

Proyecto de migración de datos Sistema de Gestión Electoral

Silvio Serra, Marcelo Marciszack, Sergio Quinteros, Felipe Steffolani, Adrea Dalhul Uez {marciszack, ser.quinteros, ingsilvioserra, [fsteffolani](mailto:fsteffolani@gmail.com)}@gmail.com, Andrea.uez@pjn.gov.ar

Centro de investigación y desarrollo de Sistemas (CIDS) - Departamento de Ingeniería en Sistemas de Información -Facultad Regional Córdoba - Universidad Tecnológica Nacional
Maestro López esq. Av. Cruz Roja Argentina. Ciudad Universitaria. Córdoba

Resumen

En nuestro país la información electoral, es decir, aquella relacionada a los electores, los trámites que modifican los datos de estos, los actos electorales, los partidos políticos y sus autoridades, etc. está administrada por la Justicia Nacional Electoral y la Cámara Nacional Electoral. Para realizar estas funciones se utilizan actualmente varios sistemas distintos y geográficamente descentralizados ya que se hallan instalados en cada jurisdicción.

Se encuentra en desarrollo el nuevo sistema de gestión electoral (SGE) que reemplazará a los anteriores y que funcionará de manera centralizada, pero el mismo no puede ponerse en marcha sin migrar la información residente en las plataformas actuales. Esto da origen a un proyecto de migración complejo y de gran volumen que debe llevarse adelante paralelamente al desarrollo del sistema de información.

El presente trabajo describe los aspectos principales del proyecto de migración de datos que posibilita unir en una única base de datos centralizada, la información relativa a todos los electores de nuestro país. Para poder concretar las actividades se utiliza una metodología y una arquitectura que permiten tomar las diversas fuentes de origen y procesar eficientemente los cambios en la base de datos de destino que se encuentra en desarrollo.

1 Introducción

La Cámara Nacional Electoral (CNE) tiene competencia en todo el territorio de la Nación. Este Tribunal electoral integra el Poder Judicial de la Nación y es la autoridad superior de aplicación de la legislación político-electoral, cumpliendo un rol esencial en todo lo relativo a la organización de los procesos electorales. Dentro de sus funciones registrales y de fiscalización específicas se encuentran: El Registro Nacional de Electores (RNE), el Registro de Cartas de Ciudadanía, el Registro de Electores Residentes en el Exterior, el Registro de Electores Privados de Libertad e Inhabilitados, el Registro de Infractores al deber de votar, el Registro Nacional de Partidos Políticos, el Registro Nacional de Afiliados a los Partidos Políticos, el Registro de

Empresas de Encuestas y Sondeos de Opinión, el Registro Público de Postulantes a Autoridades de Mesa, el Registro Nacional de Divisiones Electorales, el Registro de Delegados de la Justicia Nacional Electoral y el Cuerpo de Auditores Contadores. [1]

Para desempeñar muchas de las funciones mencionadas, actualmente se utilizan varios sistemas que podrían llegar en los próximos años al límite de su capacidad y que, en algunos casos, operan de manera descentralizada en cada una de las 24 delegaciones del Juzgado Electoral. Por estas razones en el año 2014 la CNE encargó a la Universidad Tecnológica Nacional Facultad Regional Córdoba que por medio de su Centro de Investigación y Desarrollo de Sistemas (CIDS) realice una reingeniería de sus procesos y posteriormente desarrolle un nuevo sistema que se denomina Sistema de Gestión Electoral (SGE).

El conjunto de información actualmente administrada representa un gran volumen de datos y ha sido generada y almacenada por los sistemas que van a ser reemplazados cuando SEG se implemente. Naturalmente, dicha información responde a la estructura propia de estos sistemas y debe ser incorporada a la nueva base de datos antes de la puesta en marcha. Esto da origen a un proyecto de migración de datos con transformaciones complejas, alto volumen, diversidad en la estructura de las bases de origen, alto grado de importancia en la información que manipula e inestabilidad en la base de destino por encontrarse esta sujeta a los cambios del sistema en desarrollo.

El objetivo del proyecto de migración es lograr centralizar en una única base de datos toda la información proveniente de los sistemas anteriores, adaptándola a las necesidades de SGE y antes de su implementación.

2 Situación – problema u oportunidad

En general, los proyectos de migración de datos suelen ser más complejos de lo que aparentan y frecuentemente atraviesan dificultades para ser concluidos dentro de los plazos y costos estimados, Como afirma Robert Alan “la migración de datos de sistemas heredados es un proceso complejo” [2].

La figura siguiente [Figura 1] revela un estudio realizado por la firma Oracle [3] según el cual, más del 80% de los proyectos de migración exceden el costo y el tiempo estimado.

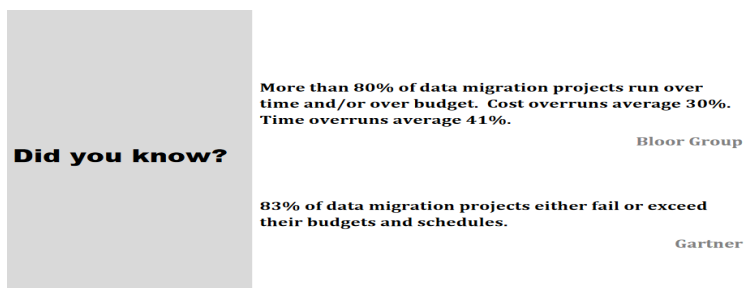


Figura 1. Resultados en los procesos de migración de datos según Oracle [3].

Según publicación de IBM [4], el 48% de las migraciones de datos se realizan en circunstancias de actualización de tecnología. Más aún, Según la firma Oracle[5], hasta el 75 por ciento de los nuevos sistemas no cumplen con las expectativas, a menudo por fallas en el proceso de migración. El mismo artículo afirma que las migraciones de datos generalmente resultan de la introducción de un nuevo sistema.

Distintos fabricantes y compañías especializadas proponen metodologías para abordar las migraciones, Así por ejemplo puede mencionarse a AIM (Oracle Application Implementation Methodology) que es una metodología desarrollada por Oracle para realizar implementaciones de aplicaciones y que da cobertura a la migración de datos en una de sus fases [6] o la metodología Oracle relational migration maps [7] o Softek – IBM Data migration methodology de la firma Softek e IBM [8] o Microsoft Data Migration Methodology [9] de Microsoft entre muchas otras. Todas ellas constituyen buenas opciones, son metodologías muy probadas provenientes de las firmas más reconocidas y con mayor experiencia, pero también todas ellas parten de un origen y un destino estable y justamente esto es lo que no ocurre en el contexto particular de la migración de datos de SGE.

También deben tenerse en cuenta las situaciones que tienen como centro el cambio de tecnología, pasar datos soportados por un producto de un fabricante a otros. [10][11][12]

Más allá de las particularidades de cada enfoque, se pueden identificar los siguientes componentes mínimos y comunes: Base de origen, proceso de transformación y base de destino tal como lo muestra la siguiente figura [Figura 2] [13].

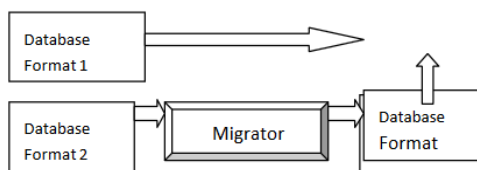


Figura 2. Componentes mínimos de una migración de datos [13].

Observando estos componentes básicos, cuando se desarrolla un nuevo sistema, las bases de origen normalmente son estables. Se trata de repositorios de datos que se encuentran en uso por los sistemas actuales (los que van a ser reemplazados), por el contrario, las bases de destino (las que se pretenden llenar con la información migrada para que sean utilizadas por el nuevo sistema) están en construcción, son inestables en su estructura y pueden cambiar conforme se produzcan modificaciones en los requerimientos o se detecten errores de diseño o de construcción. Como lo afirma Ian Sommerville, “La evolución de los requerimientos es inevitable” [14].

Los enfoques clásicos hacen un fuerte foco en esta estructura, pero no en la volatilidad de la base de destino, simplemente porque ese es un factor que solo se encuentra presente en algunos casos.

Entonces, si el proceso de migración de datos heredados es complejo [2] y si a la vez los requerimientos evolucionan [14] y esto puede afectar la composición de la base de destino, se está en presencia de una situación compleja e inestable.

En estos casos se obtienen dos procesos de desarrollo que corren en paralelo, uno es el que involucra la construcción del sistema en sí y otro es el que debe resolver el proceso de migración. Cada uno tiene sus propios desafíos técnicos y funcionales, cambios de requerimientos en el sistema pueden implicar cambios en el proceso de migración y a la inversa.

Mantener ambos procesos coordinados y avanzando a la par con la menor cantidad de retrabajo posible puede resultar difícil. Las metodologías clásicas se centran naturalmente en la complejidad inherente de la migración, pero no en este paralelismo.

En el caso particular de SGE, no resulta factible ponerlo en funcionamiento sin migrar la información residente en las bases actuales. Como la migración debe estar concluida al momento de la puesta en marcha, es necesario comenzar con ella de manera anticipada, sin que el desarrollo del sistema esté concluido. Esto da origen a un proyecto de migración de datos que se lleva adelante en paralelo al desarrollo del sistema y que debe adaptarse constantemente a la evolución de la base de datos de destino. Desde ese punto de vista hubiera sido más conveniente demorar el comienzo de la migración hasta un momento en el que el desarrollo del sistema se encuentre más avanzado, pero dada la complejidad de las transformaciones necesarias, el gran volumen de datos y la necesidad de contar con una base de datos migrada al momento de la puesta en marcha del sistema, es fundamental comenzar las tareas propias de la migración en forma anticipada.

Las bases de datos de origen, en las que reside la información a ser migrada, responden a varios sistemas y son de una tecnología diversa, pero a los efectos de la migración se transforman en archivos de texto plano y son tomadas de esa manera como entrada, aun así, su estructura sigue siendo diversa.

La base de datos de destino es única y responde a la plataforma Oracle 11G, debiendo centralizarse en ella toda la información. Este es un aspecto muy relevante del proyecto ya que los sistemas actuales funcionan de manera descentralizada en las 24 jurisdicciones electorales y debe unificarse toda la información en una base de datos centralizada.

De este modo, la migración debe hacer una fuerte transformación sintáctica porque las tecnologías que soportan las bases de datos actuales son muy diferentes a la que va a implementarse y a la vez también realizar una importante transformación semántica porque existen grandes diferencias en la nueva lógica de negocio que plantea un funcionamiento centralizado. Además, El volumen de datos a migrar es grande y el tiempo de ejecución tiene que ser mínimo pues los sistemas deben permanecer fuera de línea lo menos posible y como la base de destino cambia conforme evoluciona el desarrollo del sistema, este complejo proceso debe adaptarse de manera constante. Todo esto imposibilita el uso de una metodología clásica y plantea gran desafío.

3 Solución propuesta

Como solución se propone el uso de una arquitectura que plantea tres niveles de división de los datos separando las transformaciones sintácticas de las semánticas y una metodología que mediante dos procesos iterativos permite gestionar los cambios frecuentes en la base de destino de manera ordenada y a bajo costo. A continuación, se describen ambos elementos.

3.1 Arquitectura dividida

Como ya se ha dicho, existen dos grandes problemas a resolver dentro de una migración. Uno relacionado con las conversiones sintácticas y otro con las semánticas. Por un lado, si las plataformas de origen y destino son diferentes seguramente hay que convertir tipos de datos, longitudes, formatos etc. Estas son conversiones sintácticas. Al mismo tiempo, si los datos son los mismos, pero responden a sistemas distintos, sin duda también hay que adaptar su significado, su estructura y relaciones. Estas son conversiones semánticas.

Las metodologías y herramientas clásicas enfocan ambos aspectos como uno solo, es decir, consideran la existencia de un único proceso de transformación que debe resolver todo el problema [13], sin embargo, pueden abarcarse estos aspectos por separado y así obtener dos procesos de transformación, uno de transformación sintáctica y otro de transformación semántica.

En un escenario de cambios frecuentes esta división resulta muy conveniente. El modelo propuesto a continuación parte de dividir al proceso de migración físicamente en dos y utiliza dos bases de destino para realizar la migración en dos etapas. De esta manera ya no se tiene un único bloque complejo donde interactúan todos los agentes intervinientes y donde impactan todos los cambios, consecuentemente se disminuye la complejidad y se agiliza la gestión de las modificaciones.

La figura siguiente [Figura 3] ilustra el modelo de bases y procesos divididos.

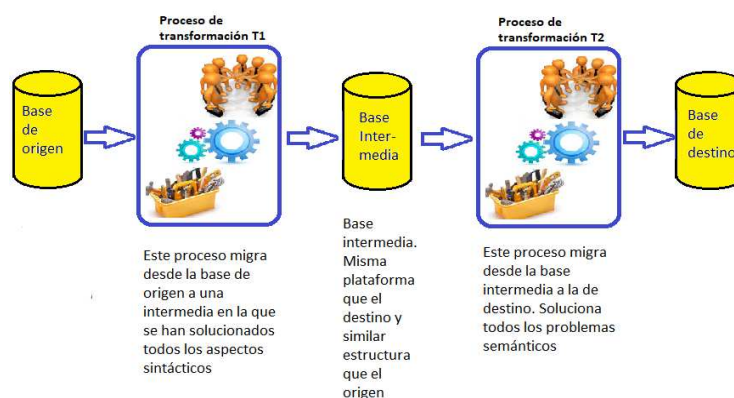


Figura 3. Modelo de bases y procesos divididos.

Como puede advertirse en la imagen, la migración parte de una base de origen, se ha ilustrado solo una, pero podrían ser varias, el planteo sería idéntico respecto de la arquitectura planteada y empleo de la metodología, aunque seguramente más complejo desde el punto de vista de la implementación técnica.

Un proceso llamado T1 realiza una primera transformación, pero haciendo solamente foco en las adaptaciones sintácticas. Su objetivo es el de adaptar tipos de datos, longitudes, formatos y todo elemento sintáctico para obtener una única base de salida intermedia montada sobre la misma plataforma que la base de destino final, pero con las estructuras semánticas lo más similares que resulte posible a las bases de origen. De este modo el proceso T1 contiene solo elementos de transformación sintáctica, todas las herramientas y recursos intervinientes tienen ese único objetivo.

Un segundo proceso llamado T2 toma la base intermedia generada por T1 y abstraído de cualquier diferencia de sintaxis (porque ya no existen) realiza todas las transformaciones estructurales para lograr que los datos de origen se adapten al diseño de la base de salida. En este módulo todos los recursos intervinientes se orientan a analizar y adaptar los datos estructuralmente.

A partir de esta división se disminuye la interdependencia entre las bases de origen y la de destino ya que al desarrollar T1 todos los recursos se orientan a resolver un problema de plataforma, teniendo en cuenta la tecnología de la base de destino, pero no su diseño ni su estructura ni ninguno de sus aspectos funcionales. Al trabajar sobre T2 ocurre lo contrario, todo cambio en la estructura de la base de datos de destino impacta en forma directa, pero como se tiene una única plataforma y una sola fuente de origen, es menos costoso realizar las modificaciones.

Al momento de procesar un cambio de requerimiento, dependiendo del tipo de cambio de que se trate, es probable que solo impacte en uno de los procesos ya sea T1 o T2 e incluso si impacta en ambos, es más fácil medir su incidencia y procesarlo, ya que cada módulo es individualmente más sencillo.

Es importante destacar que una división física y un enfoque de arquitectura no alcanzan para resolver los problemas típicos del contexto de estudio. Se requiere, sin duda, un planteo metodológico que permita organizar toda la migración en torno de esta estructura, pero hasta el momento (y solo por hacer una división física) ya pueden advertirse algunas ventajas importantes.

3.2 Metodología de tres niveles iterativa

A continuación, se propone una metodología que, utilizando la arquitectura física propuesta en el apartado anterior, define cinco fases que permiten gestionar todo el proceso. Debe tenerse presente que está especialmente diseñada para gestionar un proceso de migración en forma paralela al desarrollo del sistema y de la base de datos destino.

Fase 1: Planificación y Análisis general

En esta etapa se realiza la planificación general de todo el proceso de migración y también se hace un primer análisis global, es decir, independiente de la división de los

de procesos de transformación. Los documentos que se construyen en esta fase son de alcance general y sirven como punto de partida del proyecto.

Más allá de que la base de destino se encuentre sujeta a cambios y de que seguramente estos afectan el proceso de migración haciendo variar su ciclo de vida y los recursos afectados, resulta de suma importancia alcanzar una visión general y obtener una estimación de los recursos que serán afectados en cada etapa. También es muy importante desarrollar una primera planificación donde pueda estimarse el tiempo necesario para el desarrollo de toda la migración general.

Fase 2: Análisis y Diseño de T1

En esta fase se toma contacto con las bases de origen que constituyen los datos de entrada de la migración y se diseña el proceso que tiene como salida la base intermedia. Las tareas de esta fase son las siguientes.

- Realizar un análisis de la estructura física de los datos de entrada. Esto es, identificar cuáles son las bases de origen y sobre qué tecnología están soportadas.
- Realizar un análisis de la estructura lógica de los datos de entrada. Resulta importante para tomar decisiones posteriores conocer como están estructurados los datos de origen. Cómo son sus objetos (tablas, archivos, relaciones, funciones, procedimientos, etc.) y a qué lógica responden.
- Establecer la manera de conectarse con los servidores que dan soporte a las bases de origen para poder hacer la extracción. Debe tenerse en cuenta que esta puede no ser una tarea sencilla pues responden a sistemas en producción y pueden ser de varia tecnología.
- Identificar el subconjunto de elementos a incluir como entrada de T1. No todas las tablas o archivos contenidos en las bases de origen son objeto de migración, el nuevo sistema puede requerir el traspaso solo de algunas entidades. Debe tenerse aquí un criterio amplio porque es probable que luego se presenten modificaciones que impliquen incorporar nuevas entidades, en este sentido es mejor tenerlas en cuenta desde un comienzo e incluirlas en la base intermedia, pero al mismo tiempo no es conveniente incorporar en T1 un gran volumen de transformaciones que no tengan utilidad real.
- Diseñar la base de datos intermedia. Respetando la misma tecnología que la base de destino final, pero manteniendo una concepción lógica lo más similar posible a los datos de origen, se debe establecer la estructura de la base intermedia.
- Especificar las herramientas a utilizar en el desarrollo de T1. Se deben determinar los elementos tanto de hardware como de software para realizar la programación y en general toda la construcción del primer proceso.
- Ajustar la planificación y asignación de recursos para T1. En este punto se tiene un nivel de conocimiento mucho más detallado y profundo no solo de las estructuras de origen sino del conjunto de entidades objeto de la migración y del alcance del primer proceso, por ese motivo puede realizarse una planificación mucho más certera que la inicial.

Fase 3: Análisis y Diseño de T2

Con una visión clara de los alcances de T1 y el desarrollo de este en marcha puede realizarse un análisis y diseño del segundo proceso. Si bien T2 toma como entrada la

base intermedia que es la salida de T1, no es necesario esperar a que T1 se culmine y se ponga en marcha para comenzar con su análisis y diseño o incluso con su construcción.

Debe tenerse en cuenta que en T2 la dificultad técnica es menor porque su base de origen (la base intermedia) y la de destino, ahora se encuentran soportadas sobre la misma tecnología, pero la dificultad en las transformaciones lógicas puede ser muy significativa. Dentro de esta fase pueden aparecer procesos complejos desde el punto de vista lógico y sobre ellos debe tenerse especial atención.

Las tareas de esta fase son las siguientes:

- Estudiar la estructura de la base de datos de salida teniendo en cuenta que la misma está sujeta a cambios por la propia evolución del desarrollo de la aplicación que la utiliza. Se debe tomar contacto con su estructura, estudiar sus entidades e identificar los aspectos que aún están propensos a modificaciones y aquellos que resultan más estables.
- Identificar el conjunto de entidades de la base de salida que deben poblarse con datos provenientes de la migración. Analizar principalmente su estructura, tipos de datos, restricciones y relaciones.
- Diseñar los procesos que van a realizar las transformaciones leyendo datos de la base intermedia y escribiendo en las entidades de la base de destino. Esta es la tarea central de la fase, nuevamente vale decir, que aquí es donde se enfrenta la verdadera transformación lógica de los datos provenientes de las fuentes originales para poder ser incluidos en la nueva aplicación. Esta tarea es el centro de la migración.
- Realizar una planificación detallada de las actividades de construcción de T2. Si bien en la primera fase se hace una planificación general, al igual que en el caso de T1, ahora al tener un diseño detallado y una visión más clara de los alcances de T2 es posible elaborar una planificación más concreta y certera.

Fase 4: Fase Iterativa. Construcción Ejecución y Prueba de T1

Una vez diseñado y planeado T1 puede comenzarse con su construcción. El desarrollo inicialmente debe centrarse en alcanzar una primer línea base, o sea, su primera versión ejecutable para mantenerlo luego dentro de un ciclo iterativo de mejora y modificación.

La primera versión puede incluir todo el proceso de importación y generación de la base intermedia y todas las pruebas e informes de resultados planeados, en este caso el proceso iterativo posterior se limita a incluir modificaciones cuando surjan cambios en la base de salida que afecten a T1.

También puede definirse la primer línea base como la migración de un subconjunto de entidades, este grupo puede estar compuesto por elementos que resulten centrales en el resto del proceso, o que tengan un interés particular por su volumen o por su lógica u otro factor. Las características de cada proyecto son las que determinan si la primera versión de T1 incluye solo un subgrupo de entidades y cuáles son las que componen dicho conjunto.

El desarrollo de pruebas y la obtención de los informes también pueden realizarse por partes. Lo natural es incluir las pruebas y los informes relacionados a las entidades alcanzadas en la línea base.

Si para la primera versión de T1 se define un subconjunto, entonces el proceso iterativo posterior ya no puede limitarse a incorporar cambios, sino que debe abarcar gradualmente al resto de los objetos de origen para así hacer crecer la estructura de la base intermedia.

Lo mismo ocurre con las pruebas y los informes, a medida que se incluyen más entidades estas deben contemplarse en los test.

A partir de lo dicho anteriormente, dentro de esta fase pueden identificarse las siguientes tareas:

- Determinar el grupo de entidades que conforman la primer línea base.
- Construir la primera versión de T1. Esta tarea puede incluir actividades de mapeo con herramientas gráficas, actividades de programación en lenguajes específicos y toda acción que lleve al desarrollo del proceso.
- Ejecutar la primera versión de T1 para obtener la primera base intermedia.
- Definir y Construir los informes de migración.
- Mantener un proceso iterativo en el que se actualice T1 para incorporar modificaciones o realizar correcciones en función de los errores detectados en los informes y según la evolución del desarrollo del sistema.

Fase 5: Fase Iterativa. Construcción Ejecución y Prueba de T2

A partir de la finalización de la Fase 3, se tiene un diseño del proceso T2 por lo tanto puede comenzarse con su construcción.

Uno de los principales aspectos a definir es el momento en el que resulta más conveniente iniciar esta etapa. Este proceso realiza principalmente la conversión semántica más compleja, dentro de T2 ocurre la verdadera adaptación de los datos partiendo de las reglas de los sistemas de origen para llegar a las reglas del nuevo sistema. Desde este punto de vista conviene comenzar la etapa lo más temprano posible, en rigor, la fase 5 puede comenzar ni bien se finalice la fase 3, o sea, se puede dar inicio a la construcción, ejecución y prueba de T2 ni bien se termine con su diseño.

A la vez, debe tenerse en cuenta que la principal entrada de este proceso es la base intermedia generada por T1 y que, si no se dispone de la misma, no se puede ejecutar ni probar ningún aspecto de T2. Desde ese punto de vista, resulta conveniente comenzar con T2 luego de tener al menos una primera versión ejecutable de T1.

La metodología no establece un momento específico, en este aspecto, las características propias de cada proyecto son las que determinen el punto de inicio correcto para esta etapa. En todos los casos, se trata de alcanzar un equilibrio entre la necesidad de anticipar el desarrollo de T2 y al mismo tiempo contar con una base de entrada para realizar pruebas en firme.

Al igual que con T1, la primera versión ejecutable de T2 no tiene que contener necesariamente todas las entidades que componen la base de salida. Puede alcanzarse solo un subgrupo en función de algunos criterios, por ejemplo, se le puede dar prioridad a las entidades más relevantes por su complejidad o por su volumen o incluir aquellas que resultan más importantes para realizar pruebas sobre el desempeño de la aplicación cuando opera con datos migrados, o simplemente incorporar las entidades que se encuentren disponibles en la base intermedia generada por T1.

A partir de lo dicho anteriormente, dentro de esta fase pueden identificarse las siguientes tareas:

- Determinar el grupo de entidades de la base de datos intermedia que conforman la primera línea base.
- Construir la primera versión de T2.
- Ejecutar la primera versión de T2 para obtener la primera base de salida.
- Definir y Construir los informes de migración.
- Mantener un proceso iterativo en el que se actualice T2 para incorporar modificaciones o realizar correcciones en función de los errores detectados en los informes y de la evolución del desarrollo del sistema.

La figura siguiente [Figura 4] es un esquema de las fases de la metodología y su relación con la arquitectura dividida.

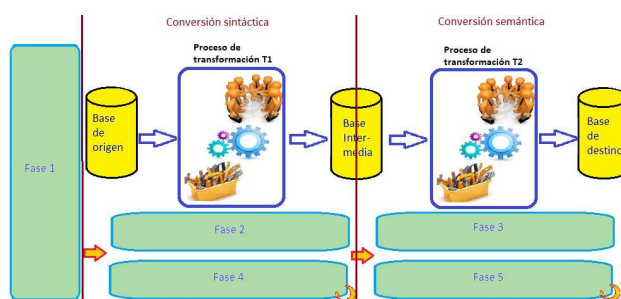


Figura 4. Fases de la metodología en relación a la arquitectura dividida.

3.3 Aplicación de la arquitectura y la metodología al proyecto de migración de SGE

Tanto la arquitectura como la metodología expuesta en los párrafos anteriores, se emplean con éxito en el proyecto de migración de SGE desde enero de 2017.

Las fuentes de datos de origen están compuestas por cuatro sistemas diferentes, donde uno de ellos se ejecuta de manera descentralizada en las 24 provincias de Argentina, de manera que cada una tiene su propia base de datos. Todas estas fuentes diversas se nuclean en una sola base intermedia Oracle 11G obtenida como salida de T1 y compuesta por 507 tablas. T2 toma esta estructura intermedia y realiza las conversiones semánticas para obtener la salida final soportada también por Oracle 11G y compuesta por 456 tablas de las cuales 148 están bajo el alcance de la migración. Esta base cuenta también con 813 procedimientos almacenados y funciones, 470 secuencias, 1338 índices entre otras entidades.

Durante la última iteración realizada se logran migrar con éxito, usando la metodología propuesta, más de 38 millones de electores con toda su información periférica como por ejemplo sus estados, tipos de documentos, situaciones electorales, datos de inhabilitaciones etc. También se migran más de 127 millones de domicilios, más de 90 millones de trámites, más de 250 millones de registros de información de historiales de electores entre otras tablas. Todas ellas con información periférica que también impacta en otras entidades. La base de salida también cuenta con información de todos los partidos políticos del país, sus autoridades y afiliados.

La figura siguiente [Figura 5] muestra un diagrama de arquitectura con las características generales de esta implementación, las tecnologías empleadas y las fases de la metodología relacionadas a dicha arquitectura.

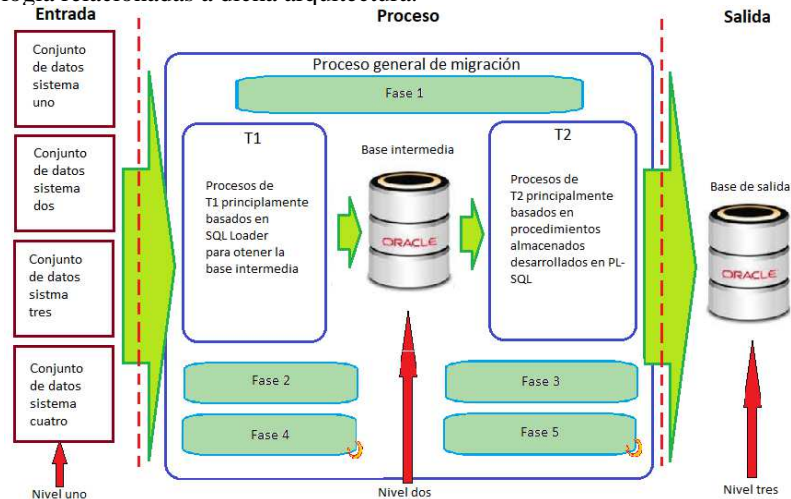


Figura 5. Diagrama general de implementación.

Los cuatro conjuntos de datos de entrada que se representan en la figura corresponden a los sistemas desde los cuales se toma información, cada uno está compuesto por varios archivos de texto que son tomados por T1 utilizando principalmente la herramienta SQL Loader de Oracle [15].

Como salida de T1 se obtiene la base intermedia que constituye el segundo nivel de datos. En este nivel se han resuelto todas las transformaciones sintácticas de manera que puede emplearse como entrada de T2. Para la construcción del segundo proceso se desarrollan 41 procedimientos almacenados divididos en 5 grupos, estos son:

- Procedimientos de migración de tablas satélites.
- Procedimientos de migración de electores.
- Procedimientos de migración de historiales.
- Procedimientos de migración de partidos políticos.
- Procedimientos de migración de trámites en curso.

También se desarrollan 13 funciones de uso general para realizar tareas comunes a diversos procesos.

Si bien a la fecha el sistema de información no se encuentra implementado y su desarrollo continúa, se ha obtenido una base de datos totalmente poblada con datos migrados de los sistemas anteriores, en otras palabras, usando la metodología de tres niveles iterativa se ha realizado la migración completa de los datos nucleando en una única fuente la información proveniente de varios sistemas y de las 24 provincias que componen el país.

Esta base se encuentra bajo operaciones de testing y paralelamente se utiliza para tareas de desarrollo del sistema, para realizar análisis de rendimiento, ensayos de implementación y capacitaciones a los usuarios finales. El proceso iterativo continúa en

marcha incorporando las correcciones de los defectos reportados por el área de testing y las modificaciones que puedan presentarse producto del desarrollo del sistema. Cuando el mismo se encuentre listo para su puesta en marcha, se planea ejecutar la última versión para obtener la base de datos final.

Es la primera vez que en Argentina se cuenta con esta información centralizada en una única base de datos. No se registraron desviaciones respecto del tiempo y de los recursos planeados para llevar adelante el proceso de migración atribuibles a la arquitectura o a la metodología. Sí existieron desfasajes relacionados, en todos los casos, a dificultades en la programación de algunos procesos complejos de T2 y a la alta diversidad de las estructuras de datos de entrada de T1.

4 Innovaciones del proyecto

Este proyecto de migración presenta una importante innovación: la aplicación de una arquitectura y una metodología de trabajo especialmente diseñadas para casos de migraciones que ocurren en paralelo al desarrollo de nuevos sistemas.

La metodología de tres niveles iterativa y la arquitectura física que la complementa han sido empleadas con anterioridad, pero nunca a una escala de estas proporciones, constituyendo una innovación la prueba en sí ya que los resultados obtenidos permiten recomendar su uso en entornos de alta complejidad y volumen.

La unificación de los datos de fuentes tan diversas y tan descentralizadas en una única base de datos, tal como se explica en el apartado anterior, es la primera vez que logra en nuestro país. Esto permite administrar la información de una manera mucho más segura y eficiente, sin duplicaciones ni tareas de mantenimiento ni sincronización y fija las bases para realizar análisis que resultarían imposibles de otro modo.

5 Relevancia para el interés público

Este proyecto de migración tiene un alto impacto en el interés público ya que permite poner en funcionamiento el nuevo Sistema de Gestión Electoral heredando la totalidad de la información generada y almacenada por los sistemas que reemplaza.

De este modo, la nueva herramienta informática centralizada y de alcance nacional, y los nuevos procedimientos diseñados para aumentar la eficiencia en la gestión de la información electoral, no se ponen en marcha solo para las nuevas acciones que se realicen a partir de su implementación, sino que continúan con cualquier gestión previa, contribuyendo al mejor desempeño de la justicia electoral de todo el país.

Toda la información cargada, almacenada y gestionada durante años por los sistemas descentralizados actuales referida a los electores, sus trámites, domicilios, información de historiales, lugar de residencia, etc. y todo lo relativo a los partidos políticos, sus afiliados, autoridades y composición estará disponible en el nuevo sistema al momento de su puesta en marcha. Incluso resultará posible iniciar un trámite en un juzgado electoral utilizando alguno de los sistemas actuales y continuarlo mediante

SGE de manera totalmente transparente para el ciudadano y sin que se deba repetir ninguna carga manual de información.

6 Viabilidad

Actualmente el desarrollo del sistema se encuentra en sus últimas etapas y se ha podido obtener una base de datos íntegramente poblada con información proveniente de las aplicaciones a ser reemplazadas, es decir, una migración completa. En este aspecto se puede considerar totalmente viable la implementación de la migración ya que solo resta aguardar el momento de la puesta en marcha para ejecutar los procesos de transformación de manera definitiva en los servidores de producción y eventualmente, procesar algún nuevo cambio que pueda surgir.

En otras palabras, las diversas iteraciones y ejecuciones realizadas y los resultados obtenidos demuestran la viabilidad de su implementación. Para su ejecución final no se requieren nuevos recursos técnicos ni humanos porque ya se han realizado pruebas de migraciones completas en los servidores de datos de producción con resultados exitosos.

7 Conclusiones

En toda la implementación, la dinámica del desarrollo del sistema implica la realización de numerosos cambios en la base de destino lo que impacta de manera directa en el proceso de migración. Sin embargo, estos cambios pueden procesarse en su totalidad y sin un empleo excesivo de recursos.

La posibilidad de realizar una planificación general en la primera etapa, para luego realizar los correspondientes ajustes conforme se avanza en las siguientes, es un factor desequilibrante que las metodologías clásicas no proveen. Sin dudas, el éxito en este proyecto tiene mucho que ver con lo antedicho ya que se hace posible planificar y administrar los recursos gradualmente.

Otro factor muy relevante es la flexibilidad para incorporar diversas herramientas tanto de programación como para la elaboración de informes y la realización de pruebas. Durante esta implementación las herramientas no son muchas ni diversas, pero se eligen conforme a las necesidades y características del proyecto y no en función de la metodología lo cual es muy positivo. Visto más ampliamente, la metodología no condiciona en ningún aspecto las herramientas empleadas en el desarrollo del proceso de migración ni el del sistema de información.

El empleo de la metodología permite contar con una base de datos poblada completamente con registros migrados en una etapa muy temprana del desarrollo del sistema de información. Más concretamente, a la fecha, el mismo no se encuentra concluido y

sin embargo ya se cuenta con una migración completa que brinda la salida esperada. Esto no solo admite utilizarla en diversas tareas que van desde el desarrollo hasta las pruebas de implementación pasando por el análisis de rendimiento (tal como se menciona en apartados anteriores) sino que posibilita realizar tareas de testing con el sistema y la base de datos migrada al mismo tiempo. De este modo, se prueba el sistema y la migración simultáneamente lo que redundará en un significativo ahorro de recursos y aumenta las posibilidades de descubrir defectos, tanto en el sistema como en la migración, en forma temprana.

La arquitectura física planteada también es un elemento muy positivo durante esta implementación. Tener los procesos divididos no resulta un factor complejo ni difícil de administrar y a la vez permite realizar las modificaciones y cambios solicitados en forma gradual y paralela asignando recursos diferentes.

Más allá de los aspectos positivos mencionados en los párrafos anteriores, sin ninguna duda, la característica iterativa de la metodología es la de mayor relevancia. Es gracias a su iteratividad que los cambios pueden procesarse gradualmente sin condicionar el desarrollo del sistema de información e incorporando modificaciones, correcciones de errores y nuevas funcionalidades de manera eficiente. Durante toda la implementación no se encuentra ningún caso en el que un cambio, del tipo que sea, no pueda procesarse o resulte en un efecto cascada que afecte toda la migración y lleve a un uso excesivo de recursos.

Por último, debe destacarse que las conclusiones obtenidas a partir de esta implementación no pueden verse como algo absolutamente general. Naturalmente, obtener los resultados mencionados durante este proyecto en particular no puede llevar a una generalización porque cada implementación tiene sus propias características. No obstante, debe destacarse que son resultados muy positivos y que indican un buen desenvolvimiento de la metodología. En los hechos ocurre que tratándose de una migración compleja (por la coexistencia de factores que ya se mencionaron), inserta en el desarrollo de un sistema de información también complejo, desde marzo a diciembre de 2017 se realiza con éxito una migración que no había podido llevarse adelante de manera positiva anteriormente. La metodología es general y puede aplicarse en otros casos de similares características.

Referencias

1. L. Barale, T. Torres Hansen, O. Medina, S. Quinteros, "SEDICI - Sistema de Gestión Electoral". Sociedad Argentina de Informática (SADIO) (45JAIIO). [Online] Setiembre 2016. [Cited: Agosto 29, 2017.] <http://sedici.unlp.edu.ar/handle/10915/58344>.
2. R. Alan, "The Serials Data Migration Dilemma". [Online] Julio 17, 2002. [Cited: Julio 25, 2017.] http://dx.doi.org/10.1300/J124v20n01_03.
3. ORACLE Corporation. "Put Your Data First or Your Migration Will Come Last". [Online] [Cited: Julio 15, 2016.] <http://www.oracle.com/us/products/middleware/data-integration/enterprise-data-quality/data-migration-wp-2345281.pdf>.

4. C. Kengchon, "Helping clients deal with economic uncertainty: Getting more from existing hardware". [Online] Febrero 2009. [Cited: Setiembre 15, 2016.] http://www-07.ibm.com/th/events/ibmuni2009/downloads/Other/GTS_2009_Thailand_University_gts.pdf.
5. Oracle Corporation, "Successful Data Migration". [Online] Octubre 2011. [Cited: Octubre 15, 2017.] <http://www.oracle.com/technetwork/middleware/oedq/successful-data-migration-wp-1555708.pdf>.
6. Crum, J. Using Oracle 11i. United States of America : BOSS Corporation, 2002.
7. Gonzalez Lau, C. and Aristiabal Moreno, C. ESTADO DEL ARTE DE PRUEBAS DE BASES DE DATOS PARA LA MIGRACIÓN Y VALIDACIÓN DE DATOS. Tesis de grado Ingeniería de Sistemas, Escuela de Ingeniería Departamento de Informática y Sistemas, Universidad EAFIT, Medellín Colombia. eafit.edu.co. [Online] 2007. [Cited: Enero 30, 2016.] https://repository.eafit.edu.co/xmlui/bitstream/handle/10784/2399/GonzalezLau_ConsueloInes_2007.pdf?sequence=1&isAllowed=y.
8. IBM Global Technology Services. Best practices for data migration. Methodologies for planning, designing, migrating and validating data migration. [Online] Julio 2007. [Cited: Marzo 09, 2017.] <https://www-935.ibm.com/services/us/gts/pdf/softek-best-practices-data-migration.pdf>.
9. Gajre, S., Bordia, B. and Soni, V. Competitive Database Migration Tools and Methodology. [Online] Marzo 2015. [Cited: Marzo 29, 2017.] <https://social.technet.microsoft.com/Search/es-AR?query=database%20migration%20methodology&ac=5>.
10. IBM Corporation. Migración a otra base de datos. [Online] 2013. [Cited: Julio 15, 2016.] http://www.ibm.com/support/knowledgecenter/SS69YH_6.0.0/com.ibm.spss.cads.config.doc/model_management/thin/migration_db_platfrom_switch.html?lang=es.
11. Lee, B. SQL Server: Administrar la Migración. [Online] 2011. [Cited: Julio 15, 2016.] <https://technet.microsoft.com/es-es/magazine/hh334645.aspx>.
12. ORACLE Corporation. Migrate to Oracle Database with SQL Developer. [Online] [Cited: Julio 15, 2016.] <http://www.oracle.com/technetwork/database/migration/index-084442.html>.
13. S. Vitthal, T. Baban, R. Warade, K. Chaudhari, "Data Migration System in Heterogeneous Database". [Online] Marzo 2013. [Cited: Julio 10, 2016.] http://www.ijesit.com/Volume%202/Issue%202/IJESIT201302_14.pdf.
14. I. Sommerville, "Ingeniería del software séptima edición". Madrid : Pearson Educación S.A., 2005.
15. Oracle Corporation. Database Utilities - SQL*Loader. [Online] 2017. [Cited: Septiembre 14, 2017.] http://docs.oracle.com/cd/B19306_01/server.102/b14215/part_ldr.htm.