



Desarrollo de contenido

Unidad 1

Bases de Datos II

Ingeniería de *Software* y Datos

Presentación general de la asignatura

Te damos la bienvenida al curso *Base de Datos II*, donde exploraremos los fundamentos para comprender y gestionar de manera efectiva la información en el entorno empresarial moderno. En este camino nos sumergiremos en tres áreas claves en el aprovechamiento estratégico de los datos: inteligencia de negocios, diseño de un *Data Warehouse* y procesos ETL.

En la primera unidad, Introducción a la Inteligencia de Negocios, exploraremos las definiciones y conceptos que hacen parte de esta disciplina, comprendiendo cómo se puede transformar los datos en información valiosa para la toma de decisiones empresariales. Además, abordaremos conceptos como la minería de datos y el impacto del *Big Data* en la inteligencia empresarial.

En la segunda unidad nos adentraremos en el diseño de un *Data Warehouse*, en donde convergen y se almacenan los datos cruciales para el análisis estratégico. Examinaremos los elementos esenciales de un DWH, los diferentes tipos y la arquitectura que sustenta su funcionalidad. Comprenderemos cómo un *Data Warehouse* bien diseñado se convierte en la columna vertebral de la toma de decisiones informada.

Por último, en la tercera unidad, conoceremos qué son los procesos ETL (*Extract, Transform, Load*). Aquí, aprenderemos cómo integrar datos de diversas fuentes, aplicar técnicas de transformación y cargarlos eficientemente en nuestro *Data Warehouse*. Este proceso es esencial para garantizar la calidad y coherencia de los datos, elementos fundamentales para obtener *insights* o hallazgos valiosos.

En cada unidad exploraremos conceptos teóricos, estudios de casos prácticos y desafíos del mundo real.

Comencemos este emocionante recorrido juntos, descubriendo cómo la inteligencia de negocios y los fundamentos del *Data Warehouse* pueden transformar la forma en que las empresas utilizan sus datos para el éxito estratégico y afrontar los desafíos de la gestión de datos en la era digital.

Objetivos de aprendizaje/ Resultados de Aprendizaje

Objetivo

Desarrollar sentencias SQL avanzadas, permitiendo una correcta y eficiente administración, gestión y manipulación de la información en bodegas de datos.

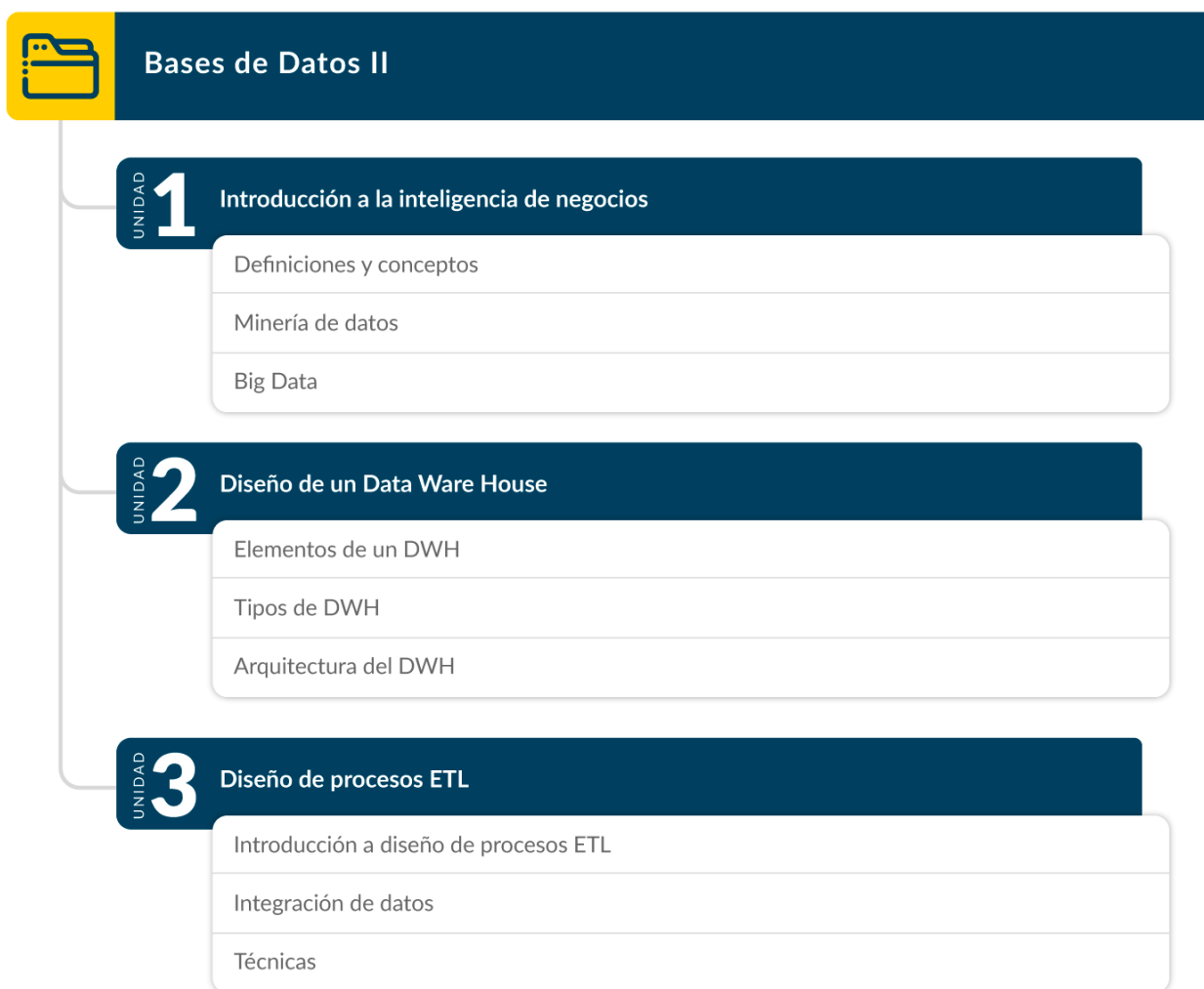
Resultado de aprendizaje

Desarrollo sentencias para avanzadas la gestión y administración de los datos, conociendo las capacidades y funcionalidades de los sistemas de gestión de bases de datos aplicando correctamente los diferentes elementos que hacen parte de la gestión y administración de la información en Bodegas de Datos con la incorporación de sentencias existentes.

Pregunta orientadora

¿Cómo la inteligencia de negocios, el diseño de un *Data Warehouse* y los procesos ETL, pueden mejorar la forma en que las empresas gestionan y utilizan sus datos para tomar decisiones más informadas?

Mapa de la asignatura



Cronograma

Conoce las actividades que realizarás en este curso.

Actividad de aprendizaje	Evidencia de aprendizaje	Semana	Ponderación
AA. Actividad de conocimientos previos	EA. Prueba de conocimientos previos	Semana 1	0 %
AA1. Introducción a la inteligencia de negocios	EA1. Modelo estrella para un <i>Data Mart</i>	Semana 2 Y 3	20%
AA2. Diseño de un <i>Data Warehouse</i>	EA2. Creación de una base de datos de <i>Staging</i>	Semana 4 y 5	30%
AA3. Diseño de procesos ETL	EA3. Proceso de transformación de datos y carga en el <i>Data Mart</i> final	Semana 6 y 7	30%
AA4. Proyecto integrador	EA4. Repositorio de todas las actividades	Semana 8	20%
Total			100%
Actividad de refuerzo			

Unidad 1. Introducción a la inteligencia de negocios (*Business Intelligence*)

Introducción *Unidad 1*

En los últimos años, el mundo de la inteligencia de negocios, también conocido como *Business Intelligence*, ha cambiado rápidamente y ahora está muy avanzado. A continuación, te contamos algunos aspectos importantes relacionados con esta transformación.

- Grandes empresas como *SAP*, *IBM*, *Microsoft* y *Oracle* han comprado compañías más pequeñas, creando un mercado más sólido.
- También hay opciones gratuitas que cubren todas las cosas que una organización necesita para usar su información de manera inteligente.
- Han aparecido nuevas compañías que se centran en ideas innovadoras, como hacer que la información se vea de manera atractiva, predecir eventos importantes y otros hechos emocionantes.
- Aunque el mundo ha tenido problemas económicos desde 2008, la inteligencia de negocios sigue creciendo porque se ha vuelto fundamental para todas las organizaciones (Curto Díaz & Conesa Caralt, 2010).

Resultados de aprendizaje de la *Unidad 1*

- Comprende los conceptos fundamentales de la inteligencia de negocios, incluyendo el proceso de análisis de datos para la toma de decisiones empresariales.
- Reconoce las técnicas de minería de datos y su aplicación en la extracción a partir de grandes conjuntos de datos.
- Aplica los conceptos aprendidos en casos prácticos y proyectos para resolver problemas empresariales reales utilizando técnicas de inteligencia de negocios.

Tema 1. Definiciones y conceptos

¿Qué es la inteligencia de negocios?

En la actualidad vivimos en una sociedad con demasiada información, motivo por el que las empresas necesitan mejores maneras, más rápidas y eficientes, para convertir sus datos en información útil y compartirla en toda la organización.

Ahí es donde entra en juego la Inteligencia de Negocios (o *Business Intelligence*). Podemos pensar en esta como una evolución de los sistemas que ayudan en la toma de decisiones.

Aunque hoy en día es muy importante para la mayoría de las empresas, la idea no es nueva. En 1958 un investigador de IBM, llamado Hans Peter Luhn, acuñó el término, describiéndolo como la habilidad de entender las relaciones entre hechos para tomar acciones hacia un objetivo deseado.

En 1989 Howard Dresden, un analista de *Gartner*, ofreció una definición más formal, diciendo que es el conjunto de conceptos y métodos para mejorar las decisiones de negocio, mediante el uso de sistemas basados en hechos.

Desde entonces, el concepto ha evolucionado, incorporando diferentes tecnologías y prácticas. En este curso utilizamos la siguiente definición: **la Inteligencia de Negocios es el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información, que permite tomar mejores decisiones a los usuarios de una organización** (Curto Díaz & Conesa Caralt, 2010).

Beneficios de un sistema de inteligencia de negocios

La introducción de estos sistemas de información trae ventajas significativas, algunas son:

- Se forma como un ciclo bueno de información, donde los datos se vuelven información, luego conocimiento, lo que nos ayuda a tomar decisiones más acertadas, consiguiendo mejores resultados que generan nuevos datos.
- Nos permite ver toda la información de manera única, completa, histórica, constante y de calidad.
- Ayuda a crear, manejar y mantener métricas e indicadores clave de rendimiento (o KPI) y metas (o KGI) que son superimportantes para la empresa.
- Ofrece información actualizada tanto en resumen como en detalle.
- Reduce la brecha entre el lenguaje de la tecnología y lo que necesita la organización.
- Facilita la comprensión y documentación de los sistemas de información en el contexto de la organización.
- Hace que la organización sea más competitiva porque podemos: distinguir lo importante de lo superficial, acceder más rápido a la información y tomar decisiones más rápidas y ágiles.

Conceptos Importantes en la Inteligencia de negocios (Curto Díaz & Conesa Caralt, 2010)

- **Cuadro de Mando Integral (CMI):**

Imagina que tienes una herramienta que te ayuda a medir qué tanto está haciendo una empresa para alcanzar sus metas, esto se llama *herramienta de control de gestión* y funciona midiendo cosas importantes que la empresa quiere lograr, como en un juego donde tienes que alcanzar ciertos objetivos.

La herramienta se fija en cuatro áreas claves: dinero (**financiera**); **clientes**, cómo hacen las cosas (**procesos**) y cómo están innovando y aprendiendo cosas nuevas (**innovación y conocimiento**). En resumen, te ayuda a ver cómo van las cosas en esos aspectos y si la empresa está acertando, o dónde puede mejorar. Es como un mapa que te muestra si estás yendo por el camino correcto o si necesitas ajustar tu ruta para llegar a donde quieres.

- **Sistemas de Soporte a la Decisión (DSS):**

La toma de decisiones puede ser un proceso complejo para los gerentes, ¿verdad? Bueno, ahí es donde entran los sistemas de apoyo a la decisión. Básicamente, son conjuntos de pasos y modelos que usan programas de computadora para ayudar a los jefes a tomar decisiones importantes.

Estos sistemas recopilan datos, los analizan y luego crean tablas, gráficos y otros elementos que facilitan el proceso hacia la decisión. ¿Quieres ver números específicos o gráficos? Este sistema te ayuda a entender toda la información de manera fácil y rápida. Además, te permite decirle qué información necesitas y cómo quieres que se muestre. Es como tener un asistente personal para la toma de decisiones.

- Un **Data Warehouse** es como un gran almacén de datos para empresas. Imagina que tienes muchas fuentes diferentes de información, como archivos, bases de datos, etc. El *Data Warehouse* toma todos esos datos, los organiza y los limpia, con el fin de que estén listos para ser analizados.

El *Data Warehouse* es como una gran biblioteca que guarda información histórica sobre diferentes temas. Esta información se muestra de una manera para que sea fácil de entender, incluso, tiene algo llamado metadatos, etiquetas que te dicen de dónde viene la información y cuánto puedes confiar en ella.

Con lo anterior, es más rápido obtener información importante para tomar decisiones, ya que automatiza el proceso de obtener datos de los sistemas operativos y los adapta en un formato que sea fácil de analizar.

- El **Data Mart** es como un hermano menor del *Data Warehouse*. Piensa en que el *Data Warehouse* es una gran biblioteca para toda la empresa, pero a veces solo necesitas información específica de un departamento o para una tarea más simple. Ahí es donde entra en juego el *Data Mart*.

Es como tener una mini biblioteca, pero solo con los libros que necesitas para tu tarea. Por ejemplo, si trabajas en el departamento de ventas, tu *Data Mart* contendría información relevante solo para las ventas. Es más pequeño y específico, lo que facilita encontrar y entender la información. Además, esta información puede usarse en herramientas visuales o de análisis de datos para obtener mejores *insights*. Se presenta minería de datos.

- **KPIs**: piensa que estás al mando de un negocio y necesitas saber cómo están funcionando tus acciones. Aquí entran los **Indicadores Básicos de Rendimiento (Key Performance Indicators KPIs)**, herramientas que te muestran lo efectivas que han sido tus decisiones y acciones.

Estos indicadores son las señales que te dicen si estás yendo en la dirección correcta para cumplir tus objetivos. Son muy útiles para comparar información importante y decidir qué hacer para mejorar constantemente. Para que sean útiles, deben ser específicos, medibles, cuantificables y relevantes en lo que estás tratando de lograr.

El *Business Intelligence (BI)* te ayuda a entender rápidamente cómo está tu negocio. Es como un médico para tu empresa, te da un diagnóstico rápido y te permite estar siempre al tanto de cómo van tus indicadores, para que así tengas un buen control de todo lo que está sucediendo.

- **BBDD relacionales:**

El modelo entidad-relación es el lenguaje común que utilizan las bases de datos, tanto para las cosas prácticas como para las analíticas. En dicho modelo las bases de datos se dividen en entidades, las cuales podemos comparar con las carpetas que guardan información sobre diferentes aspectos del negocio: usuarios, productos, ventas o clientes.

Cada entidad es como una tabla o conjunto de datos específicos. Y aquí está lo interesante: estas se conectan de una manera lógica o física.

Por ejemplo: podemos unir clientes según los productos que compran y la cantidad de dinero que gastan. Esto hace que sea más fácil entender cómo están relacionadas las diferentes partes de un negocio.

- **BBDD no relacionales:**

Piensa que tenemos una plataforma informática y recopilamos un montón de eventos que suceden en ella como clics, compras y más, sin dividirlos en partes específicas. Esto nos da demasiada información. Ahora tenemos nuevas oportunidades para entenderla y usarla.

Un ejemplo común es el de los sitios *web* donde, dependiendo de cómo te comportes allí, te recomiendan productos que podrían gustarte. Esto es posible gracias a bases de datos que recopilan eventos y nos ayudan a personalizar recomendaciones según tu perfil de usuario. Es como tener un asistente personal en línea que te sugiere cosas para comprar.

- **Data Governance:**

Cuando recopilamos información en los nuevos sistemas es importante ser muy cuidadosos para garantizar que la información sea buena y completa. Para garantizar esto ha surgido un concepto llamado *Data Governance* o *vigilancia del dato*. Es como tener a alguien o algo que se encarga de asegurar que la información sea de alta calidad. Este *Data Governance* es como el guardián final que vela por la calidad y la integridad de los datos que recopilamos.

- **ETL (*Extract, Transform and Load*):**

Es un proceso que usa una herramienta de *software* para recopilar información de las bases de datos de una empresa como las de negocios, CRM (*Customer Relationship Management*) o ERP (*Enterprise Resource Planning*). Luego transforma esa información y la adapta para, finalmente, cargarla en una base de datos analítica.

En otras palabras, con ETL creamos “almacenes especiales” (*datamarts* y *datawarehouses*) con la información que queremos estudiar y analizar en detalle.

- **Facts & dimensions (hechos y dimensiones):**

Los *datamarts* y los *datawarehouses* se asemejan a bibliotecas especiales para almacenar información importante de un negocio. Están hechos de dos tipos de tablas: las de hechos y las de dimensiones. Las tablas de hechos registran números o datos claves del negocio, como las ventas. Mientras que las tablas de dimensiones permiten medir esos datos con relación a otras cosas, por ejemplo: ver cómo van las ventas durante los diferentes meses del año.

En resumen, imagina que estás revisando las ventas (datos claves) y cómo se relacionan con los diferentes meses del año (variable de distribución) en estas bibliotecas especiales.

- **Modelado y Reporting:**

Imagina que las herramientas especiales como *OBIEE*, *Business Object* o *QlikView* convierten datos en cuadros fáciles de entender. Estas herramientas toman los datos y los presentan en paneles o informes llamados *dashboards*.

En dichos informes verás tablas con variables importantes, indicadores estadísticos, gráficos que muestran tendencias y más. Todo esto ayuda a los analistas de negocios, o a los responsables, a tomar decisiones inteligentes para hacer crecer su negocio.



Lecturas y Material Complementario

Consulta el siguiente video para reforzar el concepto de inteligencia de negocios. Haz clic sobre el título a continuación.

Título: *Introducción al Business Intelligence*. Alex Rayón.

Autor: Universidad de Deusto .

URL: https://www.youtube.com/watch?v=ElgUy_7eYLQ

También podrás conocer más sobre los beneficios del uso de inteligencia de negocios en el siguiente video.

Título: Beneficios de la inteligencia de negocios.

Autor: CorponetTV: Gold Partner.

URL: <https://www.youtube.com/watch?v=MUZbN0go47U>

Tema 2. Minería de datos

Evolución

Aunque las técnicas clave de minería de datos existen desde hace tiempo en áreas como la inteligencia artificial y la estadística, ahora están siendo reconocidas como maduras y útiles, especialmente con los avances en motores de bases de datos y herramientas de integración de información.

La historia de la minería de datos se remonta a los años 50, cuando los departamentos de informática creaban resúmenes de información para facilitar la gestión empresarial. Sin embargo, estos sistemas eran voluminosos y difíciles de usar. En los años 60, los sistemas de gestión de bases de datos no eran flexibles y los informes eran laboriosos de preparar. Además, la diversidad de bases de datos no integradas en diferentes departamentos de una organización eran un problema.

El *Data Warehouse*, surgido a finales de los años 80, resolvió estos problemas al proporcionar una solución centralizada, lo que impulsó la minería de datos, donde las tareas de análisis se automatizan, permitiendo extraer conocimiento útil de manera más eficiente.

¿Qué es minería de datos? Hasta hace poco, las empresas usaban sus datos principalmente para contabilidad, inventarios y procesos internos. Pero con la creciente competencia global, los datos se han vuelto cruciales.

Gracias a avances tecnológicos, las empresas cuentan con herramientas más avanzadas para almacenar y analizar grandes cantidades de datos. En este sentido, las empresas requieren de una mejora constantemente en sus formas de administrar y tomar decisiones, aprovechando la información como una fuente clave de competitividad.

Una técnica importante para aprovechar los datos es la minería de datos, la cual combina inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica, con el fin de descubrir información no evidente a simple vista. La minería de datos encuentra relaciones, tendencias, patrones y comportamientos ocultos, para respaldar así la toma de decisiones con un mayor entendimiento.

La minería de datos se considera la cúspide de la evolución en el análisis de datos. Su nombre se debe a la analogía con la minería de una montaña: dentro de los datos almacenados en una empresa hay información valiosa que, como los diamantes en una montaña, puede descubrirse y aprovecharse con las herramientas adecuadas (Rodríguez & Martínez, 2018).



Lecturas y Material Complementario

A través del siguiente video podrás ampliar más información sobre el tema:

Título: ¿Qué es Data Mining?

Autor: ComputerHoy.com

URL: <https://www.youtube.com/watch?v=ueAalErOPY4>

Tema 3. *Big Data*

¿Qué es *Big Data*?

Cuando hablamos de *Big Data*, nos referimos a conjuntos de datos enormes y complejos que son difíciles de manejar con las tecnologías convencionales. Estos conjuntos de datos son tan grandes, complejos y crecen tan rápidamente, que las herramientas tradicionales, como bases de datos y estadísticas comunes, no pueden procesarlos o analizarlos a tiempo (PowerData, 2024).

Aunque no hay un tamaño específico para definir *Big Data*, la mayoría de los expertos consideran que va desde 30-50 *Terabytes* hasta varios *Petabytes*. La complejidad del *Big Data* proviene, principalmente, de la naturaleza no estructurada de muchos datos generados, entre otras cosas, por tecnologías modernas como registros *web*, identificación por radiofrecuencia (RFID), sensores en dispositivos y redes sociales.

Para aprovechar eficientemente el *Big Data*, a menudo se combina con datos estructurados de aplicaciones comerciales convencionales como ERP o CRM. Esto permite obtener información valiosa de los datos enormes y complejos, combinándolos con datos más organizados y estructurados.

¿Por qué el *Big Data* es tan importante?

El *Big Data* ayuda a las empresas a responder preguntas que ni sabían que tenían. Esto lo podemos pensar como un montón de información con la cual podemos jugar para entender mejor los problemas que se presenten. Cuando las empresas usan *Big Data*, se vuelven ágiles y pueden resolver dificultades antes de que se conviertan en grandes líos. Incluso toman decisiones más inteligentes.

A continuación, te presentamos algunos beneficios del *Big Data* para las empresas:

- Ahorro de dinero:
Gracias a tecnologías como *Hadoop*, pueden almacenar grandes cantidades de datos sin gastar tanto dinero.
- Decisiones más rápidas y mejores:

Con herramientas como *Hadoop* pueden analizar datos velozmente, tomando decisiones basadas en lo que encuentran.

- Nuevos productos y servicios:

Al entender lo que quieren los clientes, las empresas pueden crear cosas nuevas que les interesen y les gusten.

Por ejemplo:

- En la industria del turismo, pueden saber qué hace felices a los clientes o qué les disgusta, antes de que algo salga mal.
- En el cuidado de la salud se puede diagnosticar rápidamente gracias a los datos encontrados.
- En el comercio minorista las empresas pueden saber lo que quieren los clientes y ofrecerlo antes de que lo pidan.

Pero *Big Data* no es solo para grandes empresas, también se utiliza en tecnología para resolver problemas rápidamente, detectar fraudes en compras *online* y entender lo que expresa la gente en redes sociales sobre productos y servicios (Powerdata, 2024).

Desafíos de la calidad de datos en *Big Data*

Las características especiales del *Big Data* hacen que sea un desafío garantizar que los datos sean buenos. Se conocen como las 5 V:

- Volumen: cantidad enorme de datos.
- Velocidad: rapidez en que llegan.
- Variedad: diferentes tipos de datos.
- Veracidad: si son datos realmente confiables.
- Valor: si son datos útiles.

Estas cinco cosas definen los problemas del *Big Data*.

Los conjuntos de datos son enormes, siempre están cambiando y son complicados de manejar. Es como tratar de sacar información útil de una algo enorme y en constante movimiento.

Antes, con ETL (*Extract, Transform and Load*), podíamos ingresar en nuestros sistemas la información ordenada que teníamos, como en el ERP (*Enterprise Resources Planning*) y CRM (*Customer Relationship Management*). Pero ahora, con el *Big Data*, podemos añadir

más datos del exterior de la empresa: comentarios en redes sociales, *likes*, resultados de campañas de *marketing* y datos estadísticos de otras fuentes. Todos estos datos nos dan información para ver si lo que ofrecemos está funcionando bien o si hay problemas.

Algunos desafíos a los que se enfrenta la calidad de datos de *Big Data* son:

- Hay muchos lugares de dónde vienen los datos y muchos tipos, razón por la que es complicado juntarlos.

Los datos de *Big Data* pueden venir de:

- Internet y teléfonos móviles.
- Internet de las Cosas.
- Empresas especializadas que recopilan datos de diferentes áreas.
- Datos de experimentos.

Los tipos de datos también son muy variados:

- Datos no estructurados: documentos, videos, audios y más.
- Datos semi-estructurado: *software*, hojas de cálculo o informes.
- Datos estructurados.
- Solo el 20% de la información es ordenada, lo que puede causar errores si no cuidamos la calidad de los datos. Es como tratar de organizar un montón de cosas diferentes que no siempre encajan bien.

Imagínate que tienes una montaña de datos gigante, el problema es que es tan grande que llevaría mucho tiempo organizarla y limpiarla para que sea útil. Transformar esos datos desordenados en algo organizado y procesarlos lleva mucho tiempo (Powerdata, 2024).

¡Felicidades! Has completado con éxito la *Unidad 1*. Aprendiste sobre conceptos relacionados con la inteligencia de negocios, minería de datos y *Big Data*. En la próxima unidad nos centraremos en el diseño de un *Data Warehouse*.

¡Continúa tu proceso de aprendizaje!



Lecturas y Material Complementario

Consulta el siguiente video para reforzar los conceptos asociados a *Big Data*.

Autor: Universidad de Deusto / Deustuko Unibertsitatea.

Título: Conceptos básicos del Business Intelligence. Alex Rayón.

URL: <https://www.youtube.com/watch?v=09Knu34bYdw>

Lecturas y material complementario

Para profundizar en el concepto de inteligencia de negocios, observa:

Autor: Grow Up.

Título: ¿Qué es Inteligencia de Negocios?

URL: <https://www.youtube.com/watch?v=ARr7RzdOLUI>



Referencias bibliográficas

Computerhoy.com, (4 de noviembre de 2017). *¿Qué es el data mining?* [Video].
YouTube. <https://www.youtube.com/watch?v=ueAaIErOPY4>

CorponetTV: Gold Partner de SAP Business One, (12 de septiembre de 2022). Beneficios de la inteligencia de negocios. [Video]. YouTube.
<https://www.youtube.com/watch?v=MUZbN0go47U>

Curto Díaz, Josep., & Conesa Caralt, Jordi. (2010). *Introducción al Business Intelligence*. Editorial UOC.

Grow Up, (22 de enero de 2021). ¿Qué es Inteligencia de Negocios? [Video]. YouTube.
<https://www.youtube.com/watch?v=ARr7RzdOLUI>

Porwer Data (2024). *Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad*. <https://www.powerdata.es/big-data>

Rodriguez & Martinez (2018). *ANALISIS DE HERRAMIENTAS DEL ENFOQUE DE INTELIGENCIA DE NEGOCIOS: CASO DE ESTUDIO DATOS DE LA BANCA CORPORATIVA Y DE INVERSION UNIVERSIDAD PILOTO DE COLOMBIA FACULTAD DE INGENIERIA DE SISTEMAS BOGOTA 2018*.

Universidad de Deusto / Deustuko Unibertsitatea, (26 de noviembre de 2015). *Introducción al Business Intelligence*. Alex Rayón. [Video]. YouTube.
https://www.youtube.com/watch?v=ElgUy_7eYLO



IU Digital

de Antioquia

INSTITUCIÓN UNIVERSITARIA
DIGITAL DE ANTIOQUIA



Esta licencia permite a otros distribuir, remezclar, retocar, y crear a partir de esta obra de manera no comercial y, a pesar que sus nuevas obras deben siempre mencionar a la **IU Digital** y mantenerse sin fines comerciales, no están obligados a licenciar obras derivadas bajo las mismas condiciones.