# Gestion de Datos con la Guia DAMA DMBOK

IGP – Peru 40h

Instructor: Ing. Carlos Carreno ccarrenovi@Gmail.com

## Capitulo 5 Modelamiento y Diseño (Desarrollo de Datos).

- Introduccion
- Equipo de Desarrollo de Datos
- Actividades de la Función Desarrollo de datos
- SLDC
- Estilos de Modelado de Datos
- Modelo Conceptual
- Modelo de Datos Lógico
- Modelo de Datos Físico
- Modelo de Datos Detallado
- Principios Rectores

#### Introduccion

- El desarrollo de datos es el análisis, diseño, implementación, despliegue y mantenimiento de soluciones de datos para maximizar el valor de los recursos de datos a la empresa.
- El desarrollo de datos es el subconjunto de actividades del proyecto dentro del Ciclo de Vida de Desarrollo de Sistemas (SDLC) centrado en la definición de los requisitos de datos, el diseño de los componentes de la solución de datos y la implantación de estos componentes.
- Los componentes primarios de la solución de datos son las bases de datos y otras estructuras de datos. Otros componentes de la solución de datos incluyen los productos de información (pantallas e informes) y las interfaces de acceso a los datos.

#### Equipo de Desarrollo de Datos

- Los administradores de datos de negocios y expertos en la materia (SMEs) proporcionan los *requerimientos del negocio* de datos e información, incluyendo las reglas de negocio y la calidad de datos esperada, y luego validan que se hayan cumplido estos requisitos.
- Los arquitectos de datos, analistas y administradores de bases de datos tienen la responsabilidad principal de diseño de base de datos. Los administradores de bases de datos colaboran con los desarrolladores de software para definir los servicios de acceso a datos en implementaciones de arquitecturas orientadas a servicios (SOA) en capas.
- Los arquitectos y desarrolladores (ambos especialistas en integración de aplicaciones y datos) de software tienen la responsabilidad principal de captura de datos y diseño de uso dentro de los programas, así como el diseño de la interfaz de usuario para productos de información (pantallas e informes impresos).

#### Diagrama de Contexto de Desarrollo de Datos

Definición: Diseño, implementación, y soluciones de mantenimientos para satisfacer las necesidades de datos de la empresa.

#### Metas:

- Identificar y definir los requerimientos de datos.
- 2. Diseño de estructura de datos y otras soluciones para los requerimientos.
- Implementar y mantener componentes de soluciones para satisfacer los requerimientos.
- Garantizar solución de conformidad de arquitectura de datos y apropiados estándares.
- Garantizar la integridad, seguridad, usabilidad, y sustentabilidad de datos activos estructurados.

#### \_\_\_>

#### **Entradas**

- Metas y Estrategias de Negocio
- Estrategias y Necesidades de Datos
- Estándares de Datos
- Arquitectura de Datos
- Procesos de Arquitectura
- Aplicación de Arquitectura
- Arquitectura Técnica

#### Proveedores:

- Administradores de Datos
- · Expertos en la Materia
- Comité Directivo de TI
- · Consejo de Gobierno de Datos
- · Analistas y Arquitectos de Datos
- Desarrolladores de Software
- Productores de Datos
- Información de Clientes

#### Participantes:

- Administradores de Datos y SMEs
- · Analistas y Arquitectos de Datos
- · Administradores de Base de Datos
- Administradores de Modeladores de Datos
- Desarrolladores de Software
- Administradores de Proyectos
- Ejecutivos DM y Otros TI Administradores

#### Actividades:

- Requerimientos de Información
- Modelado de Datos, Diseño de Análisis y Soluciones (D)
  - Analizar Desarrollar y Mantener Modelos de Datos Conceptuales
  - 2. Desarrollar y Mantener Modelos de Datos Lógicos
  - 3. Desarrollar y Mantener Modelos de Datos Físicos
- 3. Diseño Detallado de Datos (D)
  - 1. Diseñar Bases de Datos Físicas
  - 2. Diseñar Productos de Información
  - 3. Diseñar Servicios de Accesos de Datos
  - 4. Diseñar Servicios de integración de Datos
- 4. Modelado de Datos y Diseño de Gestión de Calidad
  - 1. Desarrollo de Modelado de Datos y Diseño de estándares (P)
  - Revisión de Modelado de Datos y Diseño de Base de Datos de Calidad (C)
  - Gestión de Modelado de Versiones de Datos e Integración
    (C)
- Implementación de Datos (D)
  - Implementar el Desarrollo/ Pruebas de Cambios de Base de Datos
  - 2. Crear y Mantener Pruebas de Datos
  - 3. Emigrar y Convertir Datos
  - 4. Construir y Probar Información de Productos
  - 5. Construir y Probar el Acceso de Servicios
  - 6. Validar Información de Requerimientos
  - 7. Preparar Desarrollo de Datos

#### Herramientas:

- Herramientas de Modelado de Datos
- Gestión de Sistemas de Base de Datos
- Herramientas de Desarrollo de Software
- Herramientas de Prueba

- Herramientas de Perfilado de Datos
- Herramientas de Gestión de Modelado de Datos
- Herramientas de Gestión de Configuración
- Herramientas de Producción de Oficina

#### Entregas primarias:

- · Requerimiento de Datos y Reglas de Negocio
- Modelos Conceptuales de Datos
- Modelos Lógicos de Datos y Especificaciones
- · Modelos de Datos Físicos y Especificaciones
- Metadatos (Negocios y Técnica)
- Modelado de Datos y Estándares de Diseño DB
- Modelo de Datos y Revisión de Diseños
- Versión Controlada de Modelos de Datos
- Pruebas de Datos
- Desarrollo y Prueba de Base de Datos
- Información de Productos
- Servicios de Acceso de Datos
- Servicios de Integración de Datos
- · Emigración y Conversión de Datos

#### Consumidores:

- Productores de Datos
- Trabajadores de Conocimiento
- · Administradores y Ejecutivos
- Clientes
- Profesionales de Datos
- Otros profesionales de TI

## Ciclo de Vida del Desarrollo de Software (SDLC)

- Planificación de proyectos, incluyendo la definición del alcance y la justificación del caso de negocio.
- Análisis de Requerimientos.
- Diseño de soluciones.
- Diseño Detallado.
- Construcción de componentes.
- Pruebas, incluyendo las unitarias, de integración, de sistema, de rendimiento y de aceptación.
- Preparación de la implementación, incluyendo el desarrollo de documentación y capacitación.
- Instalación e implementación, incluyendo prueba piloto y puesta en marcha.

#### **SDLC**

• Los esfuerzos de mantenimiento del sistema generalmente siguen también los mismos procesos de alto nivel del SDLC en una secuencia muy rápida, realizando análisis, diseño, codificación, prueba y despliegue en pequeñas cantidades.

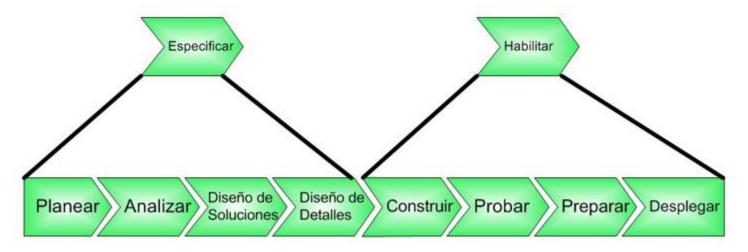


Figura 5.2 El Ciclo de Vida de Desarrollo de Sistemas (SDLC)

#### Estilos de Modelado de Datos

- IE
- IDEF1X
- ORM
- UML

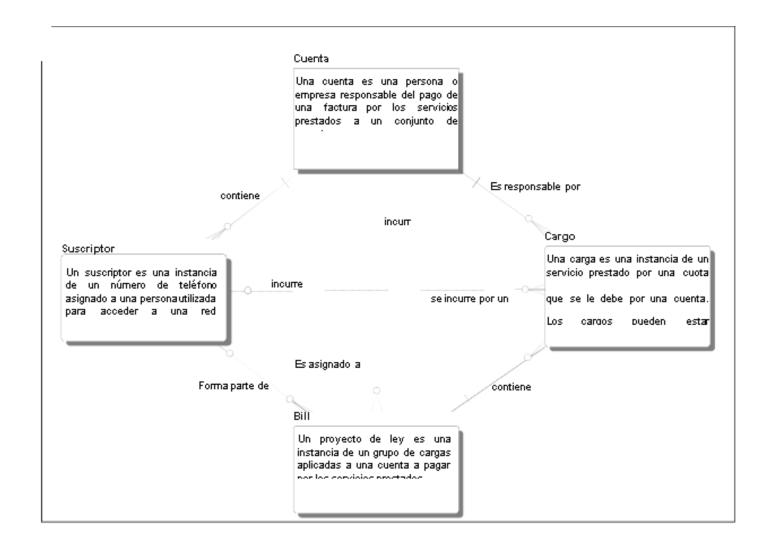
### Analizar los requisitos de información

- La información son datos en contexto que tienen relevancia y son oportunos. Para identificar las necesidades de información, tenemos que identificar primero las necesidades de información de negocios, a menudo en el contexto de uno o más procesos de negocio.
- Los procesos de negocio consumen como entrada, los productos de información de salida de otros procesos de negocio. Los nombres de estos productos de información a menudo identifican un vocabulario de negocios esencial que sirve de base para el modelado de datos.
- Independientemente si los procesos o datos se modelan de forma secuencial (en cualquier orden), o al mismo tiempo, el análisis y el diseño eficaz debe garantizar una visión relativamente equilibrada de los datos (nombres) y procesos (verbos), con el mismo énfasis en el proceso y el modelado de datos.

## Desarrollar y Mantener Modelos de Datos Conceptuales

- Un modelo conceptual de datos es una perspectiva visual de alto nivel sobre un tema de importancia para el negocio. Contiene sólo las entidades empresariales básicas y críticas dentro de un dominio y función dada, con una descripción de cada entidad y las relaciones entre las entidades.
- Los modelos de datos conceptuales definen la semántica (sustantivos y verbos) del vocabulario esencial del negocio. Las áreas temáticas de modelo conceptual de datos siempre son representativas de los datos asociados a un proceso de negocio o función de la aplicación.
- Un modelo conceptual de datos es independiente de la tecnología (base de datos, archivos, etc.) y del contexto de uso (si la entidad está en un sistema de facturación o un almacén de datos).

### Modelo Conceptual: Ejemplo



#### Entidades de Negocio

- Una entidad de negocio es algo de interés para la organización, un objeto o un evento. Una entidad de datos es una colección de datos acerca de algo que el negocio considera importante y digno de captura. Una entidad es un sustantivo:
  - ✓ Un quién: persona, organización, papel, empleado, cliente, proveedor, estudiante, partido, departamento, organismo regulador, competidor, socio, filial, equipo, familia, hogar.
  - ✓ Un qué: El producto, servicio, recursos, materia prima, producto terminado, por supuesto, la clase.
  - ✓ Un cuándo: Evento, periodo fiscal.
  - ✓ Un dónde: Lugar, dirección, web, nodo de red.
  - ✓ Un por qué: Política, norma, solicitud, queja, el retorno, la indagación.
  - ✓ Una forma: Mecanismo, herramienta, documento, factura, contrato, convenio, estándar, cuenta.

#### Tipos de Entidades Dependientes

- Entidad por atributo/característica: Una entidad que depende de una sola entidad padre, tal como Beneficiario del Empleado que depende de Beneficiario.
- Entidad Asociativa/Mapeo: Una entidad que depende de dos o más entidades, tales como Registro, que depende de un Estudiante en particular y de un Curso.
- Entidad de Categoría subtipo/supertipo: Una entidad que es "una especie de" otra entidad. Subtipos y supertipos son ejemplos de generalización y herencia. Una entidad de tipo Super es una generalización de todos sus subtipos y cada subtipo hereda los atributos de su supertipo. Por ejemplo, un entidad Supertipo Individuo tiene enlaces los susbtipos a Persona y Organización. Los subtipos pueden ser solapados (no exclusiva) o no solapados (exclusivo). Una instancia de la entidad subtipo no solapada tique que ser un sub-tipo u otro, pero no ambos.

## Reglas de Negocio (Relaciones Entre Entidades)

- Las reglas de negocio definir restricciones sobre lo que puede y no puede hacer. Las Reglas de Negocio se dividen en dos categorías principales:
  - ✓ Reglas de datos restringen cómo los datos se relaciona con otros datos. Por ejemplo, "los estudiantes de primer año pueden inscribirse por un máximo de 18 créditos por semestre." Los modelos de datos se enfocan de reglas de negocio.
  - ✓ Las reglas de acción son instrucciones sobre qué hacer cuando los elementos de datos contienen ciertos valores. Las reglas de acción son difíciles de definir en un modelo de datos. Las reglas de negocio para la calidad de los datos son reglas de acción y las aplicaciones las implementan como edición y validación de entrada de datos.

### Tipos de Reglas de Datos

- Los modelos de datos expresan dos tipos principales de reglas de datos:
  - ✓ Las reglas de cardinalidad definen la cantidad de instancias de la entidad que puede participar en una relación entre dos entidades. Por ejemplo, "Cada empresa puede emplear a muchas personas."
  - ✓ Las reglas de integridad referencial garantizan valores válidos. Por ejemplo, "Una persona puede existir sin trabajar para una empresa, pero una empresa no puede existir a menos que una persona este empleada por la empresa."

#### Relación entre dos entidades

- Una relación uno-a-uno, dice que la entidad padre puede tener una y sólo una entidad niño.
- Una relación uno-a-muchos, dice que la entidad padre puede tener una o más entidades secundarias. Las relaciones uno-a-muchos son las relaciones más comunes. En algunas relaciones uno-a-muchos, una entidad niño debe tener un padre, pero en otras relaciones, la relación con uno de los padres es opcional. En algunas relaciones uno-a-muchos, una entidad padre debe tener al menos una entidad niño, mientras que en otras relaciones uno-a-muchos, la relación a cualquier niño es opcional.
- Una relación de muchos a muchos, dice que una instancia de cada entidad puede estar asociada con cero a muchas instancias de la otra entidad y viceversa.
- Una relación recursiva relaciona instancias de una entidad a otras instancias de la misma entidad. Relaciones recursivas pueden ser de uno a uno, uno-a-muchos o muchos-a-muchos.

## Desarrollar y Mantener Modelo de Datos Lógicos

- Un modelo de datos lógico es una representación detallada de los requerimientos de datos y las reglas de negocio que rigen la calidad de datos, por lo general en apoyo de un contexto de uso específico (requisitos de la aplicación).
- Los modelos de datos lógicos seguirían siendo independientes de cualquier tecnología o de las limitaciones técnicas de implementación específicas.
- Un modelo de datos lógicos a menudo comienza como una extensión de un modelo conceptual de datos, añadiendo atributos de datos para cada entidad.
- Las organizaciones deben tener estándares de nomenclatura para orientar la asignación de nombres de objetos de datos lógicos. Los modelos de datos lógicos transforman estructuras de modelos de datos conceptuales mediante la aplicación de dos técnicas: la normalización y la abstracción.

#### La Normalización

- La *normalización* es el proceso de aplicación de normas para organizar la complejidad del negocio en estructuras de datos estables
- Niveles de Normalización.
- □Primer forma normal (1NF): Asegura que cada entidad tenga una clave principal válida, cada elemento de datos depende de la clave principal y elimina grupos de repetición y garantiza que cada elemento de datos es atómica (no multivalorado).
- □Segunda forma normal (2NF): Asegura que cada entidad tenga una clave principal mínima y que cada elemento de datos dependa de la clave primaria completa
- ☐ Tercera forma normal (3NF): Asegura que cada entidad no tenga claves primarias ocultas y que cada elemento de datos no dependa de elemento de datos fuera de la clave ("la clave, la clave de todo y nada más que la clave").

#### ... continua

- La forma normal Boyce / Codd (BCNF): Resuelve la superposición de claves candidatas compuestas. Una clave candidata es una clave primaria o bien una clave alternativa. 'Compuesto' significa más de uno (por ejemplo, dos elementos de datos en la clave principal de una entidad) y 'superposición' significa que se hay reglas de negocio ocultas entre las claves.
- La cuarta forma normal (4NF): Resuelve todas las relaciones muchos-a-muchos (y más allá) en pares hasta que no puedan desglosarse en partes más pequeñas.
- La quinta forma normal (5NF): Resuelve dependencias entre las entidades en pares básicos y todas las uniones de dependencia utilizan partes de las claves primarias.
- La sexta forma normal (6NF): Añade objetos temporales a las claves principales, con el fin de permitir la presentación de informes y análisis histórico sobre los plazos.

### Modelo de Datos Lógico: Ejemplo

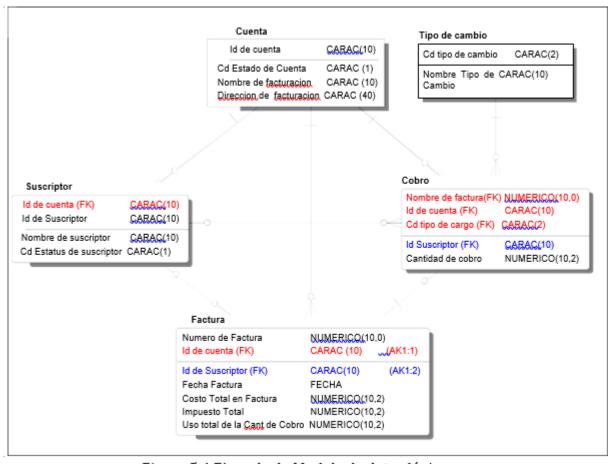


Figura 5.4 Ejemplo de Modelo de datos lógicos

La abstracción es la redefinición de las entidades de datos, elementos y relaciones mediante la eliminación de los detalles para ampliar la aplicabilidad de las estructuras de datos a una clase más amplia de situaciones, a menudo mediante la implementación de súper-tipos en lugar de subtipos

### Atributos y Dominios

- Un atributo es una propiedad de una entidad; un tipo de dato importante para la empresa cuyos valores ayudan a identificar o describir una instancia de entidad.
- El conjunto completo de todos los valores posibles para un atributo es un dominio. Un atributo no puede contener valores fuera de su dominio asignado.
- Algunos dominios tienen un número limitado de valores definidos específicos, o límites mínimos o máximos para los números. Las reglas de negocio también puede restringir los dominios.

#### Claves

- Los atributos asignados a las entidades pueden ser o no claves. Un elemento de datos que es clave ayuda a identificar una instancia única de una entidad de todas los demás ya sea totalmente (por sí mismo) o parcialmente (en combinación con otros elementos clave). Los elementos de datos que no son clave describen la instancia de la entidad, pero no ayudan a identificarlo de forma única.
- Una llave (o clave candidata) representa uno o más atributos cuyos valores identifican de forma exclusiva una instancia de la entidad. Una clave compuesta es una clave que contiene dos o más atributos. Una de estas claves candidatas se convierte en la clave principal. Sólo debe haber una clave primaria. Todas las demás claves candidatas se vuelven claves alternativas.

#### Desarrollar y Mantener Modelos Físicos

- Un modelo de datos físico optimiza la implementación de los requerimientos de datos detallados y las reglas de negocio en vista de las limitaciones de la tecnología, el uso de las aplicaciones, los requisitos de desempeño y estándares de modelado.
- Diseñe las bases de datos relacionales teniendo en mente las capacidades específicas de un sistema de gestión de base de datos específico (IBM DB2 UDB o, Oracle, Teradata, Sybase o Microsoft SQL Server o Access).
- Las organizaciones deben tener estándares de nomenclatura para orientar la asignación de nombres de objetos de datos físicos.

### Modelo Físico: Ejemplo

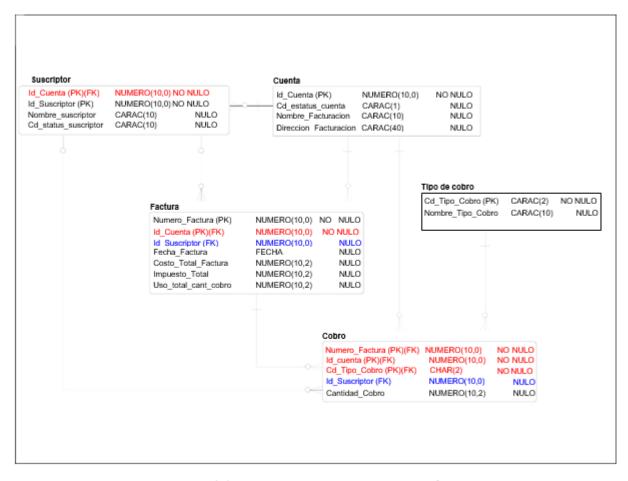


Figura 5.5 Ejemplo de modelo de datos físicos

### Modelo Físico: Implementación

- El diseño del modelo de datos físico incluye tomar decisiones sobre:
  - ✓ El nombre técnico de cada tabla y columna (bases de datos relacionales), o archivo y campo (bases de datos no relacionales), o esquema y elemento (bases de datos XML).
  - ✓ El dominio lógico, tipo de datos físico, longitud y anulabilidad de cada columna o campo.
  - ✓ Cualquier valor predeterminado para las columnas o campos, especialmente para las restricciones NOT NULL.
  - ✓ Las claves primarias y alternativas únicas e índices, incluyendo la forma de asignar las claves.
  - ✓ Implementación de pequeños conjuntos de valores de datos de referencia en el modelo lógico, tales como a) tablas separadas de código, b) una tabla de códigos principal compartida, o c) simplemente como reglas o restricciones.
  - ✓ Implementación entidades del modelo lógico de supertipo / subtipo en el diseño de base de datos físicos donde los atributos de las entidades sub-tipo "se fusionaron en una tabla que representa la entidad supertipo como columnas anulables, o colapsando los atributos de la entidad supertipo en una tabla para cada subtipo.

#### Modelo de Datos Físico: Transformaciones

- El modelado de datos físico transforma el modelo de datos lógico utilizando varias técnicas, incluyendo:
  - ✓ Desnormalización: viola de forma selectiva y con razón, las reglas de normalización, la reintroducción de redundancia en el modelo de datos para reducir el tiempo de recuperación de datos, potencialmente a expensas de ocupar espacio adicional, tiempo adicional de inserción y actualización y reducción de la calidad de datos.
  - ✓ Claves sustitutas: claves suplentes no visibles para el negocio.
  - ✓ Indexación: Crear archivos de índice adicionales para optimizar determinados tipos de consultas.
  - ✓ Partición: Romper una tabla o archivo verticalmente (separando columnas en grupos) u horizontalmente (separando filas en grupos).
  - ✓ Vistas: Tablas virtuales utilizadas para simplificar las consultas, controlar acceso a los datos y cambiar el nombre de las columnas, sin la pérdida de integridad debida a la desnormalización.
  - ✓ **Dimensionalidad**: Creación de tablas de hechos con sus dimensiones asociadas, estructurados como esquemas de estrella y esquemas copo de nieve, para la inteligencia de negocios

#### Diseño de Datos Detallado

• El diseño detallado incluye especificaciones de implementación de base de datos. Un diseño de base de datos física puede tomar ventaja de las funciones y capacidades de un sistema de gestión de base de datos específica, que puede o no estar incluido en el modelo de datos en sí únicas.

### Diseño de Base de Datos Física: Recomendaciones

- 1. Para las bases de datos relacionales que soportan el procesamiento de aplicaciones transacciones (OLTP), utilice un diseño normalizado para promover la integridad de datos, la reutilización, el buen rendimiento de la actualización y extensibilidad de datos.
- 2. Al mismo tiempo, utilizar vistas, funciones y procedimientos almacenados para crear no normalizado, aplicación específica, orientado a objetos, vistas conceptuales (virtuales) de datos. No fuerce a los desarrolladores a trabajar a nivel de base de datos físicos, ni esquemas de bases para las aplicaciones. El objetivo es abstraer la funcionalidad de los datos de su estructura física y que sea lo más fácil posible para trabajar.
- 3. Utilice convenciones estándar de nomenclatura y nombres significativos y descriptivos en todas las bases de datos y objetos de base de datos para facilitar el mantenimiento, sobre todo si las abreviaturas son necesarias.
- 4. Hacer cumplir la seguridad y la integridad de los datos a nivel de base de datos, no en la aplicación. Esto permite la fácil reutilización de los datos, mientras que le ahorra a los desarrolladores el trabajo de tener que escribir y probar las restricciones a nivel de código en cada aplicación que utilice una determinada pieza de datos.

- 5. Trate de mantener el procesamiento de base de datos en el servidor de base de datos tanto como sea posible, para un máximo rendimiento, facilidad de mantenimiento, seguridad, escalabilidad, reducción de tráfico en la red y un menor costo de desarrollo. Por ejemplo, aplicar todas las actualizaciones de bases de datos y consultas SQL complejas como procedimientos almacenados en
- 6. la base de datos, en lugar de incrustar en el código de la aplicación y utilizar los cursores del lado del servidor (en lugar de en el cliente). Usando procedimientos almacenados hace que sea fácil de aislar y corregir los errores y problemas de rendimiento, mejora el rendimiento y reduce una gran medida el tráfico de red.
- 7. Otorgar permisos en objetos de base (tablas, vistas, procedimientos almacenados, funciones, etc.) sólo a grupos o funciones de aplicación, no a individuos. Esto mejora la seguridad y la facilidad de mantenimiento.
- 8. No permita cualquier adaptación directa, ad-hoc de la base de datos; hacer todas las actualizaciones de una manera controlada, a través de procedimientos predefinidos

#### Modificaciones de Rendimiento

- La *indexación* puede mejorar el rendimiento de consulta en muchos casos.
- La *desnormalización* es la transformación deliberada de un modelo lógico de datos normalizado en las tablas con los datos redundantes

#### Técnicas de Des normalización

- Colapsar jerarquías (roll-up): Para reducir las combinaciones, combinar el acceso directo de relaciones padre / hijo en una misma tabla, repitiendo las columnas de los padres en cada fila.
- Divida las jerarquías (push-down): Para reducir conjuntos de consulta, donde las tablas de padres se dividen en múltiples tablas secundarias por tipo.
- División vertical: Para reducir conjuntos de consultas, cree tablas que contienen sub conjunto de columnas.
- División horizontal: Para reducir conjuntos de consultas, crear sub-conjuntos de tablas utilizando el valor de una columna como el diferenciador.
- Combinar y unir tablas: Para reducir uniones donde dos tablas se combinan en un número significativo de consultas, considere la creación de una tabla que ya cuenta con el conjunto de resultados de la combinación de ambas tablas.

#### ... continua

- Repetir las columnas en una fila: Para reducir el recuento de filas o para permitir comparaciones entre las filas, crear una tabla con filas repetidas
- Obtener datos a partir de datos almacenados: Para reducir costo de cálculo en tiempo de consulta, especialmente cálculos que requieren datos de varias tablas, pre-calcular columnas y almacenar los resultados en una tabla ya sea una nueva tabla o uno de las que participa en el cálculo.
- Crear copias de informes: Para mejorar el rendimiento informe, cree una tabla que contiene todos los elementos necesarios para la presentación de informes ya calculado y unidos y actualícelos periódicamente.
- Duplicar (espejos): Para mejorar el rendimiento cuando se utilizan con frecuencia ciertos conjuntos de datos y que se encuentran frecuentemente en disputa, crear versiones duplicadas para grupos de usuarios independientes, o para la carga vs. la consulta.

## Documentación de diseño de base de datos física

- El documento de diseño de base de datos física se orienta a la ejecución y mantenimiento. Se revisa para detectar y corregir los errores en el diseño antes de crear o actualizar la base de datos. Se modifica para facilitar la aplicación de las futuras iteraciones del diseño. Un documento de diseño de base de datos física consta de los siguientes componentes:
  - Una descripción introductoria de la función de negocio del diseño de bases de datos; por ejemplo, ¿qué aspecto o subconjunto de los datos de negocio abarca este diseño de base de datos?
  - ☐ Un modelo gráfico del diseño, realizado en formato ER para un diseño relacional, o un UML para un diseño orientado a objetos.
  - Declaraciones de especificación del lenguaje de base de datos. En Lenguaje de consultas estructurado (SQL), están las especificaciones de Lenguaje de Definición de Datos (DDL) para todos los objetos de base de datos (tabla espacios, tablas, índices, espacios de índices, vistas, secuencias, etc. y XML espacio de nombres).
  - Documentación de los metadatos técnicos, incluyendo el tipo de datos, la longitud de dominio, el origen y uso de cada columna y la estructura de las claves y los índices relativos a cada tabla.
  - Los casos de uso o datos de ejemplo, que muestra cómo se verán los datos reales.

#### Diseño de Productos de Información

- Mientras que el diseño de bases de datos es el objetivo principal de desarrollo de datos, los profesionales de datos también deben participar en el diseño de los entregables de relacionados con los datos.
- Los analistas de datos pueden ayudar a los diseñadores y desarrolladores de software en el diseño de productos de información, incluidas las pantallas e informes, para satisfacer las necesidades de datos empresariales.

#### Servicios de acceso de diseño de datos

- Será a menudo necesario (y deseable) acceder a los datos en bases de datos remotas y combinar esos datos con los datos en la base de datos locales.
- Existen varios mecanismos para hacer esto y el DBA debe estar familiarizado con las fortalezas y debilidades de cada uno.

#### Diseño de integración de datos

- Los analistas de datos y especialistas en integración de datos también definen las asignaciones de origen-destino y los diseños de transformación de datos para los programas de extracción, transformación y carga de datos (ETL) y otras tecnologías para el continuo movimiento de datos, la limpieza y la integración. El DBA puede colaborar en esta actividad de diseño.
- Los analistas de datos, especialistas en integración de datos y administradores de bases también diseñan programas y utilidades para la migración y conversión de datos de las estructuras de datos antiguos a las nuevas estructuras de datos.

### Modelo de Datos y Diseño de Gestión de Calidad

- Analistas y diseñadores de datos actúan como un intermediario entre los consumidores de información (las personas con los requerimientos del negocio para los datos) y los productores de datos que capturan los datos en forma utilizable.
- Los profesionales de datos deben hacer malabares con los requisitos de datos de negocios de los consumidores de información, incluidos los ejecutivos y los requisitos de las aplicaciones de los productores de datos.
- Los requisitos del sistema documentan requisitos de datos de aplicación en forma de casos de uso, un modelo de clase de aplicación y acuerdos de nivel de servicio (SLAs).

## Desarrollar modelado de datos y normas de diseño

- Los estándares de modelado de datos y diseño de bases de datos sirven como principios básicos para responder eficazmente a las necesidades de información de negocios, se ajustan a la arquitectura de datos y garantizar la calidad de los datos.
- Los arquitectos de datos, analistas de datos y administradores de bases de datos deben desarrollar conjuntamente estas normas. Ellos deben complementar y no entrar en conflicto con los estándares de TI relacionados.

## Revisión del modelo de datos y calidad de diseño de base de datos

- Los equipos de proyecto deben llevar a cabo los requisitos y las revisiones de diseño según sea el caso.
- Estas revisiones deben incluir una revisión conceptual del modelo de datos, una revisión lógica del modelo de datos y una revisión del diseño de base de datos física.
- Las revisiones de diseño deben llevarse a cabo con un grupo de expertos en la materia que representan diferentes orígenes, habilidades, expectativas y opiniones.
  - ☐ Revisiones de modelo de datos lógicos y conceptuales
  - □Validación de Modelo de Datos

## Gestionar versiones de modelado de datos e integración

• Los modelos de datos y otras especificaciones de diseño requieren un control cuidadoso de cambio, al igual que las especificaciones de requisitos y demás prestaciones del SDLC. Nota cada cambio a un modelo de datos para preservar de donde vienen los cambios en el tiempo. Si un cambio implica el modelo lógico, tales como el requisito de datos de negocios nuevos o modificados, el analista de datos o arquitecto deben revisar y aprobar el cambio.

#### Control de Cambios

#### Cada cambio debe tener en cuenta:

- ¿Por qué el proyecto o situación requiere el cambio?
- ¿Qué y cómo cambió el objeto (s), incluyendo las tablas en al que se tiene columnas agregadas, modificadas o eliminadas, etc?
- ¿Cuándo se aprobó el cambio y cuando se hizo el cambio en el modelo? Esto no es necesariamente cuando se implementó el cambio en un sistema.
- ¿Quién hizo el cambio?
- ¿Donde se hizo el cambio; para cuales modelos?

#### Implementación de Datos

Implementación de Datos consiste en actividades de gestión de datos que apoyan la construcción del sistema, pruebas y despliegue, incluyendo:

- Aplicación de base de datos y la gestión de cambios en los entornos de desarrollo y pruebas.
- La creación de datos de prueba, incluyendo cualquier procedimiento de seguridad, como la ofuscación.
- Desarrollo de programas de migración de datos y de conversión, tanto para el desarrollo del proyecto a través de la SDLC y para situaciones de negocios como consolidaciones o desinversiones.
- Validación de los requisitos de calidad de datos.
- Creación y entrega de la formación de usuarios.
- Contribución al desarrollo de la documentación eficaz.

## Implementar el Desarrollo / probar los cambios de base de datos

- Los desarrolladores pueden tener la capacidad de crear y actualización de base de datos de objetos directamente, tales como vistas, funciones y procedimientos almacenados y luego notifican a los DBAs y modeladores de datos para la revisión y actualización del modelo de datos.
- El equipo de desarrollo puede tener su propio "desarrollador DBA" que se le da permiso para hacer cambios de esquema, con la condición de que estos cambios sean revisados con el DBA y modelador de datos.
- Los desarrolladores pueden trabajar con los modeladores de datos, que hacen el cambio en la herramienta de modelado de datos y luego generan el "cambio DDL" para los administradores de base de datos lo revisen y lo pongan en práctica.
- Los desarrolladores pueden trabajar con los modeladores de datos, que de forma interactiva "empujan" los cambios en el entorno de desarrollo, utilizando la funcionalidad de la herramienta de modelado de datos, después de la revisión y aprobación por parte de los administradores de bases de datos.

#### Otras Actividades

- ✓ Crear y mantener los datos de prueba
- ✓ Migrar y convertir los datos
- ✓ Construir y probar información de productos
- ✓ Construir y servicios de acceso de datos de prueba
- ✓ Construir y servicios de integración de datos de prueba
- √ Validación de Requerimientos de información
- ✓ Preparación la implementación de datos

#### Principios Rectores

La implementación de la función de desarrollo de datos en una organización sigue nueve principios básicos:

- 1. Las actividades de desarrollo de datos son una parte integral del ciclo de vida del desarrollo de software (SDLC).
- 2. El modelado de datos es una técnica esencial para la gestión eficaz de los datos y el diseño del sistema.
- 3. El modelado de datos conceptual y lógico expresan los requerimientos del negocio y de la aplicación, mientras que el modelado de datos físicos representa diseño de la solución. El modelado de datos y diseño de bases de datos definen las especificaciones de los componentes de solución en detalle.
- 4. El modelado de datos y el diseño de bases de datos equilibra compensaciones y necesidades.

#### ... continua

- Los profesionales de datos deben colaborar con otros miembros del equipo del proyecto para diseñar productos de información y acceso a los datos e interfaces de integración.
- El modelado de datos y diseño de bases de datos deben seguir los estándares documentados.
- 3. Las revisiones de diseño deben revisar todos los modelos de datos y diseños, con el fin de garantizar que cumplan con los requerimientos del negocio y igan las normas de diseño.
- 4. Los modelos de datos representan valiosos recursos de conocimiento (metadatos). Se deben manejar con cuidado y controlarlos a través de la biblioteca, la configuración y gestión del cambio para garantizar la calidad del modelo de datos y la disponibilidad.
- 5. Los administradores de bases de datos (DBAs) y otros profesionales de datos juegan un papel importante en la construcción, prueba y despliegue de bases de datos y sistemas de aplicación relacionados.

¿Alguna pregunta, Aporte o Comentario?

