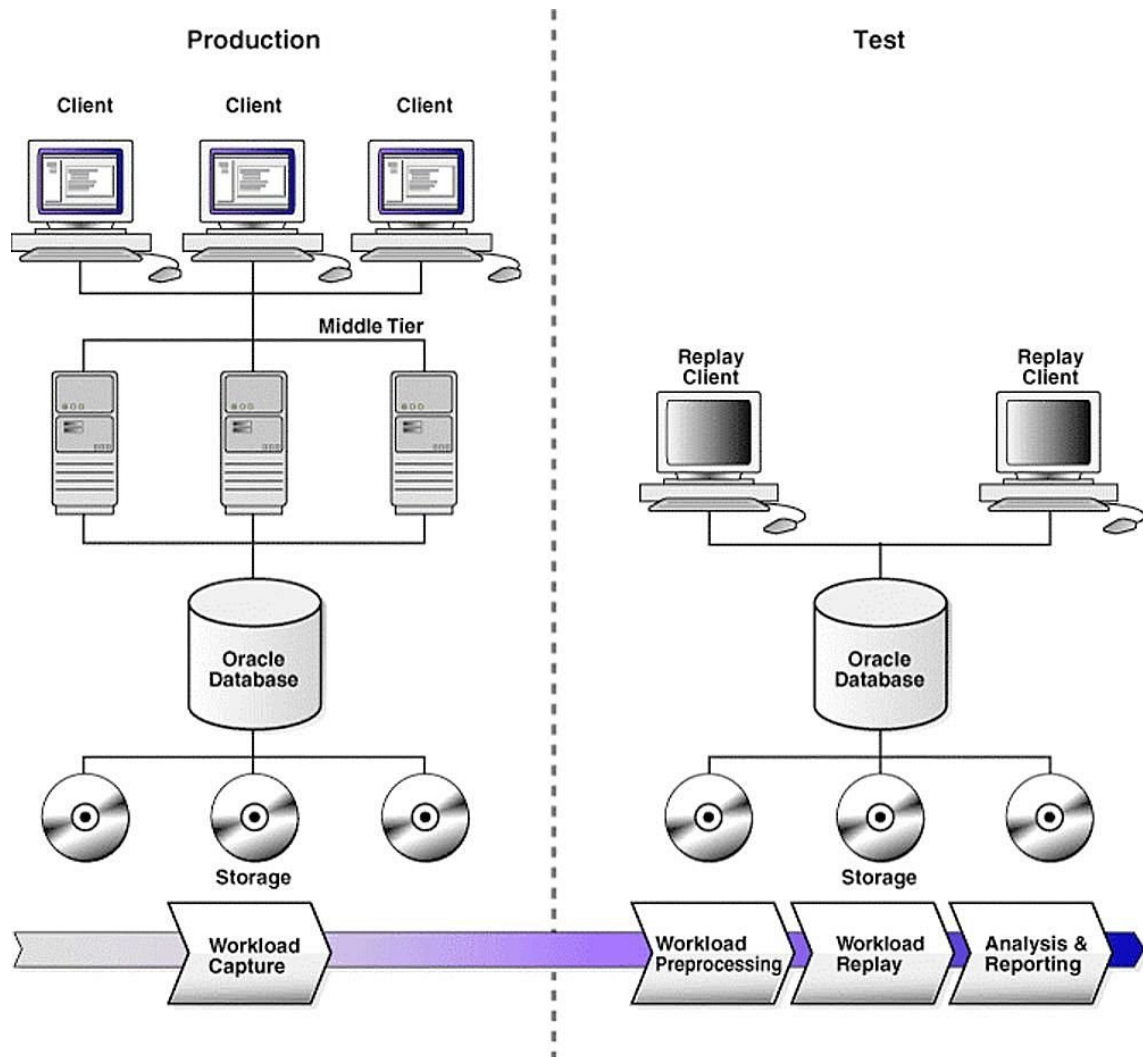


# INGENIRIA DE BASE DE DATOS



## FASE DE EXTRCCIÓN DE DATOS

---

# Jhon Alarcón V.

**Tema: Extracción de las estructuras de datos.**

**Alumno: Jhon Alarcón V.**

**Profesor: Ingeniero Francisco Tixi.**

**Fecha: 16 de Junio del 2010**

## **Introducción.**

La reingeniería es una alternativa útil para la industria del software, ya que se trata de una actividad que permita incrementar la facilidad de mantenimiento, reutilización y evolución de sistemas software. En concreto, su aplicación a las bases de datos es una necesidad que surge muy a menudo.

El proceso de reingeniería de bases de datos, consiste en la recuperación mediante distintos métodos de toda la información de las distintas vistas (física, conceptual y lógica) de la base de datos actual (LDB - Base de datos legada) para en posteriores etapas conseguir modificar y rediseñar el esquema conceptual, transformando la base de datos anterior (LDB) en otra base de datos (NDB – Base de datos nueva). Este proceso conlleva entre otros aspectos de vital importancia, la migración de datos de la LDB a la NDB.

Se trata de una etapa dentro de una metodología general de reingeniería de sistemas de información, para conseguir especificaciones que permitan un alto grado de satisfacción y obtener un sistema de información basado en un sistema anterior.

Nos basamos en la construcción de una herramienta semi-automática, en la que los

usuarios aportan su conocimiento de las bases de datos existentes, así como de mejoras que pretenden realizar. Dicha información es enfrentada con la obtenida en un proceso de ingeniería inversa.

La metodología para Reingeniería de Sistemas facilita la participación de usuarios, para los cuáles estamos diseñando herramientas lo más potentes posibles, pero siempre teniendo como objetivo fundamental la facilidad de uso, para que el usuario no se sienta forzado a participar, sino que se vea como un elemento fundamental en el proceso de reingeniería.

Metodología para la obtención de modelos de datos en sistemas de información  
Dentro de la metodología, se encuentra en primer lugar una etapa de ingeniería inversa, que podemos dividir en dos actividades que se denominan Análisis de datos y Abstracción Conceptual y posteriormente una etapa de reingeniería.

**La primera fase de análisis de datos:** pretende recuperar un esquema de datos lógico completo, que esté lo suficientemente bien documentado y con su semántica correctamente identificada. El esquema lógico que se obtiene en esta fase será la entrada de la siguiente fase, que como hemos comentado anteriormente, se denomina Abstracción conceptual, la cual transformará el esquema lógico a un esquema conceptual equivalente basado en modelo entidad-relación, el cual proporcionará una rica base en la fase de reingeniería de la base de datos (RBD).

Tanto en la etapa de ingeniería inversa como en la etapa de reingeniería, los ingenieros o diseñadores van a contar con una importante ayuda que es la que proporcionará el trabajo cooperativo de los usuarios que, por medio de unas herramientas sencillas, conseguirán acumular más información tanto de la LDB como de la NDB. El trabajo de los usuarios se basará en una utilización de un trabajo automatizado, interactivo e iterativo para de esa forma conseguir un modelo lo más correcto y eficiente posible, de tal forma que no solo se obtendrá un nuevo modelo sino que sería muy eficiente obtener un modelo que permitiera de forma automatizada realizar una migración de los datos.

## **Marco teórico.**

### **Ingeniería inversa de bases de datos.**

El proceso de Ingeniería Inversa sobre las bases de datos, es comúnmente dividido en dos fases claramente diferenciadas:

a) Extracción de las estructuras de datos, obteniendo como resultado el esquema lógico.(Fase I)

---

b) Conceptualización de las estructuras de datos, obteniendo como resultado el esquema conceptual. (Fase II)

### **Fase I.**

En esta fase se va a realizar una extracción de las estructuras existentes actualmente en el sistema de información, dividiéndose en dos etapas de extracción de información.

#### **Etapas 1: Extracción automática.**

Inicialmente debemos extraer mediante herramientas automáticas todas las estructuras de la base de datos cómo fueron diseñadas inicialmente por los desarrolladores. Se trata por tanto de una etapa típica de traducción inmediata del código para así extraer las estructuras de datos explícitas.

#### **Etapas 2: Extracción acumulativa.**

Se trata de una etapa en la que la participación de los usuarios del modelo de datos con el que trabaja supondrá acumular más información de la obtenida en la etapa anterior. Como hemos comentado anteriormente es posible que ciertas reglas no puedan obtenerse directamente en la etapa 1, por lo que aprovechando el conocimiento adquirido por los usuarios en el trabajo diario se podrá obtener información muy interesante.

#### **Así, aspectos importantes, sobre los que los usuarios nos pueden ayudar son:**

**a) Análisis de Nombres:** el usuario hará una descripción de aquellos campos en los que es posible que tengamos dudas acerca de su rol, tipos de datos, relación, etc.

**b) Extracción de claves externas:** sabemos que en la etapa 1, de forma sencilla se pueden obtener las claves principales, pero la obtención de claves externas a veces no es tarea sencilla, y la información aportada por el usuario es vital.

#### **Etapas 3: Unión del Esquema:**

Se trata de una etapa ciertamente compleja, y de la que su éxito depende en gran medida el proceso de reingeniería, pues consiste en unir y reconvertir las estructuras y

restricciones obtenidas en las dos fases anteriores. Así, con la información obtenida en la etapa 2 se pretende complementar la información obtenida en la etapa 1, ya sea para encontrar estructuras no explícitas o que fueron perdidas cuando la base de datos fue diseñada. Para ello se localizarán:

- a) Campos multivaluados.
- b) Tipos registros e identificadores de campos multivaluados.
- c) Campos opcionales.
- d) Claves.
- e) Redundancias.
- f) Dominios.
- g) Significado de los campos.

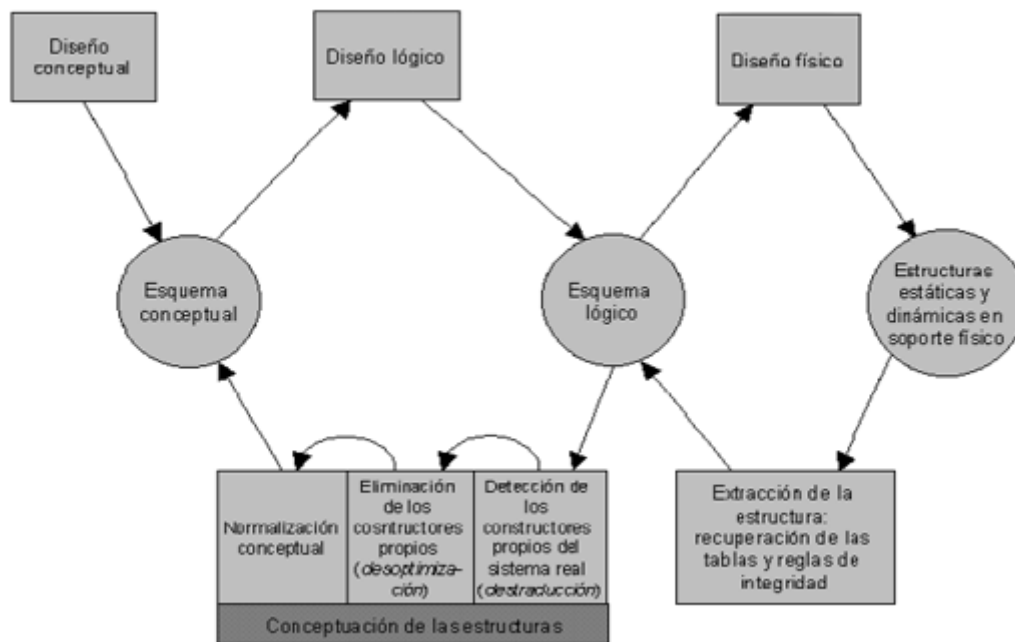
En esta misma etapa y dado que no se les puede dar a los usuarios la responsabilidad de cubrir las posibles lagunas existentes por un mal diseño, es conveniente realizar una etapa posterior para comprobar la corrección de lo realizado en esta etapa.

#### **Etapa 4: Análisis de Programas.**

En esta fase se realiza un estudio del código fuente existente, para comprobar que las restricciones, forma de procesar los datos, significado, etc. se corresponde con el estudio realizado en las fases anteriores. Es posible que esta etapa modifique el resultado obtenido en las etapas anteriores, pero sería deseable que no existieran cambios, pues de esa forma podríamos asegurarnos que lo hecho en las etapas anteriores es correcto y vamos por un buen camino.

En esta misma etapa podríamos incluir el análisis de formularios, consultas, informes, etc. que puedan existir, ya que de esa forma podemos ver si los resultados obtenidos con la LBD coinciden con el modelo extraído.

---



La ingeniería inversa de datos se aplica sobre algún código de bases de datos (aplicación, código SQL, etc.) para obtener los modelos relacionales o sobre el modelo relacional para obtener el diagrama entidad-relación. Hay que tener en cuenta que la ingeniería inversa se puede dar entre distintos productos del ciclo de vida de una aplicación.

Existen muchas técnicas para hacer ingeniería inversa de base de datos, algunos de los cuales se pueden ver resumidos en la siguiente tabla:

- **Baltín et al (1992):**
  - Suposiciones: 3FN. Consistencia en el nombrado de los atributos. Sin homónimos. Clave principal y clave candidata.
  - Entradas: Dependencias. Esquemas de relación.
  - Salidas: Entidades. Relaciones binarias. Categorías. Generalizaciones.

- Basado en el método de Navathe de Awongs.
  
- **Chiang et al (1994,1996):**
  - Suposiciones: 3FN. Consistencia en el nombrado de los atributos. Sin errores en los atributos claves.
  - Entradas: Esquema de relación. Instancia de datos. Dependencias.
  - Salidas: Entidades. Relaciones binarias. Generalizaciones. Agregaciones.
  - Características: Requiere el conocimiento acerca del nombre de los atributos. Propone un marco de trabajo para la evaluación de los métodos de ingeniería inversa. Identifica claramente los casos en los que las entradas de los humanos son necesarias.
  
- **Davids & Arora (1987):**
  - Suposiciones: 3FN. Sin homónimos ni sinónimos.
  - Entradas: Esquemas de relación. Restricciones de claves ajenas.
  - Salidas: Conjunto de entidades. Relaciones binarias. Relaciones n-aria.
  - Características: Apunta a una transformación reversible desde el esquema relacional al esquema conceptual.
  
- **Johannessin (1994):**
  - Suposiciones: 3FN. Consultas de dominio independientes.
  - Entradas: Dependencias funcionales y de inclusión. Esquemas de relación.
  - Salidas: Generalizaciones. Entidades. Relaciones binarias.
  - Características: Basado en los conceptos establecidos de la teoría de bases de datos relacionales. Proceso de mapeo simple y automático.
  
- **Markowitz & Makowsky (1990):**
  - Suposiciones: FN Boyce-Codd.
  - Entradas: Esquemas de relación. Dependencias de claves. Dependencias de integridad referencial.

- Salidas: Entidades. Relaciones binarias. Generalizaciones y agregaciones.
- Características: Requiere todas las dependencias de clave.
  
- **Navathe & Among (1987):**
  - Suposiciones: 3FN y algunos 2FN. Consistencia en el nombrado de atributos. Sin ambigüedades de clave ajena. Claves candidatas específicas.
  - Entradas: Esquemas de relación.
  - Salidas: Entidades. Relaciones binarias. Categorías. Cardinalidades.
  - Características. Es vulnerable a la ambigüedad en el reconocimiento de claves ajenas.
  
- **Petit et al (1996):**
  - Suposiciones: 1FN. Atributos únicos.
  - Entradas: Esquemas de relación. Instancia de datos y código.
  - Salidas: Entidades. Relaciones. Generalizaciones.
  - Características: Analiza las consultas en los programas de aplicación. No pone restricciones en el nombre de los atributos.
  
- **Permerlani & Blaha (1993,1994):**
  - Suposiciones: Sin F3N. Comprensión semántica de aplicación.
  - Entradas: Esquemas de relación. Observaciones de patrones de datos.
  - Salidas: Clases. Asociaciones. Generalizaciones. Multiplicidad. Agregación.
  - Características: Requiere un alto nivel de entrada de los humanos. Enfatiza en el análisis de las claves.
  
- **Sotou (1997,1998):**
  - Suposiciones: Sin nombres únicos de atributos. Dependencias desconocidas.
  - Entradas: Esquema de datos. Instancia de datos. Diccionario de datos.
  - Salidas: Cardinalidad de las restricciones de relaciones n-arias.
  - Características: Proceso automatizado completamente para SQL.

Entre las distintas técnicas de Ingeniería Inversa de datos, se propone el método de Hainaut et al () para explicarla.

El método pasa por dos fases. En la 1ª fase se realiza la extracción de estructuras y en la 2ª la conceptualización de las mismas:

#### **FASE I: Extracción de estructuras**

- Considerar cada fichero una posible tabla.
- Considerar cada campo del fichero como un posible campo de la tabla.
- Identificar las claves primarias.
- Identificar claves ajenas.
- “Filtrar” las tablas (por ejemplo, despreciar aquellos ficheros sin clave principal).
- Detección de campos obligatorios.
- Detección de asociaciones entre tablas

#### **Conclusión.**

La tecnología de información está transformando las actividades económicas y cotidianas como uno de los fenómenos sociológicos más importantes del siglo. Por esta razón, los niveles de oportunidades de trabajo se incrementan de una manera acelerada en diferentes áreas del conocimiento.

Indiscutiblemente, las computadoras han invadido ya todos y cada uno de los campos de la actividad humana:

ciencia, tecnología, arte, educación, recreación, administración, economía y de acuerdo a la tendencia actual, nuestra civilización y las venideras dependerán cada vez más de estos "cerebros" electrónicos.

Se ha venido acelerando la velocidad de cambio del medio de casi todas las organizaciones, de allí que éstas necesiten ahora más información como soporte a la toma de decisiones.

Aunque las entidades de tipo educativo se han descuidado en este aspecto, en estos momentos se percibe un cierto interés en la implantación de estrategias que logren interesar a estudiantes y profesores en el aprendizaje de técnicas que pretenden a corto plazo masificar e implementar el uso de bases de datos, redes de datos e información y

tecnología informática de punta como herramientas básicas de los actuales y nuevos profesionales del país.

Para responder a los retos planteados por la nueva situación económica y tecnológica mundial, se impulsa una dinámica tendiente a dar a conocer los elementos necesarios para estar a la vanguardia en este campo.

### **Bibliografía.**

<http://es-es.facebook.com/topic.php?uid=116578695037086&topic=60>

<http://cnx.org/content/m17435/latest/>

<http://www.docstoc.com/docs/42398701/Aplicacin-de-Tcnicas-de-Ingeniera-Inversa-y-Reingeniera-en>

<http://www.google.com.ec/search?q=Ingeniería+inversa+de+bases+de+datos+fase+1&hl=es&start=10&sa=N>

<http://www.google.com.ec/search?q=Etapa+2:+Extracción+acumulativa.&hl=es&start=10&sa=N>

<http://www.monografias.com/trabajos14/datos/datos.shtml#con>