GUÍA PARA LA MIGRACIÓN DE BASES DE DATOS

MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL
BOGOTÁ, DICIEMBRE DE 2016

TABLA DE CONTENIDO

1. OBJETIVO 3
1. ALCANCE 3
2. ÁMBITO DE APLICACIÓN 3
3. DOCUMENTOS DEL SIG ASOCIADOS A LA GUÍA 3
4. NORMATIVA Y OTROS DOCUMENTOS EXTERNOS 3
5. DEFINICIONES 3
6. METODOLOGÍA Y BUENAS PRÁCTICAS PARA MIGRACION DE BASES DE DATOS 5
7. ETAPAS A SEGUIR EN EL PROYECTO DE MIGRACIÓN DE BASE DE DATOS 10
8. BIBLIOGRAFÍA 14

1. OBJETIVO

Orientar el desarrollo de las actividades en proyectos de migración de bases de datos teniendo en cuenta las buenas prácticas descritas para estos proyectos.

1. ALCANCE

Describir el proceso metodológico a ser usado para la migración de bases de datos de acuerdo a la experiencia obtenida por la Oficina TIC del Ministerio de Salud y Protección Social.

2. ÁMBITO DE APLICACIÓN

Aplica a los procesos misionales en sus sistemas de información.

3. DOCUMENTOS DEL SIG ASOCIADOS A LA GUÍA

- Proceso CVSC01 Ciclo de Vida y Reingeniería de Sistemas de Información
- Guía CVSG01 Guía para el Ciclo de Vida de un Sistema de Información

4. NORMATIVA Y OTROS DOCUMENTOS EXTERNOS

- Decreto 4107 de 2011: ARTÍCULO 10. FUNCIONES DE LA OFICINA DE TECNOLOGÍA DE LA INFORMACIÓN Y LA COMUNICACIÓN - TIC.
- Documento externo CVSX01 Lineamientos del Manual de la política de Gobierno Digital en el marco de referencia de Arquitectura TI.
- ISO 27001:2013: estándar para la seguridad de la información (Information technology - Security techniques - Information security management systems - Requirements).

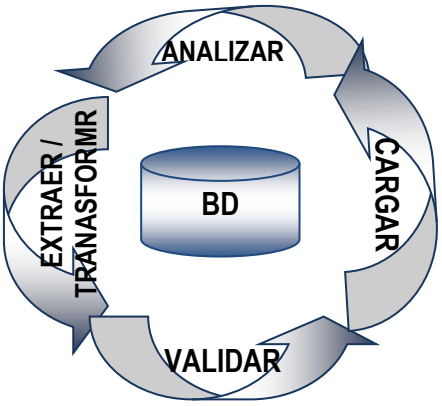
5. DEFINICIONES

- **DBA**: Data Base Administrator. Administrador de base de datos
- **DTS**: Data Transformation Services. Servicios de Transformación de Datos.
- **ETL**: Extract, Transform, Load. Extraer, Transformar, Cargar.

- **INTERCALACIÓN (Collation)**: hace referencia al patrón de bits utilizado para representar/almacenar cada carácter, y en consecuencia también se refiere a las reglas utilizadas para ordenar y comparar caracteres lo cual afecta campos de tipo texto.
- **Migración de Base de Datos:** es un proceso que se realiza para mover o trasladar los datos almacenados en una base de datos de origen a otra. Consiste en la extracción, transformación y carga de datos (ETL). Dicha migración se requiere llevar a cabo cuando es necesario mover un esquema dentro del mismo servidor, o de un servidor a otro, así como para actualizar la versión del software, y hacer un cambio de manejador de bases de datos por el de otro fabricante o para cambiarlo a una plataforma de cómputo distinta. Se considera importante tener presente: la extracción, limpieza (mediante ciclos de depuración), homologación, completitud de los datos, conversión, mapeo y carga de datos hacia la fuente destino.
- **RDBMS**: Relational Data Base Management System. Sistema de gestión de bases de datos relacionales.
- **Tuning:** Afinar la configuración de hardware y software para optimizar su rendimiento
- **Script**: conjunto de instrucciones que se encuentran escritos en un formato regular y se almacenan en un archivo de texto plano, los cuales se ejecutan para lograr alcanzar un objetivo especifico.

6. METODOLOGÍA Y BUENAS PRÁCTICAS PARA MIGRACION DE BASES DE DATOS

El proceso para la migración de bases de datos incluye cuatro pasos fundamentales, a saber:



Fuente: Autores OTIC

Analizar: En esta fase se contempla:

- Levantamiento de Información: corresponde a la definición de los datos que requiere el nuevo sistema y la identificación de la fuente de origen.
- Análisis de Calidad: Se verifica la calidad de los datos, con el propósito de establecer los tiempos reales de las etapas de Transformación, Limpieza y Mapeo de Datos. Se requiere detectar si existen datos erróneos, si un mismo concepto se encuentra definido con distintos términos, o si faltan datos.
- Identificar el total de datos a migrar.
- Determinar cuáles son las reglas de negocio que le corresponden
- Establecer las transformaciones requeridas
- Seleccionar herramientas de migración ó si el proceso se va a llevar a cabo de forma manual.

Es importante señalar, que realizar el Análisis de calidad, no necesariamente se debe suscribir a un proceso de migración, ya que puede darse en otros momentos para las necesidades de las organizaciones. Si se efectúa, se debería contemplar:

- "Limpieza de datos: momento en el que corrigen los datos que se consideran incorrectos o inconsistentes, eliminando errores, duplicados o inconsistencias detectadas.
- Homologación: implica la unificación de criterios, códigos, descripciones, etc.
- Enriquecimiento: se lleva a cabo al complementar y perfeccionar los datos maestros, para garantizar su consistencia, integridad y fiabilidad." [1]

[1] Introducción a la Migración de Datos. PowerData. Especialistas en Gestión de datos. Consultado Abril de 2014.

Extraer / Transformar: En esta fase se deben construir los scripts, programas ó seleccionar el software de ETL que apoyará el proceso de migración e integración de datos, teniendo en cuenta que deben manejar las irregularidades de datos y excepciones durante la transformación, tales como incompatibilidad de los tipos de datos.

Es necesario preparar los datos para que sean cargados en la nueva estructura y para ello hay que realizar las conversiones necesarias en base a las reglas de negocio identificadas y a la configuración de los sistemas actual y futuro.

Idealmente la herramienta debe generar un reporte de violaciones de integridad e irregularidades, ya que los datos de la fuente deberán ser corregidos por los usuarios finales o los administradores de tales sistemas.

Validar: En esta etapa se debe asegurar la validez, calidad, operatividad y controlar la ejecución del proceso de migración en los distintos entornos que se vean afectados, asegurando una carga correcta de datos.

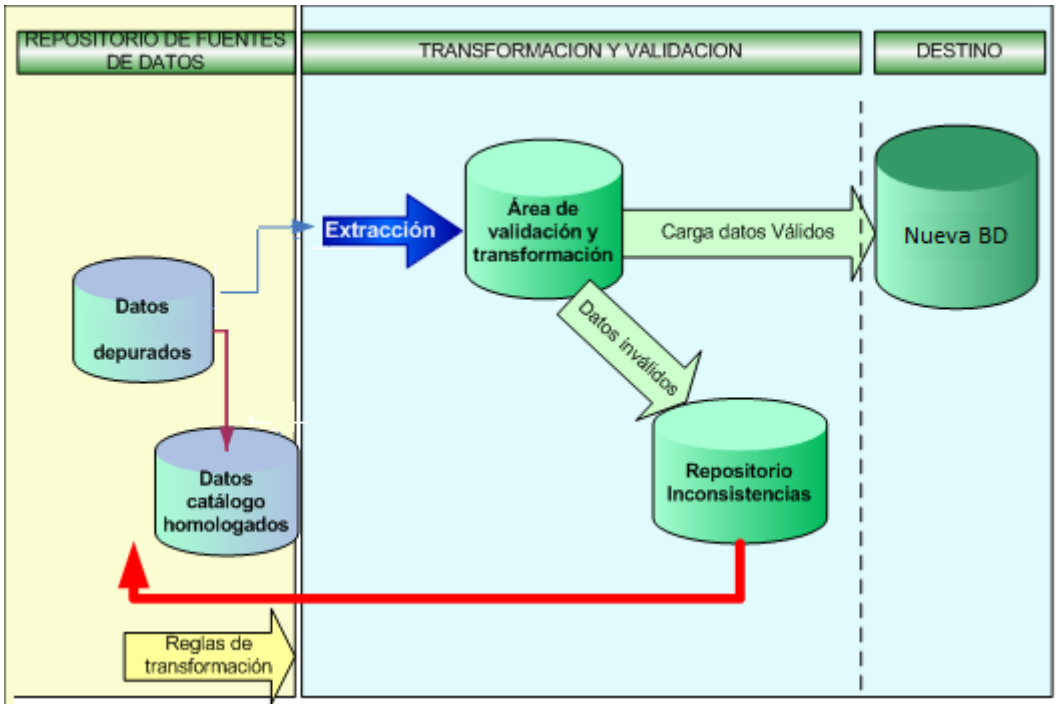
- Se recomienda emplear datos reales, ya que los datos provenientes de producción superan en fiabilidad y realismo a los datos simulados
- Realizar simulaciones de Cargas: para identificar los inconvenientes que se pueden tener en las cargas reales y proceder a ajustar.
- Realizar pruebas funcionales para verificar la integridad de los datos migrados a través de validaciones realizadas a la medida y comprobar que estos datos son correctos desde el punto de vista funcional.

Cargar: En esta etapa se procede a realizar el cargue de los datos ya limpios y transformados al nuevo sistema. Este proceso se puede hacer directamente o empleando áreas intermedias (repositorios) donde se pueda validar y someter a los datos a ciclos de pruebas y simulaciones de carga, de tal manera que se aumenten las garantías de éxito en el momento de ejecutar la carga al destino final.[2]

Una vez realizada la carga final, se recomienda efectuar una revisión final de la información cargada al nuevo sistema.

Entre las **Buenas prácticas** para los procesos de migración de bases de datos tenemos:

[2] Ibid.



Fuente: Migración Base de datos Oracle http://www.geocities.ws/rescala29/fase3/t4/t4_teoria.html

- Tres repositorios: i) Uno para las fuentes de datos, ii) un segundo para el proceso de extracción de datos en el cual se generan tanto los datos válidos como los datos inválidos y iii) un tercero para la base de datos nueva.

- El proceso debe iniciarse realizando respaldos de la base de datos antes y después del proceso de migración; El primer respaldo asegura que si en el proceso de migración, el sistema de base de datos o el servidor fallan la información se puede recuperar y continuar operando tal y como se hacía antes de iniciar el proceso.

  Un respaldo al finalizar el proceso, sirve para asegurar que exista el modo de recuperar la información ya migrada, en caso de que exista alguna falla con la base de datos cuando ya se encuentra en operación con el nuevo modelo.

- Es necesario considerar en qué momento y durante cuánto tiempo se va a detener la operación de la base de datos en producción. Si esto no es posible, se debe determinar el procedimiento para identificar los datos que fueron ingresados durante el proceso de migración, para su actualización posterior.

- Se debe contar con los recursos de infraestructura necesarios bajo el dimensionamiento realizado para minimizar el riesgo en la operación. Igualmente es necesario considerar efectuar ajuste de la base de datos (tuning).

- "Contar con un área de staging del sistema origen y un área de pre-carga: proporciona la capacidad de extraer los datos de los sistemas de producción y utilizarlos para las actividades relacionadas con procesos de calidad de datos (limpieza, homologación y enriquecimiento). Contener la información en estas áreas, permite ver los datos reales de producción sin afectar a la operación de los sistemas actuales.

- Disponer de un área intermedia con una estructura igual a la del origen para realizar la carga de los datos de producción y poder llevar a cabo las tres **fases que implica el proceso de calidad de datos (limpieza, homologación y enriquecimiento**): este **repositorio** proporciona la capacidad de **extraer los datos de los sistemas de producción y** cualificarlos, estando en continuo contacto con los datos reales de producción pero sin afectar a la operación de los sistemas actuales.

- Contar con un área de pre-carga: una práctica muy recomendable para toda migración de datos ya que esta área presenta una estructura similar a la de destino y es en ella donde se realizará la carga de los datos en el momento en que finalice su conversión. El área de pre - carga permite practicar las validaciones aplicando varios ciclos sin afectar el sistema destino.

- Utilizar tablas de referencias y archivos de parámetros: en el transcurso de la migración puede surgir la necesidad de cambiar un código, por ejemplo, y es en casos como éste en los que la utilización de tablas de referencia y parámetros asegurará una mayor flexibilidad y un menor mantenimiento, ya que los códigos en origen siempre deben corresponderse con uno en destino, dentro de unos parámetros configurables que el desarrollador debe conseguir.

- Diseño modular: hace referencia a la necesaria definición de estándares para el desarrollo de múltiples procesos simples, en vez de uno único más complejo y de mayores dimensiones. Actuar de este modo mejora el desarrollo, aumenta la eficiencia en las pruebas y facilita el mantenimiento, procurando mayor agilidad y simplicidad en el desarrollo, así como unos menores costes, junto con un incremento en la eficacia general del proceso de migración de datos.

- Reutilización: implica reutilizar las asignaciones / estrategias / procesos / scripts / documentos de prueba siempre y cuando sea posible, en procesos de extracción, transformación y/o carga; para optimizar el tiempo de proceso y armonizar resultados.[3]

- Si se lleva a cabo la elección de una herramienta de ETL - se recomienda tomar en cuenta las consideraciones dada por el cuadrante de Gartner en cuanto a las cualidades que la herramienta debe tener:
    - Adaptabilidad a distintos sistemas operativos y plataformas de hardware
    - Capacidades de gestión de datos

[3] ibid

- Habilidades para monitoreo, gestión y control de procesos de integración de datos, en lo relativo a gestión de errores, recolección de estadísticas de ejecución controles de seguridad, entre otros.
- Capacidades SOA
- Capacidades de diseño
- Capacidades de diseño y entorno de desarrollo: como soporte para test, representaciones gráficas, modelos y flujos de datos, etc.

- Ya para el proceso de carga, se requiere deshabilitar los Triggers y/o restricciones que puedan generar errores al momento que el DBMS ejecute el proceso de escritura de los datos.

- Codificación de Caracteres, cuando el copiado se realiza de forma automática, es necesario identificar la codificación de caracteres que la BD destino espera, pues así se evita el reemplazo automático de caracteres o en su caso, pérdida de los mismos.

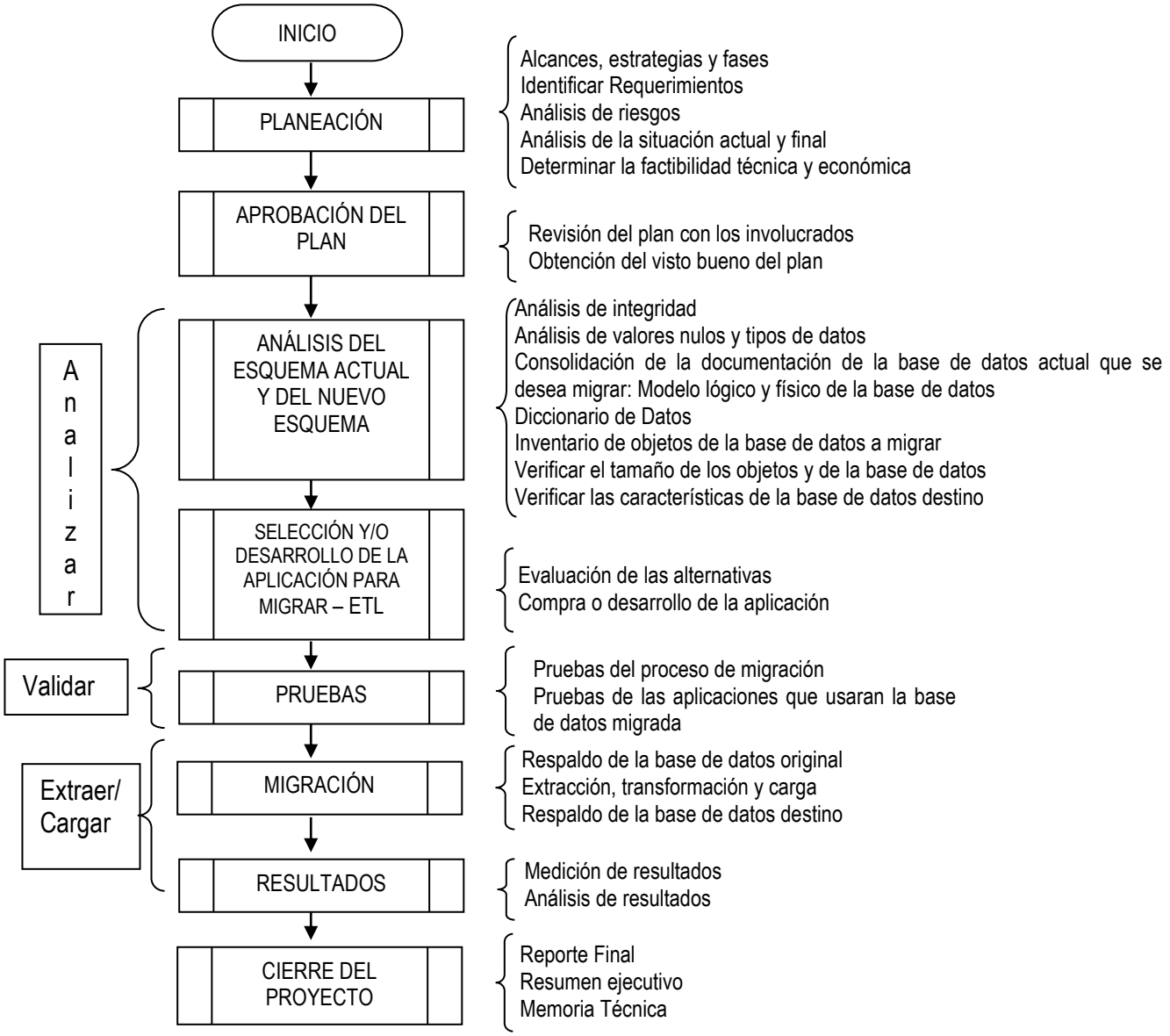
- Generar un reporte de violaciones de integridad e irregularidades, ya que los datos de la fuente deberán ser corregidos por los usuarios finales o los administradores de tales sistemas.

- Los proyectos de migración de datos deben contar con especificaciones de mapeo columna a columna, deben incluir las reglas del negocio, de limpieza de datos, las validaciones, etc.

- Estos proyectos requieren definir tiempo y el perfil de las personas idóneas para el desarrollo del mismo y que tengan conocimiento del negocio. Es recomendable contar con el apoyo de personas con experiencia en este tipo de proyectos. Entre los perfiles a considerar se encuentran:

  - Especialistas de los sistemas de información
  - Administrador de las bases de datos
  - Usuarios que trabajen el sistema
  - Líder funcional que apoyará la definición de las soluciones en el proceso de migración como también en las pruebas a realizar al sistema.

7. ETAPAS A SEGUIR EN EL PROYECTO DE MIGRACIÓN DE BASE DE DATOS



Planeación
Previo a cualquier etapa se debe identificar la causa de la migración que puede ser debido a diferentes factores tales como: i) Cambios de plataforma, ii) Cambios en la versión del motor del manejador de bases de datos, iii) Actualización del software, iv) estandarización a nivel organizacional, v) nuevos procesos de la operación o vi) por una necesidad de un usuario final. Lo anterior permite **"Establecer si existe la Necesidad"** de realizar una migración de bases de datos.

Se lleva a cabo durante esta fase la elaboración del documento: **Plan de Migración**, mediante el cual se debe describir el proceso de movimiento de datos desde los sistemas actuales hacia el nuevo Sistema de Información y debe contemplar los siguientes ítems:

- Necesidad de la migración. Dejar documentado el primer paso de la metodología.
- Alcances
- Estrategias y fases según la magnitud de la información que se va a migrar
- Identificar Requerimientos de la migración
- Definición de las reglas de aceptación de la migración
- Establecer los factores críticos a tener en cuenta y realizar análisis de los riesgos.
- Análisis de la situación actual y final basado en los documentos que se realizaron al establecer la necesidad
- Determinar la factibilidad técnica y económica
- Ventajas y justificación de la migración
- Grupo de Trabajo y responsables
- Cronograma de Trabajo

Aprobación del Plan de Migración
Se debe validar tanto por la parte funcional como la parte técnica. Una vez sea aprobado se genera acta de aprobación del Plan de Migración.

Análisis del Esquema Actual y del Nuevo Esquema:
Se debe realizar un análisis de la migración a ser ejecutada, la cual debe quedar en un documento de análisis de migración que contenga lo siguiente:

a) Análisis de integridad
b) Análisis de valores nulos y tipos de datos
c) Verificar el tamaño de los objetos y de la base de datos
d) Verificar las características de la base de datos destino
e) Consolidación de la documentación de la base de datos actual que se desea migrar:
    - Modelo lógico y físico de la base de datos
    - Diccionario de Datos
    - Inventario de objetos de la base de datos a migrar:
        1. Tablas
        2. Índices

3. Vistas
4. Triggers
5. Procedimientos almacenados
6. Funciones
7. Idioma.
8. Formato de fecha y hora
9. Formato de moneda

f) Elaborar documento de usuarios, permisos de perfiles con acceso a la base de datos actual y a la base de datos destino
g) Documentar la fuente y entradas de información a la Base de Datos
h) Documentar las salidas de información de la Base de Datos
i) Definir el usuario y rol que controla la calidad de los datos de la base de datos actual y la base de datos destino
j) Definir, diseñar y documentar el modelo de Transporte y Transformación.
k) Documentación de la base de datos destino:
   - Modelo lógico y físico de la base de datos, si es igual al origen aclararlo o documentar las diferencias
   - Diccionario de Datos si el igual al origen aclararlo o documentar las diferencias
   - Inventario de objetos de la base de datos destino si el igual al origen aclararlo o documentar las diferencias a nivel del Inventario de objetos de la base de datos que a migrar
     1. Tablas
     2. Índices
     3. Vistas
     4. Triggers
     5. Procedimientos almacenados
     6. Funciones
     7. Idioma (COLLATE).
     8. Formato de fecha y hora
     9. Formato de moneda
l) Requisitos previos de la migración, tareas previas a la migración, tareas posteriores a la migración.
m) Definir, diseñar y documentar el procedimiento de verificación y aceptación de la migración y el plan de pruebas.

ETL - Extracción – Transformación – Carga

Corresponde a la generación del proceso de extracción de datos. Se requiere:

- Configurar los usuarios, permisos de perfiles con acceso a la base de datos destino
- Configurar el usuario y rol que controla la calidad de los datos de la base de datos destino
- Desarrollar el modelo de Transporte y Transformación.
- Desarrollar el procedimiento de verificación y aceptación de la migración

Pruebas:
Creación de un ambiente de pruebas, donde se ejecutaran los artefactos de migración y validación de la misma.

- Ejecutar el modelo de Transporte y Transformación
- Ejecución del plan de pruebas
- Ejecución del procedimiento de verificación y aceptación.

Migración
Corresponde a la ejecución de la migración en el ambiente de producción:

- Primero, suspender la entrada de datos a la base de datos actual.
- Ejecutar el modelo de Transporte y Transformación
- Ejecución del procedimiento de verificación y aceptación
- Inactivar el acceso a la Base de datos origen
- Activar el acceso a la base de datos destino
- Última actividad, habilitar la entrada de datos a la base de datos destino.

Resultados y cierre de la Migración
Una vez terminado el proceso se deben medir los resultados y entregar un reporte global del trabajo realizado, mencionando cuáles son los productos que se entregan, cuántas tablas u otros objetos fueron migrados, cuántos registros se migraron exitosamente, cuántos no fueron migrados y cuál fue la causa de ello

- Elaborar informe final de migración, donde se brinda el detalle de la cantidad de registros por cada objeto de tipo tabla, así como el detalle de los objetos debidamente migrados.
- Entrega de la documentación final relacionada con el proceso de migración.
- Elaboración de acta de cierre de la migración, debidamente firmada por el equipo de trabajo involucrado en el proceso de migración.

8. BIBLIOGRAFÍA

http://es.wikipedia.org/wiki/Migraci%C3%B3n_de_datos

http://www.oracle.com/technetwork/es/articles/sql/migrar-base-de-datos-a-asm-1891814-esa.html

http://ccdoc-automatizacion.blogspot.com/2008/03/10-la-migracin-de-datos-exportacin-e.html

http://www.informatica.com/es/solutions/enterprise-data-integration-and-management/data-migration/

http://www.slideshare.net/carlosgruiz.arahat/mejores-prcticas-para-migracin-de-bases-de-datos

http://www.monografias.com/trabajos27/bases-datos/bases-datos.shtml

http://hexaware.mx/migracion-de-datos.htm

Introducción a la Migración de Datos. PowerData. Especialistas en Gestión de datos.

ELABORADO POR:	REVISADO POR:	APROBADO POR:
Nombre y Cargo: Marcela Mosquera, Profesional de la Oficina de TIC. **Fecha:** *30 de agosto de 2016*	**Nombre y Cargo: Herly Suárez,** Contratista de la Oficina de TIC. **Fecha:** *02 de diciembre de 2016*	**Nombre y Cargo: Dolly Esperanza Ovalle Carranza,** Jefe de la Oficina de TIC. **Fecha:** *02 de diciembre de 2016*