**Módulo 1: Introducción al mundo de los datos**

**Introducción:**

Una base de datos es la organización estructurada de un conjunto de información con al menos una característica en común que permite su agrupación. Además, permite la recopilación de la información para su posterior recuperación, para lo cual generalmente ofrece un motor de búsqueda interno que utiliza características especiales de cada artículo, con el fin de lograr una rápida y eficaz ubicación. Son denominadas fuentes secundarias, porque contienen información elaborada que hace referencia a fuentes primarias.

Existen bases de datos que recopilan información de varios países del mundo, mientras que otras son exclusivas de un país o una región determinada, la cual es otra diferencia fundamental al momento de realizar una búsqueda.

Existe evidencia que señala que la comunicación se usa como una estrategia de poder cuando se utiliza para incidir en el compromiso, en la moral, en el comportamiento y en lograr un mejor desempeño para alcanzar el cambio organizacional (Trahant, 2006). Entendida la estrategia como los medios a través de los cuales se busca alcanzar aquello que se ha propuesto, es decir, cómo se va a lograr algo. Cabe señalar que las estrategias o los medios de poder de acuerdo con Foucault (1979), se despliegan por medio de acumulación y uso de información, de manera que las estrategias se pueden vincular con la comunicación.

Este curso presenta el universo amplio y ambiguo de los sistemas de gestión del conocimiento de los datos, individual, colectivo y su utilización estratégica dentro de las organizaciones: las aplicaciones de gestión de contenidos, el trabajo colaborativo, las herramientas más o menos complejas de ayuda a la consulta de información y el concepto de expansión en conocimiento de los procesos de negocio.

En este módulo se abordarán las estructuras y funciones de los datos para su uso y consulta. El objetivo es que los estudiantes adquieran el conocimiento necesario para elegir el lenguaje de datos más apropiado para cada problema.

También se estudiará el tema de la comunicación en las empresas, la cual es un elemento esencial en la vida de las organizaciones, pues con ella es posible que las personas realicen acciones conjuntas, generen acuerdos, compromisos y participen en el cambio organizacional. En este módulo se busca analizar la influencia que tienen las comunicaciones como estrategias de poder en el cambio organizacional.

**1.1. Datos: estructura, tipos y funciones**

#### Introducción:

Las estructuras de datos representan modelos para gestionar y manipular grandes cantidades de datos de manera rápida y eficiente. En particular, las estructuras de datos funcionales son la llave para diseñar algoritmos eficientes. Las estructuras de datos se conforman de diferentes modelos, dependiendo de la estrategia para su almacenamiento y recuperación o según la secuencia en la que se presentan para realizar recorridos en las bases de datos.

La importancia de la informática y los sistemas computacionales radica fundamentalmente en su capacidad para procesar información, esta característica les permite realizar actividades que antes exigían mucho tiempo. Con el propósito de que la información sea procesada, se requiere que esta sea almacenada de acuerdo con la forma en que los datos se clasifican, ya sea simples o estructurados.

En la investigación, los datos presentan una parte importante para el desarrollo de la ciencia, ya que a partir de ellos los investigadores obtienen resultados tras su utilización y análisis. Es relevante que los datos sean compartidos con otras investigaciones a través de tecnologías de información y comunicación (TIC). Mediante la reutilización de los datos, puede asegurarse los resultados a nuevas conclusiones. Para que los datos sean de utilidad deben no solo compartirse, sino también publicarse.

#### Objetivo:

Identificar la estructura, los tipos y las funciones de los datos, con el fin de entender estos sistemas y sus propiedades.

Los temas a revisar para responder al objetivo de esta sesión son:

* Datos: Estructura y funciones
* Tipos de datos y sus propiedades

#### Consultas en programas de aplicación

Los sistemas de ficheros tuvieron un gran auge comparado a los sistemas manuales, como consecuencia de esto, aumentó la necesidad de realizar distintos tipos consultas de datos. No obstante, los sistemas de ficheros son dependientes del programador de aplicaciones, cualquier búsqueda o modificación que se requiera debe ser programada.

En algunas consultorías solo fijan el tipo de consulta, siendo imposible realizar otro movimiento en la base de datos que no fue considerada al escribir en el programa de aplicación.

#### Arquitectura de sistemas de bases de datos

El propósito de una arquitectura de base de datos es dividir, en niveles de abstracción, el esquema de una base de datos. Existen tres formas de ver o representar una base de datos:

* **Nivel interno:**describe la estructura física de la base de datos mediante un esquema detallado para el almacenamiento, así como los métodos de acceso.
* **Nivel global:**describe toda la base de datos para todos los usuarios mediante un esquema conceptual que oculta detalles de las estructuras de almacenamiento y se concentra en definir tablas y registros.
* **Nivel externo:**describe varios esquemas o visitas de usuario, cada uno detalla la parte de la base de datos que le interesa a un grupo de usuarios determinados y oculta el resto de las tablas y registros.

#### Modelos de datos

Las bases de datos están compuestas de registros precisos referentes a un sistema o parte de un todo que modelamos. Estos datos son fáciles de consultar y manipular cuando son limitados, pero cuando su volumen crece se requiere del uso de distintos sistemas de gestión.

Existen muchos modelos de diferente tipo, un modelo nos brinda herramientas de abstracción para representar una parte de los registros, cuyos datos son del interés del usuario. A esta representación se le conoce como esquema de los datos.

#### Tipos de base de datos

Existen distintos tipos de bases de datos:

**Modelo plano:** En un modelo de base de datos plano hay dos dimensiones de conjunto de datos. Existe una columna de información y dentro de esta columna cada dato estará asociado con la columna.

**Modelo jerárquico:** Este modelo de base de datos se asemeja a la estructura de un árbol. Busca conservar los datos organizados en un orden particular en un mismo nivel de lista.

**Modelo relacional:** Este modelo ayuda a organizar y estructurar datos de distintos tipos. El modelo relacional es una herramienta muy poderosa, no solo para almacenar información sino también para acceder a ella.

**Modelo de red:** La principal característica de un modelo de red es que se almacena un registro con un enlace a otros registros conformando una red. Estas redes pueden ser de diferentes tipos de información.

**1.2. Gestión de la información**

**Introducción:**

Las organizaciones con mayor éxito a nivel mundial son aquellas capaces de extraer los datos sensibles, para crear nuevo conocimiento de forma consistente, continua y extenderlo a través de la compañía, introduciéndose de forma rápida en nuevos productos y servicios existentes. Muchos autores refieren a estas organizaciones como empresas que crean conocimiento. Peter Druker, codificador y gestor de empresas, predijo que las organizaciones del siglo XXI estarán basadas en la información.

Las tecnologías de información constituyen, cada vez más, un acervo fundamental de las organizaciones. La información puede ser capturada en todos los procesos de la cadena de valor y gestionada a lo largo de esta. Aun así, también puede ser analizada y empaquetada para construir productos o servicios de información o para el surgimiento de nuevas líneas de negocio en la compañía. La información dispone de características, que la hacen ir más allá de los datos. La información permite interpretar tendencias y patrones y, por lo tanto, implica el juicio sobre los datos y la toma de decisiones.

Para alcanzar el éxito de la organización, es necesaria la implantación de una "cultura de la información", que se puede definir como los valores, los comportamientos y las actitudes que influyen en la manera en cómo los profesionales de una empresa perciben, recogen, organizan, procesan, comunican y utilizan la información. El objetivo principal de la gestión de la información en la empresa es satisfacer la demanda de información de la organización y proporcionar valor de una manera alineada con los objetivos de negocio.

**Objetivo:**

Analizar los objetivos principales de la gestión de la información, con el fin de entender el proceso del conocimiento sobre los distintos modelos de datos y fuentes de investigación.

Los temas a revisar para responder al objetivo de esta sesión son:

* Gestión de la información y del conocimiento
* Sistematización y automatización: información vs procesos

**Modelo de datos relacional**

Los datos relacionales se basan en un concepto matemático de relación, dicha teoría sugiere para su completa implantación que el ingreso a la memoria sea por contenido y no por dirección, como ocurre actualmente en los ordenadores, de hecho, esto se está implantando y desarrollando en la mayoría de los equipos.

Las principales ventajas de la utilización de bases de datos relacionadas son:

* Trabajan sobre las tablas en su conjunto, en lugar de hacerlo sobre registros.
* Permite realizar consultas complejas.
* Son de uso amigable, la utilización de los datos es independiente.

**Modelo de datos orientado a objetos**

Este modelo de base de datos trabaja en el concepto de encapsulamiento y código que opera sobre cada objeto. La interfaz entre un objeto y el resto del sistema se define mediante un conjunto de mensajes.

Un objeto tiene asociado:

* Un conjunto de variables que contienen los datos del objeto.
* Un conjunto de mensajes a los que el objeto responde.
* Un método, que es parte de un código para implementar cada mensaje.

**Modelo de datos en red**

Corresponde a una estructura de grafo, donde existe más de una conexión entre los nodos de diversos niveles. Este modelo se extiende al modelo jerárquico permitiendo que un registro participe en múltiples relaciones padre/hijo (Conjunto de modelos de red).

El inconveniente de esta estructura es el tener que utilizar mucha memoria al tener que almacenar en cada nodo las posiciones de los campos siguientes, mediante apuntadores.

**Fuentes de información**

Existen diferentes fuentes de información. Entre ellas se encuentran:

* **Fuentes de información informativas:** se emplean para localizar otros tipos de información.
* **Fuentes de información personales:** se utilizan para conocer datos referentes a personas.
* **Fuentes de información institucionales:** brindan datos de cualquier tipo sobre una determinada empresa o entidad.
* **Fuentes de información bibliográficas:** permiten encontrar información sobre publicaciones seriadas, libros o monografías.
* **Fuentes de información documentales:** muestran información de un documento específico y, si es posible, algún tipo de consulta.
* **Fuentes de información estadísticas:** aportan datos a diferentes niveles de exhaustividad, con o sin interpretación de los mismos, de una actividad o institución.

**1.3. Panorama global sobre el rol de los datos**

**Introducción:**

Los avances tecnológicos en gestión de datos cambian constantemente, por eso es importante buscar un asesoramiento actualizado antes de seleccionar un sistema de aplicación. La idea al desarrollar una base de datos no solo es la gestión, sino también el sistema de búsqueda, se debe considerar una base de datos descentralizada para facilitar su validación y uso de los datos. En un sistema distribuido, los datos se introducen y se validan localmente, pero se vinculan a otros sistemas de análisis.

El objetivo de esta unidad de aprendizaje es comprender que una base de datos digital tiene la capacidad de almacenar la cantidad de datos necesarios para cada organización, ya sean diez o un millón de registros, esto implica que una base de datos almacena prácticamente infinitos datos de una empresa, que en un futuro puedan ser consultados de forma confiable, rápida y segura.

Existen diversos modelos para el almacenamiento de datos, siendo el modelo relacional el más habitual en la actualidad. En el modelo relacional, la información se organiza en tablas relacionadas entre sí, cada fila de una base de datos conforma una tupla que contiene la información correspondiente a una entidad dada.

El diseño de la base de datos es de gran importancia, este conlleva el diseño de un modelo conceptual, el diseño de un modelo físico, la implementación y el mantenimiento. Herramientas como los diagramas E--R son de ayuda en las fases de diseño, cuyo principal objetivo es crear una estructura de la base de datos que facilite la interpretación de la información contenida y permita sacar el máximo rendimiento de esta.

En la actualidad la mayoría de las empresas manejan una base de datos digital, esto se debe a que no es necesario tener mucho trabajo manual, el sistema se encarga directamente de realizar las tareas, para así poder tener la información de forma ordenada y segura. Toda base de datos incluye un ingreso, proceso y salida de datos, en donde los comandos trabajan internamente de forma independiente, aunque en la mayoría de las ejecuciones las sentencias de escritura sean similares.

**Objetivo:**

Obtener un panorama general de la administración de las bases de datos, con el fin de promover sus ventajas en la empresa u organización.

Los temas a revisar para responder al objetivo de esta sesión son:

* Panorama global sobre el rol de los datos

**Modelos de datos conceptuales**

Los modelos de datos conceptuales están orientados a la descripción de estructura de datos y restricciones de integridad, estos se utilizan en la etapa de análisis en un problema y están orientados a exponer los elementos que intervienen en ese problema y sus relaciones.

Un ejemplo es el diagrama o modelo entidad-relación, el cual es una herramienta para el modelado de datos de un sistema de información. Estos modelos expresan entidades relevantes para un sistema de información, así como sus interrelaciones y propiedades.

**Modelos de datos lógicos**

Los modelos de datos lógicos están orientados a las operaciones más que a la descripción de una realidad, normalmente están implementados en algún Gestor de Base de Datos.

Un ejemplo es el Modelo Relacional, que cuenta con buenas características conceptuales. El modelo relacional para la gestión de una base de datos es un modelo basado en la lógica de predicados y en la teoría de conjuntos, es el modelo más utilizado en la actualidad para modelar problemas reales y administrar datos dinámicamente.

**Modelo de datos físicos**

El modelo de datos físicos es un modelo específico de bases de datos que representa objetos de datos relacionales y sus relaciones. Un modelo de datos físico se puede utilizar para generar sentencias DDL que después se puede desplegar en un servidor de base de datos.

Son estructuras de bajo nivel implementadas dentro del propio gestor de base de datos. Ejemplos típicos de estas estructuras son los árboles B+ y las estructuras Hash.

**Estructura de datos**

Las estructuras de datos son la forma de organización de los datos. Las mismas se clasifican en internas y externas.

Dentro de las estructuras de datos internas se encuentran:

* **Estáticas:** almacenan arreglos y matices y no cambian su tamaño.
* **Dinámicas:** pueden ser lineales, es decir, abarcan líneas, pilas y colas y no lineales, las cuales abarcan árboles y grafos.

Cada una de estas estructuras de datos tienen una forma diferente de trabajar. Por ejemplo, los arreglos y matices se recorren con índices.