

CAPITULO 9

Autor: Ing. Emilio Lorenzon

EL DATO Y LA ORGANIZACION

1. Introducción

Las organizaciones como sistema dinámico, durante su evolución y funcionamiento generan datos que de alguna manera representan sus estadios en los distintos instantes de tiempo de su ciclo de vida. Es decir que los datos son un sinónimo de conocimiento y por lo tanto deben tratarse como un patrimonio, un activo como pueden ser los automóviles de la empresa, un recurso, que obliga a las organizaciones a establecer, respecto de los mismos, un sentido de propiedad y una administración que incluya su cuidado. (**Fig. 9.1**).

Cada uno de los estados de la Base de Datos, almacena todos los datos de la realidad de la Organización en un instante de tiempo determinado. Esta definición de estado garantiza la consistencia de la Base de Datos. La reserva histórica de los estados de la base de datos, conforman la base del conocimiento.

Los datos son valores acerca de hechos o eventos. La cantidad de dinero o cantidad de artículos en una transacción son "datos". La base de datos es esencial para recoger y almacenar los datos y de esta manera poder organizar informes. Estos informes están constituidos por datos formateados, y por el valor añadido de ese formateo pueden ser considerados "información". Los datos son importantes para medir los costos, velocidades, cantidades, totales, capacidades, etc, de la organización. Sin la interpretación del observador, disponer de más datos no hacen más que tener menos datos. Las decisiones requieren datos que han sido convertidos en información útil, es decir en un "consumible".

La información se constituye de datos recopilados, organizados e interpretados. Si los datos son letras, organizar las letras en palabras inteligibles, es información. Si los datos representan clientes, organizar la lista en un formato utilizable, puede ser la información necesitada.

La información se desplaza en las organizaciones en forma de papel, voz y documento digital. Los ejemplos son el correo en papel, las páginas html, revistas, correo electrónico, correo vocal

y otras transmisiones por computadora. La información puede ser generada por personas o por máquinas. El almacenamiento y transmisión de la información dependen mucho de la tecnología.

Por otro lado, el concepto de conocimiento individual, es el resultado de un proceso de aprendizaje y cambio en el comportamiento que ocurre en una persona después de interiorizar la información. El conocimiento es una integración de valores, experiencias e información como parte de un modelo mental.

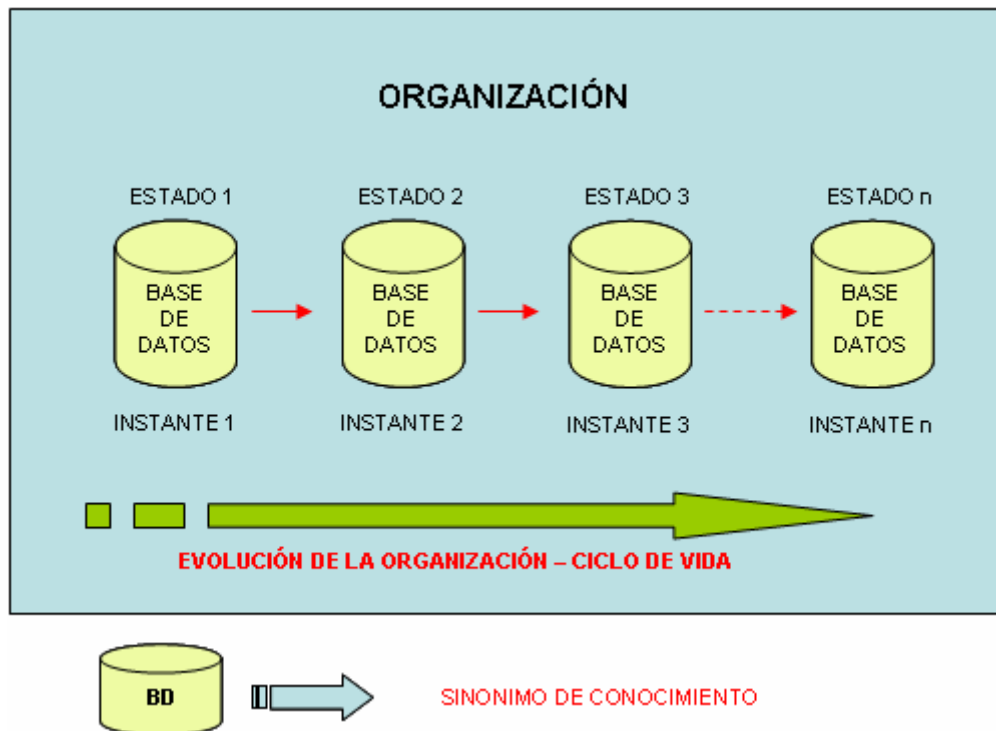


Fig . 9.1 - La Organización como sistema dinámico

2. Datos, información, conocimiento

¿En qué se diferencia el conocimiento de los datos y de la información? En una conversación informal, los tres términos suelen utilizarse indistintamente y esto puede llevar a una interpretación libre del concepto de conocimiento. Quizás la forma más sencilla de diferenciar los términos sea pensar que los datos están localizados en el mundo y el conocimiento está localizado en agentes de cualquier tipo (personas, empresas, máquinas...), mientras que la información adopta un papel mediador entre ambos.

Los conceptos que se muestran a continuación se basan en las definiciones de Davenport y Prusak (1999).



9.2 - Pirámide de datos, información y conocimiento

2.1 Datos

Introducción

Los datos son la mínima unidad semántica, y se corresponden con elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones. También se pueden ver como un conjunto discreto de valores, que no dicen nada sobre el por qué de las cosas y no son orientativos para la acción.

Un número telefónico o un nombre de una persona, por ejemplo, son datos que, sin un propósito, una utilidad o un contexto no sirven como base para apoyar la toma de una decisión. Los datos pueden ser una colección de hechos almacenados en algún lugar físico como un papel, un dispositivo electrónico (CD, DVD, disco duro...), o la mente de una persona. En este sentido las tecnologías de la información han aportado mucho a recopilación de datos.

Como cabe suponer, los datos pueden provenir de fuentes externas o internas a la organización, pudiendo ser de carácter objetivo o subjetivo, o de tipo cualitativo o cuantitativo, etc.

Definición

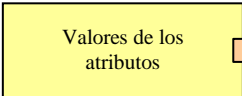
El término "Dato", puede ser definido como un elemento, efectivo o supuesto, que representa un término de información, y puede ser utilizado como base para el procesamiento.

El hecho que un "Dato" se considere como un "elemento", implica que puede ser identificado y descrito. Esto no significa que necesariamente debe ser un elemento físico y medible, sino también abstracto o simplemente de tipo lógico.

Que representa un "término de información", significa que esta relacionado con las propiedades de los objetos (**Fig. 9.3**). Veamos cuáles son los términos de información necesarios para definir un objeto. Se necesitan cuatro:

- Identidad del objeto
- Clase de propiedad que deseamos especificar para un objeto. (o: un atributo del objeto)
- La especificación de esa propiedad para ese objeto. Este es el "Valor" asociado al atributo, denominado valor de la propiedad.
- El punto en el tiempo.

Puede observarse que los primeros tres términos de información están asociados con conjuntos. Por ejemplo, para suministrar información acerca de la identidad del objeto hombre, es necesario señalar un elemento en una lista de individuos. Asimismo, para indicar de qué clase de propiedad nos proponemos hablar, seleccionamos una clase de una lista de un conjunto (dominio) de propiedades. Por ejemplo la profesión. Finalmente para indicar el valor de la clase de propiedad tenemos nuevamente que identificar en una lista de valores. Por ejemplo ingeniero. Por lo tanto la información acerca de un objeto está dada por un conjunto de valores, que corresponden al dato para cada uno de los términos seleccionados.



Identidad	Atributo 1	Atributo 2	Atributo 3	Atributo 4
Apellido	Profesión	Edad	Altura	Nacionalidad
Pérez	Ingeniero	35	1.80	Argentino

FIG. 9.3 - Representación del objeto hombre

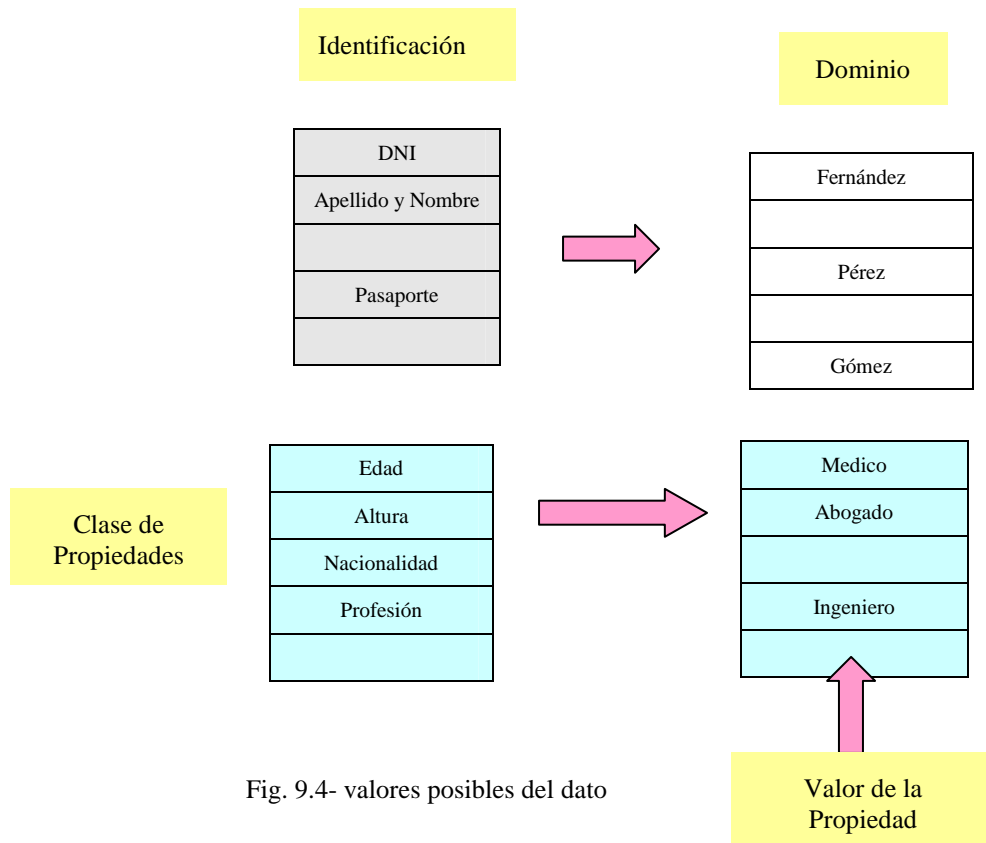


Fig. 9.4- valores posibles del dato

En conclusión, aquí empleamos la palabra datos para referirnos a los medios de representar información. Más específicamente, consideraremos que Datos es el plural de Dato y que el "Dato" representa una parte del objeto, el valor del término.

Consistencia de la Base de Datos

En el caso del almacenamiento de los datos en una base de datos relacional, la representación podría ser la siguiente:

- El objeto puede representarse como una fila de una tabla.
- Las clases de propiedades como las columnas de la tabla.
- El dato (Valor de un término) puede representarse como el valor de la intersección entre una fila y columna.

Podemos (Dentro de la base de datos que se actualiza por un sistema determinado) considerar a cada término de información, como un valor de un conjunto posible (dominio) en un intervalo de tiempo. Por consiguiente, se esta "seleccionando un punto en el tiempo" para la actualización de los atributos de cada objeto, y de esta manera poder controlar la consistencia de nuestra base de datos. Si bien se considera al tiempo como una variable continua, establecemos intervalos de partición apropiados a los requerimientos del sistema en análisis. (**Fig. 9.5**)

Transportando lo expuesto a una base de datos corporativa, todos los sistemas tendrían que tener los intervalos de actualización compatibles, a los efectos de mantener la consistencia de la base.

Finalmente, el dato es un elemento "Base para el procesamiento", porque representa un valor cuantitativo, usado para llegar al conocimiento de un "echo".

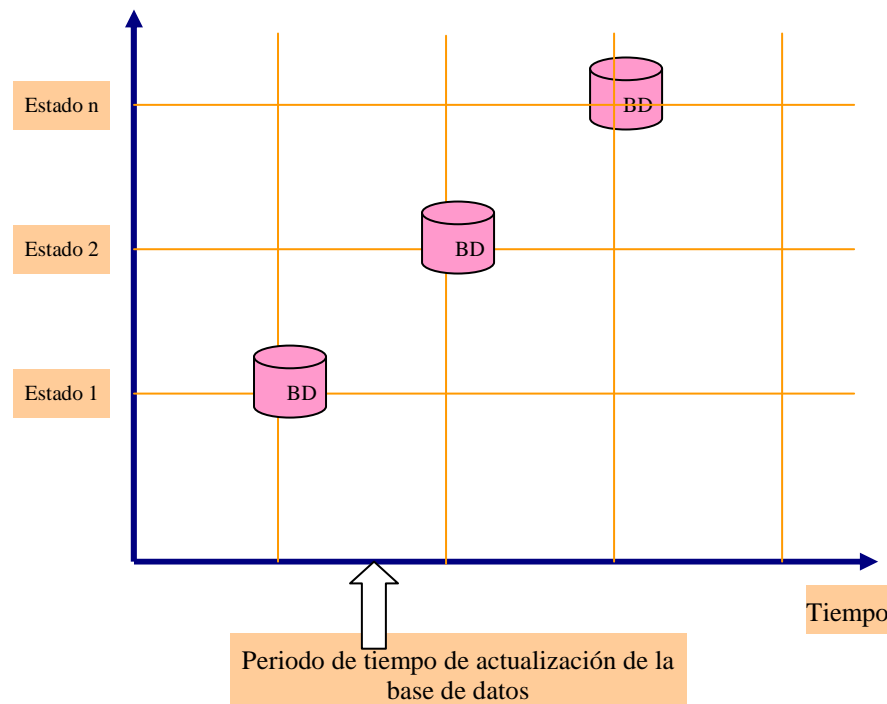


Fig. 9.5- Estados de una base de datos corporativa en el tiempo

2.2. Información

La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre. Los datos se pueden transformar en información añadiéndoles valor:

- Contextualizando: se sabe en qué contexto y para qué propósito se generaron.
- Categorizando: se conocen las unidades de medida que ayudan a interpretarlos.
- Calculando: los datos pueden haber sido procesados matemática o estadísticamente.
- Corrigiendo: se han eliminado errores e inconsistencias de los datos.
- Condensando: los datos se han podido resumir de forma más concisa (agregación).

Por tanto, la información es la *comunicación de conocimientos* o inteligencia, y es capaz de cambiar la forma en que el receptor percibe algo, impactando sobre sus juicios de valor y sus comportamientos.

Información = Datos + Contexto (añadir valor) + Utilidad (disminuir la incertidumbre)

Definimos como "Información" al aumento del conocimiento proporcionado por los datos, que puede usarse para mejorar o posibilitar una decisión. Por lo tanto, que un dato informe de algo o no lo haga, depende del receptor y de sus exclusivos y personales conocimientos e interpretación.

Es importante tener en cuenta, que los sectores de la organización poseen una visión parcial de la base de datos y por lo tanto el resto de la misma no la considera como información. Esta situación es solucionada por la existencia del "Área informática" que es la responsable de tener la visión integradora de la base de datos corporativa.

2.3. Conocimiento

Si bien este tema lo trataremos específicamente y en profundidad en el Capítulo 12 de este libro, adelantaremos algunos conceptos básicos para entender el conocimiento y su relación con los datos y la información. El conocimiento es una mezcla de experiencia, valores, información y know-how que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción. Se origina y aplica en la mente de los conocedores. En las organizaciones con frecuencia no sólo se encuentra dentro de documentos o almacenes de datos, sino que también está en rutinas organizativas, procesos, prácticas, y normas.

El conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierta en conocimiento es necesario realizar acciones como:

- Comparación con otros elementos.
- Predicción de consecuencias.
- Búsqueda de conexiones.
- Conversación con otros portadores de conocimiento

En cuanto al "Conocimiento", podemos definirlo en líneas generales como información retenida. Esto significa que, para obtener información, es necesario hacer algo con los datos, pero también significa que existe al menos dos tipos de información: La que se retiene, que luego se transformara en conocimiento, y la que no se retiene.

Con esta definición primaria podemos decir que, para el caso de las organizaciones, parte de su conocimiento no es otra cosa que la información almacenada en su propia Base de Datos. Si bien esto no es todo el conocimiento de la Organización, podemos decir que la base de datos corporativa es una "Base del conocimiento de la Organización".

3. El ciclo de vida de los datos.

Desde el punto de vista informático los datos existen independientemente de la tecnología que los gestiona, si esta cambia, a lo sumo será necesario un proceso migratorio para su nueva representación y su posterior procesamiento.

Es importante aclarar también como es la relación del ciclo de vida de los sistemas informáticos respecto del ciclo de vida de los datos. En primer lugar estamos aceptando que cada uno tiene su propio ciclo de vida, esto es así desde el momento que si bien los datos constituyen la materia prima de la información, su ciclo de vida puede ser superior al del propio sistema que los originó, y en algunos casos logran trascender a la organización misma.

Podemos establecer los siguientes ciclos de vida dentro de una organización, los cuales son totalmente asincrónicos entre si. Estos son:

- Ciclo de vida de la Organización.
- Ciclo de vida de los sectores de la Organización.
- Ciclo de vida de los sistemas de la Organización.

- Ciclo de vida de los datos de la Organización.

El área informática tiene la indelegable responsabilidad de administrar este importante recurso de la organización

El ciclo de vida de los datos es paralelo al ciclo de vida de su entorno. Si este entorno fuera un sistema, los datos nacerían al nacer el sistema y morirían, como ya lo dijimos, después que el mismo sistema que los origina. En cierto sentido, durante su vida dentro de una organización, los datos "evolucionan", tienen un crecimiento.

Puede explicarse el concepto de crecimiento de los datos presentando una idea que llamaremos "Propiedad de los datos". A los fines de este análisis convengamos en que solo existen tres tipos de propietarios de los datos en una organización, a saber:

- Los datos privados.
- Los datos compartidos (sectoriales).
- Los datos comunes (de la organización).

Los datos privados lo utiliza una persona. Si esta muere, sus datos mueren con ella. Los datos compartidos lo utilizan varias personas. Si una muere, los datos siguen siendo usados por las demás. Los datos comunes son utilizados por muchas personas. A medida que los datos tienen mas penetración dentro de la organización, son necesarios controles de calidad más rigurosos.

En un comienzo, todos los datos eran privados. Los mejores de esos datos llegaron a ser compartidos. Los mejores de los mejores se transformaron en comunes. Es la supervivencia de los mejores, pues resulta muy costoso mantener datos irrelevantes.

La demanda de datos comunes proviene del crecimiento y de los cambios en la organización. Estos pueden relacionarse directamente con el ciclo de vida de la organización. Los datos comunes deben tratarse como un bien de la organización, y ella deberá desarrollar su propio sentido de la propiedad y del cuidado de los datos. La administración de esta evolución es el factor más importante en la prevención del "caos" en la información.

El movimiento de los datos de un estado privado a otro compartido, y luego al común, debe ser acompañado por tecnologías y responsabilidades organizativas adecuadas a las necesidades de la organización. Estas tecnologías se relacionan con los siguientes puntos:

- Independencia de los datos de los programas de aplicación.
- Confiabilidad. (Relacionada con el comportamiento de la base de datos ante desperfectos de Hardware y Software)
- Facilidades de modificación. (Posibilidad de modificar el modelo de datos de la organización)
- Integridad. (Relacionado con la posibilidad de recuperación de datos ante transacciones equivocadas)
- Seguridad. (Relacionado con la violación dolosa de la consistencia de los datos)
- Privacidad. (Relacionada con la protección de la información ante accesos o consultas no deseadas)
- Posibilidad de acceso y uso compartido. (Relacionado con el acceso simultáneos desde varios puntos)

- Rendimiento y administración. (Relacionado con la administración de la base de datos, control de la performance, tiempos de acceso al dato, etc.)

Los datos comunes deben ser administrados por una tecnología que tenga un alto rendimiento en los puntos precedentes.

En contraste, los entornos de datos privados poseen bajos rendimientos en cada punto. El problema se plantea al administrar la transición de los datos de un estado a otro y, por lo tanto, de un nivel de tecnología a otro. Estos niveles de tecnología deben definirse de manera consistente con las reglas de propiedad de los datos, y los pasos de la migración empleando metodologías y herramientas para crear bases de datos y software para el usuario final.

4. Estructura de datos.

Como lo dijimos en los puntos anteriores, el nivel de datos esencial es el elemento dato. (También se puede llamar por otros nombres: Campo, dato o parte elemental). No hay ninguna unidad más pequeña que tenga significado para el observador de la realidad; por ejemplo, el número de factura, la fecha de la factura, la cantidad de productos, etc. son elementos dato incluidos en el objeto factura.

Los elementos dato son los bloques que construyen todos los otros datos (objetos) del sistema y por sí mismos no proporcionan suficiente información a ningún usuario. Por ejemplo, el significado del dato FECHA de una factura puede ser bien entendido: significa la fecha en que se emitió la factura; sin embargo, fuera de este contexto no tiene significado; puede pertenecer a una fecha de pago, a una de nacimiento, de inicio o de factura.

En consecuencia entendemos por una estructura de datos como un conjunto de datos que están relacionados entre sí y que describen en forma colectiva un componente del sistema (objeto de la realidad). Por ejemplo la estructura de datos FACTURA está definida como algo que consiste en datos que incluyen la fecha de factura, proveedor, dirección y detalle de los artículos. (Fig. 9.6)

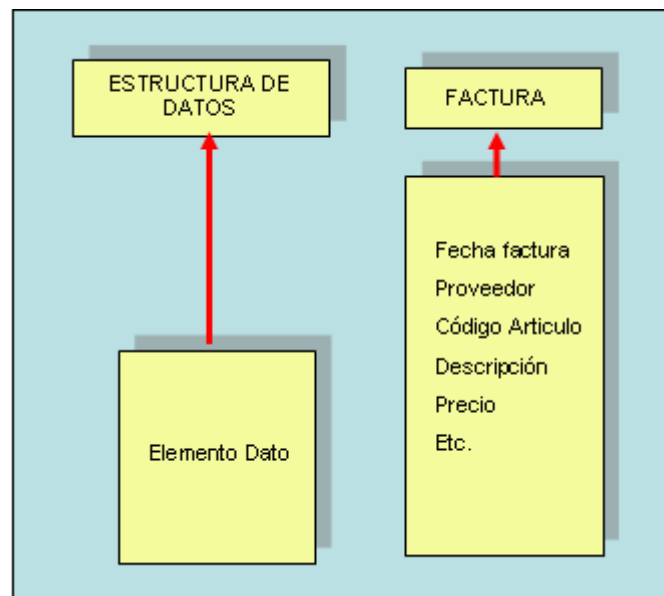


Fig. 9.6 - Concepto de estructura de datos

5. Concepto del modelo de datos de la realidad.

Como hemos señalado en los capítulos anteriores, el observador capta los datos la realidad bajo estudio y luego los organiza de manera que interpreten dicha realidad. El objetivo es conseguir captar la semántica del mundo real, y es por ello que examinaremos en este punto el concepto de "modelo de datos", que sirve de soporte a esta interpretación.

En una primera acepción al término modelo es usada para indicar la representación de una cierta realidad. Pensemos por ejemplo a un mapa que reproduce, según ciertas convenciones, una porción de territorio. Según esta acepción, el mapa constituye un modelo del territorio. En este caso, el modelo viene a ser utilizado como una imagen de la realidad. Este significado lo hemos empleado en todos los capítulos anteriores al referirnos que modelábamos la realidad.

En una segunda acepción el término modelo significa un conjunto de estructuras simbólicas utilizadas para describir la representación de una cierta realidad de interés. En este caso, por lo tanto, el modelo es el lenguaje de descripción utilizado para construir una representación de la realidad, representación que generalmente se denomina "esquema". Este segundo significado es el que generalmente es usado en el campo de las Bases de Datos y que haremos referencia en los próximos puntos. Así, por ejemplo en el ámbito de los sistemas de gestión de base de datos, podemos hablar del modelo jerárquico, el modelo de redes y el modelo relacional para indicar los tres formalismos que vienen típicamente utilizados para la descripción lógica de los datos.

En la **Fig. 9.7**, viene representado lo explicado en el párrafo precedente.

CONCEPTO DE MODELO

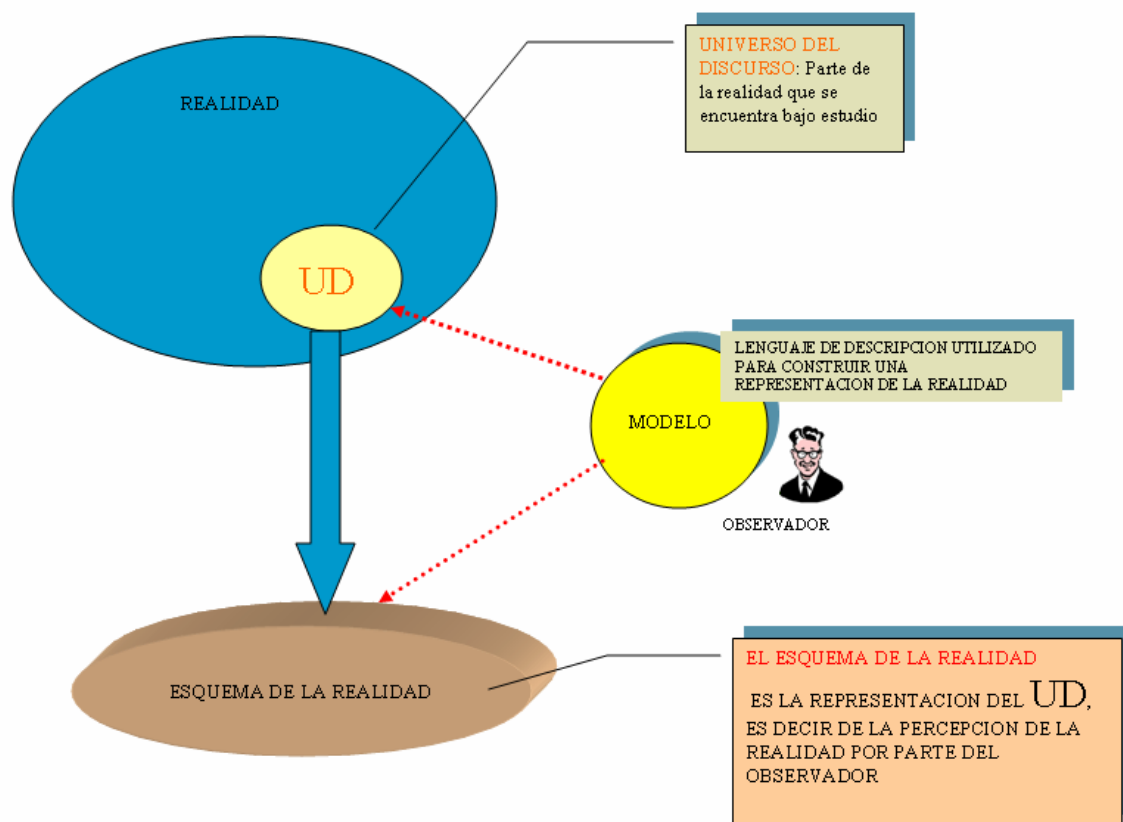


Fig. 9.7 - Concepto de modelo de datos de la realidad

Los objetivos que persigue todo modelo de datos son:

- **Formalización**, ya que el modelo de datos permite definir formalmente las estructuras permitidas y las restricciones a fin de representar los datos de una realidad sesgada por los filtros de un observador. Un ejemplo de lo expuesto lo constituye el modelo entidad (ER) interrelación que nos permite definir el esquema conceptual de una base de datos.
- **Diseño**, ya que el modelo de datos es uno de los elementos básicos (junto con los lenguajes, documentación y otras herramientas) para el desarrollo de una metodología de diseño de la base de datos, que almacena los datos representativos del comportamiento de la realidad bajo estudio. Tomamos como ejemplo de dicha metodología, los tres niveles de abstracción para el diseño de una base de datos: El esquema conceptual, el esquema lógico y el esquema físico.

En síntesis y considerando las anteriores precisiones, podemos definir un "modelo de datos" como un conjunto de conceptos, reglas y convenciones que nos permiten describir los datos de la realidad percibida, constituyendo, como hemos indicado, una herramienta intelectual que facilita la interpretación de nuestra visión de la realidad y su representación en forma de datos en el diseño de la Base de Datos.

6. Paradigma de desarrollo accionado por los Datos y los procesos. La tecnología BPM.

En el desarrollo de los capítulos anteriores, ya se mencionó que datos y procesos comienzan a reconocerse como dos cosas diferentes. Empleando este concepto, se identifican los procesos y sus objetos (físicos, humanos, de información, etc.) sobre los que la empresa recopila datos. Por ejemplo, una organización, en alguno de sus procesos puede recopilar datos sobre clientes y productos. Estos objetos tienen interrelaciones. La relación entre los clientes y los productos, es que los clientes compran productos. Una vez definidos los objetos y sus relaciones, se definen todos los atributos que describen los objetos. Los clientes tienen un nombre, un domicilio, etc. La determinación de los objetos, sus relaciones y los datos que describen a los objetos constituyen el esquema de datos de la Organización. Obsérvese que para construir el esquema, no se ha hecho mención alguna de las aplicaciones informáticas de soporte. El esquema vale por sí mismo, es decir el resultado sería una estructura más independiente de las estructuras de los procesos y de las aplicaciones que los sirven, y por lo tanto también de los cambios que pudieran surgir en las mismas. Veremos más adelante como la tecnología BPM (Business Process Management o Gestión de procesos de negocio), contribuye a este modelo.

Esta estrategia del desarrollo de una estructura de datos basada en los procesos de la Organización, permite la construcción de un sistema de información gerencial, en la cual la alta gerencia puede participar, si tiene la herramienta adecuada, como un usuario del procesamiento de datos. Sus palabras son usualmente parecidas a éstas: "Quiero averiguar dónde viven nuestros mejores 100 clientes y cuáles son nuestros 50 productos que se venden con mayor frecuencia en esas áreas. La pregunta está orientada a obtener datos sobre los objetos, es decir al mismo tema del que tratan los esquemas de datos. Si existen los esquemas, es fácil ir a los objetos a los datos que lo describen, y finalmente a las bases de datos que contienen esos datos.

En un nivel organizacional, los procesos de negocio son esenciales para comprender cómo opera una organización. Aunque también son importantes para el diseño e implementación de sistemas de información flexibles. Estos sistemas proveen la base para la creación rápida de nueva funcionalidad que cree nuevos productos, y también para adaptar rápidamente funcionalidad existente a requerimientos del negocio. BPM es entonces una estrategia para gestionar y mejorar el rendimiento de un negocio optimizando sus procesos a través de la modelización, ejecución y medida de rendimiento dentro de un ciclo de mejora continua. Es una disciplina que va mucho más allá de la arquitectura de software que la implementa, pero éste es, sin lugar a dudas, un elemento muy importante.

En los orígenes de la ciencia de la computación todo el énfasis estuvo puesto en desarrollar sistemas que automatizaran tareas que se hacían manualmente. Este era suficiente desafío. En la actualidad, se han automatizado muchas tareas y el gran desafío es cómo mejorar la capacidad de los sistemas para alcanzar nuevos requerimientos: agregar nuevas interfaces, combinar múltiples fuentes de datos en una sola, interactuar con dispositivos móviles y reemplazar viejas aplicaciones con nuevas.

La tendencia hoy es hacia un paradigma orientado a procesos, donde las aplicaciones cubren la actividad global de la empresa y las herramientas son los BPMS (Business Process Management Systems o Sistemas de Gestión de Procesos de Negocio).

Los modelos actuales resultan insuficientes porque son pobres en su integración y se orientan a describir datos y transacciones. El cambio de enfoque en el modo de diseñar aplicaciones e implementar soluciones radica en:

- Explicitar el conocimiento de un proceso de negocio ayudando a documentarlo, a definirlo y a implementarlo.
- Proveer interoperabilidad de las soluciones
- Resolver la dinámica de los problemas en términos declarativos y cubriendo todas las etapas del ciclo de vida del software.

BPM (Business Process Management o Gestión de procesos de negocio) ha adquirido una atención considerable recientemente tanto por las comunidades de administración de negocios como las de ciencia de la computación. Los miembros de estas comunidades están identificados por diferentes soportes e intereses educacionales; entre ellos encontramos los analistas de negocios, que están interesados en mejorar las operaciones de las compañías. Hechos como incrementar la satisfacción de los clientes, reducir los costos de operación, y establecer nuevos productos y servicios a bajo costo, son aspectos importantes del manejo de procesos de negocio desde el punto de vista de un analista del negocio.

La comunidad de software, por su parte, está interesada en proveer sistemas robustos y escalables. Como los procesos de negocios son realizados en espacios de información tecnológica compleja, la integración de los sistemas de información existentes se convierte en una base importante para la implementación técnica de los procesos de negocio.

En una Organización, el BPM se basa en la observación de cada producto que la compañía provee al mercado, lo cual genera como resultado un número de actividades ejecutadas. Los procesos de negocio son la clave para organizar estas actividades, y mejorar el entendimiento de sus interrelaciones.

La tecnología de la información en general y los sistemas de información en particular, merecen un rol importante en BPM, porque cada vez más actividades que realizan las compañías son soportadas por sistemas de información. Las actividades de los procesos pueden ser ejecutadas manualmente por empleados de la compañía, o con la asistencia de aplicaciones dedicadas a tal fin. También puede ocurrir que las actividades sean directamente ejecutadas por sistemas sin intervención humana.

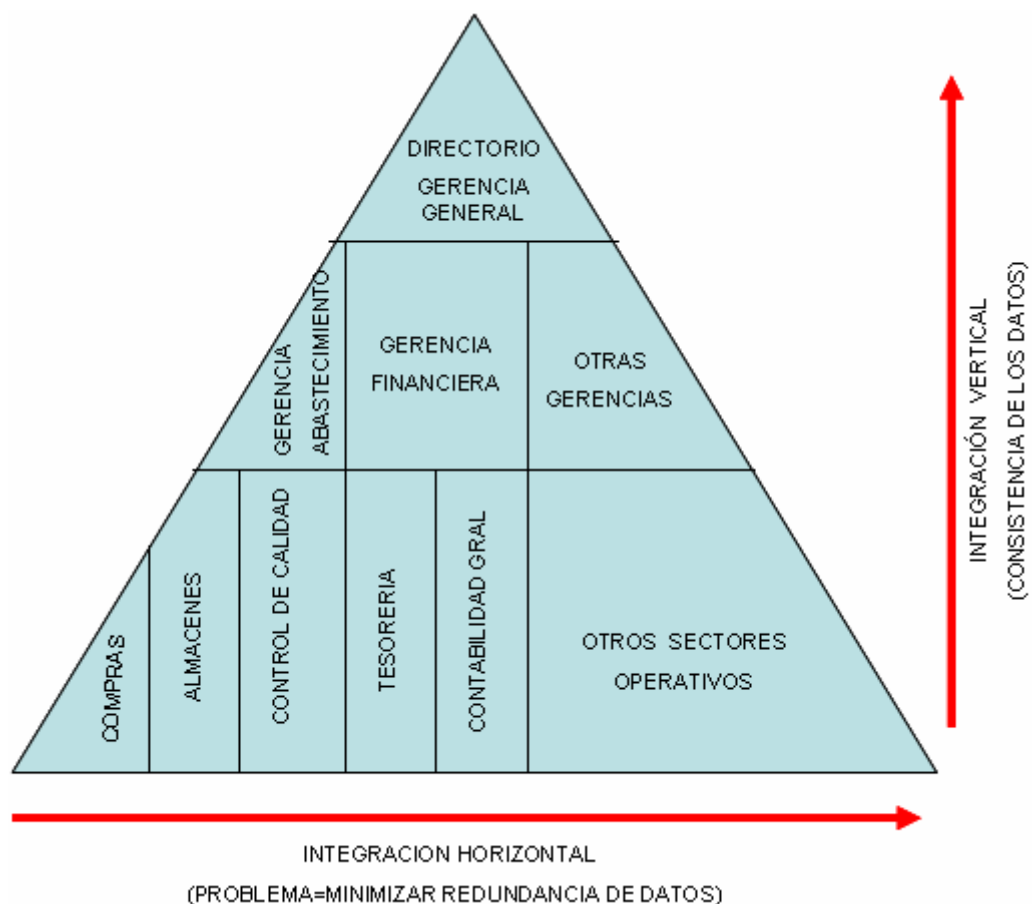
7. Integración vertical de los Datos (Valor del Dato).

La administración de datos está dedicando sus mayores esfuerzos a la integración vertical de los datos, apartándose de la horizontal. La **Fig. 9.8** explica qué se entiende por integración de datos, tanto horizontal como vertical.

Si las aplicaciones se ven en el contexto del modelo piramidal de las organizaciones, en el cual la relación entre dichas aplicaciones y los procesos no es tomada en cuenta, pueden agruparse en términos de los niveles a las cuales apoya. Definitivamente, la mayor parte de las aplicaciones

de una cartera de sistemas empresarios actual, serán de tipo transaccional, y darán apoyo al nivel operacional (a los procesos). En muchas organizaciones son los únicos sistemas que existen. Pero crece el interés en los de apoyo a los niveles de control y estratégicos. Crece el desarrollo de sistemas de información gerencial, y de soporte a las decisiones. Con este crecimiento está cambiando el punto focal de la administración de datos.

La integración horizontal ve a los sistemas operacionales casi como una línea de montaje en la que los datos se transfieren de un sistema a otro. La responsabilidad de la administración de datos, si el objetivo es la integración horizontal, consiste en desarrollar e implementar normas para minimizar la redundancia de datos. La interfase de mayor importancia de la administración de datos es la que tiene con el personal del área informática.



9.8- Tipos de integraciones en una Organización

Esto cambia con la integración vertical de los datos, en la que el punto central consiste en proporcionar información sobre las operaciones a las gerencias de control, y a las estratégicas. Aquí, el requerimiento es desarrollar e implementar, en los sistemas de soporte operacional, normas para los datos que son requeridos a nivel gerencial (ver calidad de los datos). El objetivo es permitir que la organización se comunique. La redundancia, como preocupación, se ve reemplazada por la consistencia. Por ejemplo; concuerdan y pueden ser interrelacionados los datos de varios sistemas de soporte operacional?. (un ejemplo de lo expresado es si la actualización de los datos de la base se realizaron a una misma fecha)

La integración vertical de los datos está produciendo un impacto de importancia en su administración. Ahora que la alta gerencia comienza a participar directamente en sus propios datos, se comprende mejor el concepto de que debe haber una perspectiva empresarial de los

datos, y de que éstos son un recurso. La alta gerencia está llegando a conocer muy bien el volumen de los datos, y los problemas de calidad que se le plantean cuando intenta acceder directamente a ellos. El papel de la administración de datos está, por lo tanto, dejando de ser sólo una función técnica cuya interfase principal es el personal del área informática. Esta sumando a estas responsabilidades las interfases con las gerencias intermedias y altas, que se interesan en una perspectiva empresarial de los datos. Esto da realce al conocimiento de cuál es el significado de los datos.

En consecuencia, la administración de datos está articulando la necesidad de controlar toda la dimensión de los mismos dentro de la "Organización". Las normas del pasado han tratado sobre la forma de nombrar los datos, definirlos, fijar reglas para su uso en un programa de computadora, y seguir la pista del lugar en que se usan. Esta es la mayor parte del total de responsabilidades de la administración de datos en un entorno de integración horizontal. Pero, con la integración vertical, llegan a la superficie nuevas responsabilidades. La más importante de ellas es la definición y puesta en vigor de las políticas de datos dentro de la organización. Poco es lo que tienen que ver estas políticas con las aplicaciones de computadora, pero mucho con los datos de esas aplicaciones. Los gerentes que dirigen a quienes ingresan los datos en los sistemas de nivel operacional, utilizan estas políticas. En la integración horizontal, el efecto de no seguir políticas uniformes de datos queda en general oculto mientras los datos fluyen por el sistema.

Pero la naturaleza especial de las consultas de las gerencias intermedias y alta con respecto a los datos, hace que se pongan de manifiesto muchos problemas de inconsistencia en la política de datos. Con la integración horizontal de los datos era casi imposible identificar su utilidad. Debido a ello, se asignó el mismo valor a todos los datos, y los procedimientos de control los trataron de la misma manera. La integración vertical introduce otro patrón; varios niveles gerenciales utilizan algunos de los datos, mientras que otros son utilizados a nivel operacional, quizás en tareas de poca importancia. Los datos que son usados por más de un nivel de la organización, o por quienes toman decisiones estratégicas, son por lo tanto mas importantes. Es necesario normalizar esos datos. Sabiendo esto, el administrador de datos puede concentrar sus recursos en las áreas que tienen mayor impacto.

8. La calidad de los Datos.

8.1. Introducción

Cada minuto en la operación diaria de una Organización se toman decisiones, la mayor parte de esas decisiones son tomadas con información o cierta intuición de los decidores, mientras más usamos la información y los hechos, y menos utilizamos la intuición, se puede estar más seguro de haber tomado una buena decisión.

Entramos a otro dilema de cuál información utilizamos para tomar una decisión, ¿es una fuente única con la que todos en la Organización tomamos una decisión? , ¿ Hay un sistema formal que apoye el proceso de toma de decisiones en la Organización? ¿ Existen otras fuentes de información disponibles para validar las existencias de inventarios, los precios de los productos, los pedidos de los clientes, los gastos de un centro de costos, las ventas netas de la compañía?.

Desgraciadamente en la mayoría de las Organizaciones la respuesta es no hay una sola fuente de información, existen diversos sistemas informales en los que muchas personas se basan para tomar decisiones que inciden en la operación de sus actividades, las cuales en caso de ser erróneas pueden tener un fuerte impacto económico en la Organización.

El primer paso para tener una administración basada en información es tener una sola fuente, si existen diversas fuentes siempre tendremos que dedicar inútilmente recursos a que la información cuadre, o tendremos que dedicar un doble esfuerzo para mantener actualizados

ambos sistemas, tanto los formales como los informales, es conveniente tener una sola fuente de información, con información buena o mala, pero que al menos es la misma para todos.

Una vez logrado este primer paso, pasamos a una segunda instancia, la cual es asegurarnos que esta fuente única de información, este sistema formal de información tenga la información correcta, dicho de otra manera, que la información de nuestro sistema refleje la realidad de lo que esta sucediendo en nuestra Organización.

Lograr esto no es nada fácil, ya que no forma parte de la cultura de la Organización que su personal perciban como parte de su trabajo, el mantener una información confiable. El problema de calidad de datos es una realidad, la exactitud promedio de los inventarios esta por debajo del 90%, las rutas de fabricación no reflejan todos los pasos que se requieren para fabricar un producto o los tiempos definidos no son los que suceden en la realidad, las listas de materiales están incompletas, no reflejan las cantidades correctas o no contemplan las mermas inherentes al proceso, las ordenes de compra no son actualizadas al momento de haber cambios en la fecha o en la cantidad a entregar, la lista de precios no es la misma en diversos almacenes de la republica, los limites de crédito son obsoletos y mandan todos los pedidos a estatus "detenidos por crédito" generando grandes cargas de trabajo al personal de Crédito. En fin podríamos mencionar más ejemplos, pero creo que es claro que la pobre calidad de datos va en contra de la buena administración de los recursos humanos, financieros y materiales de una Organización, va en contra de la creación de valor y de la excelencia operacional, incrementa los costos, impacta los niveles de servicio y reduce la eficiencia de nuestras operaciones.

8.2. Que es la calidad de los Datos?

Podemos decir que la calidad de datos implica que los datos capturados, procesados, almacenados y entregados son un fiel reflejo de la realidad que se desea tratar mediante sistemas informáticos. Esto supone que los datos no contengan errores, sean veraces y estén actualizados.

La definición más aceptada de calidad de datos es “aptitud para su uso” (“fitness for use”): los datos tienen calidad si sirven para el propósito que se necesita. Podemos decir entonces, que la calidad depende tanto de los datos como del uso de los mismos. A la vista de esta definición, se deduce que la calidad de datos es un concepto subjetivo (depende del uso que vaya a hacer cada persona) y multidimensional (es factible entender la calidad desde distintos puntos de vista – al estilo del software con ISO 9126). A los diferentes puntos de vista se les llama “*dimensiones de calidad de datos*”. Estas dimensiones son ampliamente discutidas y explicadas en la literatura (ver punto 8.3 de este capítulo).

Para poder hacer una gestión adecuada de la calidad de los datos, es necesario definir medidas para cada una de las dimensiones requeridas para un escenario. La forma en la que se definen las medidas depende fuertemente del método de medición que cada parte interesada necesite emplear según sus necesidades: así una misma dimensión de calidad puede tener tantas medidas diferentes como partes interesadas haya en un escenario. En cualquier caso, las medidas de calidad de los conjuntos de datos pueden clasificarse como orientadas al modelo de datos (suelen ser objetivas) o las orientadas a los propios datos (pueden requerir metadatos que complementen el significado del dato con respecto a la dimensión de calidad de datos según la parte interesada). En (Lee et al., 2006) y en (Batini and Scannapieco, 2006) se dan indicaciones oportunas sobre cómo definir estas medidas. Estas medidas serán la base para la creación del sistema de reglas.

8.3. Dimensiones de la calidad de los Datos

Como dijimos anteriormente, la Calidad de Datos tiene varias dimensiones, relacionadas pero distintas, a continuación se describen algunas:

Exactitud: Mide el grado en que la información refleja lo que esta pasando en la realidad de la Organización bajo estudio (ej. En una empresa podría ser: Exactitud de inventarios, exactitud de rutas de fabricación, de listas de materiales, etc.).

Totalidad: Medición que refleja el grado en que las bases de datos de la Organización cuentan con toda la información crítica para el negocio.

Oportunidad: Medición de que la información este disponible cuando se requiere para tomar una decisión.

Relevancia: Que la información le sea útil al observador al cual se la estamos proporcionando.

Nivel de detalle: Que la información tenga el nivel de detalle requerido, dependiendo del nivel Organizacional y al tipo de decisión al cual este destinada la información.

Consistencia: Que la información sea la misma en todas las áreas o sistemas utilizados por la Organización.

En fin, hay que validar trabajar sobre varias de estas dimensiones hasta hacer que nuestro objetivo no sea tan solo proveer información de calidad, sino el de apoyar a la organización a hacer buen uso de la información para apoyar y mejorar el uso de los recursos y las operaciones del negocio.

8.4. Método para mejorar la Calidad de Datos en una Organización

Las Organizaciones, especialmente las dedicadas al giro industrial han pasado años comentando que es imposible llegar a niveles arriba del 95 % de exactitud de inventarios, sin embargo esto es una cuestión cultural y de disciplina, podemos imaginarnos el impacto de una liquidación de sueldos con el 90 % de exactitud, o que la aplicación de medicamentos a los pacientes en un hospital estuviera en un 90 %, o que los aviones tuvieran un 98 % de exactitud en los aterrizajes, o que nuestra cuenta de cheques del banco tuviera un nivel por debajo del 95 %, sencillamente estos porcentajes en los bancos, hospitales, aeropuertos y nominas son inconcebibles, el cliente/ciudadano no podría vivir con ellos.

Los bancos tienen gente de bajo nivel de remuneración en las cajas, al igual que los mecánicos de los aviones o las enfermeras de un Hospital, sin embargo ellos pueden hacer su trabajo con perfección, pero en las industrias el pretexto eterno, "como quieren que este correcta la información si mis almacenistas, o operarios, o compradores tienen un ingreso bajo y no tienen la capacidad adecuada". Esto es cuestión de incorporar una cultura de calidad dentro de la Organización, como responsables debemos crear una alta expectativa para la calidad de la información de nuestra área, ya que esta no solo nos impacta a nuestra área sino a toda la organización.

- Una vez creada la cultura Organizacional, y reconocidos los beneficios de mejorar la calidad de la información, se debe aplicar una metodología para mejorar la calidad de la información en nuestra Organización. Existen muchas metodologías, a continuación se explica brevemente una al solo efecto de brindar un ejemplo:
- Paso 1: Identificar la información crítica para el negocio. Existe tanta información en una organización que difícilmente podemos dedicar suficientes recursos para mejorar la calidad de toda la información, por lo que hay que identificar cual es la información que tiene un mayor impacto en las operaciones del negocio.
- Paso 2: Definir criterios de Calidad de Datos. Esto es definir cuando un dato es exacto para su organización.

- Paso 3: Realizar mediciones iniciales para detectar posibles problemas de Calidad de Datos. Este paso es un diagnóstico que nos apoya a medir la calidad actual de la información crítica definida en el Paso 1, identificando donde se encuentran los mayores problemas y priorizando las áreas con las que hay que iniciar el esfuerzo.
- Paso 4: Automatizar Indicadores de Calidad de Información. En este paso se realizan programas que apoyen a medir periódicamente la Calidad de la Información, *“lo que no se puede medir no se puede administrar”* y por lo tanto no se puede mejorar. Estos medidores deben estar al alcance de las personas que serán las responsables de monitorear y mejorar la calidad de la información.
- Paso 5: Definir responsables de Calidad de Datos. Una de los factores críticos de éxito de un proyecto de Calidad de Datos es definir un responsable de cada indicador, esta persona debe monitorear las tendencias del indicador y realizar planes de acción encaminados a la mejora de los indicadores.
- Paso 6: Diagnósticos de calidad de Datos. En estos diagnósticos se determinan las posibles causas de la mala calidad de Datos y se definen planes de acción con responsables para mejorar el indicador. Entre las planes de acción normalmente se incluye el establecer controles preventivos y correctivos para la mejora de la calidad de Datos.
- Paso 7: Monitoreo de los indicadores por parte de la Gerencia. Si los empleados no perciben las altas expectativas por parte de la gerencia, el proyecto tendrá resultados limitados, es importante que los indicadores de Calidad de Datos se revisen periódicamente, asegurando su seguimiento y mejora continua.
- Paso 8: Finalmente, una vez estabilizados los indicadores de Calidad de Datos, se regresa al punto uno para identificar información que tenga impactos en el negocio y que sea necesario su medición y mejora.

Nota Importante.

En una Organización con una gestión por procesos, las mediciones de la “Calidad de los datos”, deben estar incluidas dentro del mismo proceso.

8.5. Actividades desfavorables dentro de una Organización que impactan en la calidad de los Datos.

La propia dinámica del sistema Organización, obliga a realizar actividades que impactan directamente en la calidad de los datos que luego son gestionados por la misma. Estas actividades son:

- Nuevos usos de los datos por las nuevas aplicaciones informáticas.
- Replicación (o duplicación).
- Integración.

Ya conocemos las necesidades de las Organizaciones en cuanto a que los sistemas informáticos y de información, evolucionan y cambian constantemente para adaptarse a las variaciones de contexto. Estos cambios conducen, en forma inexorable a nuevos usos de los datos existentes. Estos nuevos usos, en general, son diferentes de los previstos en el diseño original, causando

una degradación en la calidad de la base de datos. Las razones del cambio de calidad son las siguientes:

- El diseño puede no incluir todos los campos (Atributos) necesarios.
- Se acomodan los datos a un diseño inadecuado.
- Las aplicaciones y los datos están fuertemente acoplados.
- Con frecuencia existe duplicación de datos.
- Etc.

Por las razones expresadas, hemos insistido en distintos puntos de este libro, la importancia del observador en identificar las fronteras del sistema y modelar los datos de la realidad considerada sin tener en cuenta los procesos. Las simplificaciones y los filtros propios del observador atentan contra la calidad de los datos. La propia limitación del observador, hace muy difícil anticipar todos los usos futuros de los datos al construir la base de datos de la Organización.

En la nueva manera de utilizar los datos, existe la tendencia a replicar (o duplicar) los datos para satisfacer las nuevas necesidades. Replicación incluye agrupamiento de datos, combinación de diferentes fuentes, migración a estructuras de datos diferentes de las originales y la incorporación de series históricas o de tiempo. Todo esto es una fuente de error que atenta contra la calidad de los datos.

De la misma manera, en un proceso de integración, es necesario integrar los datos de diferentes bases de datos, las cuales son alimentadas por aplicaciones interactivas. Normalmente en el proceso de integración de datos, debemos trasladar los datos a una estructura de datos diferente, correspondiente a la base de datos integradora. Por ejemplo en el caso de un proyecto de “Data warehouse”.

En resumen, el problema de calidad de datos es muy común en las Organizaciones y resulta de la naturaleza cambiante de la actividad dentro de la misma. Nuevas aplicaciones, nuevos usos, integración y replicación afectan a la calidad de los datos. Por otro lado, los errores de datos se trasladan a los sistemas de decisiones, produciendo efectos impredecibles.

8.6. Calidad de datos y la definición de exactitud

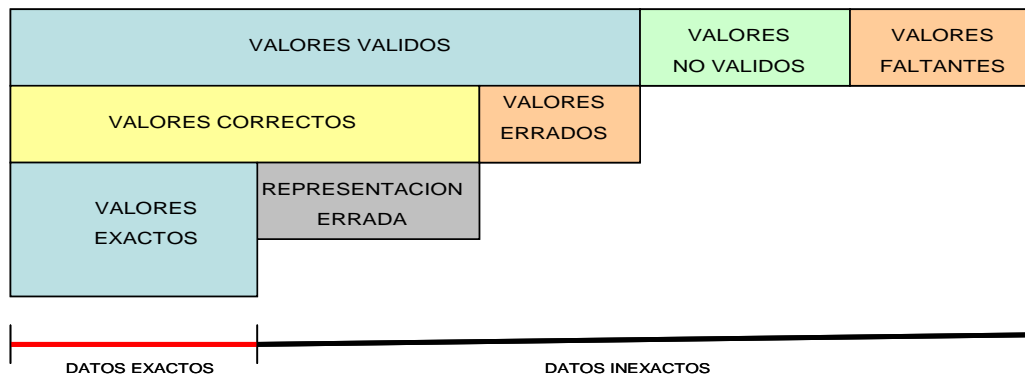
Como vimos en el punto 8.3, la exactitud es solo una de las dimensiones de la calidad de datos, es decir es una condición necesaria pero no suficiente, aunque la consideramos como la componente más importante de la calidad de datos. Cualquier programa de mejoramiento de la calidad de datos debe comenzar asegurando “la exactitud de los datos”. Para ser exacto, un dato debe tener el valor correcto y estar representado de manera consistente e inequívoca. La exactitud tiene dos características:

- Forma y
- Contenido

La forma es importante, porque elimina ambigüedades sobre el contenido. Por otro lado, un valor no es exacto si el usuario del valor no puede determinar que es o el significado del mismo.

La representación del valor debe tener consistencia, esto es, dos valores diferentes no pueden representar lo mismo. La consistencia del valor, forma parte de la exactitud del dato. Un valor es válido, si es un elemento del conjunto de posibles valores correctos (dominio). Un valor válido no necesariamente es correcto. Pero el valor correcto siempre es válido. Por otro lado los valores faltantes, también son causa de errores en los datos, puesto que su significado es

ambiguo. Es correcto y necesario distinguir entre blanco “no hay valor” y nulo “no se conoce valor”.



9.9- Exactitud del dato

8.7. Como identificar los valores errados

Luego de lo expuesto en los puntos anteriores sobre la calidad de los datos, la pregunta es: como podemos, en una Organización, identificar los valores errados??. Podemos decir que existen técnicas que nos permitirían determinar la mayoría de los errores, muy difícilmente la totalidad. Estas técnicas son las siguientes:

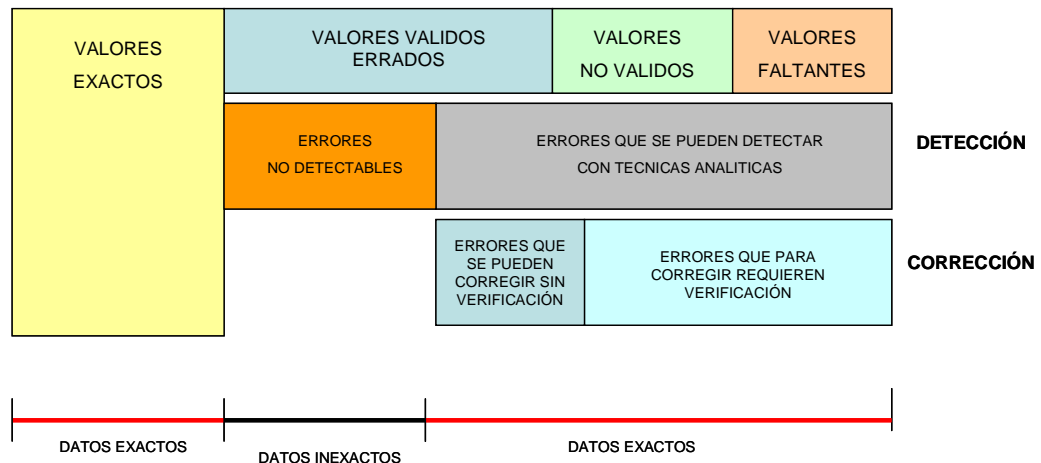
- Verificación manual
- Análisis automático

El proceso manual consiste, en base a la fuente original de la información, se verifican todos y cada uno de los valores. Por supuesto que este mecanismo, además de no garantizar la detección total de los errores, es poco práctico, muy largo y costoso. La aplicación de este método es posible solamente si nos limitamos a un universo reducido, como un subconjunto de datos críticos de la Organización.

Las técnicas analíticas (formalmente denominada *Data Profiling*) utilizan un software y la habilidad del responsable de la calidad de datos para detectar los datos inexactos. Existen cuatro categorías de análisis que se pueden aplicar a los datos:

- Análisis de elementos (Análisis de información sobre los Datos. “Metadatos”. Esto es, los tipos, longitud, rangos, valores discretos, patrones, formatos, etc.)
- Análisis estructural. (Errores o inexactitudes en una base de datos como dependencias funcionales, sinónimos, reglas de integridad, etc.)
- Análisis de reglas del negocio. (análisis de los datos que impactan mas fuertemente en las reglas principales sobre determinados objetos del negocio, simples y compuestas)
- Análisis estadístico. (Este análisis es aplicable cuando cuando no es posible formular una regla concreta. Entonces se puede utilizar una base estadística en base a Distribución de frecuencias, conteos, sumas, promedios, etc.)

De acuerdo a lo anterior, el análisis de elementos nos permite detectar valores no validos, el análisis estructural, análisis de reglas del negocio y análisis estadísticos, nos permite detectar inexactitudes entre valores validos.



9.10- Análisis con técnicas analíticas automático de datos

En conclusión, los responsables de la calidad de datos han utilizado durante mucho tiempo métodos ad-hoc, no formales, para evaluar y examinar datos. En la actualidad existen herramientas y métodos formales, como el “Data Profiling” el cual ha evolucionado y madurado a una tecnología formal y efectiva que utiliza para la evaluación de la calidad de datos.

9. Conclusiones

Como dijimos anteriormente los datos se han convertido en uno de los activos mas importantes de las Organizaciones, ya que son la fuente de información para la toma de decisiones a nivel estratégico y a nivel operativo. Es por ello que las Organizaciones recopilan una gran cantidad de datos para ser más competitivas. Esto, en algunas ocasiones, yo diría en general, conlleva a una situación caótica en la gestión de los mismos, puesto que la Organización suele llegar a la administración de datos inútiles e innecesarios, tener redundancia incontrolada, y disponer de gran cantidad de datos históricos caducados. Es decir, los datos pueden convertirse en una fuente de problemas, y estos problemas afectar negativamente al rendimiento de los procesos del negocio de la Organización.

Podemos decir, que en este capítulo planteamos una solución a este tipo de problema, típico de las Organizaciones. Esto es, “mejorar la calidad de los datos y de la información a través de una gestión de la calidad de los mismos”.