**ACTIVIDAD EVALUABLE**

1. **Responde a las siguientes preguntas cortas:**
2. **Define con tus palabras el concepto de base de datos**

Las bases de datos se definen como una forma de almacenar información de forma electrónica en un sistema usando un tipo de gestor de esos datos recopilados para su uso futuro.

1. **Explica las diferencias entre una base de datos y una base de información**

Base de datos → es donde se almacenan todos los datos que se quieren consultar o de información importante para poder organizarla y consultarla en un futuro

Base de información →

1. **Define las principales características de una base de datos relacional.**

Las bases de datos relacionales son un tipo de base de datos que almacena y organiza datos con relaciones definidas para un acceso rápido. En una base de datos relacional, los datos se organizan en tablas que contienen información sobre cada entidad y representan categorías predefinidas mediante filas y columnas.

Características →

* Uso de Claves primarias: Cada tabla tiene un campo o una combinación de campos que identifica de manera única cada registro en la tabla
* Normalización: proceso para diseñar bases de datos relacionales de manera eficiente y libre de **redundancias**. Consiste en eliminar la redundancia de información.
* Consultas SQL: utilizan el lenguaje de consulta estructurado para realizar consultas y manipular los datos almacenados en las tablas (*filtrar datos, realizar agregaciones y combinar datos de varias tablas*).

1. **Explica las fases para el desarrollo de una base de datos¿Qué organismo oficial se encarga de proponer y definir dichas fases?**

Las bases de datos tienen 3 fases de diseño:

* **Diseño conceptual:**

Se obtiene una estructura de la información de la futura BD independiente de la tecnología que hay que emplear. No se tiene en cuenta todavía qué tipo de base de datos se utilizará –relacional, orientada a objetos, jerárquica. Tampoco se tiene en cuenta con qué SGBD ni con qué lenguaje concreto se implementará la base de datos.

Para representarla se usa el modelo ER (*entidad relación*)

* **Diseño Lógico:**

Se parte del resultado del diseño conceptual, que se transforma de forma que se adapte a la tecnología que se debe emplear. Si se trata de un SGBD relacional, esta etapa obtendrá un conjunto de relaciones con sus atributos, claves primarias y claves foráneas.

* **Diseño Físico:**

Se transforma la estructura obtenida en la etapa del diseño lógico, con el objetivo de conseguir una mayor eficiencia; además, se completa con aspectos de implementación física que dependen del SGBD.

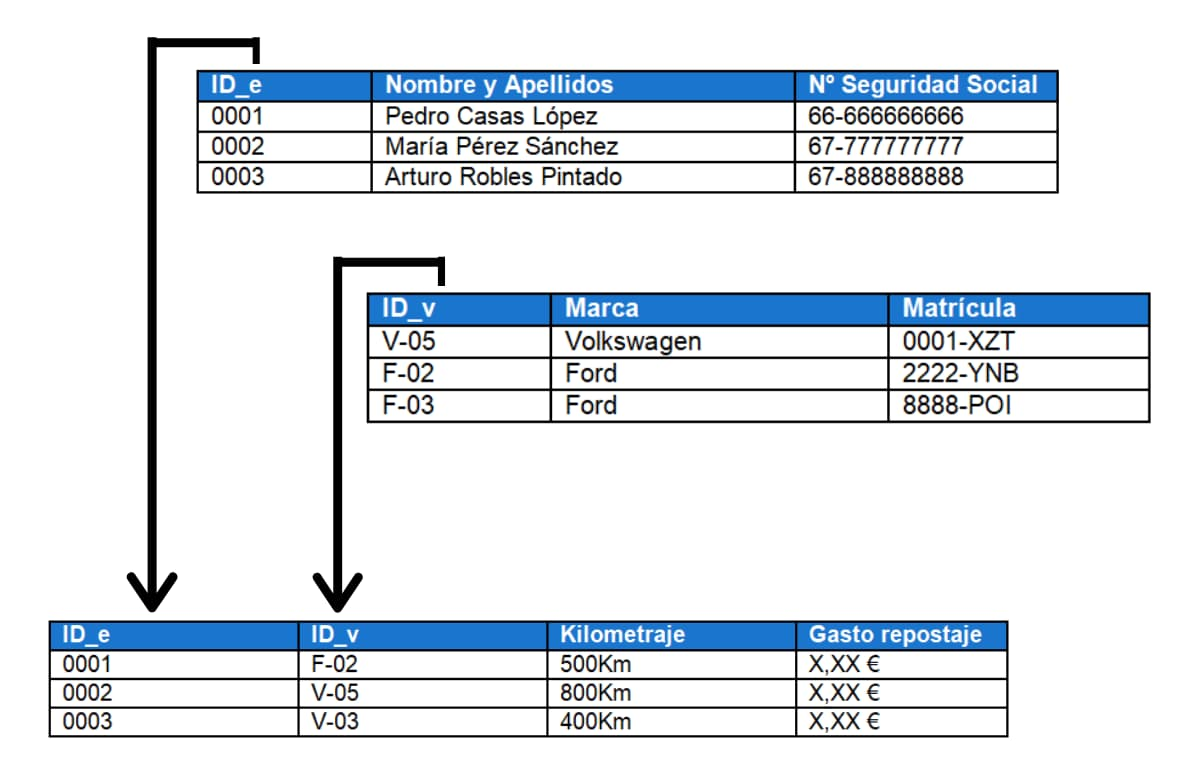
Se deben tener en cuenta las características de los procesos que consultan y actualizan la base de datos, como por ejemplo los caminos de acceso que utilizan y las frecuencias de ejecución

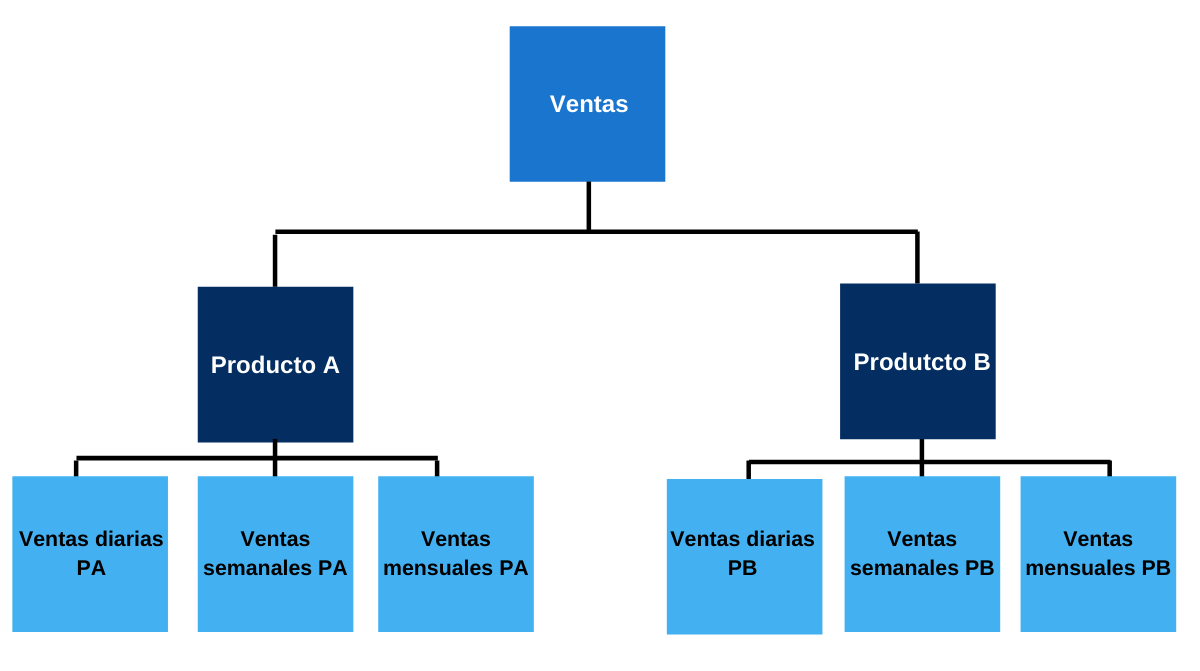
1. **Identifica los 4 tipos o clasificaciones de bases de datos según el modelo de datos que hemos trabajado en el tema y defínelos**

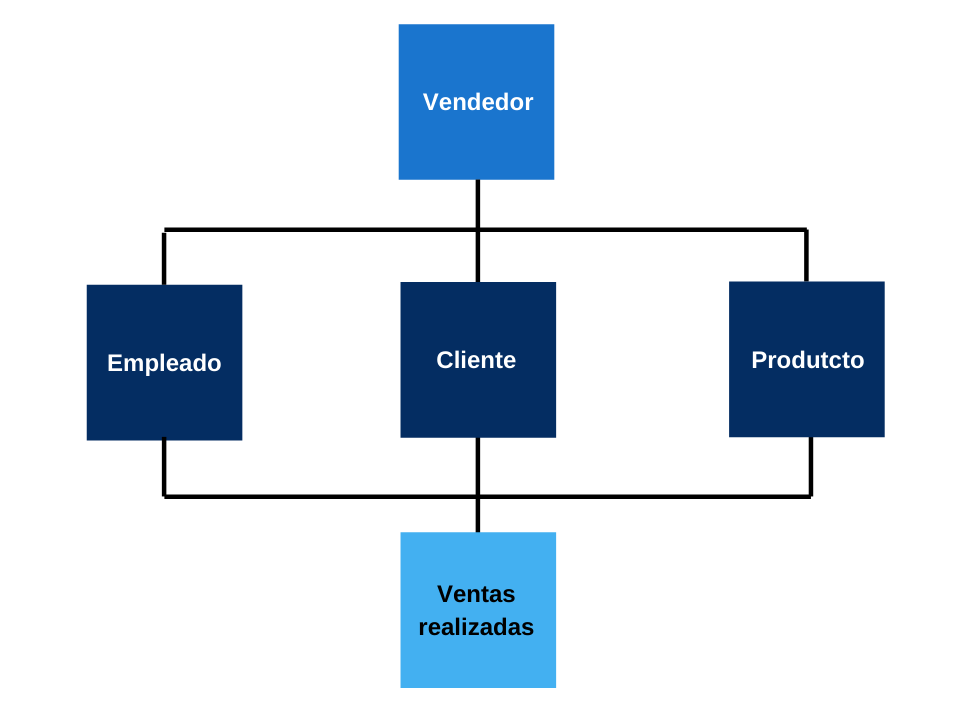
* Base de datos relacional: ordena los datos en tablas (relaciones) compuestas por columnas y filas.

Cada columna alberga un atributo de la entidad (nombre, dirección, fecha de nacimiento…).

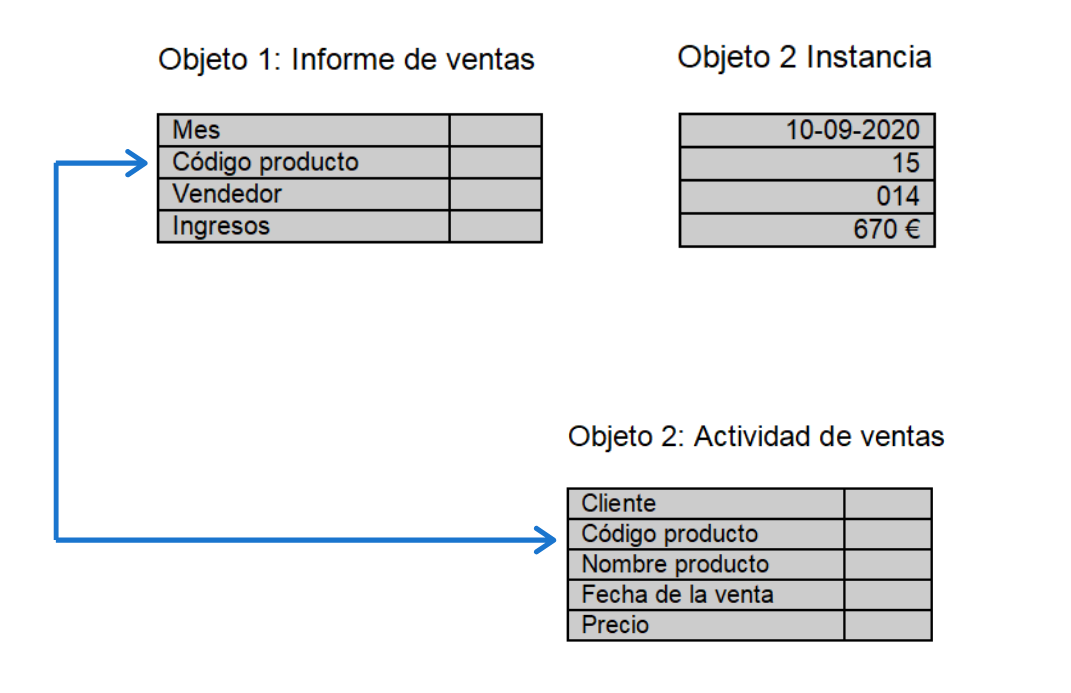
Escogiendo un atributo en concreto o una combinación de varios tenemos una clave primaria, a la que se puede hacer referencia en otras tablas, en las que será un clave externa.



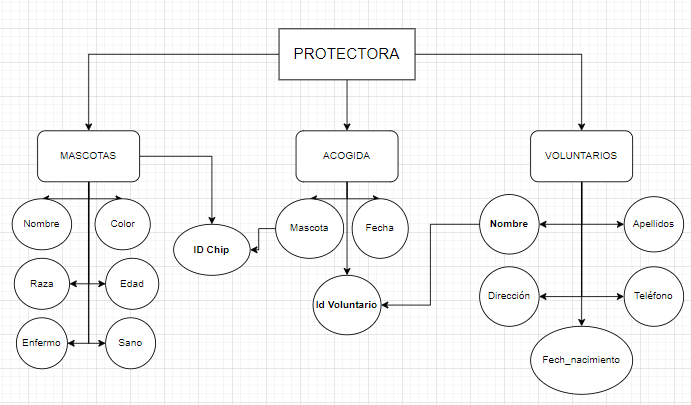
* Base de datos jerárquica: presenta los datos en una estructura de árbol invertido, donde cada registro tiene un único **nodo raíz,** del que surgen otros nodos (registros); los nodos en un mismo nivel son **nodos padre,** cada nodo padre tiene el mismo nodo raíz, y puede tener **nodos hijos,** pero los nodos hijos solo pueden tener un nodo padre.
* Base de datos en red: se permiten las relaciones de uno a muchos o de muchos a muchos entre registros vinculados, teniendo registros principales múltiples. Cada uno de estos conjuntos consiste en un registro propietario o principal y uno o más registros miembros o secundarios. En este modelo se permite que los nodos hijos tengan más de uno nodo padre, de manera que se pueden representar relaciones más complejas.



* Base de datos orientada a objetos: es una colección de objetos utilizados en la programación orientada a objetos (es decir, que emplea lenguajes como C + + o Java, por ejemplo).

Este modelo de base de datos utiliza tablas también, donde los objetos se dotan de un conjunto de características propias, que a su vez les diferencian de objetos similares. Los objetos similares pueden agruparse en una clase y cada objeto de esta es una instancia. 

1. **Se requiere realizar una aplicación web que almacene información acerca de una organización protectora de animales. Se quiere almacenar todas las mascotas que llegan al refugio para ello se guardarán su identificación del chip, su nombre, su raza, su color, su edad y si padece o no de alguna enfermedad. Además, se desea conocer todo lo relativo a las acogidas, para ello se quiere almacenar para cada acogida, la fecha en la que se realiza, el voluntario que realiza y las mascotas que fueron rescatadas en dicha acogida. Por último hay un registro de voluntarios de los que se conoce su nombre, apellidos, dirección y teléfono además de su fecha de nacimiento. Realiza un esquema donde identifiques todas las tablas necesarias que van a intervenir en la base de datos, junto a sus atributos y las relaciones que existen entre ellas. Para la realización del esquema puedes usar la app draw.io**

****

1. **Investiga acerca de la minería de datos. Qué es, para que se utiliza y en qué casos, identifica el software necesario para trabajar con minería de datos y pon algún ejemplo real aplicable a una empresa.**

**Definición de *minería de datos:*** denominada descubrimiento de conocimiento en datos (KDD, en inglés), es el proceso de descubrir patrones y otra información valiosa en grandes conjuntos de datos sin procesar esta misma.

La minería de datos ***transforma datos en bruto en conocimiento práctico***.

**Usos de la minería de datos:** actualmente se usan en gran variedad de sectores para obtener mejores resultados.

* Ventas y marketing → recopilan una cantidad masiva de datos sobre sus clientes actuales y potenciales. Al observar la demografía de los consumidores y el comportamiento del usuario en línea, las empresas pueden utilizar los datos para optimizar sus campañas de marketing mediante mejoras de segmentación, ofertas de venta cruzada y programas de fidelidad del cliente
* Educación → para evaluar a los estudiantes, personalizar las lecciones y hacer que el aprendizaje sea más lúdico. Una visión unificada y basada en datos del avance de los alumnos puede ayudar a los educadores a determinar qué necesitan estos y a apoyarlos mejor.
* Bancos y seguros → resolver problemas complejos de fraude, cumplimiento, gestión de riesgos y deserción de clientes. Por ejemplo, las compañías de seguros pueden descubrir el precio óptimo del producto al comparar el rendimiento del producto en el pasado con los precios de la competencia.

Estos son algunos de los usos que se le da, existen muchos más.

**Software para trabajar con él:** para su poder realizar la minería de datos, podemos usar múltiples aplicaciones, que viene a ser el software para trabajar.

Entre los que existen algunos de ellos son:

* **Xplenty**

Esta plataforma ayuda a las empresas a preparar y recibir datos analíticos. Es una gran fuente para utilizar debido a sus muchas características y a que cuenta con un hardware y un software excepcionales.

Además tiene la opción de construir su propio ***pipeline de datos*** e incorpora un gran sistema de soporte al cliente. Debido a la variedad de características que puedes personalizar a tu gusto.

* **Rapid Miner**

Como su nombre deja entrever, se posiciona como una de las principales plataformas de minería de datos. Creada a partir del lenguaje de programación **Java**, cuenta con un sinfín de opciones que se centran en el análisis de negocio.

Permite crear conexiones «apuntar y hacer clic» con almacenes de datos empresariales, almacenamiento en la nube, redes sociales, aplicaciones empresariales, lagos de datos y bases de datos.

* **Teradata**

También llamada base de datos *teradata,* es otra opción de data mining ideal para empresas. Esta plataforma recoge información específica sobre las ventas, las preferencias de los clientes o la colocación de productos.

Esta plataforma multi-nube le da la portabilidad y flexibilidad para desplegar en cualquier lugar, como en las instalaciones y nubes públicas (Azure, AWS, Google Cloud).

Teradata es una de las mejores plataformas de data mining enfocadas a las ventas

**Ejemplo real →** Actualmente hay multitud de empresas que usan la minería de datos ya sea para mejorar la calidad de sus servicios, las ventas o ambas cosas. Para demostrar ese uso, podemos usar como ejemplo real la empresa a **NETFLIX**

1. **Qué errores encuentras en la siguiente imagen y cómo lo mejorarías haciendo uso de una base de datos relacional. Realiza un diseño basado en tablas de cómo quedaría tu solución.**

****

A simple vista podemos observar como nos haría falta más de una tabla para la correcta representación de los datos, es decir, una tabla principal de las películas que se relacione con actores y directores por separado cada una.

Otra cosa que debemos cambiar es la redundancia de los datos, en el que debemos de evitar poner datos irrelevantes o no repetir los mismos dentro de una sola tabla.

Tenemos que saber que en cada película actúan más de un actor por lo tanto esta debe de ir anidada según el nombre de la película a cada actor que trabaje en ella según el ID de la película.

**Mi solución propuesta sería la siguiente:**

| **ID Película** | Nombre | Año | Duración (min) | Género | Director | Compañía |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | Hulk | 2005 | 128 | Acción | Stan lee | Fox |
| **2** | Spider | 2008 | 58 | Terror | David Cronenberg | MTV |
| **3** | 300 | 2015 | 110 | Acción | Zack Snyder | Warner |
| **4** | Zohan | 2001 | 68 | Comedia | Dennis Dugan | Warner |

| **ID Actor** | Nombre | Fech-Nacimiento |
| --- | --- | --- |
| **1** | Tom | 19-8-1985 |
| **2** | Mabel | 21-3-1995 |
| **3** | Caprio | 5-6-2001 |
| **4** | Henry Cavil | 5-5-1983 |

| **ID Director** | Nombre | Fech-Nacimiento |
| --- | --- | --- |
| **1** | Stan Lee | 28-12-1922 |
| **2** | David Cronenberg | 15-3-1943 |
| **3** | Zack Snyder | 1-3-1966 |
| **4** | Dennis Dugan | 5-9-1946 |

| **ID Compañía** | Nombre | Dirección |
| --- | --- | --- |
| **1** | Fox | Los Ángeles, California |
| **2** | MTV | Nueva York |
| **3** | Warner Bros | Hollywood, Los Ángeles |

1. **El Sistema gestor de base de datos.**
2. **Define qué es, cómo interactúa con una base de datos, las operaciones que permite**

**¿Qué es un sistema gestor de base de datos?**

Son un conjunto de programas invisibles para el usuario final con el que se administra y gestiona la información que incluye una base de datos.

Estos permiten administrar todo acceso a la base de datos, pues tienen el objetivo de servir de interfaz entre esta, el usuario y las aplicaciones.

Entre sus funciones se encuentran la de permitir a los usuarios de negocio almacenar la información, modificar datos y acceder a los archivos de conocimiento de la organización. También se ocupa de realizar consultas y hacer análisis para generar informes.

1. **Identifica los 4 sistemas gestores de bases de datos de la actualidad más extendidos y utilizados. Realiza una tabla como esta:**

| **Sistema** | **Año 1º versión** | **Aplicaciones conocidas que lo utilizan** | **¿Software libre?** | **Cliente de BD para su explotación** | **Modelo de datos\*** |
| --- | --- | --- | --- | --- | --- |
| **MySQL** | 1995 | Airbnb, Uber, Netflix, Twitter | **SÍ** | Mysql Workbench | Base de datos relacional |
| **Maria DB** | 2009 | Google, Wikipedia y Alibaba | **SÍ** | Php My Admin | Base de datos relacional |
| **Oracle** | 1970 | Burger King, Coca Cola y Peugeot | **SÍ** | Oracle Instant Client | Base de datos relacional |
| **Postgresql** | 1989 | IKEA, BMW, UBER, Netflix o Instagram | **SÍ** | pgAdmin4 | Base de datos relacional y objeto-relacional |

1. **Un cliente nos solicita una base de datos para llevar el inventario y facturación de una empresa. No nos indican ningún tipo de preferencia en cuanto al modelo de datos a utilizar ni entra muy en profundidad en cuanto a los requisitos de usuario. Sin embargo, muestra especial preocupación por la integridad de sus datos. No quiere bajo ningún concepto que su información se pierda por completo y necesita que el sistema esté funcionando en “Alta disponibilidad”. Plantea con tus palabras una posible solución a su necesidad.**

Teniendo en cuenta las especificaciones que ha indicado el cliente en cuestión a la base de datos que desea, la solución que más se adecua a sus necesidades sería la siguiente:

Como modelo de base de datos, lo mejor sería relacional, que es ideal para garantizar la **integridad de los datos.** Al ser un sistema para inventario y facturación, se recomienda un esquema relacional que incluya tablas relacionadas para productos, stock, clientes, facturas y transacciones. Con ello, nuestro cliente ya asegura los datos de su empresa sabiendo que están en un lugar seguro.

En cuanto a la **alta disponibilidad** que desea tener el cliente, debemos de usar un *sistema gestor de base de datos* por ejemplo → Mysql; que nos permitirá gestionar todas las consultas y procesamiento de los datos que se requieran y el cliente quiera realizar.

1. **Investiga acerca del Big Data y el puesto específico existente en numerosas empresas llamado Business Intelligence. Indica sus tareas fundamentales e indica si te gustaría o no participar en dicho puesto razonando tus respuestas.**

**Definición →** nos referimos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño, complejidad y velocidad de crecimiento, dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales.

Se caracteriza por las “5 V”:

* **Volumen:** Grandes cantidades de datos.
* **Variedad:** Diversidad en los formatos de datos (estructurados, no estructurados, semiestructurados).
* **Velocidad:** Alta velocidad en la generación y procesamiento de datos.
* **Veracidad:** Calidad y exactitud de los datos.
* **Valor:** Extraer valor significativo de los datos.

La naturaleza compleja del Big Data se debe principalmente a la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas, como los weblogs, la identificación por radiofrecuencia (RFID), los sensores incorporados en dispositivos, la maquinaria, los vehículos, etc.

En cuanto a las empresas, el llamado **Business Intelligence (**BI**),** se refiere al uso de estrategias y herramientas que sirven para transformar información en conocimiento, con el objetivo de mejorar el proceso de toma de decisiones en una empresa.

Las tareas que se desarrollan son:

* **Minería de datos:** usar bases de datos, estadísticas y aprendizaje automático para descubrir tendencias en conjuntos de datos más grandes.
* **Generación de informes:** compartir análisis de datos con las partes interesadas para que todos puedan sacar sus propias conclusiones y tomar decisiones.
* **Métricas de rendimiento y valores de referencia:** comparar los datos del rendimiento actual con los datos históricos para hacer un seguimiento del rendimiento frente a los objetivos. En general, esto se lleva a cabo con dashboards personalizados.
* **Consultas:** el usuario realiza preguntas específicas relacionadas con los datos y se extrae las respuestas de los conjuntos de datos.
* **Análisis estadístico:** a partir de los resultados de análisis descriptivos, se exploran aún más los datos a través de estadísticas, por ejemplo, para determinar cómo ocurrió una tendencia y por qué.
* **Visualización de datos:** transformar el análisis de datos en representaciones visuales, como gráficos e histogramas, a fin de consumir más fácilmente los datos.
* **Análisis visual:** explorar los datos a través de la narración visual de historias para compartir información sobre la marcha y permanecer en el flujo de análisis.
* **Preparación de datos:** recopilar varias fuentes de datos, identificar las dimensiones y las medidas y preparar los datos para el análisis.

Sabiendo todo lo anterior, trabajar en este puesto en la empresa conlleva una gran responsabilidad, además de multitud de tareas en el que serías un trabajador fundamental. Desde mi más sincera opinión a mi me gustaría participar en este puesto, aunque conllevase bastantes quebraderos de cabeza o simplemente la dificultad de optar por ese puesto, pero creo que sería una meta personal el poder llegar a conseguirlo y con ello impulsar tu carrera y aprendizaje personal.