

Capítulo 1

Introducción a las bases de datos relacionales

Al finalizar el capítulo, el alumno podrá:

- Analizar la visión actual de las bases de datos.
- Reconocer la arquitectura de una base de datos.
- Reconocer el rol que juegan las bases de datos en el desarrollo de las aplicaciones empresariales.
- Identificar la diferencia entre un gestor de base de datos y una base de datos.

Temas:

1. El enfoque actual
2. Diferencias entre gestor de base de datos y base de datos
3. Ventajas y desventajas de un gestor de base de datos
4. Arquitectura de una base de datos.
5. El entorno de desarrollo

1. El enfoque actual

El enfoque actual
¿Qué es un dato?

- Representación simbólica de un atributo o variable.
- Datos aislados que no contienen información relevante.
- Aplicándole una hipótesis se transforma en información.
- Por sí solo, no constituye información.
- El procesamiento de los datos proporciona la información.

1 - 4 Copyright © Todos los Derechos Reservados - Cibertec Perú S.A.C.

1.1 Dato

El **dato** es una representación simbólica (numérica, alfabética, algorítmica, entre otros), un atributo o característica de una entidad. Los datos describen hechos empíricos, sucesos y entidades.

Los datos aisladamente pueden no contener información humanamente relevante, solo cuando un conjunto de datos se examinan conjuntamente a la luz de un enfoque, hipótesis o teoría se puede apreciar la información contenida en dichos datos.

Los datos pueden consistir en números, estadísticas o proposiciones descriptivas, los datos convenientemente agrupados, estructurados e interpretados se consideran que son la base de la información humanamente relevante que se pueden utilizar en la toma de decisiones, la reducción de la incertidumbre o la realización de cálculos es de ejemplo muy común en el ámbito informático y, en general. Prácticamente en cualquier disciplina científica.

1.2 Información

La **información** es un conjunto de datos con un significado, o sea, que reduce la incertidumbre o que aumenta el conocimiento de algo. En verdad, la información es un mensaje con significado en un determinado contexto

La **información** reduce la incertidumbre en las tres etapas de la toma de decisiones.

- **Investigación:** incrementa el conocimiento sobre la organización.
- **Diseño:** facilita la búsqueda y evaluación de alternativas de solución.
- **Elección:** permite la evaluación de los resultados (conocidos o probables).

Sus características más importantes son:

- Confiable (procedencia).
- Presentación adecuada a las necesidades. Concisa y, de ser necesario, gráfica.
- Exacta (sin errores).
- Oportuna (tiempo de espera).
- Frecuencia (tiempo de procesamiento).

1.3 **Sistemas de información**

Un **sistema de información** se puede definir técnicamente como un conjunto de componentes relacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar la toma de decisiones y el control de una organización.

Hay cuatro actividades en un **sistema de información** que producen la información que estas organizaciones necesitan para tomar decisiones, controlar operaciones, analizar problemas y crear nuevos productos o servicios. Estas actividades son:

- **Entrada:** captura o recolecta datos en bruto, tanto del interior de la organización como de su entorno externo.
- **Almacenamiento de información:** puede hacerse por computadora o archivos físicos para conservar la información.
- **Procesamiento:** convierte esa entrada de datos en una forma más significativa.
- **Salida:** transfiere la información procesada a la gente que la usará o a las actividades para las que se utilizará.

Los **sistemas de información** también requieren de retroalimentación, que es la salida que se devuelve al personal adecuado de la organización para ayudarle a evaluar o corregir la etapa de la entrada.

1.4 Características que cumple un sistema de información

CAPACIDAD	CONTROL	COSTOS	COMUNICACIÓN	COMPETENCIA
Mayor volumen de información.	Mejor manejo de la información.	Saber en qué se invierte.	Integración de las áreas.	Ser más fuertes ante la competencia.
Llegar a más clientes y usuarios.	Más exactitud en los procesos, minimizando los errores.	Saber en qué se gasta.	Mejores conductos para compartir información.	Brindar mejores servicios en el mercado.
Crecer a medida de las necesidades.	Conocer mejor los procesos de información.	Optimizar recursos y reducir pérdidas por falta de control.	Mantener la información siempre actualizada y compartida para todos aquellos que la requieran.	Ser vistos como una empresa moderna y que emplea tecnologías de punta.
Ser más eficientes en los procesos.	Más orden, menos caos.	Reducir tiempos de trabajo y por consiguiente, costos de recursos humanos.	No interesa dónde estén ubicados los clientes o usuarios, la información va hacia ellos.	Ingresar a mercados complejos.

1.5 La ingeniería de la información

Es el conjunto de técnicas automatizadas dentro de una organización compleja, que permitirán proporcionar la información necesaria a quien la requiera, en el momento oportuno. Los aspectos involucrados en la ingeniería de la información son:

- Generadores de aplicaciones
- Centros de información
- Administración de datos
- Modelamiento de datos
- Conocimiento
- Inferencia
- Prototipo
- Planeamiento estratégico
- Diseño automático de procesos
- Usuarios finales, involucrados en el proceso
- Alta gerencia involucrada en la definición de prioridades y requerimientos de información

2. Diferencias entre gestor de base de datos y base de datos



2.1 Introducción

Al estudiar el desarrollo del procesamiento automatizado de datos, en lo que se refiere al aseguramiento técnico, se habla de diferentes generaciones.

Desde el punto de vista del aseguramiento matemático y, en particular, del aseguramiento de programas, algunos autores reconocen 3 generaciones:

- Solución de tareas aisladas
- Integración de tareas aisladas en sistemas particulares
- Integración de sistemas particulares en sistemas automatizados de dirección

Este proceso de integración ocurre paralelamente, aunque no simultáneamente, en dos esferas:

A. Integración de los programas

Ha estado facilitada por el uso de lenguajes de programación, cada vez más sofisticados y de redactores que permiten el acoplamiento de módulos escritos en lenguajes diferentes.

B. Integración de los datos

En la integración de los datos se han producido dos categorías de técnicas para su manipulación:

- **Sistemas orientados a los dispositivos:** programas y archivos son diseñados, además de empleados, de acuerdo con las características físicas de la unidad central y los periféricos. Cada programa está altamente interconectado con sus archivos, por lo que la integración de datos de diferentes sistemas es prácticamente imposible.
- **Sistemas orientados a los archivos:** la lógica de los programas depende de las técnicas de organización de los archivos (secuencial, directo, etc.). Cada usuario organiza su archivo, de acuerdo con sus necesidades y además, las relaciones entre los elementos se establecen a través de los programas de aplicación.

Esta forma de trabajo implica redundancia de datos, así como, mayor gasto de memoria y complicaciones en las operaciones de actualización. Esto aumenta el tiempo de tratamiento y atenta contra la integridad de la información. Cuando se habla de integridad se hace referencia a que, en todo momento, los datos almacenados estén en correspondencia con la realidad.

Además, en la vida real se establecen relaciones entre los objetos que son muy difíciles de representar u obtener, a partir de sistemas tradicionales de archivos.

Ejemplo:

Si se tiene información sobre trabajadores y estudiantes de una facultad, las aplicaciones requeridas van a definir la manera de organizar y estructurar los archivos, si se desea obtener datos como: promedio de las calificaciones de cada estudiante, listado de estudiantes por grupo, categoría científica y docente de cada profesor, salario de cada profesor, etc.

Resulta adecuado establecer dos archivos: uno de profesores y uno de estudiantes.

Pero ¿qué ocurre si se quiere establecer vínculos entre los profesores y estudiantes?, por ejemplo:

- Los estudiantes de un profesor.
- Los profesores de un estudiante.

Se estructuraría un archivo de profesores y estudiantes que resolvería algunas demandas, pero sería ineficiente para otras. Entonces, ¿es posible representar de manera eficiente, utilizando los medios de cómputo, los fenómenos o procesos de la realidad objetiva, aunque sea, de forma esquemática, pero en la que se establezcan determinados vínculos entre los elementos u objetos que forman parte de esos procesos o fenómenos?

Ello es posible, a través de la utilización de bases de datos (BD) y de los Sistemas de Gestión de Bases de datos (SGBD) que dirigen su manipulación.

2.2 Definición de base de datos

Existen muchas definiciones de Base de datos; la mostrada a continuación es una de las más interesantes.

Se define una **base de datos** como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.

También, se puede definir a una **base de datos** como una colección de información organizada de forma que un programa pueda seleccionar rápidamente los fragmentos de datos que necesite.

2.3 Componentes y objetivos de una base de datos

Existen muchas formas de organizar las bases de datos, pero hay un conjunto de objetivos generales que deben ser cumplidos previamente, de modo que faciliten el proceso de diseño de aplicaciones y los tratamientos sean más eficientes y rápidos, dando la mayor flexibilidad posible a los usuarios.

Los objetivos fundamentales de la base de datos son las que se describen a continuación.

A. Independencia de los datos y los programas de aplicación

En los archivos tradicionales, la lógica de la aplicación contempla la organización de los archivos y el método de acceso. Por ejemplo, si por razones de eficiencia se utiliza un archivo secuencial indexado, el programa de aplicación deberá considerar la existencia de los índices y la secuencia del archivo. Entonces, es imposible modificar la estructura de almacenamiento o la estrategia de acceso, sin afectar el programa de aplicación, pero si afecta son aquellas partes de dicho programa, que tratan los archivos.

En una BD sería indeseable la existencia de aplicaciones y datos dependientes entre sí, por dos razones:

- La independencia de los datos se define como la inmunidad de las aplicaciones a los cambios en la estructura de almacenamiento.
- En la estrategia de acceso a los datos, no debe interesar la herramienta de desarrollo de la aplicación, lo cual constituye el objetivo fundamental de las BD.

B. Minimización de la redundancia

Uno de los objetivos de las BD es minimizar la redundancia de los datos, no eliminarla, pues, aunque se definen las BD como no redundantes, en realidad existe redundancia en un grado no significativo para disminuir el tiempo de acceso a los datos o para simplificar el método de direccionamiento. Lo que se trata de lograr es la eliminación de la redundancia superflua.

C. Integración y sincronización de las bases de datos

La integración consiste en garantizar una respuesta a los requerimientos de diferentes aspectos, de los mismos datos, pero por diferentes usuarios, de tal forma que, aunque el sistema almacene la información con cierta estructura y cierto tipo de representación, garantice la entrega al programa de aplicación los datos que solicita y en la forma en que los solicita.

Además, está vinculada a la sincronización, que consiste en la necesidad de garantizar el acceso múltiple y simultáneo a la BD, de modo que los datos puedan ser compartidos por diferentes usuarios a la vez.

D. Integridad de los datos

Consiste en garantizar la no contradicción entre los datos almacenados, de modo que en cualquier momento del tiempo, los datos almacenados sean correctos, es decir, que no se detecte inconsistencia entre los datos. Asimismo,

se encuentra relacionada con la minimización de la redundancia, ya que es más fácil garantizar la integridad, si se elimina la redundancia.

E. Seguridad y recuperación

Seguridad (también llamada protección): garantiza el acceso autorizado a los datos, de forma que se interrumpa cualquier intento de acceso no autorizado, ya sea por error del usuario o por mala intención.

Recuperación: permite que el sistema de bases de datos disponga de métodos que garanticen la restauración de las bases de datos al producirse alguna falla técnica, interrupción de la energía eléctrica, etc.

F. Facilidad de manipulación de la información

Los usuarios de una BD pueden acceder a ella con solicitudes para resolver muchos problemas diferentes. El **SGBD** debe contar con la capacidad de una búsqueda rápida por diferentes criterios, así como, permitir que los usuarios planteen sus demandas de una forma simple, aislándolo de las complejidades del tratamiento de los archivos y del direccionamiento de los datos. Además, los SGBD actuales brindan sub-lenguajes de datos, de alto nivel con diferentes grados de facilidad para el usuario no programador.

G. Control centralizado

Uno de los objetivos más importantes de los SGBD es garantizar el control centralizado de la información. Asimismo, permite controlar, de manera sistemática y única, los datos que se almacenan en la BD, así como, el acceso a ella.

Esto implica que debe existir una persona o conjunto de personas, que tenga la responsabilidad de los datos operacionales: el administrador de la BD, que puede considerarse parte integrante del sistema.

2.4 El software de gestión de base de datos (SGDB)

El software que permite la utilización y/o actualización de los datos almacenados en alguna(s) base(s) de datos, por uno o varios usuarios, desde diferentes puntos de vista y en simultáneo, se denomina **Software de gestión de bases de datos** (SGBD).



*Es importante diferenciar los términos **BD** y **SGBD**.*

El objetivo fundamental de un SGBD consiste en suministrar al usuario, las herramientas que le permitan manipular los datos, en términos abstractos, es decir, de forma que no le sea necesario conocer el modo de almacenamiento de los datos en la computadora, ni el método de acceso empleado.


Los programas de aplicación operan sobre los datos almacenados en la base, utilizando las facilidades que brindan los SGBD, los que, en la mayoría de los casos, poseen lenguajes especiales de manipulación de la información que facilitan el trabajo de los usuarios.


*Los más conocidos SGBD relacionales son **SQL Server**, **Oracle**, **Informix**, **MySQL**, **DB2**, **MS Access**.*

3. Ventajas y desventajas de las bases de datos

Ventajas y desventajas de las bases de datos
Principales ventajas

- Mayor eficiencia en el procesamiento de la información.
- Permite mantener normalizados los datos.
- Mayor disponibilidad y rapidez.
- Mayor consistencia en el tiempo.
- Más facilidad para compartir los datos.
- Mayor seguridad de acceso y control.



1 - 19
Copyright © Todos los Derechos Reservados - Cibertec Perú S.A.C.


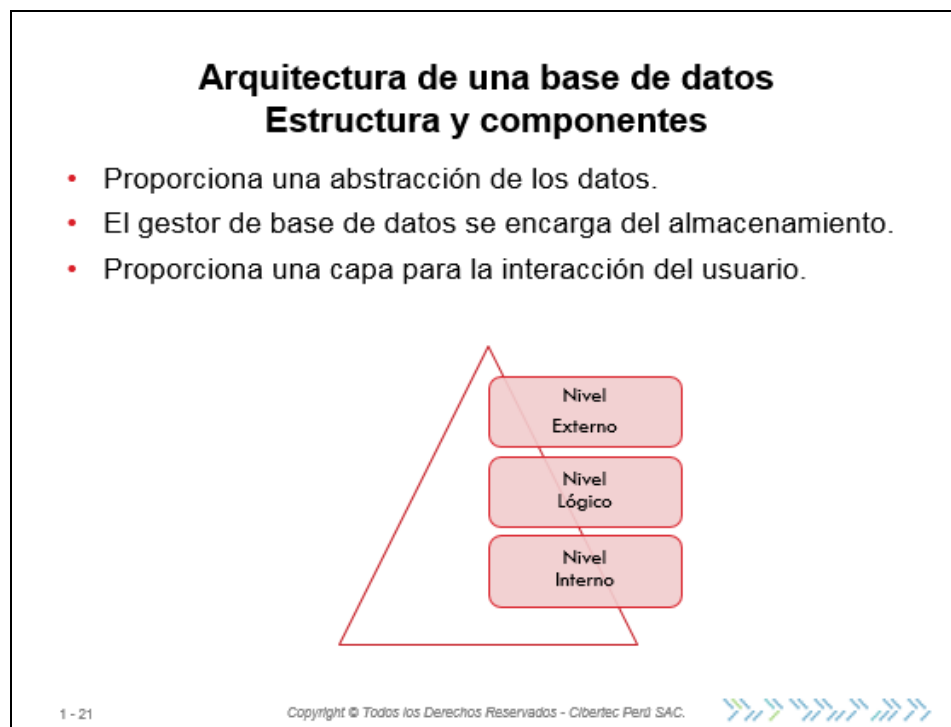
3.1 Ventajas

Ventajas de las bases de datos	
Referidas	Ventajas
Datos	<ul style="list-style-type: none"> • Independencia de estos, respecto a los tratamientos y viceversa. • Mejor disponibilidad de los mismos. • Mayor eficiencia en el recojo, codificación y entrada.
Resultados	<ul style="list-style-type: none"> • Mayor coherencia. • Mayor valor informativo. • Mejor y más normalizada documentación de la información.
Usuarios	<ul style="list-style-type: none"> • Acceso más rápido y sencillo de los usuarios finales. • Más facilidades para compartir los datos por el conjunto de los usuarios. • Mayor flexibilidad para atender a demandas cambiantes.

3.2 Desventajas

Desventajas de las bases de datos	
Referidas	Desventajas
Implantación	<ul style="list-style-type: none">• Costosa en equipos (lógico y físico).• Ausencia de estándares.• Larga y difícil puesta en marcha.• Rentabilidad a mediano plazo
Usuarios	<ul style="list-style-type: none">• Personal especializado.• Desfase entre teoría y práctica.

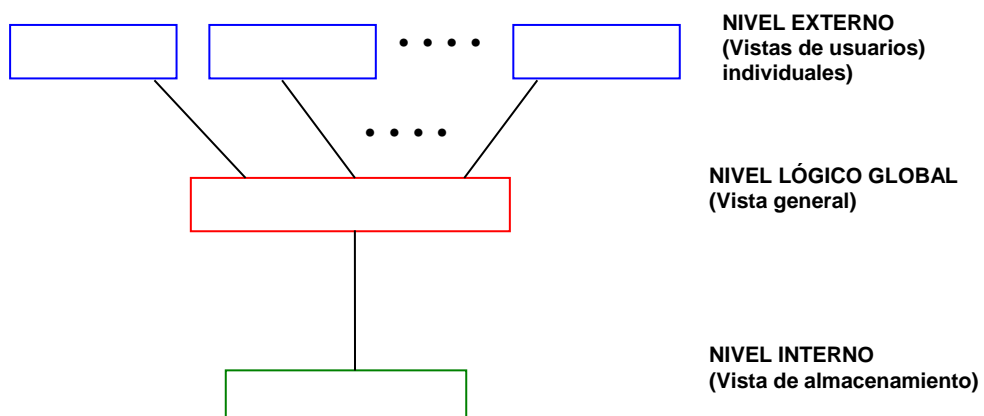
4. Arquitectura de una base de datos



A pesar de no poder asegurar que cualquier SBD corresponda exactamente con ella: esta arquitectura se corresponde suficientemente bien con un gran número de sistemas. Además, está de acuerdo con la arquitectura propuesta por el grupo ANSI/SPARC.

La arquitectura se divide en tres niveles generales:

- El **nivel interno** es el más cercano al almacenamiento físico, es decir, es el que está relacionado con la forma en que los datos se encuentran realmente almacenados.
- El **nivel externo** es el más cercano a los usuarios, es decir, es el que está relacionado con la forma en que los datos son vistos por cada usuario individualmente.
- El **nivel lógico global** es un nivel intermedio entre los dos anteriores.



Existirán varias "vistas externas" diferentes y cada una de ellas, representa alguna porción de la BD total; asimismo, existirá, únicamente, una "vista general", consistente en una representación también abstracta de la BD en su totalidad. De igual forma, existirá una única "vista interna", que representa a la BD completa, tal y como está realmente almacenada.

A continuación, se detalla cada uno de los niveles de la arquitectura y la forma en que interactúan.

A. Nivel externo

Es el nivel del usuario individual, en el que este usuario puede ser bien un programador de aplicación o un usuario final con cualquier grado de sofisticación. Cada usuario tiene un lenguaje a su disposición.

- Para el programador ese lenguaje puede ser bien un lenguaje de programación convencional, tal como Java o Visual Basic. NET.
- Para el usuario final, el lenguaje puede ser de consulta (interrogaciones, query) o de propósito especial, quizás basado en sistemas de menús y ventanas, y es construido para satisfacer los requerimientos de un usuario.

Todo lenguaje debe incluir un sub-lenguaje de datos, es decir, un subconjunto del lenguaje que trata específicamente, con los objetos de la base de datos y sus operaciones. Se dice que el sub-lenguaje de datos (DSL) está embebido dentro del correspondiente lenguaje huésped, el cual se encarga de asegurar otras facilidades ajenas a la base de datos, tales como variables locales, operaciones de cálculo, lógica if-then-else, etc.

De esta manera, un sistema dado puede soportar múltiples lenguajes huésped y múltiples sub-lenguajes de datos.

En principio, cualquier sub-lenguaje de datos es realmente una combinación de, al menos, dos lenguajes subordinados: un lenguaje de definición de datos (**DDL**), el cual garantiza la definición o descripción de los objetos de la base de datos, y un lenguaje de manipulación de datos (**DML**), que garantiza la manipulación o procesamiento de esos objetos.

Se sabe que un usuario individual estará interesado solo en cierta porción de la BD completa. Aún más, la vista de esa porción será abstracta cuando se compara con la forma en que los datos están físicamente almacenados. El término definido por el comité **ANSI/SPARC** para una vista de un usuario, es vista externa, la cual es el contenido de la BD tal y como es vista por un usuario en particular. Es decir, para ese usuario, la vista externa es la BD.

En términos generales, una vista externa consiste en múltiples ocurrencias de múltiples tipos de artículos externos. Un artículo externo no es necesariamente igual a un artículo almacenado.

El sub-lenguaje de datos del usuario se define en términos de artículos externos; por ejemplo, una operación del DML que sea recuperar artículos, recuperará una ocurrencia de artículos externos y no una ocurrencia de artículos almacenados.

Cada vista externa se define mediante un esquema externo, consistente en definiciones de cada uno de los diferentes tipos de artículos externos en esa vista. El esquema externo se escribe usando la porción del DDL del sub lenguaje de datos del usuario; además, tiene que existir una definición de la correspondencia entre el esquema externo y el esquema lógico-global.

B. Nivel lógico-global

La vista lógica es una representación del contenido informativo total de la BD. Es una forma abstracta en comparación con la forma en que los datos están almacenados físicamente. Esta vista puede ser muy diferente de la forma en la que los datos son vistos por un usuario en particular.

La vista lógica pretende ser una vista de los datos tal como son, en lugar de cómo los usuarios están forzados a verlos por las restricciones de un lenguaje particular o de un determinado hardware que utilicen. Esta vista lógica consiste en múltiples ocurrencias de múltiples tipos de artículos lógicos. Por ejemplo, puede ser una colección de ocurrencias de artículos de departamentos, más una colección de ocurrencia de artículos de empleados, etc. Un artículo lógico no es necesariamente igual a un artículo externo ni a un artículo almacenado.

Asimismo, la vista lógica se define mediante el esquema lógico que incluye las definiciones de cada uno de los diferentes tipos de artículos lógicos. El esquema lógico se describe usando otro lenguaje de definición de datos: el DDL lógico. Si se desea lograr la independencia de los datos, entonces las definiciones del DDL lógico no deben comprender ninguna consideración sobre la estructura de almacenamiento ni la estrategia de acceso. Ellas tienen que ser definiciones referentes al contenido informativo.

Si el esquema lógico logra la independencia de los datos, entonces los esquemas externos que se definen sobre el esquema lógico también lograrán la independencia de los datos. Entonces la vista lógica es una vista del contenido total de la BD y el esquema lógico es una definición de esa vista. Sin embargo, el esquema lógico no es simplemente un conjunto de definiciones como las que se encuentran, por ejemplo, en un programa Pascal. Las definiciones en el esquema lógico deben incluir una gran cantidad de aspectos adicionales, tales como los chequeos de protección y los chequeos de integridad.

C. Nivel interno

La vista interna es una representación de bajo nivel de la BD completa, que consiste en múltiples ocurrencias de múltiples tipos de artículos internos.

Artículo interno es el término definido por ANSI/SPARC para la construcción que hasta ahora se ha denominado: artículo almacenado. Entonces, la vista interna está a un paso del nivel físico, ya que ella no opera en términos de artículos físicos (páginas o bloques) ni con consideraciones específicas de los equipos, tales como tamaños de sectores o pistas. Básicamente, la vista interna asume un espacio de dirección lineal infinita. Los detalles de cómo se hace corresponder ese espacio con el almacenamiento físico son muy específicos de un sistema y se han omitido deliberadamente de la arquitectura.

La vista interna se describe mediante el esquema interno, el cual no solo define los diferentes tipos de artículos almacenados, sino que también especifica los índices que existen, la representación de los campos almacenados, la secuencia física en que están los artículos almacenados, etc. El esquema interno se describe usando el DDL interno.

D. Correspondencias entre los niveles de la arquitectura

En el esquema presentado de la arquitectura de un SBD, se observan los niveles de correspondencias: una entre los niveles externo y lógico-global, y otra, entre los niveles lógico-global e interno.

La **correspondencia lógica/interna** especifica la forma en que los artículos y campos lógicos se representan en el nivel interno. Si se cambia la estructura de la vista interna, es decir, si se hace un cambio en el esquema interno, entonces la correspondencia lógica/interna tiene que cambiar en consecuencia, de modo que el esquema lógico permanezca invariable. En otras palabras, los efectos de estos cambios deben ser aislados por debajo del nivel lógico para que se mantenga la independencia de los datos.

Existe también una **correspondencia externa/lógica** entre cada vista externa particular y la vista lógica. Las diferencias que pueden existir entre estos dos niveles son similares a las que pueden existir entre la vista lógica y la interna. Por ejemplo, los campos pueden tener diferentes tipos de datos, se pueden cambiar los nombres de artículos y campos, múltiples campos lógicos pueden ser combinados en un único campo externo, etc. Puede existir al mismo tiempo, cualquier cantidad de vistas externas, cualquier cantidad de usuarios puede compartir una vista externa dada, las diferentes vistas externas se pueden solapar. Algunos sistemas permiten la definición de una vista externa a partir de otra; esta característica es útil cuando varias vistas externas están estrechamente relacionadas entre sí.

E. El administrador de la base de datos

Un **administrador de base de datos** tiene la responsabilidad de mantener y operar las bases de datos que conforman el sistema de información.

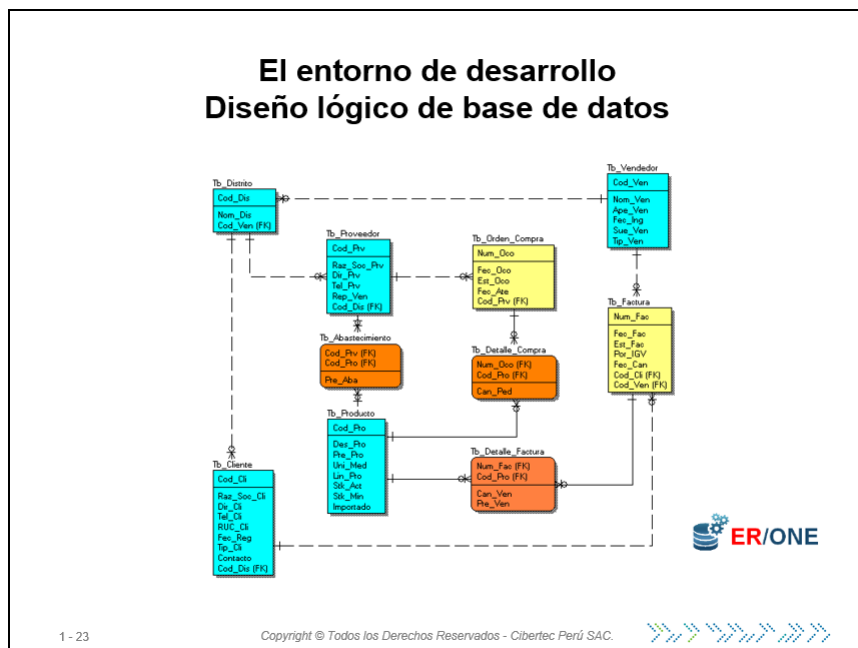
Debido a la importancia de los datos que están a su cargo, el **administrador de base de datos** debe ser experto en tecnologías de la información, teniendo particular conocimiento de sistemas de administración de base de datos (DBMS) y el lenguaje de consulta SQL. También, debe tener conocimiento de varios tipos de lenguaje de programación para poder automatizar ciertas tareas.

Una de las tareas es la de asegurar la integridad del sistema de información de la compañía. Además, es necesario que posea un buen entendimiento de DBMS para optimizar las consultas, ajustar la configuración de DBMS o para sincronizar en forma precisa las herramientas de control de acceso a las bases de datos.

Es posible que el administrador de base de datos tenga que brindar asistencia técnica a usuarios de las aplicaciones cliente o equipos de desarrollo para solucionar problemas, dar consejos o ayudar a resolver consultas complicadas.

Al trabajar con el jefe de seguridad, el administrador de base de datos debe crear copias de seguridad, planes y procedimientos de restauración para preservar los datos de los cuales es responsable.

5. El entorno de desarrollo



Un área usuaria que controle o emplee su base de datos, dispone de una excelente herramienta de información, consulta y control, y puede acudir cada vez que se precise, obteniendo datos, planificando campañas, analizando segmentos de consumidores, etc. En otras palabras, se trata de acumular toda la información necesaria, organizándola y clasificándola para que, una vez controlada, sus poseedores planifiquen sus estrategias empresariales.

Conviene recordar que el control de una base de datos exige una constante puesta al día. Las grandes empresas, sobre todo, pagan muchísimo dinero por disponer de una base de datos controlada y actualizada. Por ello, se pueden analizar el comportamiento de todos sus clientes y atacar, directamente, una serie de nichos de mercado, prefijados con agresivas estrategias de marketing directo. Así, no es extraño que todos los sistemas de información que se implementan en el sector empresarial se estructuren en relación con este poderoso elemento común.

Una base de datos correctamente actualizada permite que la organización controle, al instante, la existencia de stocks, saldos, deudas de clientes, etc., así como, una serie de temas fundamentales para el desarrollo de todo lo que es gestión de su red de usuarios.