de la utilización del data mining gestado en las grandes empresas tecnológicas.

No obstante, es a partir del **nacimiento de la Web** cuando comienza a aumentar de manera exponencial la creación de datos. Es cuando se empieza a hablar de **volumen, velocidad y variedad** para referirse a los datos, términos que posteriormente definirán las características del big data como las 3 v.



Las 3 V

Del Big Data



Las 6 V del Big Data

El Big Data empezó como el proceso de las 3 V; volumen, velocidad y variedad. Pero al ir **evolucionando** aparecieron otras V; veracidad, valor y variabilidad.



Las 6 V del Big Data

01

Volumen

Cantidad de datos que se trata.

02

Velocidad

Procurar una infraestructura y unos procesos capaces de tratar los datos de manera ágil y en el menor tiempo posible para aplicar estrategias de cambios.

03

Variedad

Disponer de distintas fuentes de recopilación de datos sobre diferentes aspectos relacionados con el negocio y los consumidores, no solo datos estructurada, sino de diferentes tipos; afinidades, comportamiento...etc.



04

Veracidad

Cuánto es de veraz el dato del que se dispone. A mayor volumen, mayor es el trabajo para organizar esos datos.

05

Valor

Aplicar un correcto tratamientos de los datos recopilados para exprimir el máximo valor de ellos.

06

Variabilidad

Diferentes interpretaciones que pueden resultar en el proceso.

Las 6 V del Big Data





¡Lo conseguiste!

Has llegado al final de la unidad