



**Facultad de  
Ciencias**  
UNAM

**Fundamentos de Bases de Datos  
Grupo 7077**

**TAREA 1. CONCEPTOS BÁSICOS DE LAS  
BASES DE DATOS**

**EQUIPO: mmm xd**

317205776 - Acosta Arazate Rubén.

317042522 - Bernal Márquez Erick.

317180321 - Cristóbal Morales Karen.

317088296 - Garcia Toxqui Demian Oswaldo.

317061521 - Robles Huerta Rosa María.

8 de Marzo de 2022



## 1. Conceptos generales

### a. ¿Qué ventajas y desventajas encuentras al trabajar con una base de datos?

Podemos encontrar algunas ventajas como

- Facilita la integridad de los datos, así como la redundancia.
- Permite almacenar una gran cantidad de datos.
- Protege el acceso a los datos de ciertos usuarios.
- Permite consultar y actualizar de manera rápida los datos.
- Los datos están organizados.
- Permiten acceso rápido.

También existen desventajas como

- Pueden llegar a crecer demasiado
  - Puede ocupar mucha memoria
  - Incrementan los costos para al contratar una persona calificada para manejarla
  - Vulnerable a fallos ya que la información está centralizada
- 

### b. ¿Qué es la independencia de datos? ¿Cuál tipo de independencia de datos es más difícil de lograr? Justifica tu respuesta.

La independencia de datos es la capacidad para modificar la definición del esquema en un nivel del sistema sin tener que modificar el esquema del nivel superior.

Hay dos tipos de independencia de datos:

**Independencia física:** Los cambios ocurren en la organización física de la base de datos y no afectan la visión conceptual o externa de los datos. Altera la organización o procedimientos de almacenamiento relacionados con los datos, en lugar de modificar a los datos en sí.

**Independencia lógica:** Los cambios ocurren en el esquema conceptual sin tener que modificar los esquemas externos ni los programas que hacen uso de los datos.

La independencia lógica es la más difícil de lograr debido a que los programas de aplicación son muy dependientes de la estructura lógica de los datos que tienen acceso. Además, su diseño debe ser desde un inicio detallado si queremos hacer la mínima cantidad de referencias en niveles superiores en caso de realizar un cambio en el esquema conceptual. La independencia física es más fácil porque sólo separa las aplicaciones de las estructuras físicas de almacenamiento.

---

- c. **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?**

Las principales diferencias son:

**Esquema externo:** Se trata del nivel más cercano al usuario (es la visión que tienen de la base de datos). Se extrae información específica del esquema lógico (donde se almacenan, mantienen y acceden a los datos).

**Esquema interno:** Se trata del nivel más cercano a la máquina, describe las estructuras de almacenamiento que guardará a los datos. En el esquema interno se definen los tipos de datos, los índices, la estructura de los ficheros almacenados, etc.

**Esquema conceptual:** Se trata del nivel de descripción central (proporciona un punto intermedio entre la visión de la base de datos del usuario y el almacenamiento físico). En el esquema conceptual se definen todos los datos de la base de datos, atributos, relaciones, etc.

La relación que tiene estos esquemas con la independencia lógica es que los usuarios no se verán afectados debido a que la independencia lógica modifica al esquema conceptual sin tener que alterar los esquemas externos. La relación con la independencia física es que los cambios en la estructura de almacenamiento no implican hacer cambios en los programas de aplicación debido a que la independencia física modifica al esquema interno sin tener que alterar al esquema conceptual.

---

- d. **Investiga qué papel juegan los analistas de bases de datos, diseñadores y desarrolladores de bases de datos en la construcción de un sistema de bases de datos.**

Como su nombre lo indica un *diseñador* diseña la base de datos indicando cómo van a estar relacionados los datos, así mismo da información sobre la base de datos, la prueba y ve de qué manera está mejor estructurado, se podría decir que el diseñador es el *arquitecto* de la obra.

El *analista* al igual que el diseñador prueba la base de datos pero con diferente objetivo, busca fallos y los corrige para asegurar la funcionalidad de la misma. Optimiza y analiza los beneficios de la base de datos.

Por último el *desarrollador* es quién implementa la base de datos asegurando las funciones de esta.

---

- e. **Describe las relaciones que existen entre una base de datos y un Sistema Manejador de Bases de Datos.**

El Sistema Manejador de Bases de Datos es la *herramienta* (software) que nos facilita el proceso de definir, construir, manipular y compartir datos en una Base de Datos. Estos dos están íntimamente relacionados, pues es el Sistema Manejador de Base de Datos el que nos permite comunicarnos con la Base de Datos.

---

- f. **Entrevista a algún usuario de sistemas de bases de datos, ¿qué características de SMBD encuentran más útiles y por qué? ¿qué instalación(es) de SMBD encuentran más/menos complicada y por qué? ¿cuáles perciben estos usuarios que son las ventajas y desventajas de un SMBD?**

Entrevistamos a la profesora de la facultad de ciencias Rosa Victoria Villa Padilla y nos dio las siguientes respuestas:

¿Qué características de SMBD encuentran más útiles y por qué?

Principalmente llaves e índices en tablas para facilitar y acelerar la búsqueda de datos en una tabla y también acelerar búsquedas conjuntas de diferentes tablas

¿Qué instalación(es) de SMBD encuentran más/menos complicada y por qué?

Creo que a la fecha el que ha sido más fácil de instalar ha sido para mí el Microsoft SQL server si el servidor es Microsoft obvio por los wizards es instaladores que tiene. En su defecto creo que MySQL en cloud lo he visto no tan complicado y el cliente de MySQL workbench es bueno. Aquí si nos metemos en no relacionales he visto que la instalación de MongoDB no es tan pesada

¿Cuáles perciben estos usuarios que son las ventajas y desventajas de un SMBD?

Yo percibo como ventaja mundo real que la velocidad y facilidad en consultas mejora los sistemas, he visto de primera mano ver cómo un sistema con queries sucios o tablas sin índices causa conflictos y timeouts en peticiones y solamente con sanitizar queries y poner índices el mismo proceso acelera fuertemente. Ventaja de SMBD es tener información bien distribuida pero localizada en lugares de fácil acceso y además la posibilidad de que se puedan tener réplicas y backups en todo momento

- 
- g. **Supón que deseas crear un sitio de videos similar a YouTube. Considera cada uno de los puntos enumerados en el documento “Purpose of Database Systems”, como desventajas de administrar los datos en un sistema de procesamiento de archivos. Discute la relevancia de cada uno de los puntos indicados con respecto al almacenamiento de datos de los videos: título, el usuario que lo subió, la fecha de carga, las etiquetas, qué usuarios lo vieron, cantidad de “Me gusta”, entre otros.**

Si quisiéramos crear un sitio para subir y ver videos y quisiéramos trabajarlo con un sistema de procesamiento de archivos, podríamos tener los siguientes problemas:

-Redundancia e inconsistencia de datos: Supongamos que tenemos una persona que esta suscrita a 10 canales, si cada canal tiene un archivo donde se almacenan sus suscriptores, en cada archivo aparecerían datos innecesarios para el canal como su numero de teléfono y su nombre completo, entre otros, aparte de que podría haber inconsistencias pues si el usuario cambio de correo los canales a los que se suscribió antes tendrían un correo inútil y en canales que tuvieran al suscriptor tendrían diferentes correos lo que causaría inconsistencia.

-Dificultad de acceso a los datos:Supongamos que un canal quisiera ver cuantos de sus suscriptores son de Latinoamérica tendría que checar lo manualmente y este problema seria aun peor entre mas popular fuera el canal, teniendo que pedir un programa que lo hiciera en caso de ser números exorbitantes.

-Aislamiento de datos: Supongamos que para recibir la información de todo el mundo se dividiera en 7 zonas y donde cada zona tendría un equipo que se encargaría de almacenar datos de su zona, pero que por la diferencia de zonas cada quien guardaría sus datos en diferentes formatos, seria un caos si se quisiera hacer un consenso general o si varias zonas tuvieran que compartir datos entre ellas.

-Problemas de integridad:Un problema de integridad podría darse al momento que un cliente se registre en el sitio o actualice sus datos, por ejemplo que un cliente ponga un numero de 9 dígitos faltandole uno por error o 11 agregando 1 mas por accidente.

-Problemas de atomicidad: Si un sistema no es atómico podría pasar que un usuario A le haga un comentario a un canal B pero después se arrepiente y lo borra, si ocurre un error podría pasar que para A si se haya eliminado pero para B si le haya llegado el mensaje y sea visible.

-Anomalías de acceso concurrente: Un problema que podría pasar sería que si un canal sube un video muy popular y muchas personas se conectan al mismo tiempo el numero de vistas y/o numero de likes podría variar mucho por el gran numero de acciones al mismo tiempo.

-Problemas de seguridad: Algo que podría pasar es que un canal al tenerle en su lista de suscriptores podría ver información personal del usuario que no debería tener acceso como su numero de teléfono y su correo.

- 
- h. **Indica las principales responsabilidades de un Sistema Manejador de Bases de Datos. Para cada responsabilidad, indica qué problemas que surgirían si la responsabilidad no se cumpliera. Justifica en cada caso tu respuesta.**

Las principales responsabilidades son mantener la seguridad, consistencia e integridad de los datos así como evitar la redundancia de los mismos.

Si no se tiene la seguridad suficiente de los datos estos podrían ser manipulados por cualquier persona provocando la pérdida de información importante.

En caso de no haber consistencia podría llegar a ser muy confuso interpretar la base de datos, del mismo modo, al no tener integridad, no podríamos confiarlos de los datos de la empresa

Por último, si permitieramos la redundancia de los datos tendríamos dos interpretaciones para un mismo dato, lo cual es indeseable pues no refleja, de manera precisa, la realidad de la empresa.

- 
- i. **Asumiendo que una base de datos es un lugar donde se almacenan datos de forma sistemática y que la información se obtiene al consultar los datos entonces, un diccionario puede considerarse como una base de datos. Imagina que vas a buscar el significado de la palabra Luminiscencia, indica cómo efectuarías la búsqueda y los problemas que enfrentarías con:**

1. **Un diccionario con palabras desordenadas.**
2. **Un diccionario con palabras ordenadas, pero sin índice.**
3. **Un diccionario con palabras ordenadas y con índice.**

1. Tendría que buscar palabra por palabra la que quiero encontrar, así como puede estar al principio también puede estar al final y tomaría mucho tiempo el encontrarla.

2. Tomaría menos tiempo el encontrarla ya que puedo filtrar la palabra primero por aquellas que empiezan con l, luego por la letra que le sigue u, después m, así sucesivamente, sin duda tomaría menos tiempo pero sigue siendo ineficiente.

3. Lo mismo que el caso anterior solo que guardo el índice de la palabra para que la siguiente vez que lo tenga que buscar solo me preocupe por el índice.

- 
- j. **Investiga por qué surgieron los sistemas NoSQL en la década de 2000 y compara a través de una tabla sus características vs. los sistemas de bases de datos tradicionales.**

En general los sistemas NoSQL surgieron para combatir las limitaciones de los sistemas de bases de datos tradicionales. Dichos sistemas fueron concebidos no por una, sino por distintas empresas

que buscaban soluciones más específicas a sus problemas. Algunas de las carencias que se buscaba subsanar son **escalabilidad**, **rendimiento** y **almacenamiento masivo**.

	Sistemas NoSQL	Sistemas de bases de datos tradicionales
Rendimiento	Al ocupar menos recursos de hardware, ofrecen un mayor rendimiento.	Menor rendimiento comparado al de los sistemas NoSQL.
Consistencia	Utilizadas cuando no hay necesidad de que los datos sean consistentes.	Son muy consistentes (nos garantizan la integridad de los datos).
Almacenamiento	Ideales cuando se desea manejar grandes volúmenes de datos.	Utilizadas cuando la cantidad de datos no son extremadamente grandes.
Escalabilidad	Dado que los sistemas NoSQL están diseñados para funcionar en entornos distribuidos es más sencillo agregar nodos o instancias conforme pasa el tiempo.	Los sistemas de bases de datos tradicionales también son escalables, pero a un costo económico más elevado.

## 2. Lectura

- a. **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).**

En docenas de empresas se desarrollan problemas como presentar errores en sus datos afectando así en todos los departamentos, sectores, los niveles y la información que maneja la empresa. Los empleados suelen corregir de forma rutinaria la gran mayoría de estos errores, pero los costos son enormes tanto por el impacto que podría ocasionar que los datos erróneos se filtren como en el tiempo, ya que se pierde hasta el 50% del tiempo buscando datos, identificando y corrigiendo errores, buscando fuentes de confirmación para los datos en los que no se confía.

La solución a este problema es muy fácil: Asegurarse que los empleados que participan en la de crear los datos comprendan el problema. Se trata de mejorar la comunicación entre los creadores y los usuarios de los datos para que los creadores sepan cómo se van a utilizar los datos.

Hay dos momentos en la vida de un dato: el momento en que se crea y el momento en que se utiliza. La calidad de los datos se fija en el momento de su creación y se juzga hasta su momento de uso. Para la mejora de la calidad de los datos no se necesita de grandes inversiones en nuevas tecnologías, se debe de conseguir que los creadores de datos se asocien con los usuarios para que puedan identificar las causas fundamentales de los errores e idear formas de mejorar la calidad en el futuro.

Cuando una empresa se da cuenta de que la calidad de sus datos está por debajo de la media, su primera reacción suele ser lanzar un esfuerzo masivo para limpiar los datos erróneos existentes, las empresas deberían centrarse en mejorar la forma en que se crean los nuevos datos, identificando y eliminando las causas raíz de los errores.

Muchos problemas de calidad de los datos tiene su origen en los metadatos (datos sobre datos). Los metadatos de alta calidad facilitan la búsqueda de los datos que se necesitan, las comunicaciones

internas, la combinación de información, la toma de mejores decisiones y la extracción de conclusiones adecuadas.

Cuando los problema de calidad se vuelven graves, la respuesta de la organización suele ser encargar al departamento de TI que los solucione, pero suelen tener poco éxito a la hora de liderar programas de calidad de datos porque la calidad de los datos se fija al momento de su creación. Por lo tanto las empresas inteligentes no atribuyen la responsabilidad de la calidad de los datos a TI, sino a los creadores de datos y a sus clientes internos.

**b. Realizar un ensayo donde expresas tus comentarios sobre la lectura,**

**Acosta Arzate Rubén**

La lectura habla principalmente de la importancia de la calidad de los datos, y el autor hace un énfasis en mostrar los problemas que puede traer no darle importancia a esto, con ejemplos de la vida real donde una calidad de datos mala, provocó desde pérdidas de tiempo y recursos hasta pérdidas millonarias de dinero, aparte de estos el autor hace incapie en mostrar alternativas para resolver y prevenir este tipo de problemas, de igual manera poniendo ejemplos de la vida real donde esto se resolvió y/o evito para que la gente ya no cometa estos errores y si ya los tiene pueda saber como arreglarlos, con el objetivo de que ya no ocurran pérdidas de este tipo, siendo muchas veces la mejor solución una buena comunicación entre los creadores de datos y los usuarios de estos para si hay un problema poder localizarlo mas fácil y poder resolverlo y evitar que problemas similares ocurran.

Este tema esta altamente relacionado con la materia de Fundamentos de Base de datos ya que las bases de datos nos sirven para recolectar y almacenar datos relacionados para con estos poder obtener información y si recibimos datos de baja calidad, eso hace que estos datos se vuelvan inútiles ya que se pierde la confiabilidad de estos. Siento que es necesario tomar una postura de conciencia sobre este tema y mas como estudiantes, ya que si queremos dedicarnos a algo relacionado al manejo de datos poder prepararnos mejor para en el futuro no tener que lidiar con estos problemas y sus costosas consecuencias y tener un desarrollo profesional mucho más productivo.

**Bernal Márquez Erick**

El autor intenta persuadirnos de cual es el impacto en la confiabilidad de lo datos exponiendo ejemplos de cómo el más mínimo error puede afectar a otros departamentos dentro de la empresa así como la calidad de la misma.

Argumenta que los datos no son algo que, aunque pareciera lógico, no se debe dejar en manos del departamento de TI sino más bien en aquellos quienes crean los datos y quienes lo consumen mejorando entre ellos la comunicación para saber en todo momento qué hacer con ellos, el porqué y para qué tenemos esos datos.

Como resultados colaterales explica que una mala calidad de los datos puede afectar tanto económicamente, tiempo de empleados (que no tienen nada que ver con los datos), costos, desiciones, etc, pero también cómo podemos mejorarlo que es justo con una mejor comunicación.

Esto nos da una nueva perspectiva de porqué llevamos una materia acerca de los datos, como el autor explica no es que seamos expertos en los mismos, no es una especialización pero sí debemos estar consientes de lo que hacemos en todo momento, saber de donde provienen los datos, a dónde van, que hacemos con ellos.

Personalmente creo que el tener buenas practicas sobre cualquier tema que estemos desarrollando es algo indiscutible, tener buenas practicas tiene sentido en sí mismo para mejorar la calidad de dicha actividad. Sin embargo es interesante saber las consecuencias que podrían desatarse al no



seguir las mismas, es claro que no deseamos el desastre pero sí conocerlo para no hacerlo, algo parecido al típico “Te voy a mostrar como **no** se hace para que no lo hagas”, estoy de acuerdo en tener buenas practicas para evitar futuros desastres, algo que se debería hacer no solo en *datos* o *programación* sino en cualquier área que uno se dedique para evitar futuros resultados no deseados.

### **Cristóbal Morales Karen**

Creo que el objetivo general que plantea el autor es contarnos lo común que llega a ser que las empresas tengan y trabajen con datos erróneos afectando así a todos los departamentos de la empresa dejando costos enormes, ya que por ejemplo se pierde hasta el 50% del tiempo buscando, corrigiendo, validando e identificando datos. El autor menciona que la solución a este problema es bastante sencillo “mejorar la comunicación entre creadores y los usuarios de los datos”.

También una de las temáticas centrales del artículo es determinar la calidad de los datos y cómo obtener datos de buena calidad. Cuando una empresa se da cuenta que sus datos no son de buena calidad regularmente recurren a hacer un esfuerzo masivo por limpiar los datos erróneos existentes, sin embargo deberían enfocarse mejor en la forma en que se crean los nuevos datos. Otro de los puntos importantes del artículos son los metadatos, donde si son de buena calidad ayudarán a la empresa a facilitar la búsqueda de datos necesarios, mejorara la planificación, comunicación interna, toma de mejores decisiones , etc.

Con el artículo aprendí que tener datos erróneos tiene costos enormes, tanto en el tiempo y esfuerzo que se usa para corregir estos datos como en la reputación y confianza que tendrá la organización o empresa. Concuerdo totalmente con el autor que la comunicación entre los creadores y usuarios de datos es una solución para disminuir el problema de tener datos erróneos, considero que las empresas deberían enfocarse en mejorar la forma en que se crean los nuevos datos para así identificar la raíz de los errores y tener en un futuro grandes frutos ya que poco a poco la cantidad de datos erróneos será mínima.

### **Garcia Toxqui Demian Oswaldo**

El texto nos habla acerca de la problematica que pueden llegar a tener las empresas si no se tiene correctud sobre el control de calidad a la hora de fabricar datos para un uso específico, el autor toma una serie de ejemplificaciones relacionadas con la importancia de mantener los datos limpios, habla acerca de que un mal manejo sobre la creacion de los datos puede tener consecuencias costosas a largo plazo.

Un punto importante a destacar es que los datos llegan a fallar debido a que no se hace el esfuerzo para que los creadores de datos tengan una interaccion con los clientes no se trata de corregir un error tras otro intentando parchar un neumático viejo cada que se estropea sino de mejorar la calidad a futuro comprando una llanta nueva para evitar problemas, el autor intenta persuadir al lector mediante problemáticas que han llegado a tener las industrias con el mal manejo de los datos, industrias desde la petrolera hasta la espacial han tenido que lidiar con este problema.

Se pone como ejemplo empresas que han logrado mejorar la eficiencia de la misma a base de esfuerzo y dedicacion pero el factor clave en el que se basa su éxito es que tomaron en cuenta estos problemas y los solucionaron desde raíz.

Este escrito es de bastante importancia puesto que resalta una problematica muy comun dentro de la industria, es muy llamativo el analizar a empresas de gran prestigio como la Nasa y saber que estos problemas pueden ocurrir si no se planea adecuadamente un buen manejo y control de calidad de los datos usados en la empresa.

La relacion que tiene con la materia es que debemos tener en cuenta que los datos son importantes, pero es mas importante el tener datos de utilidad y que mejor que estos datos aporten algo de utilidad a lo que vayamos a implementar, por que si tenemos datos erroneos, puede llegar a ser mas peligroso que la ausencia de estos, este escrito influye en mi percepcion con los proyectos a los que me enfrentare a lo largo de mi vida puesto que ahora tendre en mente que la comunicacion



y la recepción de información son importantes para la creación de un buen proyecto.

### **Robles Huerta Rosa María**

A través de múltiples ejemplos el autor nos presenta la importancia que tiene la forma en la que se crean nuevos datos y las graves consecuencias que ocurren cuando se trabaja con datos de mala calidad. La clave está en el momento en el que se crean los datos, pues es ahí cuando se determina su calidad, por lo que es bien importante asegurarse de entender la problemática que se desea resolver antes de tomar acción.

Lo anterior nos compete puesto que el reto más importante en la aplicación de una base de datos, es el diseño correcto de su estructura. Esto evitará problemas a futuro. Por otra parte, la limpieza de datos no es una buena opción ya que precisará de demasiado tiempo y atención y aún así no garantizará la credibilidad de los datos a largo plazo, por lo que seguramente será necesaria nuevamente otra limpieza de datos.

Actualmente la mala calidad de datos aqueja a numerosas empresas, según estudios, los trabajadores del conocimiento llegan a perder hasta el 50% del tiempo buscando datos, identificando y corrigiendo errores. Esto supone grandes pérdidas económicas, pero también pone en riesgo la fiabilidad de la empresa.

Comparto la opinión del autor, y me parecen muy interesantes los ejemplos que expone ya que honestamente no había pensado en la gravedad de las consecuencias que traen consigo los datos de mala calidad. Es por esto que es muy importante prestar atención al diseño de las bases de datos, para tener acceso a información que sea actualizada y precisa. Y esto se logrará en un primer momento si se mantiene una buena comunicación entre los usuarios de datos y quienes los crean.