

Metodología para el desarrollo de un sistema de migración de la base de datos del Instituto Tecnológico de Acapulco a formato XML

Ing. Géminis Lucía Mejía Layna¹, M.T.I. Jorge Carranza Gómez²,
M.C. Francisco Javier Gutiérrez Mata³, M. C. Rafael Hernández Reyna⁴

Resumen — La migración de datos es el proceso mediante el cual realizamos una transferencia de datos de unos sistemas de almacenamiento de datos a otros, generalmente se llevan a cabo para reemplazar o actualizar servidores o equipos. En el presente artículo se propone el desarrollo de un sistema que realice la migración de la base de datos del Instituto Tecnológico de Acapulco, del gestor de base de datos SYBASE a formato XML, este sistema está basado en la metodología unificada de procesos de migración, con el cual se pretende atender la necesidad de migrar la información de los institutos del TecNM que aún no actualizan sus sistemas de información.

Palabras clave — Formato XML, Base De Datos, Migración

I. Introducción

La innovación de los sistemas de información y la creación de hardware con mejores características, fomentan la migración entre diversas estructuras de datos, constantemente será recomendable realizar actualizaciones que mejoren los atributos de los programas con los que se maneja la información, o en los equipos en los que se almacena. El Instituto Tecnológico de Acapulco al ser parte del Tecnológico Nacional de México (TecNM), debe mantenerse a la vanguardia en los sistemas que le solicite su entidad superior, para ello deberá migrar sus datos a las plataformas que les sean sugeridas. Originalmente, las instituciones dependientes del TecNM utilizaron el gestor de base de datos SYBASE, ya que el Sistema Integral de Información fue diseñado para trabajar con este administrador, posteriormente se desarrolló una nueva versión, la cual era compatible con el gestor de base de datos PostgreSQL, actualmente se desarrolla otra versión de esta plataforma, La creación de un sistema que permita la migración de datos hacia un formato de transferencia que sea compatible con múltiples plataformas, permitirá que la renovación de los sistemas se realice de forma sencilla. En la primera sección de este artículo se presentan los antecedentes teóricos que fundamentan el desarrollo del migrador de datos a formato XML, a continuación, se describen las metodologías de procesos de migración de base de datos que dieron pie a la definición de una metodología adaptada a las necesidades requeridas para el sistema que se pretende desarrollar, finalmente incluye las consideraciones a trabajos futuros y las referencias consultadas.

II. Marco teórico

Formato XML

El lenguaje XML (*Extensible Markup Language*) se utiliza para realizar el intercambio de datos estructurados. Más que un formato de archivos rígido, XML es un lenguaje que define los formatos aceptados que pueden utilizar los grupos para intercambiar información. Numerosas personas, organizaciones y empresas utilizan XML para transferir información de productos, transacciones, datos de inventario y otros tipos de información empresarial. Boulanger lo menciona como “un archivo de texto especial compuesto por un determinado número de etiquetas en estructura de árbol, los elementos están vinculados entre ellos a través de las ramas que se crean” (Boulanger,

¹ Ing. Géminis Lucía Mejía Layna, alumna de la maestría en sistemas computacionales del Instituto Tecnológico de Acapulco. mejia.geminis86@gmail.com

² M.T.I. Jorge Carranza Gómez, profesor de la maestría en sistemas computacionales del Instituto Tecnológico de Acapulco. jcarranzamx@hotmail.com

³ M.C. Francisco Javier Gutiérrez Mata, profesor de la maestría en sistemas computacionales del Instituto Tecnológico de Acapulco. fcomata84@hotmail.com

⁴ M. C. Rafael Hernández Reyna, profesor de la maestría en sistemas computacionales del Instituto Tecnológico de Acapulco. gameduardo@hotmail.com

2015, pág. 21). permite jerarquizar, estructurar la información y describir los contenidos dentro del propio documento, así como la reutilización de partes del mismo.

Migración de datos

Podemos llamar migración de datos al “**proceso de transferencia** de los mismos que se desarrolla mediante la implementación de una serie de fases” (PowerData Especialistas en Gestión de Datos, 2015). Generalmente, se realiza mediante un proceso ETL, en el que se lleva a cabo la extracción, transformación y carga, cuando se logra el formato apropiado para el nuevo sistema se logra preservar la información disponible en el antiguo sistema y, por lo tanto, la migración resulta exitosa.

Según Fernández (Fernández, 2011), la migración se vuelve necesaria cuando:

- Se ha vuelto muy lenta la carga y consulta de datos
- Se debe mantener compatibilidad hacia arriba con otros productos
- El volumen de información que se maneja necesita mayor robustez de la base de datos
- Se desea mantener actualizados los productos
- Es necesario cambiar hacia un producto de base de datos sin licencia
- El actual Sistema Operativo S.O. esta desactualizado
- No es posible usar los programas nuevos
- Se necesita bajar los gastos en tecnología
- El Sistema Operativo actual no es confiable desde que se utiliza Internet
- Se desea armar una red
- Se han comprado nuevos equipos

III. Metodologías para el proceso de migración de datos

La definición de una metodología adaptable a las necesidades para la construcción de un migrador de base de datos a formato XML está basada en las siguientes metodologías:

Metodología mariposa

Desarrollado en el departamento de ciencias computacionales de *Trinity College* en Dublín, Irlanda, su foco está puesto en lo que sus autores consideran lo más importante: los datos.

Propone eliminar el problema de mantener el sistema legado y el sistema destino corriendo juntos, lo que implica mantenerlos sincronizados para poder tener coherencia entre ellos.

Con el fin de manejar el proyecto, esta metodología hace que la base de datos del legado se convierta en sólo lectura desde que se inicia el proceso de migración y hace uso de almacenamientos temporales para los datos que necesitan ser guardados antes de que el sistema destino se ponga en producción. A las bases o almacenamientos temporales se accede a través de un componente que redirecciona las peticiones a la base correcta.

La migración es apoyada también por una herramienta de transformación de datos. Durante el proceso, la base se convierte temporalmente a modo sólo lectura y luego se migra de forma incremental hasta que los datos que quedan en el legado son menores a un margen o valor establecido al comienzo del proyecto. Después de este último paso de la migración, el sistema destino estará listo para ponerse en producción.

Metodología unificada de procesos de migración

Fue propuesta por Javier Fernández en la Universidad Cooperativa de Colombia en 2010, este esquema viene representado a través de las siguientes fases:

Fase de análisis y definición de procesos de migración y reconocimiento del origen: En ella se define la estructura destino de los datos y las características técnicas de la plataforma hacia la cual se migrarán los datos. Se procede a la conexión del Migrador con la fuente de datos de destino.

Fase de migración: Implementando un algoritmo de reconocimiento de estructuras se procede al proceso de migración de los datos, En este procesamiento de las bases de datos se soporta un fuerte esquema de control y análisis de datos, en el cuál mediante técnicas propias de la administración de las bases de datos, se consolida la información que interviene en los procesos de migración. Una vez soportado el esquema de migración, se procede a la fase de “Auditoría de la Información”.

Fase de auditoría de la información: El Migrador permite generar los informes de consistencia de la información migrada, con el fin de detectar posibles errores, como duplicidades o falta de datos.

IV. Metodología para la migración de datos de Sybase a Formato XML

La migración de datos es una tarea que requiere una serie de pasos que deben garantizar la correcta transferencia de información, resguardando la calidad e integridad de los datos, la siguiente metodología engloba los puntos más importantes requeridos para afianzar un acertado proceso migratorio.

Se consideran cinco etapas, en las cuales se deben realizar determinadas acciones que permitirán que el flujo del procesamiento no se vea mermado, por lo que, se debe dar un tratamiento especial a los datos para que sean apropiados para seguir con el proceso,



Figura 1 Metodología para la migración de los datos de Sybase a Formato XML

| Fase | Procesos |
|---------------|---|
| Recolección | Seleccionar las fuentes de datos, Establecer qué campos serán migrados. Calcular una estimación de los requisitos de almacenamiento. Determinar los derechos de acceso y conexión a los datos. |
| Análisis | Aplicar la función de análisis a cada uno de los campos en función de valores mínimos y máximos, valores únicos, Valores numéricos y valores más comunes. Prestar una atención especial a los campos críticos. |
| Limpieza | Determinar el nivel de limpieza que cada campo debe alcanzar para estar en condiciones de ser migrado. Aplicar una función de limpieza de datos en caso requerido. |
| Transferencia | Determinar qué campos van a ser migrados. Aplicar una función de procesamiento de los datos para su transformación a formato XML. Auditar la transferencia de datos mostrando, al menos, el origen y destino de los datos, el número de registros procesados. |
| Validación | Comprobar que los datos se encuentran en la estructura destino según estaba previsto. Aplicar comprobaciones de calidad e datos. Determinar mediante un porcentaje de efectividad, el éxito o fracaso del proceso. |

V. Conclusión

La migración de datos no es una tarea simple, es necesario un fuerte análisis y conocimiento de los datos que se desean migrar, de la estructura final y las formas que tomarán en el proceso, así como las acciones que requieren antes de ser migrados tales como la extracción, transformación y limpieza de los datos antes de moverlos. También implica implementar políticas de migración de datos que aseguren que los datos serán transferidos de una manera ordenada. Someter los datos, una vez finalizada la acción de migración de datos a tests y pruebas de validación que garanticen su idoneidad y exactitud. Auditar y documentar todo el proceso de migración de datos.

VI. Trabajo a futuro

Esta última metodología descrita es la que se implementara para el trabajo a futuro que es el desarrollo de un sistema que desarrolle esta migración de Sybase a XML. Actualmente en el Instituto Tecnológico de Acapulco en el SII que es el sistema encargado de llevar a cabo de gestión de los alumnos y su historial académico está programado en PHP y tiene como gestor de base de datos a Sybase. Debido a que ya se han realizado anteriormente actualizaciones al sistema y en este proceso se cambia de gestor de base de datos a pesar de que se concluye el proceso de migración de datos de forma exitosa, ya que cada actualización trae un nuevo gestor se han observado perdida de información o inconsistencia en los datos y debido a esto se tienen que agregar de manera manual, lo cual es una perdida de tiempo y esto genera costos extra aparte de la compra de este nuevo sistema, lo que se pretende con la elaboración de este sistema es evitar estos detalles.

VII. Referencias

- Boulanger, T. (2015). *XML práctico: Bases esenciales, conceptos y casos prácticos (2ª edición)*. Ediciones ENI.
- Ceballos, F. (2013). *Enciclopedia de Microsoft® Visual C# Interfaces gráficas y aplicaciones para Internet con Windows Forms y ASP.NET 4.ª edición*. Madrid: RA-MA Editorial.
- Celis, A. (s.f.). Metodología para la migración de forma segura de Sistemas de Gestión de Bases de Datos Relacionales a software libre y estándares abiertos. *REVISTA PGI - INVESTIGACIÓN, CIENCIA Y TECNOLOGÍA*, 140-142.
- Cerezo, Y. (2007). *Iniciación a la programación en C#: un enfoque práctico*. Madrid: Delta Publicaciones.
- Fernández, J. (2011). METODOLÓGIA PARA LA CONSTRUCCIÓN DE UN MIGRADOR UNIVERSAL DE BASES DE DATOS. *DYNA*, 266-275.
- Gonzales, J. A. (2001). *el lenguaje de programacion c# jose antonio gonzalez seco*. "Documento en formato digital".
- Henríquez, N. (s.f.). Postgresql una alternativa efectiva en las empresas. 1-5.
- Microsoft. (17 de 09 de 2015). Obtenido de <https://docs.microsoft.com/es-es/dotnet/csharp/getting-started/introduction-to-the-csharp-language-and-the-net-framework>
- POSTGRESQL TUTORIAL*. (08 de 2012). Obtenido de <http://www.postgresqltutorial.com/what-is-postgresql/>
- PowerData Especialistas en Gestión de Datos. (12 de Agosto de 2015). *PowerData*. Obtenido de <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/406547/tipos-y-funci-n-de-los-gestores-de-bases-de-datos>
- Tecnológico Nacional de México. (s.f.). *TecNM*. Obtenido de <http://www.tecnm.mx/telecomunicaciones/sistema-integral-de-informacion-dgest>
- Universidad Politécnica de Puebla. (s.f.). *UPPuebla*. Obtenido de <http://repositorio.uppuebla.edu.mx:8080/xmlui/bitstream/handle/123456789/338/OA-UPPue-SGBD.pdf?sequence=1>