

# Fundamentos de Bases de Datos

## Tarea 1

Profesor: M.I. Gerardo Avilés Rosas

Ayudantes: Ailyn Rebollar Pérez y Rodrigo Alejandro Sánchez Morales

Alumnos: Calvario González Berenice, Navarrete Baltazar Mario, Reyes Ordóñez Marilyn Midori, Serratos Ramirez Brian, Solis Chavez Arnold

Fecha de entrega: Junio 28

### 1. Conceptos generales:

- a. Describe las principales características del **enfoque de bases de datos** y contrástalo con el enfoque basado en **archivos**. ¿En qué casos **no tendría sentido** utilizar una base de datos?

**RESPUESTA:** En el enfoque basado en archivos, cada usuario define e implementa los archivos necesarios para una aplicación concreta como parte de la programación de esa aplicación. Es decir, dos usuarios pueden almacenar en distintos archivos distintos datos sobre un mismo objeto. Esta redundancia en la definición y el almacenamiento de datos derrocha espacio de almacenamiento e implica esfuerzos redundantes por mantener al día datos comunes.

En el enfoque de bases de datos los datos se mantiene en un único lugar, que se define una sola vez, y al que acceden varios usuarios. En los sistemas de archivos cada aplicación tiene libertad para asignar un nombre independientemente a los elementos de datos. Mientras que en una base de datos, los nombres o etiquetas de los datos se definen una vez, y son utilizados por consultas, transacciones y aplicaciones. Las principales características del enfoque de bases de datos frente al enfoque basado en archivos son las siguientes:

- Naturaleza autodescriptiva de un sistema de bases de datos: El sistema de bases de datos no sólo contiene la propia base de datos, sino también la definición o descripción de la estructura de la base de datos y sus restricciones.
- Aislamiento entre programas y datos, y abstracción de datos: En el enfoque basado en archivos si se realiza un cambio en la estructura de un archivo se puede realizar cambios en todos los programas que acceden a ese archivo. En el caso del enfoque de las bases de datos, en la mayoría de los casos los programas que acceden al SMBD no necesitan esos cambios. La estructura de los archivos de datos se almacena en SMBD, independientemente de los programas de acceso. A esto se le llama independencia programa-datos.
- Soporte de varias vistas de los datos.
- Compartición de datos y procesamiento de transacciones multiusuario: Permitir que varios usuarios puedan acceder a la base de datos al mismo tiempo. Además es necesario el control de la concurrencia para que esos varios usuarios que intentan actualizar los mismos datos, lo hagan de un modo controlado para que el resultado de la actualización sea correcto. El SMBD debe garantizar que las transacciones concurrentes operen correcta y eficazmente.

No tendría sentido utilizar bases de datos cuando tenemos información que no necesita ser modificada o actualizada, que será sólo para consultar, los datos que no seguirán creciendo y son limitados, y no necesitan ser consultados por muchos usuarios. Para ese caso, el enfoque basado en archivos es útil.

- b. ¿Qué **ventajas** y **desventajas** encuentras al trabajar con una **bases de datos**?

**RESPUESTA:** Las bases de datos son colecciones de datos interrelacionados con un objetivo en común.

Ventajas	Desventajas
-Reducción de tiempo de actualización de los datos	-El tamaño aumenta de manera creciente, lo que provoca que pueda desestabilizar la base de datos.
-Facilitación de la consulta de información	-Si el tamaño incrementa es necesario ampliar el espacio de almacenamiento o cambiar el motor de la base de datos, pues puede que el anterior entorpezca el manejo de los datos.
-Integridad de datos, es decir, la información no se repite ni se pierde el almacenamiento	-Incremento de costos en caso que se den las situaciones anteriores.
-Seguridad, ya que la información sólo puede ser consultada por personas con acceso autorizado.	-Necesitan de actualizaciones constantes tanto de servidores como de aplicaciones.
-Independencia de los datos, programas y procesos.	-En caso de algún fallo en el hardware, software o debido al ataque de un hacker las bases de datos pueden fallar dejando a los usuarios sin datos.
-Datos más documentados gracias al uso de metadatos.	
-Almacenamiento de una cantidad de datos ilimitados.	
-Compartir datos de manera global.	

- c. Describe el papel que tienen los **Sistemas Manejadores de Bases de Datos (SMBD)** en el enfoque de **bases de datos**. ¿Por qué consideras **que es importante** (o no) que un administrador de bases de datos conozca las **características** de un SMBD?

**RESPUESTA:** Un sistema manejador de bases de datos es una colección de programas que permite a los usuarios crear y mantener una base de datos. El SMBD es un sistema de software de propósito general que facilita los procesos de definición, construcción, manipulación y compartición de bases de datos entre varios usuarios y aplicaciones. La definición involucra especificar los tipos, estructuras y restricciones de los datos almacenados en la base de datos. La construcción de la base de datos es el proceso consistente en almacenar los datos en algún medio de almacenamiento controlado por el SMBD. La manipulación de una base de datos incluye funciones como la consulta de la base de datos para recuperar datos específicos, actualizar la base de datos para reflejar los cambios introducidos en el minimundo y generar informes a partir de los datos. Compartir una base de datos permite que varios usuarios y programas accedan a la base de datos de forma simultánea. Otras funciones importantes ofrecidas por el DBMS son la protección de la base de datos y su mantenimiento durante un largo periodo de tiempo. La protección incluye la protección del sistema contra el funcionamiento defectuoso del hardware o el software (caídas) y la protección de la seguridad contra el acceso no autorizado o malintencionado. Una gran base de datos típica puede tener un ciclo de vida de muchos años, por lo que el SMBD debe ser capaz de mantener el sistema de bases de datos permitiendo que el sistema evolucione según cambian los requisitos con el tiempo.

En un entorno de bases de datos, el recurso principal es la base de datos en sí misma, mientras que el recurso secundario es el SMBD y el software relacionado. La administración de estos recursos es responsabilidad del administrador de la base de datos (ABD). El ABD es responsable del acceso autorizado a la base de datos, de la coordinación y monitorización de su uso, y de adquirir los recursos software y hardware necesarios. El ABD también es responsable de problemas como las brechas de seguridad o de unos tiempos de respuesta pobres. Debido a estas responsabilidades es que el ABD debe conocer las características de los SMBD.

- d. ¿Qué es la **independencia de datos**? ¿Cuál **tipo de independencia** de datos es **más difícil** de lograr? Justifica tu respuesta.

**RESPUESTA:** La arquitectura de tres esquemas (ANSI/SPARK) se puede utilizar para explicar el concepto de independencia de los datos, que puede definirse como la capacidad de cambiar el esquema en un nivel de un sistema de bases de datos sin tener que cambiar el esquema en el siguiente nivel más alto. Se pueden definir dos tipos de independencia de datos:

- Independencia física de datos: Es la capacidad de cambiar el esquema interno sin que haya que cambiar el esquema conceptual. Por tanto, tampoco es necesario cambiar los esquemas externos.
- Independencia lógica de datos. Es la capacidad de cambiar el esquema conceptual sin tener que cambiar los esquemas externos o los programas de aplicación. Es posible cambiar el esquema conceptual para expandir la base de datos, para cambiar las restricciones o para reducir la base de datos.

La independencia física de datos existe en la mayoría de las bases de datos y de los entornos de archivos en los que al usuario se le ocultan la ubicación exacta de los datos en el disco, los detalles hardware de la codificación del almacenamiento, la colocación, la compresión, la división, la fusión de registros, etcétera. Las aplicaciones siguen obviando estos detalles. Mientras que la independencia lógica de datos es muy difícil de conseguir porque permite los cambios estructurales y restrictivos sin afectar a los programas de aplicación.

- e. Explica la **diferencia** entre los **esquemas externo, interno y conceptual**. ¿Cómo se relacionan estas diferentes **capas de esquemas** con los conceptos de **independencia de datos** lógica y física?

**RESPUESTA:** El esquema interno describe la estructura de almacenamiento físico de la base de datos. El esquema interno utiliza un modelo de datos físico y describe todos los detalles del almacenamiento de datos y las rutas de acceso a la base de datos. El esquema conceptual describe la estructura de toda la base de datos para una comunidad de usuarios. El esquema conceptual se concentra en describir las entidades, los tipos de datos, las relaciones, las operaciones de los usuarios y las restricciones. Normalmente, el esquema conceptual se describe con un modelo de datos representativo cuando se implementa un sistema de bases de datos. El esquema externo describe la parte de la base de datos en la que un grupo de usuarios está interesado y le oculta el resto de la base de datos, cada esquema externo se implementa normalmente mediante un modelo de datos representativo. Como se dijo en la pregunta anterior, éstos esquemas están relacionados con los conceptos de independencia de datos lógica y física pues permiten el cambio de un esquema en un nivel sin que afecte a los otros esquemas.

- f. Investiga qué es la **redundancia de datos**. ¿De **qué forma** crees que **pueda afectar** el trabajo de una base de datos?

**RESPUESTA:** La redundancia de datos es la repetición innecesaria de información en varios ficheros. También se produce cuando hay datos que no aportan información adicional, puesto que se pueden obtener o calcular a partir de otros. La redundancia se debe a la ausencia de integración de datos en ficheros procedentes de departamentos diferentes o al uso de lenguajes diferentes.

La redundancia puede traer problemas a la base de datos como la falta de integridad de los datos, el derroche del espacio de almacenamiento (coste del espacio ocupado inútilmente), la dificultad para mantener actualizadas los datos comunes y la recuperación lenta de los datos.

- g. Investiga **cuáles** son las **responsabilidades** de un **DBA**. Si asumimos que el DBA **nunca está interesado** en ejecutar sus **propias consultas**, ¿necesita entender sobre **optimización de consultas**? ¿por qué?

**RESPUESTA:** Una de las principales razones de usar SMBDs es tener un control centralizado tanto de los datos como de los programas que tienen acceso a esos datos. La persona que tiene ese control central sobre el sistema se denomina administrador de bases de datos (ABD). Las responsabilidades del ABD incluyen:

- La definición del esquema. El ABD crea el esquema original de la base de datos mediante la ejecución de un conjunto de instrucciones de definición de datos en el lenguaje para definición de datos (LDD).
- La definición de la estructura y del método de acceso.
- La modificación del esquema y de la organización física. El ABD realiza modificaciones en el esquema y en la organización física para reflejar las necesidades cambiantes de la organización, o para alterar la organización física a fin de mejorar el rendimiento.
- La concesión de autorización para el acceso a los datos. Mediante la concesión de diferentes tipos de autorización, el administrador de bases de datos puede regular las partes de la base de datos a las que puede tener acceso cada usuario. La información de autorización se guarda en una estructura especial del sistema que el SMBD consulta siempre que alguien intenta tener acceso a los datos del sistema.
- El mantenimiento rutinario. Algunos ejemplos de las actividades de mantenimiento rutinario del administrador de la base de datos son:
  - Copia de seguridad periódica de la base de datos, bien sobre cinta o sobre servidores remotos, para impedir la pérdida de datos en caso de desastres.
  - Asegurarse de que se dispone de suficiente espacio libre en disco para las operaciones normales y aumentar el espacio en disco según sea necesario.
  - Supervisar los trabajos que se ejecuten en la base de datos y asegurarse de que el rendimiento no se degrade debido a que algún usuario haya remitido tareas muy costosas.

Si se supone que el ABD no está interesado en hacer sus propias consultas, de toda forma debe saber sobre optimización de consultas, ya que una de sus responsabilidades es asegurarse que la base de datos tenga un buen rendimiento, si las consultas tienen tiempo de respuesta lento, entonces eso indica que el ABD no está realizando sus tareas de forma correcta.

- h. ¿Qué son los **datos estructurados** y los **no estructurados**? Proporciona **algunos ejemplos** para contrastarlos. ¿Cuál **tipo es más prevaleciente** en un **ambiente de negocios** típico?

**RESPUESTA:** La estructura de datos es la forma en que se encuentran organizados un conjunto de datos.

Datos estructurados: La información viene representada por un conjunto de datos atómicos elementales (datos de tipo simple, no compuestos de otras estructuras) o agrupaciones de ellos. Se conoce de antemano la organización de los datos, la estructura y el tipo de cada dato elemental, su posición y las posibles relaciones entre los datos. Por ejemplo, se encuentran en las bases de datos relacionales, en las aplicaciones operacionales (ERP, CRM, aplicaciones de RR. HH.), o en ficheros con una estructura fija en forma de tabla, como por ejemplo ficheros CSV u hojas de cálculo.

Datos no estructurados: Son aquellos donde la información no aparece representada por datos elementales, sino por una composición cohesionada de unidades estructurales de nivel superior. El valor informacional de estos datos tiende a ser mayor que el de los estructurados, pero su interpretación y manipulación resulta mucho más compleja. Ejemplos de orígenes de datos no estructurados son textos, audios, imágenes o vídeos.

El tipo de dato que más prevalece es el estructurado, pues se puede manejar sin mucha dificultad, permitiendo ser de fácil análisis en los negocios.

- i. ¿Por qué **no se puede considerar** que una **hoja de cálculo** sea una **base de datos**? Indica las **diferencias más representativas** con respecto a una base de datos.

**RESPUESTA:** Se tiene que una de las características importantes en una base de datos es que asegura la integridad de los datos, las hojas de cálculo no tienen la capacidad de poder validar información, es decir, no se puede validar si los datos han sido duplicados dentro de la hoja de cálculo. Es por ello que no se puede considerar una base de datos.

Las diferencias más representativas además de la expuesta arriba son:

- Tienen restricciones respecto al número de datos que pueden almacenar.
  - No hay control de usuarios, por ejemplo, el número de usuarios está limitado.
  - El tamaño de los archivos es limitado.
- j. Imagina que eres el **director de TI** de una **cadena de minoristas a nivel nacional**. Redacta un **informe** al presidente ejecutivo explicando **las oportunidades, beneficios y ventajas** que se pueden tener si se implementa un **sistema de base de datos** en la compañía.

---



---

## INFORME DE IMPLEMENTACIÓN DE UNA BASE DE DATOS

Cliente: Presidente ejecutivo de la compañía XX

Autor: Director TI de la compañía XX

28 de junio de 2021

### RESUMEN EJECUTIVO

#### OBJETIVOS:

- Informar sobre las oportunidades, beneficios y ventajas de implementar una base de datos.

## RESUMEN:

En el siguiente informe se enfatiza el beneficio de utilizar sistemas de bases de datos en las compañías.

A continuación se describe cómo almacena y guarda datos la compañía XX.

## ALMACENAMIENTO ACTUAL EN LA COMPAÑÍA XX.

La compañía XX utiliza el sistema enfocado a archivos, este sistema almacena los registros permanentes en varios archivos y necesita diferentes programas de aplicación para extraer y añadir a los archivos correspondientes. Lo que provoca diversas desventajas como:

- Redundancia e inconsistencia de los datos.
- Dificultad en el acceso a los datos.
- Aislamiento de datos. Como los datos están dispersos en varios archivos, y los archivos pueden estar en diferentes formatos, es difícil escribir nuevos programas de aplicación para recuperar los datos correspondientes.
- Problemas de integridad.
- Anomalías en el acceso concurrente.
- Problemas de seguridad.

## FORMAS DE MEJORAR EL ALMACENAMIENTO DE DATOS EN LA COMPAÑÍA XX.

Un sistema de base de datos es una colección de datos interrelacionados y un conjunto de programas que permiten a los usuarios tener acceso a esos datos y modificarlos. Las ventajas de utilizar éstas bases de datos en la compañía XX son:

- Almacenar los datos de la empresa en un único lugar.
- Reducción de tiempo de actualización de los datos. Es decir, se puede obtener información de clientes, proveedores, empleados, etc., de manera actualizada y veloz.
- Facilitación de la consulta de información. Esto permite consultar la información sin necesidad de revisar aplicación por aplicación para ver cuál contiene la información que se necesita.
- Integridad de datos, es decir, la información no se repite ni se pierde el almacenamiento.
- Seguridad, ya que la información sólo puede ser consultada por personas con acceso autorizado.
- Independencia de los datos, programas y procesos.
- Datos más documentados gracias al uso de metadatos.
- Compartir datos de manera global.

Las ventajas anteriores pueden beneficiar a la compañía a crecer ya que: 1. Permite que distintas áreas de trabajo compartan la misma información, evitando redundancia de datos; 2. mejor manejo de la información para juntas/negociaciones con clientes o proveedores; 3. aumento de eficiencia dado que habrá tareas que se realicen con mayor rapidez y agilidad, reduciendo tiempos de espera; 4. hacer análisis de datos y usarlo en campañas de marketing para mejorar las ventas y posicionarnos en un lugar mejor en el mercado.

## CONCLUSIONES.

Los sistemas basados en archivos son útiles cuando no se manejan grandes cantidades de datos y no es necesario el manejo y análisis de ellos, ya que si se desea esto último se derrochará espacio de almacenamiento, tiempo y surgen problemas como los mencionados anteriormente.

Al ser una compañía que está creciendo, es necesario que también crezca y se desarrolle en distintas áreas, una de ellas es la área de TI, para eso es necesario actualizar los conocimientos y herramientas que se usan normalmente. Se puede dar el primer paso usando sistemas de bases de datos.

---

---

## 2. Lectura

Leer el artículo **Data's Credibility Problem** y realizar un **resumen** del documento, destacando los puntos que a su consideración sean los **más relevantes** (no más de una cuartilla).

Realizar un **ensayo** donde expresen **sus comentarios** (esta parte la deben responder de **forma individual**) sobre la lectura, considerando los siguientes puntos:

- Deberás indicar **cuál es el objetivo** que quiso plantear el autor: *qué intenta decir, de qué intenta persuadirnos y/o convencernos, ¿cómo se relaciona con la materia de **Fundamentos de Bases de Datos**?*
- Deberás indicar **cuál es la temática central** del artículo y se deben señalar el *tema o los temas laterales* que desarrolla el mismo y cómo estos tienen relación con tu práctica profesional.
- En el caso que plantea la **ejecutiva de gestión de productos**, qué **recomendaciones** le darías para evitar que el futuro tenga que pasar la misma situación.
- En el caso de que **no fuera posible** resolver el tema de **calidad de datos** desde el punto de creación, ¿qué recomendaciones le brindarías no solo a la ejecutiva, sino a la organización.

- **Consideraciones personales:** deberás indicar una postura ante las ideas planeadas en el documento, proporcionar argumentos a favor o en contra (propios).
- Antes de leer el artículo, ¿habías escuchado o conocías sobre el término **calidad de datos**? ¿cambió tu percepción sobre el tema después de leer el artículo?

## RESPUESTA:

### RESUMEN

Existen problemas con la calidad de datos, un ejemplo de ello es: una ejecutiva de gestión de productos en ascenso se prepara para una presentación y se da cuenta de que algo se ve mal en las cifras de participación de mercado. Le pide a su asistente que verifique las cifras, él indaga y encuentra un error en los datos proporcionados por el departamento de investigación de mercado, y la ejecutiva hace las correcciones necesarias. Ella concluye que se debe establecer una política para verificar los datos cada vez. No piensan informar al departamento de Investigación de Mercado sobre el error, y tampoco trabajar con el grupo para asegurarse de que la próxima vez se proporcionen los datos adecuados. Así, los problemas de calidad de datos afectan a todos los departamentos, industrias, niveles y tipos de información.

Generalmente los empleados terminan solucionando o arreglando la mayoría de los errores a medida que van realizando su trabajo. Pero esto provoca grandes costos. Los trabajadores terminan perdiendo hasta el 50 % del tiempo buscando datos, identificando y corrigiendo errores y buscando fuentes confirmatorias de datos en los que no confían. Los errores pueden provocar consecuencias graves.

Cuando los datos no son confiables, los gerentes pierden la fe en ellos y recurren a su intuición para tomar decisiones, dirigir sus empresas e implementar estrategias.

Aún después de cincuenta años de que se acuñara la expresión "basura entra, basura sale", todavía se tiene problemas con la calidad de los datos. Solucionar el problema no es tan difícil como se puede pensar, esta solución es una mejor comunicación entre los creadores de datos y los usuarios de datos; un enfoque en mirar hacia adelante; y, sobre todo, un cambio en la responsabilidad de la calidad de los datos de la gente de TI, que no es dueña de los procesos comerciales que crean los datos, y pasa a manos de los gerentes, que están muy interesados en obtener los datos correctos.

### CONECTE A LOS CREADORES DE DATOS CON LOS CLIENTES DE DATOS

Desde una perspectiva de calidad, sólo dos momentos importan en la vida útil de un dato: el momento en que se crea y el momento en que se utiliza. La calidad de datos se fija en el momento de su creación. Pero en realidad se juzga esa calidad hasta el momento de su uso. Si se considera que la calidad es mala, la gente trabaja en torno a los datos o corrigiendo los errores ellos mismos.

Mejorar la calidad de los datos se trata de lograr que los creadores de datos se asocien con los usuarios, sus clientes, para que puedan identificar las causas fundamentales de los errores y encontrar formas de mejorar la calidad en el futuro. Por decir algo, en el ejemplo de la ejecutiva ella no informó al departamento de Investigación de Mercado del error que encontró y lo corrigió ella misma, dejó que otros fueran víctimas de los mismos datos erróneos provenientes del departamento. También se encargó de ajustar los datos a pesar de que era menos calificada para hacerlo que los creadores de datos.

Un poco de comunicación es muy útil. Asegurarse de que los creadores sepan cómo se utilizarán los datos es una de las formas más fáciles y eficaces de mejorar la calidad.

Abordar la mayoría de problemas de calidad de los datos no requiere de grandes inversiones en nuevas tecnologías o reingeniería de procesos. El primer paso decisivo es simplemente lograr que los usuarios y los creadores de datos se comuniquen entre sí.

### CONCÉNTRENSE EN OBTENER NUEVOS DATOS CORRECTAMENTE

Una vez que una empresa se da cuenta de que la calidad de sus datos es mala, su primer reacción suele ser limpiar los datos defectuosos existentes. Un mejor enfoque es centrarse en mejorar la forma en que se crean nuevos datos, identificando y eliminando las causas fundamentales del error. Una vez que se haya realizado ese trabajo, es posible que se requieran limpiezas limitadas, pero no será necesario realizar una limpieza continua.

En lugar de lanzar un esfuerzo masivo para limpiar los datos defectuosos existentes, las empresas deben centrarse en mejorar la forma en que se crean los nuevos datos.

Es necesario poner la responsabilidad de los datos en manos de los gerentes de línea. A menudo, los creadores de datos no están vinculados organizativamente con los usuarios de datos.

Cuando los problemas de calidad se vuelven generalizados o graves, la respuesta de la organización es encomendar al departamento de TI que los solucione. Pero los departamentos de TI no suelen tener éxito en la conducción de calidad de datos. Eso es porque la calidad de datos se fija en el momento de la creación y eso no ocurre en TI, excepto raras veces. Todo lo que pueden hacer es encontrar y corregir errores, lo que no es una solución a largo plazo.

Las empresas inteligentes asignan la responsabilidad de la calidad de los datos a los creadores de datos y sus clientes de datos internos. Para la mayoría de las empresas, la verdadera barrera es que se niegan a admitir que sus datos no son buenos y otros no saben cómo corregir los datos de mala calidad. Superar esa meseta requiere el compromiso de los altos directivos.

## ENSAYOS INDIVIDUALES

### 1. Ensayo de Calvario González Berenice

El artículo nos habla sobre como es muy fácil llegar a tener errores en los datos de las bases de datos y el gran impacto que pueden llegar a tener los muchos errores que se filtran, entre ellos perder mucho tiempo buscando datos, identificando y corrigiendo errores. También nos plantea la idea de que es posible hacer algo respecto a la calidad de datos, que una mejor comunicación entre los usuarios y los creadores podría mejorar mucho la calidad de los datos. Nos dice que el problema no radica en el área de TI, si no en los creadores de datos y en sus clientes de datos internos.

En el vídeo de Conceptos Básicos del curso se hace mucho incapie en el hecho de que una base de datos debe estar bien planeada desde un principio porque si no luego se vuelve un animal difícil de domar.

Tener datos erróneos puede ser algo muy costoso de corregir y que podría llevar mucho tiempo corregirlo, por lo cual la idea de ser mucho más cuidadoso a la hora de crear datos y estar constantemente revisando que los datos ingresados anteriormente sean correctos ya no suena como una tarea pesada a comparación con lo que se tendría que hacer y el gran gasto y tiempo que implicaría si no se tuviera el cuidado necesario.

Personalmente también creo que la tarea de mejorar la calidad de datos se encuentra entre los usuarios y los creadores ya que ahí es el inicio de todo y si se hace bien desde un principio el error podría ser mínimo, igual los datos que son erróneos no tomarlos en cuenta ya que estos afectan mucho y no se pueden ver los resultados tan apegados a la realidad. Esto requiere de mucho compromiso por parte de las personas a las que les toca encargarse de esto, pero realmente es muy importante la calidad de datos ya que de esto dependen muchas cosas, como el ejemplo que mencionaban sobre datos médicos en donde se encuentra la salud e incluso vidas de por medio.

El objetivo del autor es hacernos ver los problemas que trae consigo una base de datos con mala calidad de datos, algunas de las consecuencias (ya anteriormente mencionadas) que implica y la forma en que comunmente se tratan de arreglar estos problemas los cuales no son las mejores porque simplemente tratan de corregir los problemas por uno mismo pero eso para nada es una solución porque solo es una solución temporal y lo que se busca es tener siempre todo bien hecho.

Antes de leer el artículo y del curso no había escuchado sobre el término calidad de datos pero ahora que sé sobre el puedo ver el gran impacto que puede tener el hecho de no contar con calidad de datos, y cómo es que grandes empresas llegan a tener problemas muy grandes por ello.

### 2. Ensayo de Navarrete Baltazar Mario

El artículo nos describe una serie de ejemplos donde datos corruptos han sido costosos para las empresas, y que muchas veces se debe a que los datos fueron mal planteados desde el principio y que las empresas optan por corregir estos datos que están mal en específico pero esto solo soluciona el problema al momento ya que al estar los datos mal desde la raíz hace que nuevamente resurjan datos mal interpretados. Otro punto importante que hace el artículo es el que la solución óptima está desde la comunicación con el cliente, una buena comunicación cliente asegura que los datos recolectados sean correctos, esto es lo que mejorada la calidad de los datos, no el hecho de corregir los datos de alguien más, ya que como antes mencionamos esto solo provocará que se solucione temporalmente el problema.

Uno de los ejemplos más destacados es el de la ejecutiva de la gestión de productos, el ejemplo es que ella antes de realizar la presentación de su producto ella se da cuenta que los datos que le dieron esta mal, y los corrige en ese momento que hace que su presentación sea un éxito, sin embargo, esta solución solo fue temporal, lo que ella debe hacer es ir a hablar con el equipo de marketing e informales que sus datos son erróneos y así re evalúen su proceso de recuperación de datos, ya que sin esta retroalimentación el equipo de marketing seguirán teniendo datos equivocados sin saber que están mal. Otra solución sería fijarse en los metadatos, esta solución se puede aplicar a toda la empresa, no involucra una comunicación directa con el cliente como tal pero sí es una solución más técnica en cómo se están planteando los datos.

Estoy a favor de la idea de que la solución está en la creación de los nuevos datos a partir de la comunicación con el cliente y no en dejárselo al equipo de IT que se revuelva cada dato erróneo al momento.

En conclusión el enfoque debe ser en la construcción de datos correctos desde un inicio tomando en cuenta también los metadatos ya que muchas veces la raíz del problema suele estar en esa parte. . . Algo que no se debe considerar es dejarle la resolución del problema siempre a los gerentes, los gerentes que usualmente se consideran son los que se encuentran en el nivel de senior, estos si se encuentran desmotivados o sin poder acceder a otras áreas hace que el cambio en que hicieron en esta corrección de datos hace que solo su área tenga buenos datos y el resto tenga malos datos. Personalmente no había escuchado el término calidad de datos, el artículo me hizo ver lo importante que es la buena planeación de datos ya que si estos son incorrectamente planteados, en el futuro hará que se tenga realizar correcciones que serán costosas. . . sin mencionar que estas correcciones solo resolverán el problema temporalmente antes de que los datos corruptos de la raíz vuelvan a ocasionar otro problema.

### 3. Ensayo de Reyes Ordóñez Marilyn Midori

La calidad de datos es algo importante que debe considerarse en cualquier tema, tal como: negocios, medicina, finanzas, economía, entre otros. Por medio de distintos ejemplos el autor quiere que veamos la importancia de saber sobre la calidad de datos y la solución para lograrla. Si los datos que se tienen no son confiables, entonces esto puede ocasionar consecuencias como: pérdida de dinero, una medición incorrecta en un hospital puede matar pacientes, etc., esto nos lleva a pensar en cómo aprendemos sobre el manejo de datos, es importante saber la responsabilidad que tendremos si nos dedicamos a tratar con ellos, ya sea creando, diseñando o administrando bases de datos.

Durante la clase de Fundamentos de bases de datos se señaló que: en la práctica, sólo hay dos momentos que importan en la vida útil de un dato: cuando se crea y cuando se utiliza; ésta es otra de las ideas importantes que el autor en el artículo

también nos hace saber. Justo es en esta idea que se sostiene la solución para obtener una buena calidad de datos: mejorar la comunicación entre los que crean los datos y en los que lo utilizan, muchas veces los que utilizan los datos se dan cuenta que están obteniendo datos erróneos y lo arreglan por sí mismos sin avisar a los encargados de crearlos, esto es una solución sólo para el grupo específico, los datos erróneos siguen circulando y habrá otros que tendrán que corregirlos o bien, utilizarlos sin darse cuenta de la mala calidad de los datos; éste problema se pudo haber resuelto si los que encontraron los datos erróneos lo hubiesen notificado para arreglar el problema y verificar que los datos que están siendo creados son lo que realmente se está pidiendo, y finalmente empezar a obtener datos confiables, después sería necesaria una limpieza de los datos obtenidos anteriormente. Justo esto pasa en el caso del ejemplo de la ejecutiva, las recomendaciones para evitar errores en el futuro sería que notificara al área de Investigación de Mercado para hacerle saber que los datos están mal, para que ellos corrijan el error, además de pedirles que antes de que envíen los datos los revisen.

A partir de ahora es muy probable que, en el área de ciencia y tecnología en la que nos encontramos, lidemos con grandes cantidades de datos e información, considerar la solución propuesta nos ayudará en el futuro, ya seamos nosotros los encargados de crear datos o de utilizarlos.

El autor también nos menciona las acciones que muchos gerentes pueden llegar a tomar en caso de tener mala calidad, por ejemplo, sólo hacer limpieza de los datos, ésta no es si no sólo una solución momentánea. Creo que esto se debe a que no todos conocen el hecho de que uno de los puntos importantes es la creación de datos, se le toma importancia hasta que son utilizados. Conuerdo con el autor con que se debe dejar la responsabilidad a los creadores de datos y los usuarios, pero creo que también es importante que se enseñe a los demás departamentos la importancia de los datos e información y las acciones que se pueden tomar para obtener una buena calidad de datos. También es importante que los directores de las organizaciones obtengan ésta información, pues finalmente son ellos los que se encargan de delegar tareas, si ellos no asignan de forma correcta las responsabilidades de los distintos departamentos, puede pasar el caso que menciona el autor y dejar todo en manos del departamento de TI. Imaginemos el caso donde la calidad no se resuelve desde la creación, entonces creo que la recomendación sería que hubiese un departamento específico para buscar e identificar los datos que se deben ocupar para después enviarlos a los demás departamentos, también creo que en esta situación tener un departamento dedicado a la limpieza de los datos también es una buena opción, y por sobre todas las cosas que la organización sepa comunicarse entre sí para poder resolver los problemas de manera conjunta y no de manera aislada.

Antes de leer el artículo no tenía idea de lo que significaba tener una buena calidad de datos, como lo expuse antes, muchas veces creemos que teniendo los datos es suficiente y con eso se puede trabajar, eso creía yo hasta que leyendo el artículo me pude dar cuenta que no es así, la planeación de los datos desde su creación hasta el momento en que se usarán es muy importante, pues de esa manera evitas errores y pérdida de tiempo. Además de darme cuenta que la solución es tan sencilla que me sorprende que sigan habiendo casos de mala calidad, concluyendo que es muy importante la comunicación en una organización.

#### 4. Ensayo de Serratos Ramírez Brian

5. **Ensayo de Solis Chavez Arnold** "Data credibility problem" se refiere a: qué tanto se puede confiar en los datos. Existen problemas de calidad de los datos prácticamente en todos lados y muchas veces no se solucionan en la marcha, más bien ven cómo le dan la vuelta y los problemas siguen ahí por siempre, esto es una mala práctica y tanto las consecuencias pueden ser costosas como el resolver esos problemas. Se recomienda asegurarse desde un principio de que los empleados involucrados en manipular los datos entiendan el problema en cuestión. Desde la perspectiva de la calidad de los datos nos importan dos cosas: el momento en que es creada y el momento en que es utilizada, la calidad de los datos se establece al momento de crearla, pero podemos determinar su calidad al momento de utilizar los datos. Para mejorar la calidad de los datos se tiene que trabajar entre quienes crearon los datos y quienes la utilizan. Afortunadamente, solucionar la mayoría de los problemas relacionados a la calidad de los datos no requiere de grandes inversiones o de nuevas tecnologías. Se recomienda concentrarse en mejorar la calidad de los datos que se está creando para no seguir "arrastrando" los problemas que haya en los datos. Vale la pena destacar que distintos sectores de la empresa perciben distinto cómo deben ser utilizados los datos y cuáles son sus fines, muchas veces cualquier problema relacionado a tecnología es resuelto por IT pero en esta situación, la calidad de los datos requiere de la colaboración de distintos perfiles para ser mejorada, en especial de quienes terminan utilizando los datos. En muchos casos, una de las barreras para mejorar la calidad de los datos es que existen managers que no quieren aceptar que la calidad de los datos no es suficientemente buena o no saben cómo solucionar problemas relacionados con mala calidad de datos.

El objetivo que quiso plantear el autor es que nos demos cuenta de que tener mala calidad de los datos puede ser muy costoso, tanto para la empresa en sí como para solucionarlo, y que en muchos casos, la mala calidad se podría haber evitado si se hubieran hecho las cosas bien desde un principio. Se relaciona con la materia en el sentido de que las bases de datos tienen un papel muy importante para la calidad de los datos, como en las bases de datos es donde viven los datos, cualquier error será un problema en especial a largo plazo ya que solucionarlo será más costoso, así que buscamos poner bien las bases para que después no tengamos problemas. La temática central es la calidad de los datos, se tratan temas como la credibilidad de los datos, que es un problema ya que muchas decisiones dependen de datos, si estos datos tienen mala calidad entonces la calidad de las decisiones se verá afectada y esto podría hacer que se tomen decisiones incorrectas, otro tema que trata es que hay que darle su espacio a quienes utilizan los datos para que ellos sean quienes proporcionen información útil para manejar mejor los datos, otro tema que toca es que no puede recaer en IT en su totalidad, el mejorar la calidad de los datos, esto es algo que tiene que reunir distintos perfiles. Creo que es muy útil aprender sobre esto ya que es de esas cosas que prestar atención a los detalles y trabajar con quienes son los usuarios finales de un producto nos puede brindar mucha información útil para mejorarlo, esto aplica a cualquier área. Yo recomendaría que los managers no tomen decisiones sin fundamento y que trabajen con distintos perfiles para buscar cómo tener buena calidad de datos, también sería muy útil tener equipos

especializados para esto y documentar sus experiencias para saber qué hacer en un futuro. Estoy a favor, me parece muy bien lo que propone, debe ser un problema muy complicado en la práctica en empresas grandes así que hay que hacer bien las cosas desde un principio para evitar problemas. Sí había escuchado sobre el término pero nunca lo había investigado, está interesante.

## Referencias

- [1] ELMASRI, R. AND NAVATHE, S. B., *Fundamentals of Database Systems*, Addison-Wesley Publising Company, 6ta edición, 2011.
- [2] SILBERSCHATZ ABRAHAM, *Fundamentos de bases de datos*, McGraw-Hill/Interamericana de de España, 5ta edición, 2006.
- [3] GÓMEZ GARCÍA J. L. AND CONESA I CARALT J., *Introducción al big data*, Universitat Oberta de Catalunya, 1ra edición, 2015.
- [4] RAVENTÓS MORET J., *Sistemas de base de datos*, Universitat Oberta de Catalunya, 1ra edición, 2013.