

# TÉCNICAS DE MIGRACIÓN DE DATOS

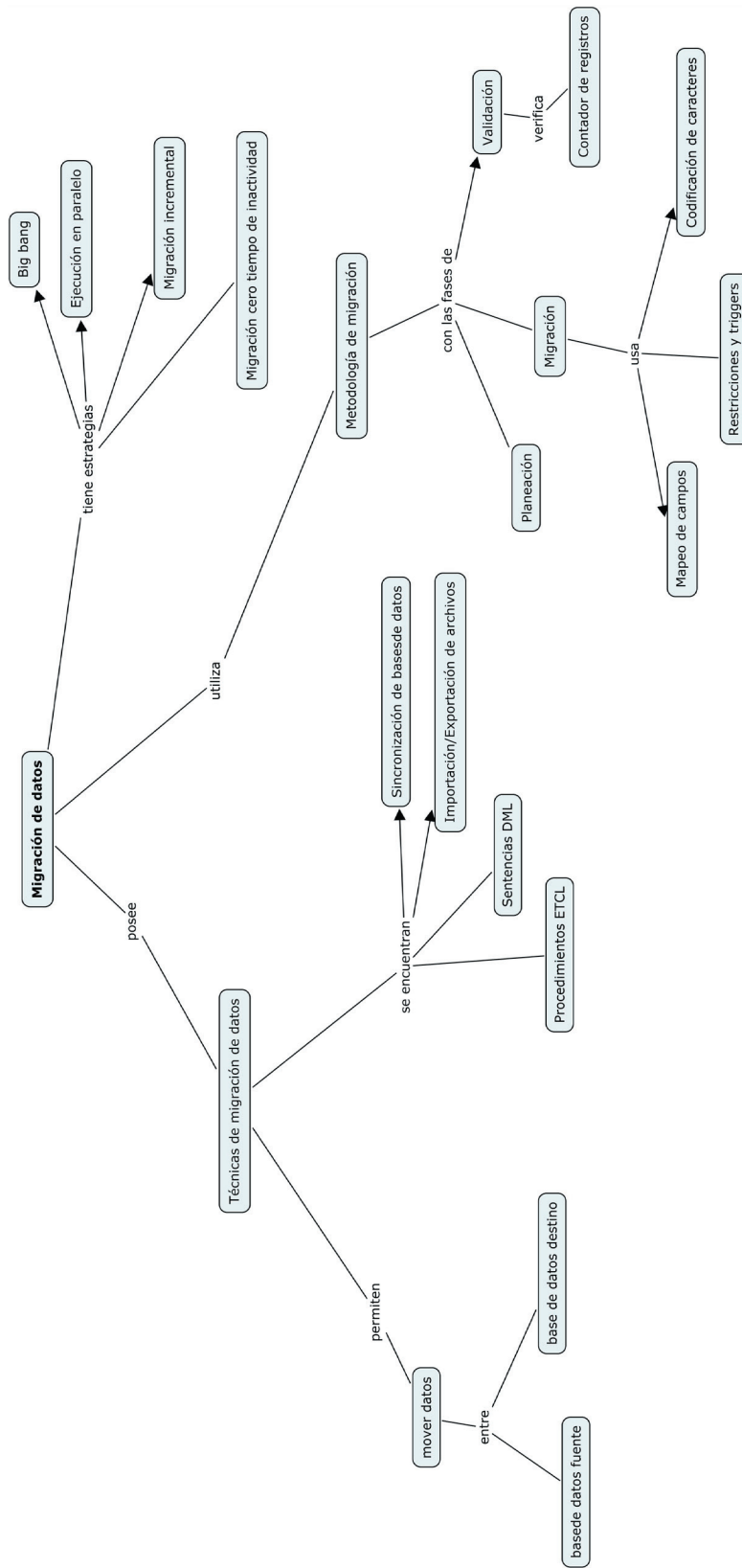
---

## Estructura de contenidos

INTRODUCCIÓN .....	3
1. MIGRACIÓN DE DATOS .....	3
1.1 Razones para realizar la Migración de datos .....	3
2. METODOLOGÍA DE MIGRACIÓN DE DATOS .....	4
2.1 Planeación .....	4
2.2 Migración de datos .....	8
2.3 Validación de datos .....	9
3. TÉCNICAS EXISTENTES .....	9
3.1 Sincronización de base de datos.....	9
3.2 Importación/Exportación de Archivos (CSV) o XML a través de comandos de consola o interfaz gráfica.....	10
3.3 Sentencias DML .....	10
3.4 Procedimientos ETCL .....	11
4. DISEÑO Y DESARROLLO DE PROCEDIMIENTOS ETCL .....	12
4.1 Extracción de los datos .....	12
4.2 Transformación de los Datos .....	13
4.3 Cargue de datos .....	14
5. ESTRATEGIAS DE MIGRACIÓN DE DATOS.....	14
5.1 Big Bang .....	15
5.2 Ejecución en paralelo.....	15
5.3 Migración incremental.....	16
5.4 Migración con cero tiempos de inactividad .....	16
GLOSARIO.....	17
BIBLIOGRAFÍA .....	18



# Mapa conceptual TÉCNICAS DE MIGRACIÓN DE DATOS



# INTRODUCCIÓN

Uno de los procesos básicos en las implementaciones de nuevos sistemas de información es lo referente a las actividades de migración de datos. Es importante establecer las verdaderas razones que se tienen para realizar una migración de datos y documentar los objetivos de la misma.

En la actualidad se está presentado muchas actividades de migración de datos de diferentes plataformas, debido por un lado a los altos costos de manejadores de base de datos privados o la capacidad técnica de otros, así como también la actualización de versiones y la estandarización de tecnología, todo esto llevando a la necesidad de realizar migraciones de bases de datos de los aplicativos más significativos en el mercado.

## 1. MIGRACIÓN DE DATOS

La migración de datos es un proceso en el cual se traslada o se mueve información entre bases de datos. Por ejemplo: en el caso que se necesite pasar de un sistema gestión de base de datos como Mysql a otro sistema similar como SQL Server. Este proceso requiere realizar una serie de actividades técnicas para traspasar los datos almacenados en las base de datos de un sistema a otro.

### 1.1 Razones para realizar la Migración de datos

Entre las principales razones que tenemos para realizar una migración de datos tenemos:

- Nuevos requerimientos por parte de los usuarios
- Fallas en el desempeño del actual SDBD
- Compatibilidad entre aplicaciones
- Actualización de versiones de los aplicativos
- Mejora en el intercambio de información de procesos
- Estandarización de Sistemas de Información
- Aplicación de políticas de seguridad



- Reducción de costos para la utilización de software libre
- Control de información

## 2. METODOLOGÍA DE MIGRACIÓN DE DATOS

Aunque existen varios enfoques para la migración de datos una propuesta metodológica fundamentalmente tendría estos tres procesos o fases que enmarcan buenas prácticas de migración:



### 2.1 Planeación

En la planeación se tienen en cuenta aspectos básicos como: los requerimientos generales de migración, el entorno actual de almacenamiento, el plan de migración, los requerimientos de diseño, la arquitectura de migración y el desarrollo de un plan de prueba.

La planeación debe especificar: cuáles son los objetivos, alcance, estrategias y fases a seguir, requerimientos y riesgos asociados, planes de contingencia y condiciones iniciales o actuales de la migración y proyectar cual será la condición final. También se debe tener en cuenta: La viabilidad técnica y la factibilidad económica de la solución a implementar.

# PLAN DE MIGRACIÓN DE DATOS

ITEMS	RESPONSABLE	ESTADO	FECHA
Establecer equipo de gestión de migración			
Determinar los cronogramas del proyecto			
Documentar los procedimientos de control de cambio para que se incorporen en el plan de migración			
Registrar las actividades y tiempo límites para cambios de hardware y migración de datos			
Informar con anticipación el inicio del proceso de migración usando un tiempo prudencial (aprox. 1 mes antes)			
Recolectar información de servidores de almacenamiento y aplicaciones			
Determinar con los proveedores de nueva tecnología los procesos de almacenamiento			
Constituir un equipo técnico encargado de las actividades de migración			
Implementar los niveles de seguridad del personal a cargo de las actividades técnicas de migración			
Realizar pruebas o actividades antes de la migración con los equipos de trabajo y datos de prueba para hacer verificación antes y posterior de la migración			
Verificar el seguimiento de los procesos de control de cambios			
Ejecutar las actividades técnicas de la migración con los pasos establecidos en el cronograma del proyecto de migración			
Verificar con listas de chequeos que todos los pasos de la migración se realicen de manera exitosa			



# CRONOGRAMA PARA MIGRACIÓN DE DATOS

ACTIVIDADES	FECHA INICIO	FECHA FINALIZACIÓN	RESPONSABLE
Planeación			
Instalación de herramientas			
Configuración de software y hardware			
Copiado de datos			
Pausa en la producción			
Gestión técnica de proveedores			
Verificación técnica de resultados			
Documentación del proceso			

Otro elemento fundamental es el referente a los requerimientos de diseño, entre los cuales tenemos: requerimientos de migración y replicación, cronograma establecido, proveedores participantes y configuración del hardware.

Además es importante determinar el tamaño de los datos participantes en la migración para lo cual hay que considerar elementos como:

- Número de servidores
- Sistemas operativos de los servidores y clientes
- Cantidad de almacenamiento manejada
- Administradores de volúmenes
- Tipos de base de datos
- Aplicaciones
- Velocidades de red
- Clusters de servidores

## LISTAS DE CHEQUEO MIGRACIÓN

ARQUITECTURA SERVIDORES Y CLIENTES	OBSERVACIONES
Proveedor Servidor	
Proveedor Clientes	
Número de CPUs	
Número de dominios y particiones lógicas	
Tipología de sistemas de archivos	
Versiones de Sistemas Operativos	
Tipos y versiones de base de datos participantes en la migración	
Tamaños de base de datos	
Requerimientos de disponibilidad de base de datos (SLA, tiempos de inactividad)	
Clusters existentes	

GESTION DE ALMACENAMIENTO	OBSERVACIONES
Proveedor de almacenamiento	
Tipo y número de canales de datos	
Utilización de RAIDs	
Cantidad de volúmenes a migrar	
Tamaño de los volúmenes	
Volúmenes destino	

REDES (cuando se requiera)	OBSERVACIONES
Topología	
Velocidades	



### 2.2 Migración de datos

La migración de datos es la parte más importante del proceso. Es cuando se realizan las tareas operativas tales como: instalación de herramientas que se requieran, se ejecuta el paso de datos y se hacen pruebas antes de la migración y después de la migración para verificar la consistencia de la información.

Algunas actividades técnicas que se ejecutan en la migración son:

#### 2.2.1 Mapeador de tipo de datos o Mapeo de Campos

El mapeo de campos o de tipo de datos se realiza tanto en la base de datos fuente como la de destino, para asegurar la consistencia de la información.

Para la aplicación de ésta técnica se deben realizar reuniones de levantamiento de información, descripción de los campos del sistema destino, análisis de campos de la base de datos destino que no se pueden mapear con la base de datos fuente.

La utilización de ésta técnica se hace usando el reverse mapping (Mapeo de Destino a Fuente) o direct mapping (Mapeo Fuentes a Destino) para la especificación de datos que requiere el sistema destino para su normal operatividad

#### 2.2.2 Restricciones y triggers

Las restricciones y/o Triggers hacen más complejo el proceso de migración. En muchos casos los proyectos de Migración de datos se realizan entre sistemas fuente y destino con muchos elementos diferentes, lo cual hace más difícil el uso de herramientas y procedimientos ETCL's. Entonces muchos administradores de Base de datos utilizan la técnica de suprimir o inhabilitar temporalmente las restricciones y/o triggers establecidos para poder hacer traslado masivo de datos entre los sistemas a migrar.

#### 2.2.3 Codificación de caracteres

La codificación de caracteres es necesaria en la migración de datos, debido a que en ese proceso caracteres específicos que debía recibir la base de





datos destino resultan siendo reemplazados en los procesos de cargue de información. Entonces se deben realizar actividades de verificación y copiado de caracteres que puedan ser perdidos o modificados en las copias masivas de datos.

### 2.3 Validación de datos

En la validación de datos se verifica la información resultante del proceso de migración de datos, comparándola con los resultados esperados de acuerdo con los sistemas de datos fuente. De igual manera se generan los reportes de estadísticas de la migración para valorar el proceso con base en los resultados y se informa a todos los usuarios del sistema para una verificación final de la consistencia de la información.

Una de las actividades técnicas que se puede utilizar es el contador de registros.

#### 2.3.1 Contador de registros

El contador de registros es una técnica que consiste en utilizar como herramienta una consulta SQL de inserción o un pequeño código de software para registrar la cantidad de registros que se están utilizando en la migración, teniendo como referencia cuántos registros son insertados en forma correcta y cuántos han fallado en su proceso de inserción.

El objetivo del contador de registros es presentar los resultados cuantificados en inserciones exitosas e inserciones fallidas, permitiendo hacer un análisis para determinar si el proceso de migración fue satisfactorio o no.

## 3. TÉCNICAS EXISTENTES

Existen varias técnicas relacionadas con los procesos de migración de datos, a continuación se relacionan las más utilizadas en los procesos actuales.

### 3.1 Sincronización de base de datos

Es una técnica muy utilizada en procesos de migración de datos, consiste en hacer que dos bases de datos sean equivalentes en el mismo tiempo.

Se logra mediante la copia de datos y metadatos de una base de datos origen a una base de datos destino mediante una herramienta tecnológica.

Los asistentes de sincronización de los SMBD son configurados con las bases de datos origen y la base de datos destino, parametrizando la ubicación de particiones y métodos de seguridad.

### 3.2 Importación/Exportación de Archivos (CSV) o XML a través de comandos de consola o interfaz gráfica.

Para las actividades de migración de datos es muy común utilizar archivos para el movimiento de los datos. Se pueden utilizar diferentes tipos de archivos pero los formatos más usados son los archivos de texto, llamados archivos planos, que guardan la información sin formato usando solo caracteres. Los sistemas de codificación más comunes para estos archivos son: ASCII, ISO-8859-1 o Latín-1 y Unicode.

Los archivos de texto pueden tener formato delimitado utilizando coma(,), punto y coma (;) o Tabulaciones para delimitar los campos definiendo de esa forma columnas y filas. También pueden tener formato de ancho fijo para los campos, que se utiliza para delimitar columnas y filas.

Un formato de archivo plano muy usado es el “.CSV” (Comma-Separated-Values, en español Valores Separados por Coma), donde el delimitador usado es una coma. Muy usado para movimientos de datos con hojas de cálculo.

Otro formato útil para la migración de datos es el “.XML” (eXtensible Markup Language), que utiliza un metalenguaje extensible de etiquetas las cuales sirven como estándar para intercambio de información estructurada entre distintas plataformas.

### 3.3 Sentencias DML

Una posibilidad de los SMBD es la utilización de sentencias DML para generar scripts SQL que permiten realizar migraciones de bases de datos existentes.

En un script SQL se pueden realizar las siguientes tareas:

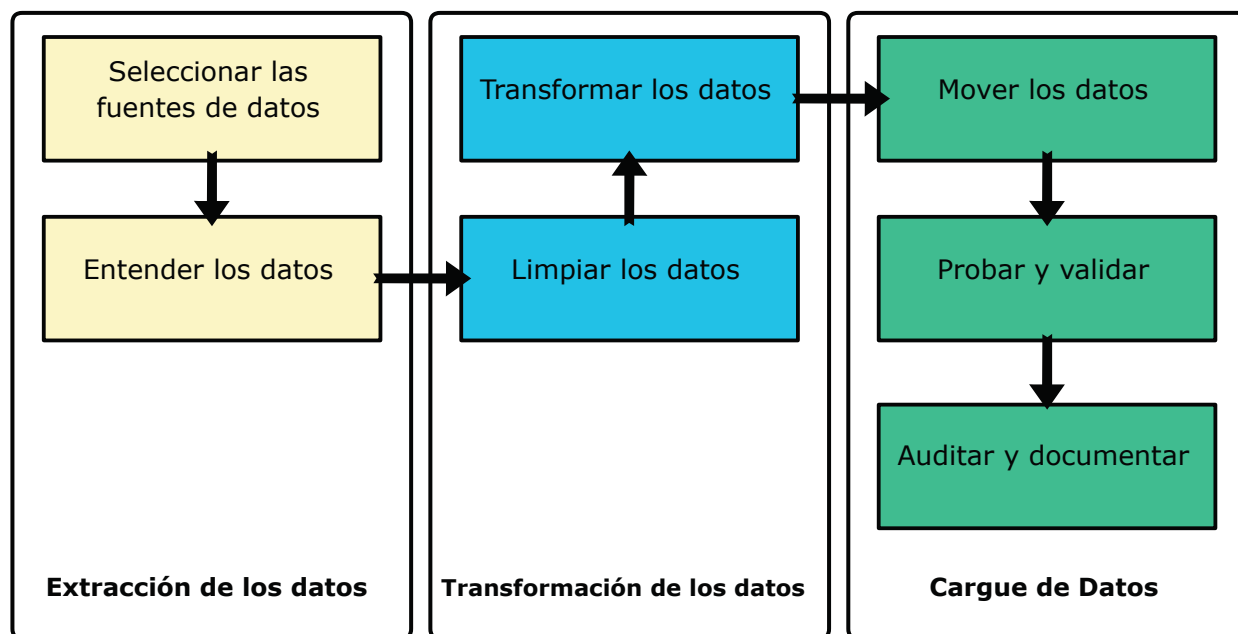
- Copia de seguridad para creación de todos los usuarios, grupos, inicios de sesión y permisos.
- Creación o actualización de código para la implementación de una base de datos.
- Creación de un entorno de pruebas.

Los procesos se orientan al uso de las sentencias DML, para realizar el paso de datos de la base de datos de origen a la base de datos destino, a través de los entornos administrativos de los SMD.

### 3.4 Procedimientos ETCL (Extracción, Transformación, Limpieza y Carga)

Esta técnica se basa en el diseño y construcción de procedimientos técnicos para realizar Extracción, Transformación, Limpieza y Carga de Datos (ETCL).

Los procedimientos o pasos necesarios para el desarrollo de un proyecto de migración de datos utilizando la metodología ETCL son:



### 4. DISEÑO Y DESARROLLO DE PROCEDIMIENTOS ETCL

El diseño y desarrollo de procedimientos ETCL son la esencia en las etapas de un proyecto de migración de datos y se refieren a obtener los datos desde el sistema origen, luego modificarlos de acuerdo con las reglas de integridad y consistencia, hacer depuración general de los datos y posteriormente montar la información en el sistema destino.

El diseño es basado en los resultados del Análisis de la Migración y el uso de un Mapeo detallado. De igual manera se debe realizar una verificación de los requerimientos del cliente para establecer los alcances respectivos y configurar la herramienta técnica de ETCL, contemplando los requerimientos de transformación de datos de acuerdo a lo establecido. Se deben configurar los reportes de conciliación para validar la migración y ejecutar las pruebas unitarias que se requieran.

#### 4.1 Extracción de los datos

En la Extracción de los datos se realizan los siguientes pasos:

##### 4.1.1 Seleccionar las fuentes de datos

En los procesos de negocios, de una empresa que va a realizar migración de su plataforma, los usuarios son los elementos claves para identificar las fuentes verdaderas de datos en cada proceso. El grupo de IT o un departamento de sistemas o de desarrollo de software, debe aportar un conocimiento técnico acerca de la plataforma destino. Por esto una migración requiere tanto del personal del negocio como de los expertos en tecnología que participan.

Una vez identificadas todas las fuentes de datos, se tiene la información para llenar las bases de datos del nuevo sistema.

##### 4.1.2 Entender los datos

Después de identificar las fuentes, se debe tener claridad sobre todos los elementos requeridos para reconocer si existen errores en los datos u omisiones presentadas. Al detectar posibles errores e inconsistencias se puede estimar los tiempos requeridos y costos generados para hacer el análisis de datos.



Un aspecto importante para entender los datos es el nivel de granularidad de los mismos, entre los que están:

- A nivel de campo: si hay correspondencia entre los campos de origen con los de destino, refiriéndose a la consistencia de la información y a los tipos de datos que se manejan.
- A nivel de tablas: que las tablas tengan las coincidencias respectivas, como por ejemplo los casos de llaves primarias y foráneas.
- A nivel de cruce de tablas: cuando existen relaciones entre tablas los procesos de conversión deben mantener la consistencia en las relaciones entre tablas para que la información no se vea afectada en la base de datos destino.

### 4.2 Transformación de los Datos

En este proceso se realizan los pasos de limpieza y transformación de los datos de acuerdo lo presentado en la base de datos de origen.

#### 4.2.1 Limpieza de los datos

Una técnica importante en la migración de datos, es poder hacer limpieza de datos para que en la base de datos destino no llegue información basura o no necesaria.

El proceso contempla realizar varias actividades complejas con el objeto de identificar datos desactualizados, información en blanco o faltante, registros con duplicados, o cualquier otro tipo de información que se deba eliminar o limpiar, de tal modo que se lleven los datos correctos al nuevo sistema o destino.

#### 4.2.2 Transformación de los datos

En este proceso se debe tener en cuenta que tipo de transformación necesitan los datos y cuál es el proceso adecuado para realizarla.

Las transformaciones de los datos se deben poder hacer en doble vía o realizar con las copias de la base de datos origen, de tal modo que si el resultado de la transformación no es consistente se pueda deshacer la transformación y corregir lo necesario de tal modo que se garantice un buen proceso de migración.

Para resultados óptimos es necesario garantizar la calidad en la transformación de datos.

### 4.3 Cargue de datos

En los procesos de cargue de datos se deben realizar actividades para mover, probar, validar y auditar los datos, además de documentar todo el proceso de acuerdo con lo planeado y según los requerimientos estipulados.

#### 4.3.1 Mover los datos

Mover los datos es un proceso que requiere utilizar herramientas técnicas de modo que la información de la base de datos destino sea consistente con respecto a la base de datos fuente.

En este proceso se pueden utilizar: programas o herramientas de migración, códigos desarrollados para la copia de los datos, procedimientos almacenados para la ejecución de la migración u otras técnicas.

#### 4.3.2 Probar y validar

La prueba y validación se realiza de acuerdo con datos de verificación que permiten verificar si los datos en los sistemas de destino concuerdan y operan de la misma manera que los datos de la fuente.

#### 4.3.3 Auditar y documentar

La documentación se realiza de manera específica de acuerdo con los requerimientos pactados y cronogramas establecidos.

Se audita con base en los procesos de los sistemas fuente y se anexan los documentos que describen los procesos con datos validados.

## 5. ESTRATEGIAS DE MIGRACIÓN DE DATOS

Hay varias estrategias de migración de datos que resultan efectivas de acuerdo con el entorno de la empresa o condiciones técnicas de los sistemas. Entre las más conocidas se tienen:

## 5.1 Big Bang

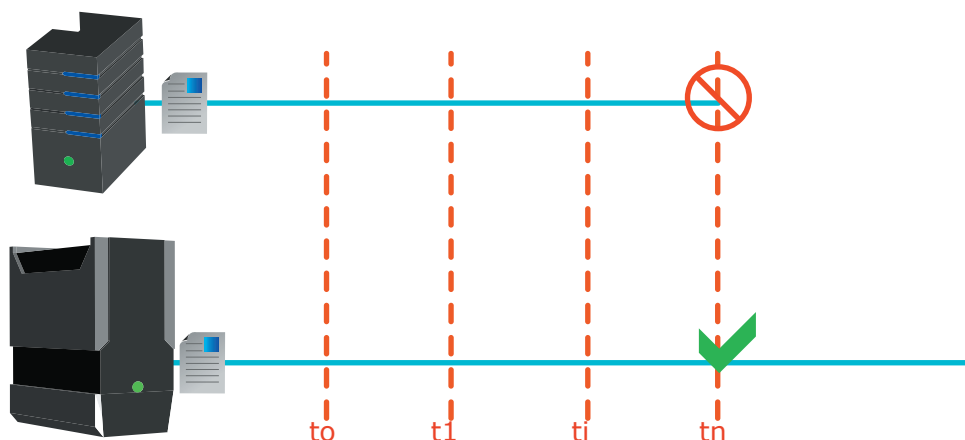
Esta estrategia se basa en seguir usando el sistema actual mientras se va implementando la migración al nuevo sistema durante todo el proyecto planteado, de modo tal que hasta que todos los datos no estén montados en el nuevo sistema el sistema de base de datos anterior no se desactiva.



En esta estrategia se requiere trabajar y hacer pruebas en el nuevo sistema con datos del viejo sistema mientras se realiza toda la migración.

## 5.2 Ejecución en paralelo

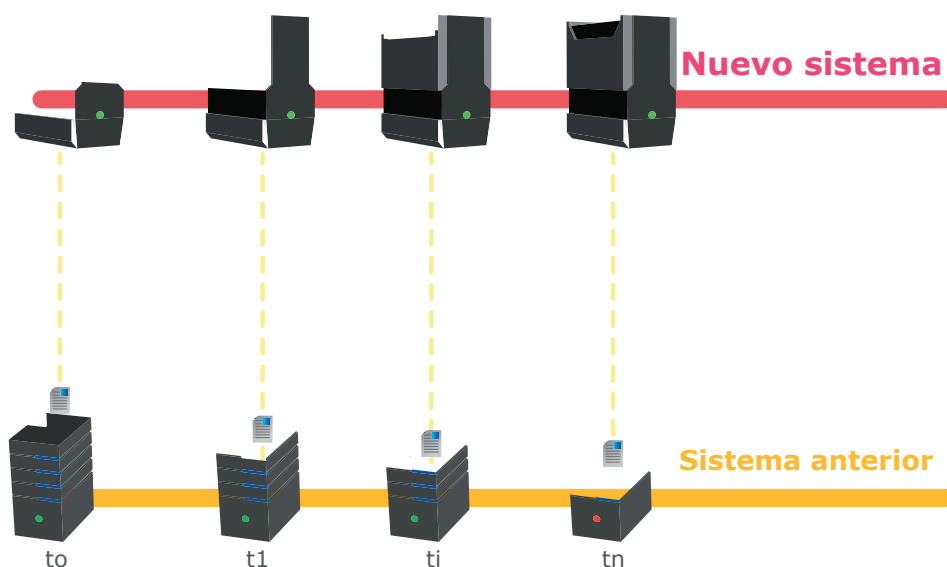
Esta estrategia se basa en validar por un tiempo estipulado los resultados que tiene el nuevo sistema comparado con el anterior, corriendo en forma paralela los dos sistemas.



Es usado en ambientes donde se requiere garantizar el funcionamiento del nuevo sistema en un entorno real, antes de desactivar el viejo sistema. Una desventaja es el costo de tener dos sistemas en funcionamiento, con personal para manejar e ingresar datos a los dos sistemas y el uso de los recursos de hardware que se consumen.

### 5.3 Migración incremental

En esta estrategia el nuevo sistema se activa en forma incremental de acuerdo a como se van haciendo las migraciones. No se espera a que todo el nuevo sistema esté listo para implementar su funcionamiento.



### 5.4 Migración con cero tiempos de inactividad

En otros tipos de proceso, que no se pueden detener o que una falla o retardo en la migración tendría drásticas consecuencias, se recomienda una opción de cero o casi cero tiempo de inactividad.

Por ejemplo, un proyecto de migración que tenga 10 meses de duración que requiere apagar todo el sistema por solo 5 minutos para el cambio de servidores, se puede considerar como una migración con un tiempo muy cercano a cero de acuerdo con el tiempo total del proyecto.

Esta estrategia se puede basar en una combinación de otras estrategias como migraciones incrementales y de ejecución en paralelo.



## GLOSARIO

---

**CAMPO:** Espacio de almacenamiento de dato. Es la mínima unidad de información que se puede acceder en la base de datos.

**CLÚSTER:** En términos de servidores de datos se refiere a un conjunto de servidores que trabajan juntos como un nodo para compartir recursos y aplicaciones de clientes.

**DML:** Data Manipulation Language. Sentencias o conjunto de comandos SQL que se usan para manipular datos con operaciones de consulta, inserción, eliminación o actualización.

**IT:** Tecnologías de la información (Information Technology)

**RAID:** Sistema de almacenamiento con múltiples discos duros que sirven para distribuir o replicar datos. Su sigla en inglés es Redundant Array of Inexpensive Disk.

**SMBD:** Sistema Manejador de Base de Datos. Es un software que tiene como objetivo facilitar la construcción y manipulación de bases de datos sirviendo como interfaz entre éstas, los usuarios y los distintos tipos de aplicaciones que las utilizan.



## BIBLIOGRAFÍA

---

Date, C. (2001). Introducción a los Sistemas de Base de Datos. Prentice Hall.

Piattini, Miguel (2006). Diseño de base de datos relacionales. Alfaomega Ra-Ma

Peter Eisentraut, Bernd Helmle (Octubre 2008). PostgreSQL-Administration

Howard Philip (2007, Diciembre). Data Migration. A White paper by Bloor Research. Consultado el 13 de Marzo de 2012, de <http://www.consultor-it.com/articulo/50308/mejores-practicas-para-la-migracion-de-datos-por-pervasive-software>

IBM Corporation (2007, Junio). Best practices for data migration. Methodologies for planning, designing, migrating and validating data migration. Consultado el 14 de Marzo de 2012 de <http://www-935.ibm.com/services/us/gts/pdf/softekt-best-practices-data-migration.pdf>

Wikipedia (2012, Febreo). Migración de datos. Consultado el 14 de Marzo de 2012 de <http://es.scribd.com/doc/51334306/Migracion-de-datos>

Corona Correa Susana. Factores críticos de éxito en el proceso de migración de base de datos relacionales. Consultado el 12 de Marzo en [http://www.paginaspersonales.unam.mx/files/35/Susana\\_Corona\\_Correa\\_Mexico\\_.pdf](http://www.paginaspersonales.unam.mx/files/35/Susana_Corona_Correa_Mexico_.pdf)



<b>Control de documento</b> <b>Construcción Objeto de Aprendizaje</b> <b>Técnicas de Migración de datos</b>	
Desarrollador de contenido Experto temático	Manuel G. Hormechea Lance
Asesor pedagógico	Rafael Neftalí Lizcano Reyes
Producción Multimedia	Luis Fernando Botero Mendoza Victor Hugo Tabares
Programadores	Daniel Eduardo Martínez Grateful Dead Montaña Sierra
Líder expertos temáticos	Ana Yaqueline Chavarro Parra
Líder línea de producción	Santiago Lozada Garcés

