

**ESTADO DEL ARTE DE PRUEBAS DE BASES DE DATOS PARA LA
MIGRACIÓN Y VALIDACIÓN DE DATOS**

**CONSUELO INÉS GONZÁLEZ LAU
CAROLINA ARISTIZÁBAL MORENO**

Proyecto de Grado para optar al título de Ingeniero de Sistemas

Asesor

**RAFAEL DAVID RINCÓN BERMUDEZ
Magíster en Matemáticas Aplicadas
Magíster en Sistemas de Calidad**

**ESCUELA DE INGENIERÍA
DEPARTAMENTO DE INFORMÁTICA Y SISTEMAS
UNIVERSIDAD EAFIT
MEDELLIN
2007**

Nota de aceptación

Presidente del Jurado

Jurado

Jurado

Medellín, 5 de Octubre de 2007

A mi Familia, mi Novio y mis Amigos que me acompañaron en este camino y me ayudaron a hacer este sueño realidad.

Carolina Aristizábal Moreno

A mi Madre, mis Tías, mi Familia, Amigos y Profesores que me acompañaron y apoyaron en esta experiencia.

Consuelo Inés González Lau

AGRADECIMIENTOS

Expresamos nuestros más sinceros agradecimientos a:

Rafael David Rincón Bermúdez, nuestro asesor de proyecto, por su constante colaboración, paciencia, guía y motivación para poder llevar a cabo este Proyecto de Grado.

Al Señor Scott Ambler de IBM Canadá por su guía, consejos y aportes para el desarrollo de este proyecto de grado.

A nuestros padres y familia por su constante apoyo y motivación en el transcurso de nuestras vidas. Gracias a ellos hemos completado una nueva etapa en nuestro proceso de crecimiento, los estudios profesionales.

A las personas y organizaciones que nos abrieron sus puertas y compartieron con nosotros sus experiencias para el desarrollo de este proyecto de grado.

A nuestros amigos que nos han apoyado, animado y vivido la experiencia con nosotras.

A los Profesores y la Universidad que han sido la fuente de nuestros conocimientos en el transcurso de la carrera y nos han guiado en este proceso de formación y crecimiento profesional y personal.

TABLA DE CONTENIDO

RESUMEN.....	11
INTRODUCCIÓN.....	12
1. OBJETIVOS.....	15
1.1. GENERAL.....	15
1.2. ESPECÍFICOS.....	15
2. GLOSARIO.....	16
3. ANTECEDENTES.....	24
4. JUSTIFICACIÓN.....	35
5. MARCO TEÓRICO.....	38
5.1. ¿QUÉ SON DATOS?.....	38
5.2. ¿QUÉ ES UNA BASE DE DATOS?.....	41
5.3. HISTORIA DE LAS BASES DE DATOS.....	42
5.4. TIPOS DE BASES DE DATOS.....	45
5.5. ¿QUÉ ES MIGRACIÓN DE DATOS?.....	49
5.6. TIPOS DE MIGRACIÓN DE DATOS.....	51
5.7. MITOS SOBRE LAS MIGRACIONES DE DATOS.....	53
5.8. ¿QUÉ ES VALIDACIÓN DE DATOS?.....	54
5.9. ¿QUÉ SON PRUEBAS DE SOFTWARE?.....	56
5.10. TIPOS DE PRUEBAS.....	57
5.10.1. PRUEBAS DE UNIDAD.....	57
5.10.2. PRUEBAS DE INTEGRACIÓN.....	58
5.10.3. PRUEBAS DE ACEPTACIÓN.....	58
5.10.4. OTROS TIPOS DE PRUEBAS.....	59

5.10.5.	PRUEBAS DE BASES DE DATOS	60
5.11.	¿QUÉ ES CALIDAD DE DATOS?.....	62
6.	ESTADO DEL ARTE DE LAS PRUEBAS PARA LA MIGRACIÓN Y VALIDACIÓN DE DATOS	69
6.1.	ENTORNO NACIONAL.....	71
6.2.	ENTORNO INTERNACIONAL	79
6.2.1.	RESULTADOS DE ENTREVISTAS Y ENCUESTA	79
6.2.2.	METODOLOGÍA DE MIGRACIÓN DE DATOS DE DULCIAN, INC....	91
6.2.3.	METODOLOGÍAS DE ORACLE.....	96
6.2.3.1.	METODOLOGÍA DE MIGRACIÓN.....	96
6.2.3.2.	METODOLOGÍA DE PRUEBAS	106
6.2.4.	METODOLOGÍA DE MIGRACIÓN DE SOFTEK.....	111
6.3.	COMPARACIÓN DE LOS ENTORNOS.....	117
7.	HERRAMIENTAS PARA MIGRACIÓN Y VALIDACIÓN DE DATOS	122
7.1.	ORACLE SQL DEVELOPER MIGRATION WORKBENCH.....	127
7.2.	ORACLE DATABASE MIGRATION VERIFIER.....	131
7.3.	ADVENTNET SWISSQL – DATA MIGRATION TOOL 4.9	136
7.4.	MYSQL MIGRATION TOOLKIT	142
7.5.	MS VISUAL STUDIO 2005 TEAM EDITION FOR DATABASE PROFESSIONALS.....	147
8.	CONCLUSIONES	152
9.	RECOMENDACIONES.....	155
9.1.	CONSIDERACIONES Y RECOMENDACIONES PARA LA MIGRACIÓN Y VALIDACIÓN DE DATOS	155
9.2.	CONSIDERACIONES PARA LAS PRUEBAS PARA MIGRACIÓN Y VALIDACIÓN DE DATOS.....	157
9.3.	FACTORES DE ÉXITO PARA UN PROYECTO DE MIGRACIÓN Y VALIDACIÓN DE DATOS.....	160
10.	TRABAJOS FUTUROS.....	163

BIBLIOGRAFÍA.....164
ANEXOS.....168

LISTA DE TABLAS

Tabla 1: Ejemplo de un Plan de Migración según Softek.....	114
Tabla 2: Ejemplo de Requerimientos de Diseño según Softek	115
Tabla 3: Factores Claves para la selección de una herramienta de migración....	124
Tabla 4: Lista de Herramientas de Prueba para Bases de Datos	126
Tabla 5: Bases de Datos soportados por Oracle SQL Developer Migration Workbench.....	127
Tabla 6: Bases de Datos soportados por Oracle Database Migration Verifier	132
Tabla 7: Tipos de Objetos y pruebas realizables sobre los mismos.	135
Tabla 8: Bases de Datos soportadas por SwisSQL Data Migration Tool.	137

LISTA DE FIGURAS

Figura 1: Estadísticas de desempeño de proyectos de migración de datos.	27
Figura 2: Tolerancia de los Tiempos de Inactividad.....	29
Figura 3: Estadísticas sobre problemas experimentados durante migraciones.	30
Figura 4: Traducción de las Áreas donde se generan datos incorrectos.	67
Figura 5: Vista general de Oracle AIM con Fases y Procesos.	84
Figura 6: Diagrama sobre qué probar en una base de datos.....	88
Figura 7: Fases de un proyecto de migración versus las actividades y duración. .	98
Figura 8: Traducción del Modelo V con Migración a Base de Datos.....	107
Figura 9: Traducción de la Metodología para la Migración de Datos.....	111
Figura 10: Modelo de Migración de Oracle SQL Developer Migration Workbench.	129
Figura 11: Vista Interfaz de Usuario para la Migración Rápida.....	131
Figura 12: Muestra del Reporte de Verificación presentado por Oracle DMV	136
Figura 13: Vista inicial del wizard de SwisSQL – Data Migration Tool 4.9	140
Figura 14: Vista Final del wizard de SwisSQL – Data Migration Tool 4.9	141
Figura 15: Bienvenido a MySQL Migration Toolkit.....	143
Figura 16: Plan de migración	144
Figura 17: Reporte Final	146
Figura 18: Interfaz gráfica del MS Visual Studio 2005 Team	148
Figura 19: Interfaz gráfica del MS Visual Studio 2005 Team	150

LISTA DE ANEXOS

ANEXO 1. PREGUNTAS DE ENTREVISTAS	168
ANEXO 2. COMPARATIVO DEL ESTADO DEL ARTE DE LAS PRUEBAS PARA LA MIGRACIÓN Y VALIDACIÓN DE DATOS CON RESPECTO A ESTANDARES INTERNACIONALES Y TMM.....	173

RESUMEN

El objetivo principal de este proyecto de grado es la identificación de las pruebas de Bases de Datos para la Migración y Validación de Datos en el entorno de las tecnologías de información en Colombia en la actualidad; también se realizará un análisis comparativo con lo que están realizando a nivel internacional algunas de las empresas líderes y firmas consultoras que se encuentran involucradas en el tema de las bases de datos y la migración de datos. Se inicia identificando en primera instancia los orígenes o razones por las cuales se llevan a cabo las migraciones de datos y los problemas a los cuales se enfrentan y las razones por las cuales se presentan dichos problemas. El propósito es poder identificar de qué manera las pruebas para las migraciones y validaciones de datos pueden ayudar a reducir las probabilidades de ocurrencia de dichos problemas, para el beneficio no solo de proyecto como tal, sino de la empresa también. De igual manera, se analizarán las herramientas empleadas y la influencia que pueden tener a la hora de realizar las pruebas.

Este proyecto no está orientado simplemente a la identificación de procesos y pruebas, sino también al reconocimiento de la importancia que los datos tienen para las decisiones en las organizaciones y el impacto que una buena calidad de los mismos tiene. Por ello, también es importante identificar aspectos que rodean el tema, tales como los datos en primera instancia, las bases de datos, la migración y la validación, la calidad de datos y las pruebas.

INTRODUCCIÓN

Es consecuencia directa del rápido crecimiento de la sociedad, sus organizaciones y las tecnologías que los apoyan, la velocidad con la cual crecen los volúmenes de datos e información dentro de las organizaciones, un crecimiento que va desde un 25% hasta posiblemente un 50% anual, porcentaje que aumenta cada año a medida que crecen las organizaciones, en las cuales la mayoría de esta información se encuentra en formato digital. Dicho crecimiento tecnológico a su vez lleva a una necesidad más frecuente de refrescamiento o cambio de dispositivos de almacenamiento, de aplicaciones empresariales o de implementaciones de nuevas tecnologías que apoyan las necesidades del negocio; estas actualizaciones y nuevas implementaciones se convierten en una necesidad para acomodar el creciente volumen de los datos y de las aplicaciones que las emplean para los procesos de la empresa, y estas necesidades son las que dan origen a los tan temidos proyectos de migración de datos para muchas organizaciones. Dichos proyectos de migración de datos se están volviendo más frecuentes cada año y va adquiriendo mayor importancia la necesidad de reducir los riesgos o problemas, que proyectos de este tipo implican.

Hoy en día, las aplicaciones críticas del negocio deben estar disponibles 24/7/365 (24 horas al día, 7 días a la semana, 365 días al año), dejando así pocas opciones por tiempos de inactividad para llevar a cabo las migraciones de datos. Con un crecimiento de los datos a una tasa anual de aproximadamente el 50%, las metodologías y tecnologías para la migración de datos que no interrumpan el proceso de negocio se están convirtiendo en una necesidad, para mantener la alta

disponibilidad de las aplicaciones¹. Dentro de las metodologías requeridas, una buena fase de planeación es uno de los aspectos más importantes; otro aspecto de mucha importancia son las pruebas a las bases de datos a emplear y pruebas a los datos a ser migrados, que pueden ayudar a reducir en un alto nivel los problemas con los cuales se enfrentan en etapas de validación y post-migración. Las necesidades tecnológicas cada vez se inclinan más hacia los lados de migraciones online, sin interferir con los procesos diarios del negocio.

Dadas las condiciones anteriores, se ve la necesidad de tener un conocimiento más claro de lo que se está realizando en el medio en lo que respecta a las pruebas para las migraciones y validaciones de datos, para poder tomar decisiones sobre este tipo de proyectos y tratar de orientarlos en la mejor forma posible, dado que el desarrollo del tema como tal aún es muy joven y apenas se está comenzando a explorar; adicionalmente, las posibles metodologías existentes son propietarias de las organizaciones que las emplean.

La primera parte de este proyecto estará dedicado a la definición y explicación de temas que guardan una relación directa con el tema principal, tales como: qué son datos, qué son las bases de datos, qué es migración y validación de datos, qué es la calidad de datos y qué son las pruebas de software y de bases de datos.

¹ “Matching Data Migration to Availability Requirements” – Softek Technical White Paper. 2005.

En la siguiente sección, dividida en dos partes, donde la primera tratará el tema acerca del estado de las pruebas para la migración y validación de datos a nivel nacional, y la segunda, el estado de las mismas a nivel internacional. En ambos casos se explicarán las metodologías para la migración y validación de datos, empleadas por diversas empresas en el medio y se analizará la fase de pruebas de las mismas, en busca de sus mejores prácticas y metodologías o prácticas en común que demarquen tendencias.

La última parte del trabajo estará compuesta por la descripción de algunas herramientas para la migración y validación de datos, se analizará cuáles son las ventajas y desventajas para los proyectos de migración de datos, los tipos de migración soportados, los diversos tipos de pruebas que les permiten a sus usuarios, etc. Además se analizarán algunos aspectos que deben tener en cuenta los Project Manager al momento de elegir herramientas para facilitar las migraciones de datos.

1. OBJETIVOS

1.1. GENERAL

Identificar, analizar y documentar las técnicas y metodologías actuales para realizar Pruebas de Bases de Datos para la Migración y Validación de Datos, incluyendo herramientas adicionales empleadas, para proporcionar así elementos que puedan servir como componentes para el desarrollo de una metodología unificada para Pruebas de Bases de Datos.

1.2. ESPECÍFICOS

- Identificar las técnicas actuales de Pruebas de Bases de Datos para la Migración y Validación de datos.
- Identificar las metodologías y mejores prácticas para las Pruebas de Bases de Datos en algunas empresas locales.
- Comparar las técnicas y/o metodologías actuales con las empleadas en las empresas.

2. GLOSARIO

A continuación se definen, en orden alfabético, algunos términos empleados a través del desarrollo de este documento, que son claves y básicos para un mejor entendimiento del mismo.

A

Aplicación (Application): Programa de software diseñado para realizar una tarea o conjunto de tareas específicas solicitados por el/los usuarios.

B

Bases de Datos (Database): Conjunto de datos relacionados que se almacenan de forma que se pueda acceder a ellos de manera sencilla, con la posibilidad de relacionarlos, ordenarlos con base en diferentes criterios, etc. La diferencia entre una BD y otro sistema de almacenamiento, es que en la base de datos los datos se almacenan de forma que cumplen tres requisitos básicos: no redundancia, independencia y concurrencia.

C

Calidad de Datos: Tiene como objetivo que la información sea la misma en todas las áreas o sistemas utilizados por la compañía, particularmente en un data

warehouse. Aseguramiento de la calidad de datos (Data Quality Assurance, DQA) es el proceso de verificar la confiabilidad y efectividad de los datos. Mantener la calidad de los datos requiere revisar los datos periódicamente y limpiarlos. Normalmente esto implica actualizarlos, estandarizarlos, y eliminar registros duplicados para crear una sola vista de los datos, aunque esté almacenado en múltiples sistemas.

Conversión de datos (Data Conversion): Cambio de un dato digital de un formato a otro para que pueda ser empleado por otro software o pueda ser empleado por un dispositivo de salida específico.

Copia de seguridad (Backup): Hacer una copia de seguridad o copia de respaldo. Se refiere a la copia de datos de tal forma que estas puedan ayudar a restaurar un sistema después de una pérdida de información. La copia de seguridad es útil por dos razones: para restaurar un ordenador a un estado operacional después de un desastre, y para restaurar un pequeño número de ficheros después de que hayan sido borrados o dañados accidentalmente.

Corrupción de datos: La corrupción de los datos se refiere a errores en los datos de la computadora, que ocurren durante la transmisión o la recuperación, introduciendo cambios involuntarios a los datos originales.

Comunidad Ágil (Agile Community): Son el grupo de personas y/o organizaciones que se unen para trabajar en el desarrollo, evolución y aplicación de metodologías ágiles, tales como SCRUM, Extreme Programming (XP), Rational Unified Process (RUP), entre otros. Se basan en los valores y principios definidos en el Manifiesto Ágil.

D

Datos (Data): Unidad mínima de información, sin sentido en sí misma, pero que adquiere significado en conjunción con otras precedentes de la aplicación que las creó. Conjunto de símbolos que unidos de cierta forma dan un significado coherente y lógico.

Data Warehouse (DW): Un almacén de datos es una colección de datos orientadas a un dominio, integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones de la empresa u organización. Se trata, sobre todo, de un expediente de una empresa, más allá de la información transaccional y operacional, almacenado en una base de datos diseñada para favorecer análisis y la divulgación eficientes de datos (especialmente OLAP, procesamiento analítico en línea).

E

ERP (Enterprise Resource Planning): Software que unifica todas las necesidades de todos y cada uno de los departamentos en un único sistema, centralizando la información de la empresa y soportando todas las necesidades particulares de cada departamento.

El software ERP podría cubrir, desde las aplicaciones del departamento financiero hasta las de Recursos Humanos, pasando por las de almacén y logística.

ETL: Son las siglas en inglés de Extraer, Transformar y Cargar (*Extract, Transform and Load*). Proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart,

o data warehouse para analizar, o en otro sistema operacional, para apoyar un proceso de negocio.

Exactitud de Datos: (o precisión) Es uno de los componentes de la calidad de los datos. Se refiere a si los valores de los datos almacenados para un objeto son los valores correctos y debe estar representado de manera consistente y sin ambigüedades.

I

Integración de Datos (Data Integration): Es el movimiento de datos entre dos sistemas coexistentes, que sean compartidos y analizados para ayudar a la toma de decisiones o para la migración a un nuevo sistema.

ISO: International Organization for Standardization - Organización Internacional para la Estandarización). Su nombre ISO significa "igual" en griego. Fue fundada en el año 1946 y unifica a más de cien países. Se encarga de crear estándares o normas internacionales.

M

Metodologías: Procesos verificados en la planeación, definición, análisis, diseño, construcción, pruebas e implementación de un sistema.

Migración (Migration): Traslado de una aplicación de un sistema a otro en condiciones de compatibilidad. Migrar es también elevar una versión de un producto software a otra de más alto nivel, o bien el movimiento de una

arquitectura a otra, por ejemplo, de un sistema centralizado a otro con una estructura basada en el modelo cliente/servidor.

Migración de datos (Data Migration): Es la transferencia de datos entre tipos de dispositivos de almacenamiento, formatos o sistemas de cómputo. La migración de datos usualmente es programática, para obtener de esta manera una migración automatizada, liberando así recursos humanos de labores tediosas de transferencia. Es requerido cuando las organizaciones cambian sus sistemas o realizan actualizaciones de las mismas.

Modelo Relacional: Para la gestión de una base de datos es un modelo de datos basado en la lógica de predicado y en la teoría de conjuntos. Éste es el modelo más utilizado en la actualidad para modelar problemas reales y administrar datos dinámicamente.

O

OLAP: (On-Line Analytical Processing o Procesamiento Analítico En Línea) Se introdujo originalmente en 1994, es una contraparte para el soporte a la toma de decisiones del OLTP. OLAP le permite a los usuarios derivar información e inteligencia de negocios de sistemas de Data Warehouse suministrando herramientas para la consulta y análisis de información.

OLAP permite la vista multidimensional y análisis de datos para procesos de soporte a la toma de decisiones. Es una categoría de aplicaciones y tecnología para la recolección, administración, procesamiento y presentación de datos multidimensionales para propósitos de análisis y de administración. Estas

aplicaciones emplean bases de datos orientados al análisis, es decir, están optimizados para permitir mejor desempeño en trabajos de consulta.

OLTP: (On-Line Transactional Processing o Procesamiento Transaccional En Línea) Sistemas operacionales para la recolección y administración de datos en una organización, tales como el procesamiento de ordenes de venta, inventario, contabilidad, etc. Normalmente ofrece poca o ninguna capacidad analítica.

OLTP se refiere a una clase de sistemas que facilitan y administran aplicaciones orientadas a las transacciones, típicamente para el procesamiento del ingreso de datos y consultas de transacciones. Estas aplicaciones emplean bases de datos orientadas a las transacciones, es decir, que están optimizados para manejar gran número de transacciones (INSERT, UPDATE, etc.).

P

Plan de Pruebas (Test Plan): Una aproximación sistemática para probar un sistema (máquina o software). El plan contiene el flujo de eventos y tareas que formarán parte de las pruebas.

Pruebas (Tests): Es la tarea de comprobar el adecuado funcionamiento de un sistema y/o los módulos que lo componen, para garantizar su adecuado funcionamiento basado en los requerimientos de los usuarios.

R

Reglas de Validación: (de datos) Una regla de validación es un criterio utilizado en el proceso de validación de datos, realizado luego de que los datos hayan sido ingresados al sistema e involucra un programa de validación, el cual es distinto a una verificación formal, donde se determina que el funcionamiento de un programa cumple con sus objetivos. El método se basa en verificar que los datos sigan los parámetros apropiados definidos por el analista de sistemas; la decisión de si el dato es válido o no es realizado por el programa de validación. Por esta razón, durante el diseño de un sistema se establecen definiciones de datos, los cuales colocan límites en lo que constituye un dato válido. Empleando estas definiciones de datos, un software de validación puede realizar las pruebas de validez.

S

Sistema Gestor de Base de Datos (Database Management System): Sistema de Gestión de Bases de Datos. Conjunto de programas que permite crear una base de datos, manipular la información que contiene y realizar todas las tareas de administración necesarias para mantenerla operativa.

SQL (Structured Query Language): Es un estándar en el lenguaje de acceso a bases de datos. Originalmente era un lenguaje de acceso al sistema de gestión de bases de datos denominado DB2 en plataformas 390 de IBM. En la actualidad está adoptado por ISO.

V

Validación: Es el proceso de asegurar que un sistema opera sobre datos limpios, correctos y útiles. Emplea rutinas para verificar los datos ingresados al sistema. Hay diversos tipos de validación, dependiendo del tipo de datos que se esté validando y el contexto en el que se encuentran esos datos en el sistema.

3. ANTECEDENTES

Antes de analizar más a fondo las migraciones de datos, sus principales obstáculos y las razones por las cuales fallan, se necesita conocer algunas de las circunstancias que llevan a la creación de un proyecto de migración de datos (ya sea como un proyecto individual o como parte de un proyecto de tecnología más complejo).

Las razones para migrar datos contenidos en uno o más sistemas pueden depender de las necesidades del negocio o inclusive de los cambios y avances tecnológicos que se estén presentando en el entorno, pero aparte de estas, hay otras razones para llevar a cabo una migración. Razones entre las que se encuentran:

- Unificación de sistemas.
- La principal razón por la cual se presentan las migraciones es el refrescamiento de las tecnologías, como la actualización de las versiones de los softwares de aplicaciones y de las bases de datos.
- Implementación de nuevos sistemas.
- Implementación de un sistema analítico (OLAP).
- Cambios de hardware.

Las migraciones no se realizan de manera esporádica o sólo una vez durante proyectos de tecnología específicos; esa es una de las percepciones erróneas sobre las migraciones de datos. En realidad, 20% de las migraciones son

semanales y otro 21% son mensuales². No son ni siquiera el 50% de todas las migraciones que se realizan, pero si se consideran todos los factores involucrados en una migración, sí es una cifra relativamente elevada.

Entre las cosas más comunes que se presentan durante una migración de datos es que en muchas ocasiones los esfuerzos están más enfocados en el sistema nuevo y los requerimientos de y para el mismo, y muy pocos esfuerzos son dedicados a los procesos de migración, pruebas y validación de los datos³.

En un estudio realizado por la empresa **Softek** en el 2003⁴, se identifican los principales dolores de cabeza de un proceso de migración de datos. Inicialmente habían encuestado 280 compañías, de las cuales 75% admitieron tener problemas con el proceso de migración de datos. Mientras que 54% identificaron que sus problemas eran de carácter técnico por incompatibilidades entre origen y destino de los datos, 58% de ellos dijeron que el tiempo de inactividad extendido o inesperado era un resultado muy común entre los problemas que se presentan durante la migración. Algo interesante que ese estudio preliminar encontró, es que 31% de los encuestados están demorando la adquisición de nuevos equipos o sistemas, simplemente por los problemas resultantes del proceso de migración.

² Según un estudio realizado por ESG (Enterprise Strategy Group) en el cual se encuestaron aproximadamente 500 personas en Norte América a finales del 2004.

³ "The Serial Data Migration Dilemma" – Robert Alan. Technical Services Quarterly. 2002.

⁴ "Data Migration Headaches Underscored by Softek Survey". – Jon William Toigo. 2005. <http://www.esj.com/news/article.aspx?EditorialsID=1266>.

Durante este primer análisis del estudio solicitado por Softek, encontraron que la mayoría de los encuestados coincidían con el hecho de que una migración de datos exitosa requiere una planeación efectiva y mucho personal, identificando así dos suministros muy escasos o muy ocupados en los departamentos de tecnología de las empresas; esos dos suministros son *Tiempo* y *Personal Técnico*. El 75% de los encuestados sostiene que el tiempo mínimo requerido para planear el proceso de migración es de dos semanas, mientras que el otro 25% sostiene que dicho tiempo debería ser de cuatro semanas, para estar del lado seguro. Casi un 50% afirma que un proyecto de migración de mediano tamaño⁵ requiere como mínimo cuatro personas trabajando en el equipo de migración de tiempo completo.

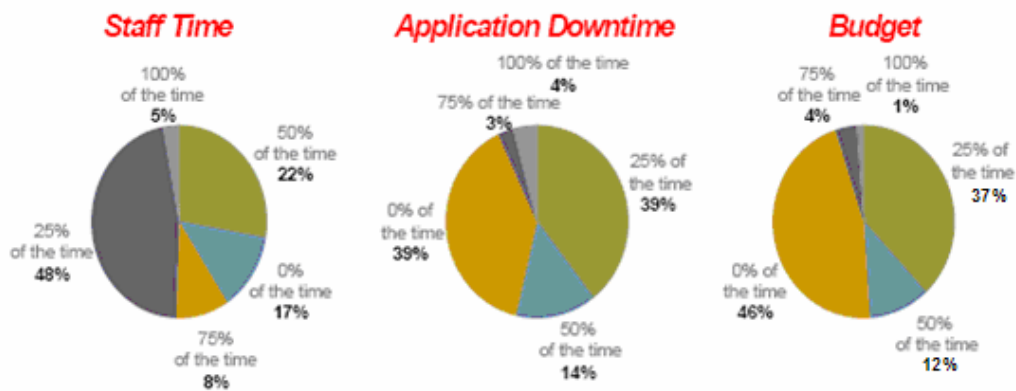
Otro aspecto interesante que se encontró, fue que el conocimiento por parte de los usuarios sobre la existencia de herramientas para automatizar la migración de datos era mínimo. Mientras un 80% de los usuarios de sistemas Mainframe z/OS empleaban herramientas de migración, solamente 50% de los usuarios de plataformas abiertas (UNIX/Windows) emplearon o emplean dichas herramientas. Mientras que otro 38% de usuarios de diversos sistemas no sabían sobre la existencia de herramientas para facilitar la migración.

Este estudio fue solicitado en un principio por Softek, pero la empresa ESG (Enterprise Strategy Group) tomó los resultados y realizó un análisis más detallado junto con su propio estudio en el campo, para presentarle a la comunidad tecnológica la situación actual y la importancia de las migraciones de datos. Los

⁵ Definiendo como migración de mediano tamaño, una migración con más de 2 y menos de 4 sistemas de origen de datos.

resultados mencionados anteriormente fueron los resultados preliminares de dicho estudio, pero de igual manera se identificaron más hechos de interés sobre los procesos de migración. A continuación se presentará el análisis y algunas de las observaciones relevantes realizadas por ESG.

Tres de los aspectos más importantes dentro de cualquier proyecto, sea de migración o implementación, son el costo, el tiempo de inactividad de las aplicaciones y el tiempo empleado del equipo de TI. En la **Figura 1** se muestran los resultados sobre el uso o el estado de esos tres factores.



83% excede el tiempo de los recursos originalmente planeado
61% excede el tiempo de inactividad originalmente planeado
54% excede el presupuesto originalmente planeado

Figura 1: Estadísticas de desempeño de proyectos de migración de datos.

Fuente: “Data Migration: Everyone is doing it – but are they doing it right?” realizado por Nancy Hurley.

En el estudio se encontró, que en lo que respecta al tiempo empleado del equipo de TI, sólo un 17% cumplió sin sobrepasar el tiempo estimado del proyecto, mientras que un 48% se excedió en un 25% del tiempo estimado; 22% se excedió

en un 50% del tiempo, 8% se excedió en un 75% del tiempo y 5% se excedió en un 100% del tiempo estimado; es decir, que de todas las compañías encuestadas, 83% de ellas excedían el tiempo de migración planeado originalmente. En gran parte ese exceso de tiempo es debido a que gran parte de las migraciones se realizan durante los fines de semana, para tratar de minimizar los demás efectos y riesgos de la migración, como el tiempo de inactividad de las aplicaciones.

El tiempo de inactividad esperado en cualquier proyecto de TI es uno de los factores más importantes a considerar. Encontraron que sólo 39% de los procesos de migración cumplen con los tiempos estimados de inactividad de las aplicaciones; eso deja un 61% de procesos que exceden el tiempo originalmente planeado y deseado. Este es un porcentaje demasiado alto, considerando la importancia de las aplicaciones tecnológicas dentro de las organizaciones hoy en día. En un estudio separado, ESG encontró que el tiempo de inactividad promedio tolerado en la mayoría de las aplicaciones es de no más de 4 horas. En la **Figura 2** se observan las tolerancias en los tiempos de inactividad entre organizaciones grandes y empresas de mediano tamaño.

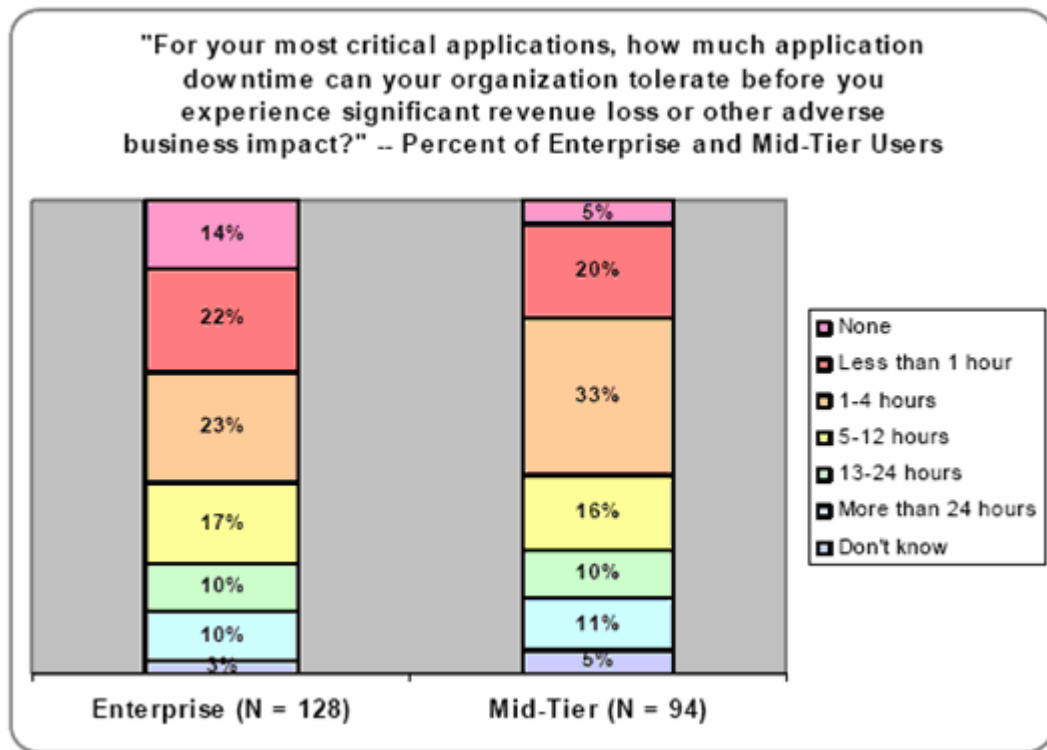


Figura 2: Tolerancia de los Tiempos de Inactividad.

Fuente: "Data Migration: Everyone is doing it – but are they doing it right?" realizado por Nancy Hurley.

De las organizaciones grandes encuestadas por ESG, sólo el 23% (aproximadamente 30) toleran un tiempo de inactividad entre 1 y 4 horas, mientras que en organizaciones de mediano tamaño, el 33% (aproximadamente 31) toleran un tiempo de inactividad entre 1 y 4 horas. De ese punto en adelante los demás porcentajes de tolerancia son similares a medida que incrementa el número de horas, es decir, su tolerancia va reduciendo. Otra gran diferencia es que las empresas más grandes prefieren no tener tiempos de inactividad (un 14% de ellas) en sus aplicaciones, en comparación con las empresas de mediano tamaño (un 5% de ellos).

El costo es el otro factor de alarma; en gran parte por los niveles administrativos de la organización. Aunque sus porcentajes son relativamente inferiores a los dos anteriormente mencionados, sigue estando por encima del 50% de los encuestados. Solamente 46% de los proyectos o procesos de migración no exceden el presupuesto asignado, dejando aun un 54% de proyectos que exceden el presupuesto asignado.

En la **Figura 3** se muestran los problemas con los cuales se enfrentan los proyectos de migración de datos, solamente un 25% de dichos proyectos no presentan problemas.

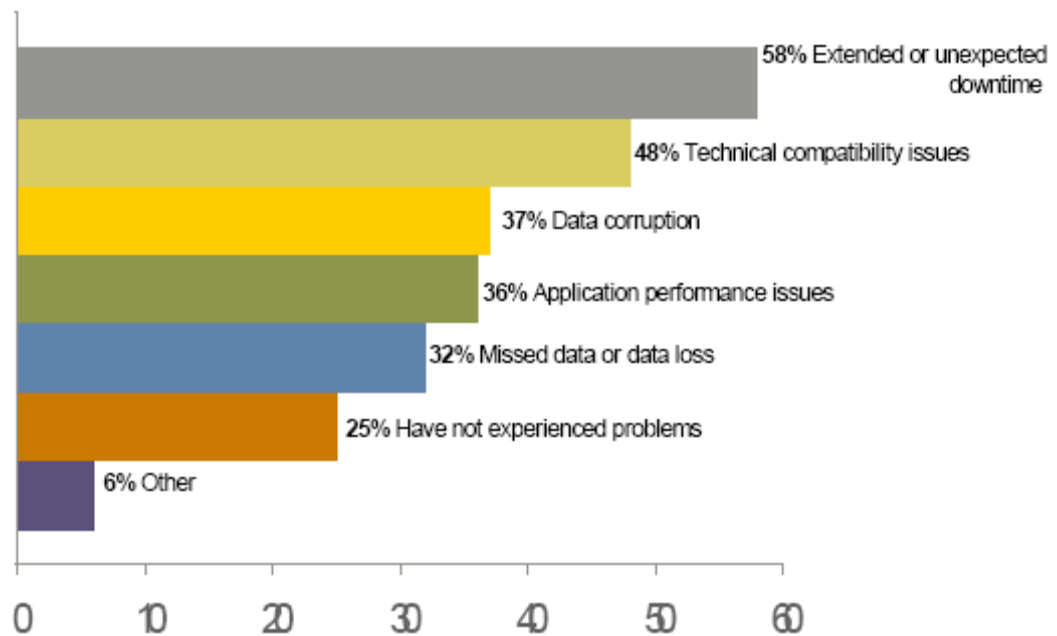


Figura 3: Estadísticas sobre problemas experimentados durante migraciones.

Fuente: "Data Migration: Everyone is doing it – but are they doing it right?" realizado por Nancy Hurley.

El problema más grande es el de los tiempo de inactividad, mostrando que el 58% se encuentra con tiempos de inactividad extendidos o inesperados, mientras que 48% con problemas de compatibilidad técnica, 37% con problemas de corrupción de datos, 36% con problemas de desempeño de la aplicación, 32% de ellos dejaron datos en el sistema antiguo o perdieron datos durante el proceso, y el 6% restante presentó problemas diversos de menor impacto con relación a las primeras mencionadas. El estudio de igual manera mostró que la mayoría de las migraciones se realizan los fines de semana.

Aproximadamente 69% de dichos problemas tienen que ver directamente con el dato, y gran parte de dichos problemas se pueden resolver hasta cierto grado si durante el proceso de planeación de la migración, las pruebas a realizar sobre los datos son bien estructuradas y orientadas.

En la actualidad, una gran parte de las migraciones se están realizando offline y en horas de poco tráfico en el sistema, es decir, durante las noches o los fines de semana. En el estudio individual que realizó ESG, encontraron que aproximadamente 80% de las aplicaciones eran colocadas fuera de servicio u offline máximo 6 horas antes de realizar la migración. En este mismo estudio encontraron que 54% de las migraciones demoraban de 1 a 6 horas, mientras que un 13% de ellas demoraban más de 6 horas. De las cosas interesantes que encontraron durante este estudio es que *“hay una correlación clara entre la*

*cantidad de tiempo de inactividad que podría ser tolerada y las implementaciones de migraciones offline vs. online*⁶.

Muchas de las organizaciones que no empleaban herramientas de migración online, sostenían que las razones por las cuales no las adoptaban eran en primera instancia para evitar más costos al proceso y en segunda instancia, porque no conocían sobre dichas herramientas, como mostró el estudio anteriormente.

Adicionalmente, ESG analizó el estado del mercado y los costos de las herramientas de migración online y encontraron que los tipos de herramientas y sus aplicaciones varían, al igual que sus precios, pero tenían una característica en común: sus costos estaban disminuyendo. La decisión ya pasaría a depender de la cantidad de datos y del impacto permitido sobre las aplicaciones.

En este punto hay que recordar cuáles son los tres factores más importantes en un proceso de migración de datos: tiempo de inactividad, horas de trabajo del personal y costos. El estudio muestra que todos estos factores son excedidos más del 50% de las veces, todos estos excesos se traducen en costos para la organización, costos distintos a los originalmente planeados para la migración, colocando en perspectiva si el costo de emplear una herramienta de migración online es superior a los costos de no tenerla.

⁶ “Data Migration: Everyone is doing it – but are they doing it right?” Nancy Hurley. – Marzo 2005.

Aunque también encontraron que el mayor costo de una migración es la falta de procedimientos y pruebas para la validación de los datos, 55% de los encuestados dependen de las “pruebas” de los usuarios en el sistema de producción para determinar si la migración fue exitosa. Podrían pasar días o semanas antes de descubrir algún problema con la migración, y eso en sí es un costo, el arreglo de dicho problema y el tiempo de inactividad de las aplicaciones mientras se realiza el arreglo. ESG sostiene que la mayoría de las aplicaciones para migraciones online realizan pruebas propias sobre los datos, para verificar lo mínimo, como integridad, y permiten realizar verificaciones programáticas a los datos, para así tratar de reducir las probabilidades de problemas después de la migración.

Unas de las principales conclusiones a las que llegaron los de ESG y Softek con estos estudios, es que las migraciones pueden ser muy riesgosas y muy costosas, particularmente para las organizaciones que tienen que realizar migraciones periódicamente.

Como se observó, en los estudios realizados y las conclusiones a las cuales llegaron las empresas que los realizaron, los procesos de migración de datos son muy costosos en tiempo y dinero. Algunos de los riesgos más importantes que corren las organizaciones cuando inician un proceso de migración son los siguientes:

- Problemas que se originan durante el tiempo en el cual el sistema permanece fuera de servicio, ya sea el sistema de origen o el sistema destino de los datos. El problema es básicamente de coordinación, estimación de los tiempos de migración y administración de los mismos.
- Problemas funcionales inesperados en el sistema destino de los datos.

Y el problema más importante y de mayor peso es la pérdida y corrupción de los datos. Este es el problema que genera las mayores pérdidas para las organizaciones en lo que respecta a tiempo y dinero.

4. JUSTIFICACIÓN

Las necesidades de almacenamiento de datos de las organizaciones han ido cambiando a medida que evolucionan los sistemas de información y las necesidades mismas de la organización para la utilización de dichos datos. Esta es posiblemente la principal razón por la cual muchas organizaciones y sus departamentos de TI se ven involucrados cada día más en proyectos o procesos de migración de datos, ya sea para la actualización o mejora de sus sistemas de almacenamiento actuales o de las aplicaciones, o para la implementación de nuevas tecnologías que le den mayor valor y utilización a los datos de la organización. Hasta hace algunos años los datos eran simplemente eso, datos, información histórica; con la evolución tecnológica y la creciente necesidad de las organizaciones de poseer una mayor ventaja competitiva con respecto a la competencia, los datos pasaron de ser simplemente eso, datos, a ser un activo valioso para la organización, luego de ser procesados y convertidos en información.

Algunas de las tecnologías disponibles para soportar las necesidades para la toma de decisiones y ayudar a obtener esa ventaja competitiva en las organizaciones y facilitar su adecuada gestión, son soluciones como ERP's, Business Intelligence, CRM's, Herramientas de reportes, etc. Son herramientas las cuales requieren que los datos sean de la mejor calidad posible.

Muchas organizaciones en la actualidad están intentando implementar mejores prácticas, tanto a nivel de procesos como en cuestiones de sistemas de información y tecnologías. Estos enfoques los están llevando a crear proyectos de

implementación de sistemas como ERP's, que funcionan sobre bases de datos; estos proyectos involucran migración de la información que tienen de sus sistemas actuales a los nuevos y la validación de los mismos. De igual manera que se prueba un software antes de hacerlo disponible a los usuarios, se necesitan realizar pruebas a las bases de datos para determinar si pueden soportar las necesidades de los clientes y se necesita probar que los datos que se trasladaron de un sistema a otro están donde deben estar y en el formato en el que deben estar. Estas pruebas son conocidas como Pruebas de Bases de Datos para la Migración y la Validación de los Datos. En la actualidad, las técnicas y/o metodologías para realizar estas pruebas no son muy conocidas, y para su efecto son poco usadas. Las pruebas que se realizan en la actualidad en este tipo de implementaciones son en su mayoría pruebas de desarrollo, tales como:

- ¿Se está trasladando la información completa? ¿Se crearon la misma cantidad de registros en el sistema nuevo que en el actual?
- ¿Se está almacenando correctamente?

Entre otras, pero aparte de eso, las pruebas son mínimas.

Es un hecho que las migraciones de datos en la actualidad son de los proyectos más costosos en una organización, pero su importancia sigue creciendo. Lo que se ha podido observar, de los estudios y encuestas realizados por Softek, ESG y las entrevistas realizadas como parte de este proyecto de grado, es la falta de metodologías estándares o procedimientos de prueba estándares para la migración y validación de datos, aunque están presentes, su estructuración y diseño, aún no están al nivel de las demás metodologías de pruebas de software. Las metodologías de pruebas que se han visto son metodologías diseñadas por las empresas proveedoras de servicios de migración, como Softek, o por

empresas como Oracle, quienes presentan una metodología para probar sus bases de datos.

Algunos de los principales problemas con los cuales se enfrentan las empresas durante y después de una migración pueden ser reducidas si las pruebas están bien orientadas para detectar ciertos problemas. Algunos de esos problemas son los siguientes⁷:

- Pérdida de datos.
- Corrupción de datos.
- Pobre calidad de datos.
- Periodos extendidos de tiempos de inactividad de los sistemas involucrados.
- Pérdida de negocio y clientes por la inactividad de los sistemas.
- Sobrecostos para arreglar problemas post-migración.

Al observar la falta de una estandarización de pruebas para los procesos de migración y validación de datos y los problemas mencionados anteriormente, se ve la necesidad de conocer qué se está haciendo en el entorno actual, a nivel de pruebas como tal, y cómo las herramientas tecnológicas ayudan en dicho proceso. Con este proyecto intentamos identificar aquellas técnicas que se han desarrollado, en su mayoría propietarias, y las mejores prácticas, o al menos las más comunes que hay en el entorno, para que pueda servir de base para trabajos futuros sobre el tema, como por ejemplo el desarrollo de una Metodología de Pruebas de Bases de Datos.

⁷ “The Hidden Cost of Data Migration” by Softek – 2006.

5. MARCO TEÓRICO

Cuando se habla de pruebas para migración y validación de datos, entre las primeras cosas que se tienen que aclarar son las definiciones de datos, bases de datos, migración, validación y pruebas.

A continuación se presentarán definiciones y explicaciones relevantes para el desarrollo posterior del tema.

5.1. ¿QUÉ SON DATOS?

Entre las definiciones más sencillas de dato existentes, algunas de las que con más frecuencia se pueden observar, son las siguientes:

*“Un dato es la unidad o cantidad mínima de información no elaborada, sin sentido en sí misma, pero que convenientemente tratada se puede utilizar en la realización de cálculos o toma de decisiones.”*⁸

*“Por dato, queremos decir hechos conocidos que pueden ser registrados y que tienen un significado implícito”.*⁹

El autor Joe Celko¹⁰ presenta una definición menos formal dentro de la cual incluye la idea de una base de datos; su definición es la siguiente:

⁸ Definición obtenida de <http://es.wikipedia.org/wiki/Dato>

⁹ Traducción de “Fundamentals of Database Systems” – R. Elmasri y S. Navathe. 1994.

¹⁰ Traducción de “Joe Celko's Data and Database: Concepts in Practice” – Joe Celko, 1999.

“Hechos que pueden ser representados con medidas empleando escalas o con un sistema formal de símbolos dentro del contexto de un modelo formal. El modelo, se supone, representa algo llamado ‘el mundo real’, de tal manera que cambios en los hechos de este mundo real se reflejan en cambios en la base de datos”.

Una definición más formal sobre qué es un dato, es la presentada por la ISO:

“Es una representación de hechos, conceptos o instrucciones de manera formalizada, adecuado para comunicación, interpretación, o procesamiento por medios humanos o automáticos”¹¹.

La mayoría de los autores y definiciones encontradas tratan de hacer una distinción entre dato e información, de tal manera que no sean confundidas y vistas como semejantes. En una de las definiciones se menciona que el *“dato es la unidad o cantidad mínima de información”*, la información es obtenida de los datos, de un recuento histórico y análisis de los mismos. Un dato es cuantitativo por naturaleza, mientras que la información es tanto cuantitativa como cualitativa. Con la información y suficiente conocimiento del campo al que corresponde, el usuario o la persona que lo esté observando puede tomar decisiones, mientras que con los datos solamente una toma de decisiones no es posible, son simples hechos estadísticos y tabulaciones.

En la actualidad, la toma de decisiones estratégicas para las organizaciones se está realizando con base en la información proveniente de los sistemas de información empleados. La información presentada a los usuarios es obtenida de

¹¹ Traducción de ISO/IEC 11179-1:1999. "Information Technology - Specification and Standardization of Data Elements. Part 1: Framework for the Specification of Data Elements".

los datos mediante un procesamiento de los mismos y el uso de tecnologías que permiten alcanzar el nivel de detalle requerido. Para realizar una buena toma de decisiones, las organizaciones requieren no sólo de la experiencia de sus tomadores de decisiones, sino también requieren cierta certeza sobre la confiabilidad de la información que se les presenta, donde la confiabilidad de la información presentada está directamente ligada a la calidad de los datos almacenados, cómo son obtenidos y las manipulaciones por las que pasan antes de ser presentados a los usuarios finales.

La calidad de los datos está obteniendo cada vez mayor importancia y atención dentro de las organizaciones. Se están dando cuenta que los problemas con la calidad de los datos están derivando en problemas con la información presentada y en consecuencia, afectando las decisiones tomadas con base en las mismas, provocando pérdida de ingresos, tiempo y oportunidades. El costo de una mala calidad de los datos está oculto y no es tan obvio si no se le está buscando.

La calidad de los datos se ve afectada por los orígenes y la manipulación de los mismos. Sus orígenes pueden ser humanos, provenientes de un sistema antiguo, de un proceso de abstracción y generación de nuevos, etc. El tener un dato incorrecto en un nivel muy bajo puede no tener un impacto inmediato, pero a medida que se reutilice o se propague el uso de ese dato en particular, se propaga con él el error y puede llegar a afectar las decisiones dentro de una organización.

El tener claro el concepto de un dato y la importancia de una buena calidad del mismo es de gran importancia, y el interés en lograr un mayor nivel de calidad seguirá creciendo a medida que hayan más avances tecnológicos que empleen dichos datos y los usuarios sigan demandando más de sus sistemas y sigan demandando mayor visibilidad del comportamiento actual, pasado y futuro de la organización, como una forma de apoyo al proceso de toma de decisiones.

5.2. ¿QUÉ ES UNA BASE DE DATOS?

Se podría decir que una base de datos es un conjunto de datos que poseen una característica en común, que se almacenan de forma que se pueda acceder a ellos de manera sencilla, con la posibilidad de relacionarlos y ordenarlos con base en diferentes criterios.

Toda base de datos se diseña, se construye y se carga con datos para un propósito específico. Las bases de datos pueden ser de cualquier tamaño y tener diversos grados de complejidad.

Las bases de datos sirven para apoyar el proceso de toma de decisiones en las empresas, ya que en esta se encuentra la información e históricos de la empresa, y si se realiza un análisis de estos, se pueden elaborar estrategias que apoyen el desarrollo del negocio y con estas lograr ventajas competitivas.

“Una base de datos es una colección de archivos relacionados que almacenan tanto una representación abstracta del dominio de un problema del mundo real,

cuyo manejo resulta de interés para una organización, como los datos correspondientes a la información acerca de sí mismo. Tanto la representación como los datos están sujetos a una serie de restricciones, las cuales forman parte del dominio del problema, cuya descripción está también almacenada en esos archivos”.

Las bases de datos computarizadas se pueden crear y mantener con un grupo de programas de aplicación escritos específicamente para esa tarea, o bien mediante un sistema de gestión de bases de datos.

Un sistema de gestión de bases de datos (SGBD) es un conjunto de programas que permite a los usuarios crear y mantener una base de datos. El SGBD es un software de propósito general que facilita el proceso de definir construir y manipular bases de datos para diversas aplicaciones. Para definir una base de datos hay que especificar los tipos de datos, las estructuras y las restricciones de los datos que se almacenarán en ella. Construir una base de datos es el proceso de guardar los datos en algún medio de almacenamiento controlado por el SGBD.

5.3. HISTORIA DE LAS BASES DE DATOS

Las bases de datos tuvieron sus orígenes en los años 60, cuando se empezaron a usar las máquinas que codificaban la información en tarjetas perforadas por medio de agujeros. Las bases de datos se crean con el objetivo de almacenar grandes cantidades de datos que antes se almacenaban en libros, lo que era lento, costoso y complejo.

Se dice que los sistemas de bases de datos tienen sus raíces en el proyecto estadounidense Apolo de mandar al hombre a la luna. En aquella época, no había ningún sistema que permitiera gestionar la inmensa cantidad de información que requería el proyecto. La primera empresa encargada del proyecto, NAA (North American Aviation), desarrolló un *software* denominado GUAM (General Update Access Method) que estaba basado en el concepto de que varias piezas pequeñas se unen para formar una pieza más grande, y así sucesivamente hasta que el producto final esté ensamblado. Esta estructura, que tiene la forma de un árbol, es lo que se denomina una estructura jerárquica. A mediados de los Sesenta, IBM se unió a NAA para desarrollar GUAM, en lo que ahora se conoce como IMS (Information Management System). El motivo por el cual IBM restringió IMS al manejo de jerarquías de registros fue el de permitir el uso de dispositivos de almacenamiento en serie, más exactamente las cintas magnéticas, ya que era un requisito del mercado por aquella época. Se desarrolló IDS (Integrated Data Store), de General Electric, a mediados de los sesenta y este fue dirigido por Charles Bachmann, quien fue uno de los pioneros en bases de datos. IDS era conocido como un sistema de bases de datos en red. Los sistemas en red se desarrollaron para representar relaciones entre los datos más complejos y para imponer un estándar en las bases de datos.

Con el objetivo de definir unas especificaciones estándar que permitieran la creación de bases de datos y el manejo de los datos, se formó un grupo denominado DBTG (Data Base Task Group), que estaba conformado por representantes del Gobierno de EEUU y representantes del mundo empresarial. El DBTG presentó un informe final, pero no fue aceptado formalmente por la ANSI (American National Standards Institute); muchos sistemas se desarrollaron siguiendo esta propuesta del DBTG. Estos sistemas se conocen como sistemas de red.

En 1970 se escribió un artículo presentando el modelo relacional; este fue realizado en los laboratorios de investigación de IBM, en 1970. Este artículo también presentaba los inconvenientes de los sistemas jerárquico y de red. Entonces se comenzaron a desarrollar muchos sistemas relacionales, apareciendo los primeros a finales de los Setenta y principios de los Ochenta. Uno de los primeros es System R, de IBM, que se desarrolló para probar la funcionalidad del modelo relacional, proporcionando una implementación de sus estructuras de datos y sus operaciones. Esto condujo a dos grandes desarrollos:

- El desarrollo de un lenguaje de consultas estructurado denominado SQL, que se ha convertido en el lenguaje estándar de los sistemas relacionales.
- La producción de varios SGBD relacionales durante los años Ochenta, como DB2 y SLQ/DS de IBM, y ORACLE de ORACLE Corporation.

Los modelos relacionales también tienen sus fallos, siendo uno de ellos su limitada capacidad de modelar datos. Se han realizado investigaciones para tratar de resolver este problema. Chen¹² en 1976 presentó el modelo entidad-relación, que es la técnica más utilizada en bases de datos. En 1979, Codd¹³ intentó subsanar algunas de las deficiencias de su modelo relacional con una versión extendida denominada RM/T (1979) y más recientemente RM/V2 (1990).

Como respuesta a la creciente complejidad de las aplicaciones que requieren bases de datos, han surgido dos nuevos modelos: El modelo de datos orientado a objetos y el modelo relacional extendido. Sin embargo, a diferencia de los modelos

¹² Chen, Peter P; "The Entity-Relationship Model: Toward a Unified View of Data"; ACM. Transactions on Database Systems, Vol.1, Nº.1, March 1976.

¹³ Científico informático británico (23 de agosto de 1923- 18 de abril de 2003) contribuyó con la teoría de bases de datos relacionales.

que los preceden, la composición de estos modelos no está clara. Esta evolución representa la tercera generación de los SGBD.

5.4. TIPOS DE BASES DE DATOS.

→ JERÁRQUICA:

Desarrollada y comercializada en los 60. Las bases de datos jerárquicas suministran una estructura de árbol para la interconexión de datos. La estructura de árbol significa que los datos son enlazados en un formato uno – a – muchos. En otras palabras, sólo existe una trayectoria desde un nodo. Esto crea una estructura rígida en la cual el sistema debe mover un registro a la vez.

Esto hace las bases de datos jerárquicas “inadecuadas” para las transacciones en línea, a causa de que no pueden acceder rápidamente los datos almacenados en lugares lógicos más bien distantes. Sin embargo, las bases de datos jerárquicas son todavía usadas para grandes transacciones en lote (batch), que requieren enlaces de datos a través de una trayectoria lógica limitada.

La jerarquía existente entre los tipos de objetos que se manipulan y las dependencias existentes, hacen que sea imposible el acceso directo a instancias de cada uno de ellos, con lo que se pierde en independencia y facilidad de uso.

→ EN RED:

Las bases de datos en red fueron desarrolladas y comercializadas en los años 70, como un medio para solucionar la baja velocidad de los productos desarrollados en el modelo jerárquico. Como se sugiere en el nombre, las estructuras en red utilizan una estructura de muchos a muchos. Esto significa que un solo archivo puede estar lógicamente enlazado a muchos archivos (en contraste con el enlace único que existía bajo el modelo jerárquico). Esto garantiza un acceso más rápido a los registros encadenados en un momento dado y por consiguiente un incremento de la productividad. La limitación de este sistema está en que todavía requiere de una estructura rígida que debe ser definida por el usuario, los registros deben ser accedidos uno a la vez, resultando en un acceso secuencial. Las bases de datos en red son usadas en transacciones en lote que requieren “poca interacción en línea”. Esta base de datos presenta una productividad relativamente baja.

→ RELACIONAL:

Fueron desarrolladas en los años 70 y comercializadas en los 80. Las bases de datos relacionales suministran una trayectoria aleatoria (random) para acceder a archivos de datos. Con las bases de datos relacionales es posible entrelazar archivos a través de registros comunes encontrados en cada archivo. Esto posibilita ingresar a cualquiera de los archivos sin tener que especificar primero una trayectoria de datos. Esto reduce la complejidad en la interacción del usuario.

Además permiten realizar consultas sin tener que conocer completamente la estructura interna de la base de datos. Por la capacidad de entrelazar trayectorias

múltiples, las bases de datos relacionales son excelentes para transacciones en línea.

Una base de datos relacional (RDBMS) es una colección de tablas (archivos) que están lógicamente conectados entre sí. Está compuesta por:

Tablas, que son conocidos también como archivos o relaciones. Las tablas se componen de: Filas, también conocidas como registros o tuplas. A su vez, los registros se componen de: columnas, también conocidas como campos o atributos.

Una ventaja de las bases de datos relacionales es que todas las entidades, aunque en realidad sean interrelaciones, se representan de la misma manera, es decir, mediante tablas.

Uno de los problemas que poseen los sistemas relacionales es el uso de los lenguajes de manipulación y definición de datos, que aunque son muy simples de usar por un usuario directamente, son difíciles de insertar en un lenguaje de programación convencional.

→ ORIENTADAS A OBJETOS:

Las Bases de datos orientados a objetos se propusieron con la idea de satisfacer las necesidades de las aplicaciones más complejas. El enfoque orientado a objetos ofrece la flexibilidad para cumplir con algunos de estos requerimientos, sin estar limitado por los tipos de datos y los lenguajes de consulta disponibles en los sistemas de bases de datos tradicionales.

“Una base de datos orientada a objetos es una base de datos que incorpora todos los conceptos importantes del paradigma de objetos:

- *Encapsulación* - Propiedad que permite ocultar la información al resto de los objetos, impidiendo así accesos incorrectos o conflictos.
- *Herencia* - Propiedad a través de la cual los objetos heredan comportamiento dentro de una jerarquía de clases.
- *Polimorfismo* - Propiedad de una operación mediante la cual puede ser aplicada a distintos tipos de objetos.”

Como cualquier Bases de Datos programable, una Base de Datos Orientada a Objetos (BDOO) proporciona un ambiente para el desarrollo de aplicaciones y un depósito persistente listo para su explotación. Una BDOO almacena y manipula información que puede ser digitalizada (presentada) como objetos, además proporciona un acceso ágil y permite una gran capacidad de manipulación.

→ DEDUCTIVAS:

Las Bases de Datos Deductivas surgieron de la aportación de la Lógica y la Inteligencia Artificial a las Bases de Datos. De hecho, tan importante es la lógica matemática en este modelo que a veces se les denomina Bases de Datos Lógicas. Nacen de la necesidad de almacenar y utilizar “conocimiento” de una manera eficiente.

“Un Sistema de Bases de Datos que tenga la capacidad de definir reglas con las cuales deducir o inferir información adicional a partir de los hechos almacenados en las bases de datos, se llama Sistema de Bases de Datos Deductivas.”¹⁴

La importancia de las bases de datos deductivas está en la posibilidad de inferir nuevo conocimiento que pueda ser usado en la toma de decisiones. Además pretende deducir qué hechos son ciertos o no, y en qué circunstancias.

Las Bases de Datos Deductivas son muy utilizadas en las áreas de: Inteligencia Artificial, Sistemas Expertos, Representación del Conocimiento, Tecnología de Agentes, Interacción de Datos e Inteligencia de Negocios.

5.5. ¿QUÉ ES MIGRACIÓN DE DATOS?

En términos generales la palabra migración puede significar:

“Movimiento masivo de animales o humanos de un lugar a otro”¹⁵.

“Movimiento periódico de seres vivos de un área a otra, en muchas ocasiones en respuesta a cambios ambientales”¹⁶.

Esas son definiciones de la palabra migración como tal. En el entorno de sistemas tiene un significado muy parecido, con diversas aplicaciones e implicaciones. Entre las definiciones de migración para el entorno sistémico, podemos citar el definido por la Universidad de Cornell, como la definición más básica de la palabra migración aplicable en el entorno de sistemas:

¹⁴ http://www.quadernsdigitals.net/datos_web/hemeroteca/r_1/nr_502/a_6850/6850.htm.

¹⁵ Traducción de <http://ecohealth101.org/glossary.html>.

¹⁶ Traducción de http://www.apsu.edu/wet/OLD_PROJECT_WET_WEBSITE/whatis.html.

“Preservación de un contenido digital, donde la información de fondo se mantiene, pero los formatos más viejos y las estructuras internas son reemplazadas por versiones más recientes”¹⁷.

En este caso en particular, lo que interesa es la migración de datos y para ello una definición próxima a lo buscado, es la siguiente:

“Movimiento de los datos de una base de datos a otra”.

“Proceso de traducir los datos de un formato a otro”.

Otra definición de migración de datos podría ser:

“Conjunto de procesos por medio de los cuales los datos son obtenidos, transformados y transferidos entre dos o más aplicaciones: Data Warehousing/Inteligencia de Negocio, esfuerzos de re-ingeniería de procesos del negocios, interfaces Inter/Intra organizacionales, etc.”¹⁸

La migración de datos por lo general es necesaria cuando una organización decide cambiar el sistema de información que venía empleando, cuando decide cambiar el sistema gestor de base de datos o cuando se actualizan los sistemas actuales y los datos en su forma actual son incompatibles con el sistema nuevo.

La migración de datos es por lo general un proceso programático, que permite un nivel de automatización alto, evitando así tener que hacer esa conversión de sistemas de forma manual, es decir, re-digitar la información nuevamente.

¹⁷ Traducción de <http://www.cs.cornell.edu/wya/DigLib/MS1999/glossary.html>

¹⁸ Traducción de "Why Data Migration Projects Fail" – Joe Hudicka.

En adición a la migración de datos, en el entorno de sistemas hay otros tipos de migración, como la migración de aplicaciones completas, migración de un hardware a otro, entre otros.

5.6. TIPOS DE MIGRACIÓN DE DATOS

Algunas empresas proveedoras de servicios de migración o inclusive algunas de las casas desarrolladoras de motores de bases de datos y aplicaciones, han clasificado los tipos de migración y las herramientas que tienen disponibles para ayudar con los procesos de migración.

Un ejemplo de esa clasificación es la realizada por Oracle Corporation, que sugiere dos tipos de migraciones:

- (a) *Migración de la base de datos*: es un proceso de migración en el cual se lleva la información de la base de datos de un proveedor distinto a ORACLE al motor de base de datos ORACLE.
- (b) *Migración de la Base de Datos y de la Aplicación*: migración de la base de datos de otro proveedor y las sentencias de acceso a la base de datos de la aplicación a la plataforma de ORACLE. Las aplicaciones pueden contener información de conexión, sentencias SQL dinámicas y llamadas a procedimientos almacenados que tendrían que ser modificados.

Pero de forma general, se pueden identificar dos tipos de migración, dependiendo de su tiempo (de uso o ejecución):

- (1) *Una Única Migración*: Es un proceso de migración de datos que se lleva a cabo una sola vez, sin importar la razón por la cual se esté realizando la migración (cambio de aplicación, cambio de bases de datos, cambio de hardware, etc.). Es un proceso que solo se realizará una vez dentro de la organización, sin incluir las pruebas del proceso que se realizaron, previo a la migración del sistema de producción¹⁹.
- (2) *Migración Continua*: Son procedimientos de migración de datos que se realizan continuamente dentro de la organización. Estas migraciones continuas se pueden presentar a través de interfaces entre sistemas que necesitan comunicar información continuamente entre ellos o los procesos ETL (Son las siglas en inglés de Extraer, Transformar y Cargar) utilizados para actualizar los Data Warehouse.

Sin importar la clasificación de los tipos de migración, siempre existirán migraciones de un tipo de base de datos a otro, es decir, puede ser una migración de una base de datos relacional a una base de datos orientado a objetos, o a una base de datos geoespacial. Sin importar el tipo de base de datos del cual se está migrando al cual se va a migrar, los principios para la migración y validación de los datos son los mismos, y las pruebas siguen los mismos fundamentos y principios, pero durante el desarrollo del plan de migración como tal, la implementación de cada uno de esos pasos varía, según las necesidades y características de la base de datos respectiva.

¹⁹ "Data migration. Database *Programming & Design*" - Moriarty, T. y Hellwege, S. 1998.

5.7. MITOS SOBRE LAS MIGRACIONES DE DATOS

1. *Una migración por fases es más seguro que una única migración.* **FALSO.** La clave para el éxito de una migración es un enfoque cuidadoso y profesional en la especificación, construcción, pruebas y verificaciones post-migración. Una migración por fases es técnicamente más compleja que una única migración:
 - Las interfaces deben ser divididas para que los dos sistemas (antiguo y nuevo) funcionen en paralelo.
 - Los grupos de datos migrados en cada fase deben ser independientes, sin relaciones entre un grupo y el otro, agregando complejidad al proceso de extracción.
 - Grupos de datos deben ser eliminados o inhabilitados en el sistema de origen a medida que son migrados.

Debido a las complejidades que agregan estos aspectos, los riesgos de tener problemas en una migración por fases incrementan en comparación a cuando se realiza una única migración, y los costos tienden a ser más altos.

2. *Todos los datos de origen deben estar limpios antes de iniciar la migración.* **FALSO.** Datos que provocarían problemas en el sistema destino deberían ser arreglados antes de ser cargados. Eso no significa que todos los datos deben ser limpiados por adelantado. Esto eleva el nivel de complejidad y de tiempo necesario para el periodo de desarrollo si se tiene que intentar detectar datos incorrectos pero válidos, por ejemplo, que un nombre esté mal escrito o que una fecha esté equivocada pero en el formato adecuado. Alrededor del 90% de los errores en los datos no pueden ser detectados. Si se quiere tener todos los errores resueltos antes de iniciar la migración, el proyecto se retrasará mientras se realiza un esfuerzo muy grande para corregir los errores (limpiar los datos).

3. *La migración de datos es una tarea puramente técnica.* **FALSO.** Hay muchos desafíos técnicos, pero el conocimiento sobre el negocio también es importante. Un enfoque puramente técnico puede ser peligroso. Habrán muchas opciones para el almacenamiento de los datos en el sistema destino, pero sin la adecuada revisión de los objetivos del negocio, procesos y procedimientos, los datos migrados podrían terminar en lugares difíciles de mantener. El enfoque técnico podría funcionar a primera vista porque los datos han sido migrados, pero cuando los usuarios comienzan a interactuar con ellos es cuando se observan los problemas.

5.8. ¿QUÉ ES VALIDACIÓN DE DATOS?

En seguridad informática, la validación de datos es una de las áreas más importantes a tener en cuenta, especialmente en el desarrollo de sistemas conectados a redes como Internet. Validar datos hace referencia a verificar, controlar o filtrar cada una de las entradas de datos que provienen desde el exterior del sistema.

Es un proceso que asegura que un programa opere sobre datos limpios, correctos y útiles. Emplea rutinas, frecuentemente conocidas como *reglas de validación*, que verifican la exactitud y la significancia del dato en el sistema.

La validación de datos más sencilla, verifica que los caracteres que forman parte del dato provengan de un conjunto de caracteres válidos. Por ejemplo, números telefónicos, contienen los dígitos numéricos y posiblemente los caracteres +, -, (y

) - signos de más, menos y paréntesis -. Una rutina de validación de datos más sofisticada verificaría que el usuario ingrese un código de país válido: el número de dígitos ingresados es igual a la convención para el área o país especificado, etc.

Una validación incorrecta de datos puede llevar a la *corrupción de datos* o a *vulnerabilidades de seguridad*.

Algunos de los métodos más utilizados para realizar validaciones son:

- ↪ *Verificaciones de formato o imagen*: Verifica que el dato esté en un formato específico.
- ↪ *Verificaciones del tipo de dato*: Verifica el tipo de dato ingresado con respecto al esperado y presenta mensajes de error cuando no se cumple.
- ↪ *Verificaciones de rango*: Verifica que el dato se encuentre entre un rango especificado de datos.
- ↪ *Verificaciones de límites*: A diferencia de la verificación de rango, este solo compara con respecto a un límite, superior o inferior.
- ↪ *Verificaciones de presencia (o de datos nulos)*: Verifica que datos de importancia no tengan valores en blanco (o nulos).
- ↪ *Verificaciones de consistencia*: Verifica los campos dentro de la base de datos para asegurar que los datos estén en sus campos correspondientes.

Hay otros tipos de verificaciones que son comunes para la validación de datos en un proceso de migración y validación de datos, pero su uso depende del tipo de dato o migración que se esté realizando. Las validaciones mencionadas anteriormente se encuentran entre las más comúnmente empleadas en dichos procesos.

5.9. ¿QUÉ SON PRUEBAS DE SOFTWARE?

“Una actividad en la cual un sistema o uno de sus componentes se ejecuta en circunstancias previamente especificadas, los resultados se observan y registran, y se realiza una evaluación de algún aspecto”.

(Software testing, prueba de software). Es el proceso empleado para identificar la correctitud, completitud, seguridad y calidad en el desarrollo de un software para computadoras.

El proceso de testing es una investigación técnica que intenta revelar información de calidad acerca del producto de software con respecto al contexto en donde operará.

El testing de un producto de software es uno de los pasos más complejos e importantes en el desarrollo de software. Esto incluye el proceso de encontrar errores en el software; pero el testeo no sólo se limita a eso. El testeo o prueba de un software se relaciona con atributos como fiabilidad, eficiencia, portabilidad, escalabilidad, mantenibilidad, compatibilidad, usabilidad y capacidad del mismo.

5.10. TIPOS DE PRUEBAS

5.10.1. PRUEBAS DE UNIDAD

La prueba de unidad se plantea a pequeña escala, y consiste en ir probando uno a uno los diferentes módulos que constituyen una aplicación.

Normalmente se realiza una fase informal antes de entrar en la fase de pruebas. La fase informal la lleva a cabo el propio desarrollador y consiste en ir ejecutando el código para convencerse de que "básicamente, funciona". Esta fase suele consistir en pequeños ejemplos que se intentan ejecutar. Si el módulo falla, se suele utilizar un depurador para observar la evolución dinámica del sistema, localizar el fallo, y repararlo.

Después de realizar esto, cuando el módulo parece presentable, se entra en una fase de prueba sistemática. En esta etapa se buscan fallos siguiendo algún criterio. Los criterios más utilizados son los denominados de caja negra y caja blanca.

Una prueba es de caja negra cuando se descartan los detalles del código y se limita a lo que se ve desde el exterior. Lo que se pretende es descubrir casos en los que el módulo no hace lo que se espera de él.

Una prueba de caja blanca es cuando se mira detenidamente el código que está escrito y se intenta determinar en que punto del código se presenta la falla.

5.10.2. PRUEBAS DE INTEGRACIÓN

Las pruebas de integración se llevan a cabo durante la construcción del sistema, involucran a un número creciente de módulos y terminan probando el sistema como conjunto.

Estas pruebas se pueden plantear desde un punto de vista estructural o funcional.

Las pruebas estructurales de integración son similares a las pruebas de caja blanca, pero en lugar de referirse a sentencias del lenguaje, se refiere a llamadas entre módulos. Se trata pues de identificar todos los posibles esquemas de llamadas y ejercitarlos para lograr una buena cobertura de segmentos o de ramas.

Las pruebas funcionales de integración son similares a las pruebas de caja negra. En esta se trata de encontrar fallos en la respuesta del módulo, cuando su operación depende de otros módulos. Así se va realizando una prueba a todo el sistema, según las especificaciones del usuario; estas también se realizan con el manual de usuario.

5.10.3. PRUEBAS DE ACEPTACIÓN

Las pruebas de aceptación son las que se plantea el cliente final, que decide qué pruebas va a aplicarle al producto antes de darlo por bueno y pagarlo. El objetivo de quien prueba es encontrar los fallos lo antes posible, en todo caso antes de pagarlo y antes de colocar el programa en producción.

Son básicamente pruebas funcionales, sobre el sistema completo, y buscan una cobertura de la especificación de requisitos y del manual del usuario. Estas pruebas no se realizan durante el desarrollo, pues sería impresentable de cara al

cliente, sino una vez pasadas todas las pruebas de integración por parte del desarrollador.

5.10.4. OTROS TIPOS DE PRUEBAS

☒ Recorridos (walkthroughs):

Es una técnica más aplicada en control de calidad que en pruebas, y consiste en sentar al lado de los desarrolladores una serie de críticos y estos leen el programa línea por línea y piden explicaciones de todo. Entonces puede suceder que simplemente falta un comentario, o que se detecte un error o que el código es muy complejo de entender.

☒ Aleatorias (random testing):

Estas se aplican porque muchos autores consideran que la probabilidad de encontrar un error es igual que si se realizan pruebas aleatorias. Por esta razón comienzan las pruebas con una serie de casos elegidos al azar. Esto mostrará los errores más evidentes. Pero esta prueba no es suficiente.

☒ Solidez (Robustness testing):

En esta se prueba la capacidad del sistema para salir de situaciones provocadas por errores en el suministro de datos. Estas pruebas son importantes en sistemas con interfaz al exterior.

☒ **Stress (Stress testing):**

En ciertos sistemas es muy conveniente realizar esta prueba para saber hasta dónde son resistentes, es decir cuántos datos podría procesar, cuánta carga de la cpu.

☒ **Desempeño (Performance testing):**

Esta prueba sirve para saber el tiempo de respuesta, para saber cuánto tiempo tarda el sistema en procesar un determinado número de datos, cuánta memoria necesita para esto, cuánto disco duro.

5.10.5. PRUEBAS DE BASES DE DATOS

A continuación se realiza una descripción de cómo realizar algunas pruebas para las bases de datos:

☒ **Pruebas de integridad:** Las pruebas de integridad de bases de datos verifican la estructura y el formato, la conformidad con limitaciones de integridad, las reglas del negocio y relaciones, corrige controles en las actualizaciones que restauran bases de datos, y la normalización de la base de datos. Se necesitan hacer seis tipos de prueba para verificar la integridad de la base de datos.

- *Integridad de la entidad:* La integridad de la entidad indica que cada fila debe tener siempre una clave primaria.
- *Integridad de la clave primaria:* El valor para cada clave primaria debe ser único y válido.

- *Integridad de la clave de la columna:* Los valores que tiene la columna deben tener unas reglas específicas.
- *Integridad del dominio:* Un dominio es un sistema de datos y las características que describen esos valores.
- *Integridad definida por el usuario:* La integridad que define el usuario se especializa en reglas de validación que van más allá de los estándares de las filas y las columnas.
- *Integridad de referencia:* La clave primaria es un identificador único de una entidad. La clave foránea hace referencia a la clave primaria de otra entidad.

Otras bases de datos incluyen otros test, como:

☒ **Test de control:** Este test incluye las siguientes pruebas:

- *Test de seguridad:* Protegen a la base de datos de accesos no autorizados.
- *Backup Testing:* Verifica la capacidad de hacer un backup del sistema.
- *Test de recuperación:* Verifica la restauración de la base de datos a un estado anterior que sea conocido, después de que una falla lo haya dejado en un estado poco confiable.
- *Pruebas de concurrencia:* Asegura que procesos paralelos tales como consultas y actualizaciones, no interfieran entre sí.
- *Control de un abrazo mortal:* Asegura que dos procesos concurrentes no formen un “gridlock” y se excluyan mutuamente de una culminación adecuada.

- ⊗ **Verificación del contenido de los datos:** Auditorias periódicas y comparaciones con referencias conocidas.
- ⊗ **Refrescar la base de datos:** Verifica los sistemas externos que refrescan la base de datos, conversiones de datos.
- ⊗ **Uso de datos:** Incluye la verificación y actualización de las bases de datos. El test compara el número de caracteres que puede tener un campo de la interfaz con el número de caracteres que puede almacenar en la base de datos ese campo.

Muchas bases de datos tienen queries que facilitan los test y verifican la actualización y la modificación de los datos en la base de datos.

- ⊗ **Procedimientos almacenados:** Estos procedimientos son almacenados e invocados cuando ocurren triggers específicos de la aplicación.

5.11. ¿QUÉ ES CALIDAD DE DATOS?

Una definición general y básica sobre qué es calidad es presentada por la ISO en el siguiente párrafo:

“La calidad es la suma de todos aquellos aspectos o características de un producto o servicio que influyen en su capacidad para satisfacer las necesidades, expresadas o implícitas” (ISO 8402).

Una definición sobre la calidad de los datos, es la siguiente:

"Un dato tiene calidad si satisface los requerimientos de uso propuestos. Le falta calidad hasta el punto en el que no satisfaga los requerimientos. En otras palabras, la calidad de los datos depende tanto del uso propuesto como del dato mismo. Para satisfacer el uso propuesto, el dato debe ser exacto, puntual o a tiempo, relevante, completo, entendible y confiable"²⁰.

Los datos se vuelven más valiosos todo el tiempo, a medida que se encuentran nuevas formas de uso de la información para hacer a la organización más exitosa.

La calidad de los datos es medida con respecto a número de dimensiones: exactitud, relevancia, puntualidad, completitud, entendible y confiabilidad. La dimensión de la exactitud o la precisión es la medida base de la calidad de los datos. Si los datos no están bien, las otras dimensiones tienen poca importancia²¹.

Las dimensiones en las cuales la calidad de los datos es medida, es una parte básica de la definición de calidad de datos, cada uno de ellos tiene una razón y una definición que los identificará y diferenciará. A continuación se presentan las definiciones de dichas dimensiones.

- ✓ **Exactitud:** Es la medida en la cual el dato es correcto para el negocio, según las definiciones del negocio mismo. La medición de esta dimensión es de las más complejas, debido a que el dato puede tener un nivel de exactitud

²⁰ "Data Quality. The Accuracy Dimension" – Jack E. Olson. Morgan Kaufmann Publishers. 2003. Pág. 24.

²¹ "Data Quality. The Accuracy Dimension" – Jack E. Olson. Morgan Kaufmann Publishers. 2003. Pág. 17

diferente dentro del mismo negocio dependiendo del enfoque que se le de o el área del negocio que este empleando el dato.

- ✓ **Relevancia:** Es la medida en la cual los datos son de importancia para las áreas del negocio que la empleen o necesiten; se puede dar el caso de que para un área no es de mucha importancia, pero para otro si lo es.
- ✓ **Puntualidad:** Es la medida en que el dato este disponible en el momento en que se le necesite para la toma de decisiones. En sistemas OLTP esta medida suele darle al dato una baja calidad en especial durante los cierres de mes y cierres fiscales, cuando datos de todas las áreas están ingresando y están siendo corregidas, hasta después de las fechas limites en muchas ocasiones. Mientras que en un sistema OLAP la dimensión de la puntualidad no tiene un efecto tan grande en la calidad de los datos debido a que las decisiones que se tiene que realizar con estos datos es a largo plazo y por lo general son con datos históricos ya procesados y verificados.
- ✓ **Compleitud:** Es la medida en la cual un dato esta “completo”, el dato no se ingreso a medias o no se actualizó cuando se presentó un cambio, en el transcurso del tiempo para las necesidades del negocio y las distintas áreas que la emplean en su proceso de toma de decisiones u operaciones diarias.
- ✓ **Entendible:** Es la medida en la cual los datos, su definición, uso y propósito, son conocido y aplicados correctamente por todos los miembros de la organización, ya sea para operaciones diarias, toma de decisiones o análisis de los mismos datos.
- ✓ **Confiabilidad:** Es la medida en la cual el dato o el contenido de la base de datos reflejan las transacciones o movimientos actuales del negocio y sus usuarios se sienten cómodos con los datos presentados y no tengan razones para desconfiar de los mismos.

La falta de atención a la calidad de los datos en sistemas organizacionales puede tener o causar mayores problemas a medida que pasa el tiempo. El intentar analizar y resolver problemas de calidad de datos en aplicaciones que ya se encuentran en producción consume mucho tiempo y es muy costoso. El problema debe ser corregido desde un principio en el origen del dato antes de que ingrese al sistema de producción. La falta de exactitud en los datos no solo afecta la confianza del usuario final en las aplicaciones de TI, sino que también puede tener un impacto financiero significativo en el negocio; los costos de una mala calidad de datos normalmente están ocultos y no son obvios a los que no lo están buscando.

Un informe del Data Warehouse Institute²² sobre la calidad de datos estima que la baja calidad de datos de los clientes le cuesta a empresas estadounidenses un aproximado de \$611 billones de dólares al año en gastos de correo (por re-envíos, etc.), impresiones y costos adicionales en personal. El mismo informe afirma que casi el 50% de las empresas encuestadas no tienen planes para manejar o mejorar la calidad de los datos. A la vez, casi la mitad de los encuestados cree que la calidad de sus datos es peor de lo que piensan los demás.

El informe del Data Warehouse Institute afirma que *"aunque algunas empresas entienden la importancia de tener una alta calidad de los datos, la mayoría no tienen conciencia del verdadero impacto al negocio de datos defectuosos o por debajo de los estándares"*. Las empresas necesitan concientizarse de este hecho e informarse más sobre las técnicas disponibles para lograr mejoras y los pasos

²² "Data Quality and the Bottom Line" - Eckerson, Wayne W. Data Warehousing Institute, January 2002.

que pueden tomar para evitar caer en graves problemas con la calidad de sus datos.

El enfoque de las empresas y la comunidad tecnológica en la calidad de los datos ha ido creciendo y seguirá creciendo, dado que la economía del mundo depende cada día más de la información de los sistemas de las diversas empresas que la conforman. Empresas con acceso a información precisa y a tiempo tienen una ventaja significativa en el negocio con respecto a la competencia.

La culpa por la baja calidad de los datos por lo general recae sobre el área de TI. Sin embargo, los datos son creados por personas fuera de TI y son utilizados por personas fuera de TI. El área de TI es responsable por la calidad de los sistemas que almacenan y mueven los datos. La mayoría de los problemas se encuentran fuera de TI. Algunos de dichos problemas son, entre otras:

- Requerimientos de baja calidad.
- Pruebas de aceptación del sistema de baja calidad.
- Procesos de creación de datos de baja calidad.

Hay otros dos factores de gran importancia que afectan la calidad de los datos, que no se encuentran dentro de las organizaciones, pero que forman parte de la evolución natural de la tecnología:

1. Las rápidas implementaciones y cambios de sistemas hace muy difícil controlar la calidad.
2. Los métodos, estándares, técnicas y herramientas para controlar la calidad han evolucionado a un paso más lento que los demás sistemas.

En general, se clasifica la fuente de datos de baja calidad o con problemas en cuatro áreas, que son:

- Ingreso inicial de los datos al sistema.
- Corrupción de los datos.
- Movimiento y Re-estructuración de los datos.
- La utilización de los datos.

La **Figura 4** muestra esas cuatro áreas y las formas en las que pueden ocurrir los problemas.

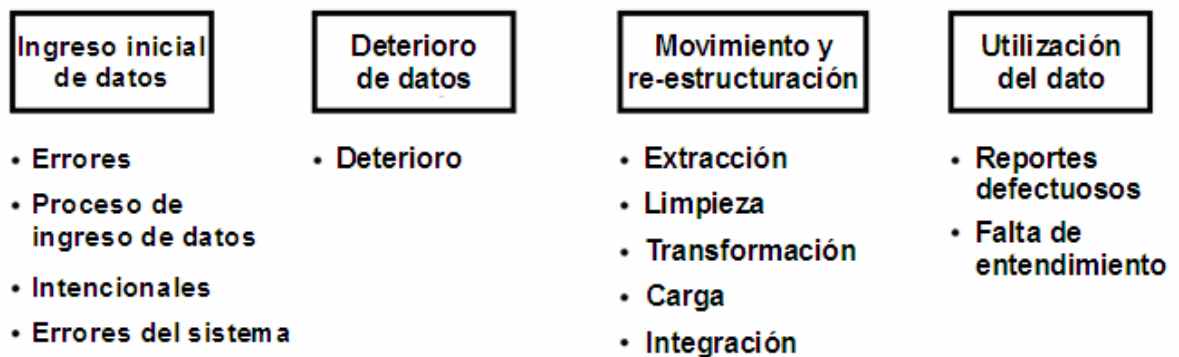


Figura 4: Traducción de las Áreas donde se generan datos incorrectos.

Fuente: "Data Quality. The Accuracy Dimension". Jack E. Olson. Morgan Kaufmann Publishers. Página 43.

En el proceso de ingreso inicial de los datos se pueden cometer errores en la digitación, en la definición del proceso de ingreso de datos, pueden ser errores intencionales o pueden ser errores del sistema al almacenar los datos.

En el deterioro de los datos, ocurre a través del tiempo, sea porque la definición del dato ha cambiado ya sea en el mundo exterior o en la base de datos.

En el movimiento y la re-estructuración, se puede presentar durante los procesos de implementación y migración a nuevos sistemas, cuando se extraen los datos, o cuando se hace una "limpieza" para adecuarlos para el nuevo sistema, cuando se les realizan transformaciones, cuando se les carga o cuando se integran diversos sistemas con diferentes estructuras.

Cuando están utilizando los datos, pueden encontrarse con que el reporte no está bien construido y está generando problemas o simplemente no se entiende el uso propuesto del dato.

La calidad de los datos es fundamental para el éxito de cualquier organización, dado que tener acceso a información exacta y completa es fundamental para la ayuda en una mejor toma de decisiones estratégicas. Datos con calidad, apoyan y fortalecen virtualmente todas las funciones de negocios y son especialmente importantes para las nuevas iniciativas del negocio.

Dedicar tiempo a calidad de datos mejora este activo y el retorno de la inversión es alto, aunque no siempre visible a simple vista.

En este caso, el área que nos interesa es el del movimiento y re-estructuración de los datos.

6. ESTADO DEL ARTE DE LAS PRUEBAS PARA LA MIGRACIÓN Y VALIDACIÓN DE DATOS

Para el desarrollo de este proyecto de grado se optó por una metodología de recolección de información que involucra entrevistas y encuestas a las empresas, en Medellín y algunas a nivel internacional, que han realizado o realizan migraciones de datos o pruebas para bases de datos.

En algunas situaciones, no fue posible concretar una entrevista con las empresas, por lo que se decidió enviar por correo electrónico una encuesta que apuntaba y ayudaba a resolver aspectos de importancia para la investigación. Luego de recibir las encuestas tramitadas, estas eran analizadas y si alguna tenía información que necesitaba ser profundizada o si se consideraba que era necesario realizar otra entrevista con dicha empresa, se contactaba a la empresa para obtener una entrevista personal.

Con este proyecto de grado se busca conocer el estado actual de las pruebas para bases de datos para la migración y validación de datos en Medellín y a Nivel Internacional. Para ello en la ciudad de Medellín se entrevistaron a 7 personas de 7 empresas, todas pertenecientes al área de tecnología en la ciudad, unos se dedican al desarrollo de software, otros a la implementación de tecnologías para terceros, pruebas de software y una institución educativa. Entre las empresas entrevistadas, se encuentran:

- Softland S.A.
- DDB Worldwide Colombia
- Intergupo S.A.
- Choucair Testing S.A.

- TREBOL Software S.A.
- Pragma S.A.
- Universidad EAFIT

Las personas encuestadas y entrevistadas en las distintas empresas se desempeñan en los siguientes cargos:

- Analistas/Programadores
- Ingenieros de Desarrollo
- Testers
- Administrador de Data Warehouse
- Ingeniero de Infraestructura

Para obtener la información sobre el entorno internacional, se enviaron encuestas y se realizaron telé conferencias con algunas de las personas. De igual forma parte de la información recolectada sobre el entorno internacional fueron white papers de la empresa, documentación en línea y otros análisis relacionados con el tema de las pruebas de bases de datos.

A nivel internacional se entrevistaron y encuestaron a personas de 4 empresas en los Estados Unidos y Canadá, todas pertenecientes al área de la tecnología como función principal del negocio. Las personas encuestadas laboran en las siguientes organizaciones:

- Computer Sciences Corporation Federal Division
- Oracle Corporation – Consulting Division
- Lucidity Consulting Group
- IBM Canadá

Entre los cargos desempeñados por las personas dentro de las empresas mencionadas anteriormente se encuentran:

- Senior IT Project Manager
- Oracle Consultant
- Practice Leader Agile Development

6.1. ENTORNO NACIONAL

Las empresas y personas entrevistadas han realizado proyectos de migración de datos, ya sea como proyectos individuales o como parte de otros esfuerzos tecnológicos.

Las preguntas realizadas durante las entrevistas y las formuladas en los cuestionarios eran de carácter general, para poder obtener una idea de qué se estaba haciendo, cómo se estaba haciendo y con qué herramientas se contaban para la automatización de las mismas. Una de las razones por las cuales las preguntas se formularon de manera tan general, fue por cuestiones de privacidad de la información de las diferentes empresas entrevistadas, porque en algunos casos las metodologías o enfoques empleados eran propietarios y confidenciales, así que la información disponible se limitaba.

Para determinar cuál es el estado del arte en Medellín, se iniciaron las entrevistas tratando de determinar qué tipos de migraciones habían realizado, si estaban participando en alguno actualmente o si habían proyecciones de proyectos futuros que involucrarán el tema. Para determinar dicho estado en la ciudad de Medellín

se realizaron las siguientes preguntas, tanto en las entrevistas como en las encuestas:

- *Si habían o no realizado migraciones de datos y qué tipos de migraciones habían realizado.*

Por la información obtenida durante las entrevistas y en los cuestionarios, se observó que en Medellín la principal razón por la cual se estaban realizando migraciones de datos en la actualidad eran para unificar los sistemas de información empresariales; la segunda razón por la cual se realizaban las migraciones era para la actualización y/o cambio de software y/o de hardware y por qué se estaban realizando migración o implementaciones a nuevos sistemas de información. Las tendencias en implementaciones de nuevas tecnologías están apuntando hacia las ERP's, como la migración hacia SAP que se iniciará en las empresas que forman parte del Grupo Empresarial Antioqueño, y las soluciones para Inteligencia de Negocios. Los procedimientos de ETL para la implementación de proyectos de Inteligencia de Negocios es la cuarta razón por la cual se realizan migraciones de datos en la ciudad de Medellín. Aunque en los próximos años se proyecta un crecimiento en las implementaciones de tecnologías para la Inteligencia de Negocios, debido a la gran acogida que el tema está teniendo con muchas directivas locales.

Lo siguiente que se necesitaba determinar era si se realizaban validaciones luego de las migraciones, con preguntas como:

- *¿Realizan validaciones de los datos migrados?*
- *¿Qué tipo de validaciones son realizadas?*

De las personas entrevistadas y que respondieron el cuestionario; solamente una persona de las entrevistadas no realiza validaciones luego de las migraciones de datos, el resto de los encuestados sí las realiza. Entre las formas de validación más comunes que se realizan, se pueden citar las siguientes:

- Integridad de datos.
- Validaciones por conteo de registros.
- Pruebas de totales en campos que contienen información financiera o numérica.
- Flujo de información desde la fuente hasta el destino donde serán cargadas.
- Validación de formatos de fecha, tildes y caracteres especiales.
- Validación del funcionamiento normal del sistema.
- Validaciones de datos faltantes.

Las validaciones que se realizan en otras empresas en Medellín son muy similares.

Otro aspecto que nos interesaba determinar era el tipo de problemas con los cuales se encontraban y si de alguna forma las validaciones y pruebas realizadas guardaban una relación con dichos problemas y la búsqueda de la solución de las mismas. Para ello, se les preguntó:

- *¿Qué tipos de problemas han tenido? Se les suministró una lista de problemas básicos como tipos de datos inconsistentes, datos incompletos, etc. Adicionalmente tenían la posibilidad de listar problemas adicionales con los cuales se habían enfrentado.*

A continuación se listan los problemas más comunes durante los procesos de migración de datos, desde los más comunes hasta los menos comunes o los específicos a un tipo de negocio.

1. Datos incompletos.
2. Tipos de datos inconsistentes.
3. Longitud de datos y datos en los campos incorrectos.

Adicionalmente, los encuestados mencionaron otros problemas que eran más comunes a su negocio, como:

- Problemas de totales, es muy común cuando los ETL's poseen cálculos muy complejos. Esto, algunos lo mitigan probando el ETL por secciones para detectar dónde se puede estar corrompiendo el dato.
- Problemas de calidad de datos.
- Problemas por el desconocimiento de las reglas del negocio.

La siguiente pregunta tenía que ver con el proceso de migración y validación de datos como tal, se buscaba determinar si se empleaban metodologías definidas para llevar a cabo los procesos de migración. Para ellos se realizaron preguntas tales como:

- *¿Tienen una metodología específica para la Migración y Validación de datos?*
- *¿La metodología es propietaria?*
- *Adicionalmente, se solicitaba la explicación de la metodología, si era posible.*

Curiosamente la mitad de las personas entrevistadas dijeron que no tenían una metodología y explicaron que la forma en que realizaban las migraciones era basado en experiencias vividas y lecciones aprendidas en proyectos similares.

Solamente en una de las empresas se comentó en forma general en qué consistía la metodología que empleaban, la cual no era una metodología propietaria; una descripción breve de la metodología es la siguiente:

“La metodología se basa principalmente en los conteos de campos y sumalización, después de cargar información por medio de ETL’s. En nuestro caso particular, la migración y carga de información la hacemos bajo los estándares de BI (Business Intelligence – Inteligencia de Negocios), y la carga masiva se certifica garantizando que el total de registros cargados es el esperado y que sus totales corresponden a los cálculos realizados.”

Las demás empresas que tenían una metodología específica para los procesos de Migración y Validación de datos, no podían compartir dicha información por cuestiones de confidencialidad.

Luego de determinar si empleaban metodologías para los procesos de migración, el siguiente punto era de vital interés para el desarrollo de este tema, se necesitaba saber si realizaban pruebas para las migraciones y qué tipos de pruebas realizaban, con preguntas como:

- *¿Durante el proceso/metodología de Migración y Validación de Datos incluye o realiza pruebas (crea planes de prueba para la migración y validación de datos)?*
- *De igual manera si realizaban pruebas se les preguntaba que tipos de pruebas y las razones por las cuales eran realizadas.*

De los entrevistados, solamente en una empresa no se realizaban pruebas, los demás sí las realizaban; a continuación serán descritas:

- Las pruebas que se realizan son muy sencillas, se buscan datos puntuales antes y después de la migración para ver qué cambios han sufrido.
- Pruebas funcionales de los procesos críticos de los diferentes aplicativos que conforman el núcleo del negocio.
- Las pruebas incluyen la verificación del origen y destino, básicamente, observando el insumo en el origen y resultado en el destino.
- Se verifican los datos, procesos, tablas, etc., dependiendo del tipo de migración que se vaya a realizar.
- Pruebas de conteo de registros.
- Pruebas de totales en campos que contienen información financiera.
- Pruebas de los usuarios al ejecutar reportes.

El objetivo principal de las pruebas citadas por uno de los entrevistados era el de *“minimizar riesgos y garantizar el normal funcionamiento de la organización, una vez realizada la migración”*.

El siguiente aspecto que se necesitaba averiguar era si habían implementados planes de calidad de los datos dentro de las organizaciones en las cuales se realizaron las migraciones. Para ello se realizó la siguiente pregunta:

- *¿Hay un plan de Calidad de Datos para verificar los datos migrados?*

Solamente una empresa contestó afirmativamente, sobre tener implementado un plan de calidad.

Entre los aspectos necesarios para definir el estado actual de las pruebas para la migración y validación de datos, se necesita conocer el abanico de herramientas

empleadas actualmente en Medellín para llevar a cabo los proyectos de migración de datos. Se les preguntó lo siguiente:

- *¿Qué tipo de herramientas son utilizadas para los procesos de Migración y Validación de Datos? Aclarando que herramientas se refiera a cualquier software y/o lenguaje de programación empleado con fines de realizar la migración y validación de datos.*

A continuación se listarán las herramientas más comúnmente empleadas en las empresas para la migración de datos:

- Scripts.
- C#
- Funciones en SQL bajo DB2.
- Las migraciones realizadas se ejecutan por medio de scripts para programas RPG.
- Para la migración se utilizan las herramientas proporcionadas por los diferentes motores de datos. En Oracle, por ejemplo EXPORT e IMPORT. En SQL Server, la DTS (Data Transfer Service).
- Se usan las herramientas propias de cada base de datos, como SQL Plus (Oracle), Management Studio (SQL Server) y otras herramientas como TOAD.

Se puede observar que ninguna de las herramientas mencionadas anteriormente es de migración, es decir, las migraciones se están realizando casi manualmente; aunque una de las personas entrevistadas mencionó que en la actualidad estaban considerando un par de herramientas diseñadas específicamente para realizar procesos de migración o de ETL.

Adicionalmente, se les preguntó si:

➤ *¿Las herramientas realizan o permiten realizar pruebas?*

Todos respondieron que no realizaban pruebas, pero que los lenguajes de programación (Scripts, SQL, etc.) les permitían desarrollar pruebas para realizarlas sobre los datos en la migración o en momentos posteriores. Uno de los entrevistados respondió: *“En el caso de los script sería posible realizar una prueba de desempeño para obtener el rendimiento esperado al momento de ejecutar el proceso de migración”*.

Al final se trató de determinar qué tipo de migración realizaban o qué les permitía realizar las herramientas que usaban:

➤ *¿Las migraciones son online u offline?*

Solamente una de las personas entrevistadas respondió que realizaba las migraciones online, pero estas migraciones eran en horarios no laborales. Los demás realizaban las migraciones offline y por lo general también en horarios no laborales o fines de semana, dependiendo del tamaño de la migración.

En síntesis, se puede decir que en la actualidad en Medellín se tiene una tendencia similar a la que se está observando y se espera que siga ocurriendo en el resto del mundo en los siguientes años, es decir, un incremento en los proyectos que involucran migraciones de datos. Lo que se alcanza a observar en la ciudad de Medellín, es que apenas se están comenzando a desarrollar metodologías propias o a emplear metodologías de terceros para realizar este tipo de proyectos. Las empresas que tienen sus propias metodologías o emplean cualquier metodología es porque llevan suficiente tiempo en el mercado, se han posicionado, y están comenzando a adoptar mejores prácticas y a mejorar la

calidad de los trabajos que realizan; las demás empresas que no emplean o no tienen una metodología, realizan las migraciones y validaciones de datos basándose en experiencias pasadas y realizan un proyecto completamente basado en necesidades. A nivel de pruebas, las que se realizan son básicas de verificación y validación de datos (formatos de datos, longitud de datos, total de datos y validaciones por parte de los usuarios) y corresponden casi exactamente a los tipos de problemas más comunes con los que se han encontrado en los proyectos. En la actualidad, en Medellín se puede decir que la mayoría de los proyectos de Migración y Validación de Datos son manuales, no emplean herramientas para facilitar el proceso, todos emplean lenguajes de programación o funcionalidades de los motores de bases de datos para extraer e importar manualmente los datos a los sistemas. Hay un desconocimiento muy grande hasta el momento sobre las herramientas disponibles en el mercado para facilitar dichos procesos. Al parecer aún no se ha presentado la necesidad de realizar migraciones completamente online, no al menos en las empresas entrevistadas, aun realizan migraciones offline.

Se podría decir que el estado actual de las Pruebas para la Migración y Validación de Datos y el Proceso de Migración como tal en Medellín, apenas se está desarrollando, tomando forma y preocupándose por la calidad de los datos y los resultados.

6.2. ENTORNO INTERNACIONAL

6.2.1. RESULTADOS DE ENTREVISTAS Y ENCUESTA

Uno de los objetivos de este proyecto era determinar el estado del arte de las pruebas para la migración y validación de datos. Para ello, una encuesta o cuestionario, similar al que se completó en algunas empresas en la ciudad de

Medellín, fue enviada a varias empresas a nivel internacional, más específicamente en Canadá y Estados Unidos. A continuación se presentarán los hallazgos de los cuestionarios.

Al igual que con los cuestionarios en Medellín, se inició tratando de identificar las principales razones por las cuales se están realizando migraciones de datos a nivel internacional, preguntándoles:

- *¿Han realizado migración de datos?*
- *¿Qué tipos de migraciones de datos han realizado? Tanto en la encuesta como en las entrevistas se presentaban varias opciones como implementación de nuevos sistemas, actualizaciones de hardware y/o software, entre otros. Pero se podían agregar o mencionar mas razones.*

Encontramos que las principales razones por las cuales se presentan las migraciones de datos son por las actualizaciones o cambios de software y/o hardware y por la migración o implementación de nuevos sistemas. Las siguientes razones en orden descendente son, por la unificación de sistemas y por las implementaciones de soluciones de Inteligencia de Negocios que requiere de procedimientos ETL para mover los datos al Data Warehouse. Adicional a estos tipos de migración, se citaron otros, como:

- El arreglo de una base de datos empleando una técnica llamada *Database Refactoring*, que está siendo empleada por la comunidad ágil en los Estados Unidos y Canadá.
- En CSC (Computer Science Corporation) Federal Division, realizaron una migración que les tomó 3 años completar, por ser en todas las divisiones. Pero esta migración no fue de bases de datos, sino una actualización e

implementación de Active Directory y otras actualizaciones de hardware y software en la organización. Se migraron y unificaron todas las estructuras, archivos y datos en cada una de las locaciones de esa división de CSC. A pesar de ser una migración de datos en principio, no fue una migración de datos de una base de datos, el cual es el interés de esta investigación.

La siguiente pregunta era para averiguar si realizaban validaciones de los datos migrados y el tipo de validaciones que realizaban. Con este fin se les preguntó:

- *¿Realizan validaciones de los datos migrados?*
- *¿Qué tipo de validaciones son realizados?*

Las personas que contestaron la encuesta respondieron afirmativamente al interrogante de si las realizaban. Entre las validaciones que realizaban, se encuentran:

- Comparar los datos migrados con la fuente para verificar que se migró el número correcto de datos.
- Verificar que cualquier transformación de los datos migrados haya ocurrido como fue planeado.
- Verificar que las tablas/columnas hayan sido migradas a las tablas/columnas correctas en la base de datos destino.
- Verificar que el software en el sistema destino sea capaz de funcionar con los datos migrados.
- Validar aplicaciones con usuarios claves.
- Verificación del dominio.

En la siguiente pregunta se trata de determinar los problemas con los cuales se encontraban en las migraciones:

- *¿Qué tipos de problemas han tenido? Se les suministró una lista de problemas básicos como tipos de datos inconsistentes, datos incompletos, etc. Adicionalmente tenían la posibilidad de listar problemas adicionales con los cuales se habían enfrentado.*

Concuerdan con que el problema más frecuente es de datos incompletos, a pesar de haber realizado trabajos de limpieza de datos en el sistema de origen antes de la migración.

Adicionalmente, algunos mencionaron otros problemas que tienen gran impacto sobre el proceso de migración. Uno de dichos problemas, estaba relacionado con la red durante la migración de los datos: cuando las migraciones se realizaban online el flujo de los datos desde el sistema origen hasta el destino, en ocasiones se volvía lento o el proceso de transmisión se interrumpía. Otros problemas mencionados giran en torno a los datos, pero en dos direcciones opuestas, uno tuvo problemas con los datos, mas no con la migración de los mismos, sino cómo lo quería ver el usuario. Otra empresa tuvo problemas con los procesos automáticos de limpieza de datos en el sistema fuente antes de la migración; esos procesos causaron problemas de invalidez de datos e in-completitud de los mismos en el sistema destino. En general, la calidad de los datos en los sistemas fuente no era la que el equipo originalmente pensaba.

Una de las preguntas de mayor importancia en el desarrollo de las encuestas era el saber si las empresas tenían una metodología para la Migración y Validación de

Datos. Con el fin de obtener una respuesta a este interrogante y poder analizar la situación actual se les plantearon las siguientes preguntas:

- *¿Tienen una metodología específica para la Migración y Validación de datos?*
- *¿La metodología es propietaria?*
- *Adicionalmente, se solicitaba la explicación de la metodología, si era posible.*

La respuesta a la primera pregunta por parte de todos fue, sí. La mayoría fueron metodologías desarrollados internamente y otra metodología es una modificación de una de las primeras.

Entre las metodologías descritas, se encuentra la empleada por la División de Consultoría de Oracle Corporation, llamada Oracle Consulting. Más adelante, en este capítulo, se explicará la metodología sugerida por Oracle a sus clientes cuando no emplean sus servicios para llevarlo a cabo, es una metodología más general y con menos detalle.

La metodología para Migración y Validación de Datos para las Aplicaciones de Oracle está basada en **Oracle Application Implementation Methodology** (Metodología de Implementación de las Aplicaciones de Oracle) **AIM**, y consta de seis fases:

1. Definición
2. Análisis de Operaciones
3. Diseño de la Solución
4. Construcción
5. Transición
6. Producción

La **Figura 5** es una vista general de los procesos que forman parte de Oracle AIM. Cada proceso está compuesto por un conjunto de objetivos, requerimientos de recursos, etc.

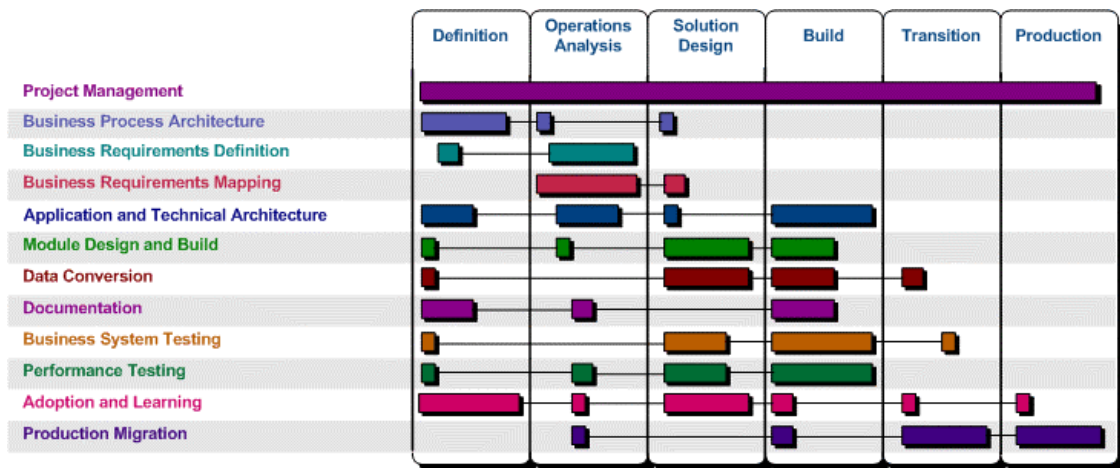


Figura 5: Vista general de Oracle AIM con Fases y Procesos.

Fuente: Oracle Consulting

Las tareas de AIM están organizadas en procesos, uno de los cuales es la **Conversión de Datos (CV)**, el cual es de gran importancia para el proceso de migración de datos. Tareas asociadas con el proceso de CV inician durante la fase de Definición, pero son ejecutadas en su mayoría durante las fases de Diseño de la Solución y la de Construcción. Otro proceso de gran importancia para la Migración y Validación de Datos es la **Migración a Producción (PM - Production Migration)**, el cual está compuesto por las tareas para crear un plan de transición y la preparación final del entorno de producción para migrar los datos y demás componentes a producción.

El primer paso en el proceso CV define los objetos del negocio que son requeridos para la conversión y los sistemas de origen que almacenan los objetos. El proceso de conversión incluye diseño, construcción y pruebas a los programas de conversión. Los entregables del proceso de conversión cuentan con tareas como: la definición de los requerimientos y estrategias de conversión de datos, mapeo de datos, definición de procedimientos manuales de conversión, definición de los procedimientos de conversión (diseño de los programas) y la preparación de los planes de prueba para la conversión de datos.

Este fue un resumen de la metodología empleada por Oracle Consulting; aunque Oracle anima a los partners y clientes a utilizar AIM con modificaciones que se ajusten a sus necesidades.

Otro de los entrevistados es una firma consultora la cual es partner de Oracle, ellos emplean AIM pero lo modifican y adaptan para cubrir las necesidades de su negocio.

Las preguntas de mayor importancia, de cierta forma, eran:

- *¿Durante el proceso/metodología de Migración y Validación de Datos incluye o realiza pruebas (crea planes de prueba para la migración y validación de datos)?*
- *Si realizaban pruebas, se les preguntaba que tipos de pruebas y las razones por las cuales eran realizadas.*

A continuación se presentarán las pruebas que realizan en las empresas entrevistadas, de manera general.

Si los datos son sencillos y simples, la comprobación que se realiza es que se mira el tamaño del dato fuente contra el tamaño del dato del destino.

Si los datos son para una aplicación complicada, se realizan pruebas de usuario, el gran problema nunca está en la corrupción de los datos sino en tener los permisos sobre los directorios o esquemas para poder conceder accesos a los usuarios. Tomar muestras al azar de los datos y mirar si los datos de origen son iguales a los datos migrados.

Scripts para las pruebas de conversión de datos: Son elementos de gran importancia para los procesos de pruebas, porque se planea cómo se va a realizar y a ejecutar la migración para así poder asegurar que los datos son válidos y completos antes de que se utilicen en el nuevo sistema.

La razón principal por la que se realiza este script es para que las pruebas puedan ser diseñadas, ejecutadas y así poder minimizar el riesgo de que los datos estén malos o incompletos.

Algunos de los tipos de prueba, son:

- *Pruebas de unidad:* Esta prueba confirma que cada módulo termina con éxito la tarea que se le diseña para realizar. Los datos en el sistema destino son iguales a los datos en el sistema de origen.
- *Prueba de objeto de negocio:* Verifica que la calidad de los datos convertidos al sistema de Oracle sean exactos y que funcionen correctamente.
- *Tipos de datos:* Verifica que los datos que se ingresaron en el sistema nuevo estén correctos.

- *Prueba de validación:* Verifica que los datos convertidos se desempeñen bien dentro de la suite de la Aplicación de Oracle
- *Prueba de regresión:* Cuando los resultados reales se comparen con los resultados previstos, se sabrá que funciona la prueba de regresión y que los otros criterios de validación se cumplieron satisfactoriamente. Si la prueba de regresión no se realiza se podría comprometer no sólo el éxito del software, sino también el del negocio.
- *Prueba de aceptación de usuario:* Esta prueba se utiliza para confirmar que los usuarios principales del sistema no encuentren problemas en la migración de los datos realizada, además esto determina si los datos ya pueden ser usados en un ambiente de producción.

Dentro de la comunidad de las metodologías ágiles, están proponiendo una metodología para la prueba de bases de datos, alguno de los aspectos de importancia que se pueden resaltar de allí, son el uso de:

1. *Database Sandboxes:* Es una de las mejores prácticas, en esta prueba se realizan varias copias de la base de datos y se realizan experimentos con dichas bases de datos, como implementar nuevas funcionalidades, agregar un nuevo factor a una funcionalidad existente, valida los cambios para probar el test. La ventaja de esta prueba es que ayuda a reducir el riesgo de errores técnicos.
2. *Prueba de escritura de la base de datos:* Este test se divide en tres procesos:

Setup de la prueba: Se debe poner la base de datos en un estado conocido antes de realizar la prueba.

Correr la prueba: Usando la herramienta de regresión de bases de datos, se corre la prueba de base de datos como se correría con una aplicación.

Verificar los resultados: Se necesitará poder hacer "table dumps" para obtener los valores actuales en la base de datos para que se puedan comparar contra los resultados esperados.

En la comunidad ágil, dentro de la propuesta para pruebas de bases de datos hacen la sugerencia de qué probar dentro de una base de datos. La **Figura 6** ilustra la estructura y los niveles en los cuales sugieren realizar las pruebas a la base de datos y algunas posibles pruebas que se podrían realizar.

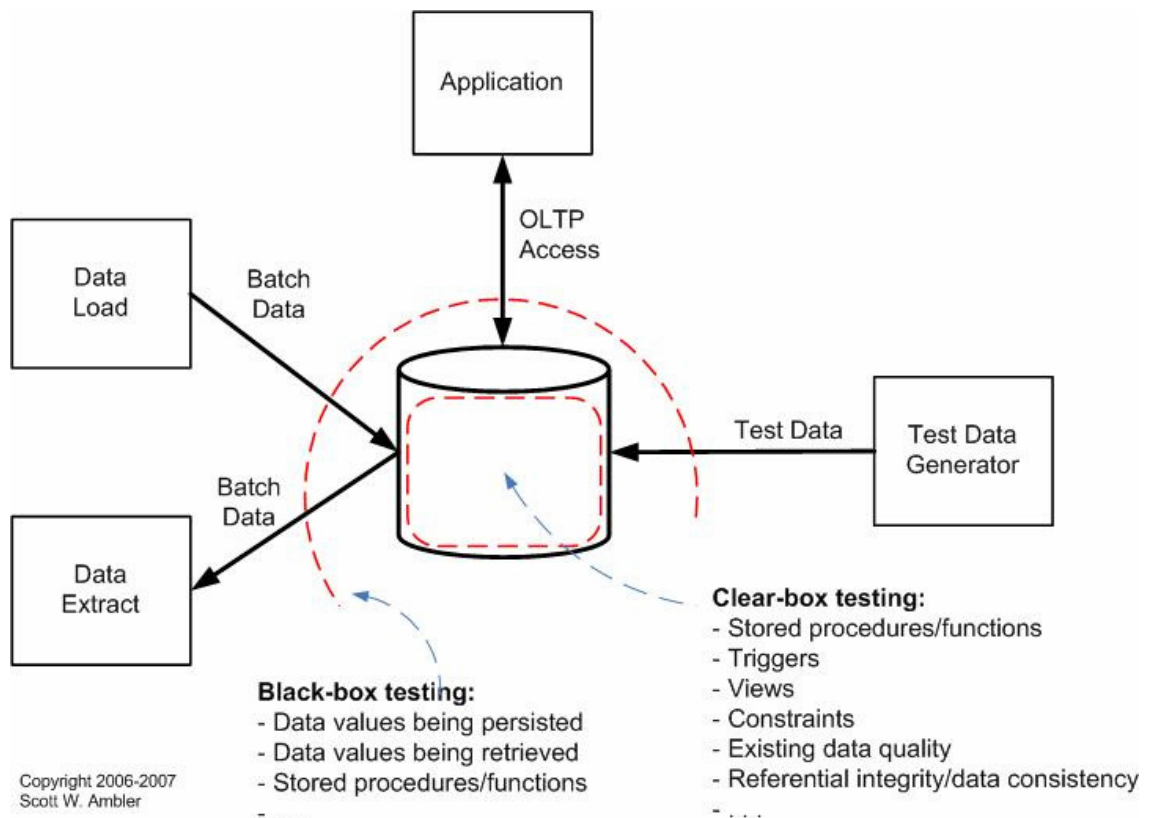


Figura 6: Diagrama sobre qué probar en una base de datos.

Fuente: Realizado por Scott W. Ambler. Sitio Web de la Comunidad Ágil

<http://www.agiledata.org/essays/databaseTesting.html>

Proponen utilizar pruebas de *caja negra* desde afuera sobre la base de datos; entre las pruebas que proponen, se encuentran: mapeo de objetos, datos que están ingresando y saliendo del sistema, funcionamiento de los procedimientos almacenados, etc.

De igual manera, proponen emplear pruebas de *caja blanca* para probar la base de datos desde adentro. Entre las pruebas a realizar, sugieren: pruebas de unidad para los procedimientos almacenados, triggers, funciones, etc., además de pruebas de existencia de los elementos de los esquemas, reglas de integridad referencial, valores por defecto de las columnas, calidad de los datos existentes, etc.

Cuando se les preguntó:

- *¿Hay un plan de Calidad de Datos para verificar los datos migrados?*

La respuesta de todos los entrevistados fue, sí.

Los siguientes interrogantes buscaban averiguar que tipos de herramientas empleaban para las migraciones y si permitían realizar pruebas. Para ellos se plantearon las preguntas:

- *¿Qué tipo de herramientas son utilizadas para los procesos de Migración y Validación de Datos? Aclarando que herramientas se refiera a cualquier*

software y/o lenguaje de programación empleado con fines de realizar la migración y validación de datos.

➤ *¿Las herramientas realizan o permiten realizar pruebas?*

Cuando se les preguntó qué herramientas empleaban para las migraciones, las respuestas en su mayoría eran de herramientas propietarias. A continuación se presentan las herramientas empleadas por las personas entrevistadas:

- SQL*Loader; es una herramienta que permite la carga desde archivos planos u otros tipos de archivos a las bases de datos Oracle, por lo general a tablas de interfaz para la aplicación.
- Data Loader; esta herramienta funciona online con la aplicación (la ERP de Oracle), la información se le suministra en hojas de cálculo, y se le “enseña” a la herramienta dónde colocar qué datos, dependiendo del orden en el cual están en la hoja, en las formas de la aplicación de Oracle.
- PL/SQL
- JUnit
- dbUnit
- Herramientas ETL

De esta lista de herramientas, solamente JUnit y dbUnit son para realizar pruebas sobre las bases de datos. Las demás realizan validaciones básicas de tipos de datos, formato, etc.

Cuando se les preguntó si:

➤ *¿Las migraciones son online u offline?*

La mayoría respondió que realizaban migraciones online, aunque en ocasiones, dependiendo de la situación, realizaban migraciones offline.

Se podría decir que el estado actual de las Migraciones y Validaciones de Datos y las pruebas correspondientes están en un nivel casi estable, por lo que se alcanzó a ver que las empresas ya cuentan con metodologías para llevar a cabo dichos procesos, ya sea propio o de terceros. Ahora se están enfocando en la estandarización y mejora de la calidad de dichas metodologías. Hay gran conciencia por la calidad de datos dentro de este tipo de proyectos, y es más común que exista ya un plan de calidad de datos para acompañar el proyecto de migración. A nivel de herramientas, tienen una mayor automatización del proceso y se comienza a observar la presencia de herramientas que ayudarán a realizar las pruebas. Se está proponiendo una nueva o mejorada técnica para realizar las pruebas de bases de datos con un enfoque en la calidad de datos; esa propuesta está siendo perfeccionada y empleada por la comunidad ágil a nivel internacional.

En las siguientes sub-secciones se describirán metodologías para realizar proyectos de Migración y Validación de Datos de varias empresas.

6.2.2. METODOLOGÍA DE MIGRACIÓN DE DATOS DE DULCIAN, INC.

Dulcian es una empresa de consultoría en varias áreas de la tecnología, pero primordialmente para en el área de bases de datos Oracle, aunque también brinda consultorías en áreas de data warehousing y migración de datos.

En 1999 Joseph R. Hudicka, uno de los miembros de la empresa sugirió una metodología para la migración de datos. Diseñó la metodología basándose en las experiencias que había tenido en proyectos anteriores y en hechos que no son ajenos a los que participan en este tipo de proyectos. Hace referencia a los problemas más comunes en una migración de datos, como la diferencia en arquitecturas entre sistemas de origen, que puede ser jerárquico, y los sistemas destino, que pueden ser relacionales. Entre otros problemas, menciona la integridad de los datos, dado que en un sistema puede ser distinto que en el otro, viendo como el problema más significativo de los proyectos de migración de datos es el que las personas subestiman la complejidad de la transformación de datos (cuando es necesario) hasta que lo experimentan en varias ocasiones.

Para Hudicka, el plan del proyecto de migración de datos debería estar dividido en 7 fases, las cuales son imagen de las fases de los proyectos en general. Estas fases son conocidas como Estrategia, Análisis, Diseño, Construcción, Pruebas/Implementación, Revisión y Mantenimiento.

- **Estrategia:** Esta fase debería ser la más fácil del proceso de planeación del proyecto. Se determina el enfoque general del proyecto, considerando que en muchas ocasiones estos proyectos no son independientes, sino que se originan de otros proyectos de tecnología, como la implementación de nuevos sistemas OLTP y/o OLAP.

En esta fase se identifica el primer error durante las fases de planeación de proyectos de tecnología. Por lo general, el enfoque está en la planeación para la implementación del nuevo sistema, y se le presta poca atención a la migración de datos. La migración puede aparecer, si se le considera, como una tarea en el plan general del proyecto, cuando en realidad debería haber un plan separado para la migración de los datos.

- **Análisis:** Recomienda que la fase de análisis de la migración de datos se lleve a cabo de manera simultánea con la fase de análisis del proyecto principal. Entre las tareas que se deberían llevar a cabo en la fase de análisis se encuentra:
 - Identificar las fuentes u orígenes de los datos a ser transportados al nuevo sistema, mas no los archivos o data files, solo los sistemas de origen. Por ejemplo: hojas de cálculo, motores de bases de datos, etc.
 - Familiarizarse con los datos a ser migrados, es decir, obtener reportes que puedan informar sobre la cantidad de filas a migrar, el número de columnas y otras estadísticas relacionadas con los datos. Esto permite realizar una estimación de cuántos datos serán migrados y una duración aproximada inicial del proceso de migración.
- **Diseño:** Luego de identificar y seleccionar las fuentes de los datos, y realizar un análisis completo de los datos, se inicia la selección de los datos a migrar. Se hace un recorrido de todos los elementos de cada una de las estructuras de las cuales se extraerán los datos, para decidir si todos serán migrados o no. Este proceso se puede iniciar durante la fase de análisis, pero en la fase de diseño es donde se realiza la selección.

La fase de diseño de la migración de datos también se lleva a cabo en paralelo con la fase de análisis del proyecto principal. Esto es porque cada elemento elegido como candidato para ser migrado puede afectar el modelo de datos.

En la fase de diseño no se identifican completamente las reglas de transformación con las cuales los datos históricos serán transformados para ser llevados al nuevo sistema.

- **Construcción:** La fase de construcción de la migración coincide con la fase de diseño del proyecto principal. La primera tarea de la fase de construcción es la generación de la nueva estructura de datos y su creación en la base de datos.

Un concepto erróneo muy común sobre el mapeo de datos es que se puede realizar con los modelos lógicos de datos. Desafortunadamente los modelos lógicos de datos representan todos los datos que surgen de relaciones sin definir los atributos en la entidad hija. Esto quiere decir; que no se puede mapear ninguna de las conexiones entre estructuras de datos mientras se esté trabajando con diseños lógicos. Por ello, es necesario realizar el mapeo de datos al modelo físico de datos.

Con la estructura de datos física definida en la base de datos se puede iniciar el proceso de mapeo. Este proceso de mapeo por lo general es llevado a cabo por un equipo de mínimo tres personas por cada área del negocio. De estas tres personas, la primera debería ser un analista del negocio, el segundo un analista de sistemas con conocimientos de ambos sistemas, y el tercero, un programador/analista, quien que se encarga de realizar las rutinas o programas, basados en los mapeos definidos por el analista de negocios y el analista de sistemas.

- **Pruebas/Implementación:** Debido a que estas dos fases son casi inseparables y se pueden trabajar en conjunto, fueron combinados en una sola fase. Las pruebas se dividen en dos áreas centrales:
 - *Errores Físicos:* son sintácticos de naturaleza y pueden ser fácilmente identificados y resueltos. Los errores físicos no guardan relación con la calidad del mapeo de los datos. Este nivel de pruebas está tratando con las semánticas del lenguaje empleado para la transformación.
 - *Errores Lógicos.*

Durante la implementación es cuando se identifican los errores lógicos. El primer paso es ejecutar el mapeo. Aunque el mapeo se complete exitosamente, aún se deberían realizar las siguientes preguntas.

- ¿Cuántos registros esperaban ser creados por el script?
- ¿Se creó el número correcto de registros? Sí no, ¿por qué?
- ¿Se han cargado los datos a los campos correctos?
- ¿Se han formateado los datos correctamente?

La verdadera prueba a los datos mapeados o migrados es darle acceso a los datos a los usuarios que participaron en el análisis y diseño del sistema principal. Comenzarán a identificar elementos que debieron ser migrados pero que no fueron identificados durante las sesiones de análisis y diseño.

Es un hecho que el mapeo de datos no tiene sentido para las personas hasta cuando no interactúen con los datos en el nuevo sistema; en estos momentos es cuando se descubren los datos faltantes. Por esta razón, es de mucha importancia permitirles ver los datos en el sistema nuevo lo antes posible.

La fase de Pruebas/Implementación debe ser alcanzada lo antes posible para garantizar que se lleve a cabo antes de las fases de diseño y construcción del proyecto principal, para evitar sobrecostos por cambios en el diseño del proyecto principal, debido a que los requerimientos para la migración de los datos han seguido cambiando.

- **Revisión:** La fase de revisión es donde se llevan a cabo labores de limpieza. Cada modificación al modelo de datos, ajuste a las reglas de transformación y modificación a los scripts, son combinados para formar la fase de revisión.
- **Mantenimiento:** La fase de mantenimiento es donde todos los mapeos son validados e implementados exitosamente en una serie de scripts que han sido

probados completamente. Esta fase se modifica, dependiendo si se está migrando a un sistema OLTP o a una OLAP.

Si se está migrando hacia un sistema OLTP, el objetivo es migrar exitosamente los datos del sistema anterior al nuevo, haciendo obsoletos los scripts de migración una vez se haya completado el proyecto.

Si se está migrando a un sistema OLAP, lo más probable es que al sistema nuevo se le alimenten datos periódicamente. A medida que se ingresen datos nuevos al sistema OLTP, estos deben ser migrados al sistema OLAP. El desempeño de los scripts es un problema crítico en las migraciones OLAP, mientras que las migraciones OLTP le prestan poca o ninguna atención al desempeño de los scripts, ya que van a ser utilizados una vez.

6.2.3. METODOLOGÍAS DE ORACLE

6.2.3.1. METODOLOGÍA DE MIGRACIÓN

Oracle, por ser una de las casas desarrolladoras de bases de datos, y una de las pocas que cuenta con diversos sistemas desarrollados alrededor de la misma, como por ejemplo, la ERP llamada Oracle E-Business Suite, las herramientas de reporte como Oracle Reports y Discoverer, ha llevado a la organización a desarrollar a través del tiempo, metodologías y encontrar mejores prácticas para realizar ciertas actividades, como migraciones, dichas metodologías son probadas y empleadas por sus equipos internos de trabajo, por los miembros de su firma de consultoría, Oracle Consulting, y por sus partners, cuando se trata de proyectos relacionados con Oracle.

Entre las metodologías desarrolladas por Oracle se encuentra el **Oracle Relational Migration Maps**, es una especie de mapa cuyo propósito principal es guiar a sus clientes y usuarios involucrados en las migraciones en el proceso de llevar sus datos de la base de datos de un tercero a la base de datos de Oracle. La metodología en sí fue diseñada para abarcar los dos tipos de migración que Oracle reconoce, que son:

- Migración de Bases de Datos
- Migración de Bases de Datos y Aplicaciones

Dentro de la metodología se identifican 6 fases, que son:

1. Definición
2. Análisis
3. Diseño
4. Migración
5. Transición
6. Producción

Cada una de esas fases puede o no presentar algunas de las siguientes actividades:

- ✓ Entrenamiento
- ✓ Evaluación del Sistema Existente
- ✓ Arquitectura técnica
- ✓ Migración de la Base de Datos
- ✓ Migración de la Aplicación
- ✓ Migración de los Datos
- ✓ Pruebas
- ✓ Transición

✓ Administración del Proyecto

La **Figura 7** muestra cuáles actividades se presentan en cuáles fases y dan una idea de qué tan importante o extendida será su desarrollo en la misma.

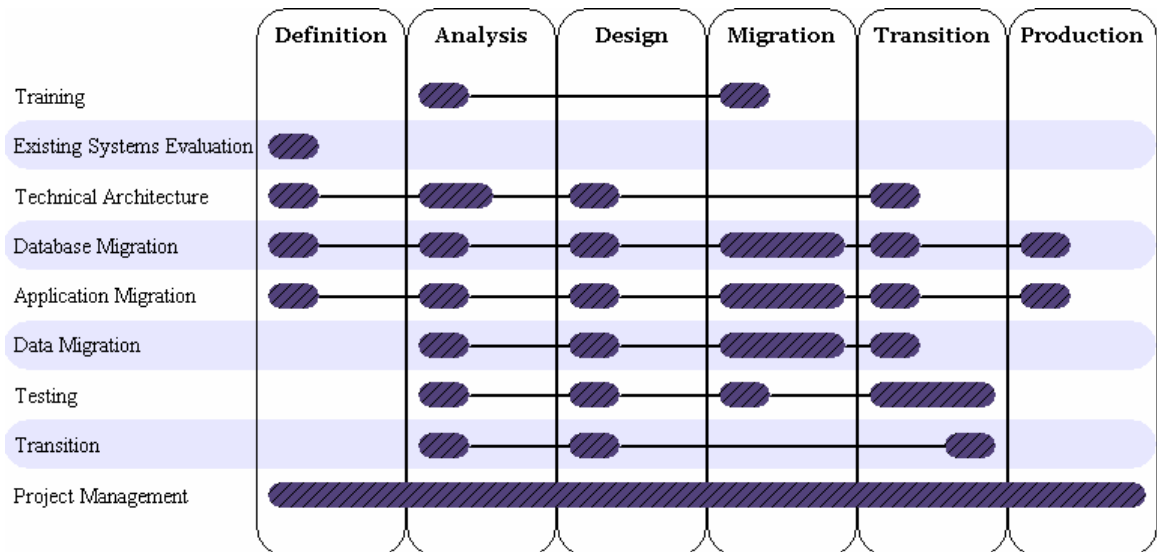


Figura 7: Fases de un proyecto de migración versus las actividades y duración.

Fuente: Oracle Relational Migration Map

<http://www.oracle.com/technology/tech/migration/maps/index.html>

Hay que tener en cuenta que esta metodología también es empleada para la migración de aplicaciones hacia Oracle, no sólo para la migración hacia la base de datos, así que pueden haber fases dentro de la metodología que no se aplicarán para la migración de datos.

A continuación se hará una descripción de las fases, en qué consisten y una descripción de las actividades que se llevan a cabo en cada una de ellas.

- **Fase de Definición:** Describe cómo reunir suficiente información sobre las aplicaciones y bases de datos del cliente para crear estimaciones de la migración. Durante esta fase entre las tareas que se deben llevar a cabo se encuentran: evaluación del sistema existente, arquitectura técnica, migración de la base de datos, migración de la aplicación y administración del proyecto.

Dentro de la tarea de evaluación del sistema existente se realiza un inventario del sistema actual, que incluye definir las bases de datos de origen, el número de objetos de la base de datos, el tipo y el tamaño. También se define el entorno existente de la base de datos y la aplicación.

En la tarea de arquitectura técnica, se definen los requerimientos técnicos de la arquitectura que se empleará, por ejemplo, requerimientos de interfaz del sistema, requerimientos de seguridad de la base de datos, de almacenamiento, de estrategias de respaldo y recuperación, estrategias de administración, arquitectura de la red, etc.

La tarea de migración de la base de datos involucra pasos más técnicos en los cuales se requiere de la definición de la estructura de la base de datos actual y el empleo de herramientas de migración que Oracle ha creado para tales propósitos, como Oracle Migration Workbench y OMWBAlyzer.

Similar al paso de migración de la base de datos, la migración de la aplicación emplea otras herramientas para realizar un análisis de la aplicación que se migrará.

Dentro de la tarea de administración del proyecto, se definen el alcance y los objetivos de la migración. Se crea un plan para el proyecto y se estiman los recursos necesarios.

- **Fase de Análisis:** Para Oracle la fase de análisis marca el verdadero inicio del proyecto, porque durante esta fase se confirman y profundizan los detalles hallados en la fase de definición, para poder obtener un mayor detalle de la planeación del proyecto. Entre las tareas que recomiendan realizar durante esta fase se encuentran: entrenamiento, arquitectura técnica, migración de la base de datos, migración de la aplicación, migración de datos, pruebas, transición y administración del proyecto.

En esta fase se identificarán las necesidades de entrenamiento del cliente, incluyendo entrenamiento sobre la migración y la base de datos.

En el aspecto de la arquitectura técnica se analizarán los requerimientos técnicos de la arquitectura en detalle, se refinará y servirá como suministro para el desarrollo del diseño para la migración de la base de datos y la aplicación.

En la migración de la base de datos se crean las diversas instancias de la base de datos necesarias, se compara la estructura de la base de datos actual con la estructura que se tendrá en Oracle para realizar un mapeo inicial e intentar identificar posibles problemas durante la migración. En la migración de la aplicación se realizan pasos similares que en la migración de la base de datos. En ambos casos, la metodología es más específica, puesto que incluye las herramientas y las tareas a realizar con dichas herramientas.

En esta fase, para la tarea de migración de datos, sólo se realizará un análisis de los datos a ser migrados y la información será documentada. Para la tarea de pruebas se analizarán los requerimientos de pruebas, tales como los ciclos de pruebas, duraciones, testers y scripts de pruebas requeridos. Para la tarea de transición, se analizarán y definirán los requerimientos para la estrategia de

transición a producción, incluyendo la programación o cronograma, el hardware y el impacto que tendrá ese corte a producción en el negocio.

En la administración del proyecto, se realizará una revisión a los costos y duración, para incluir modificaciones realizadas a los planes de migración durante la fase de análisis. Y se iniciará el proceso de control de los problemas en la migración.

- **Fase de Diseño:** Durante la fase de diseño se emplea la información obtenida en las fases de definición y análisis para realizar el diseño de la solución para la migración y una solución para los primeros problemas identificados en las fases anteriores. Dentro de esta fase se realizan las siguientes tareas: entrenamiento, arquitectura técnica, migración de la base de datos, migración de la aplicación, migración de datos, pruebas, transición y administración del proyecto.

En la tarea de entrenamiento se inicia el proceso de diseño del plan de entrenamiento para el personal del cliente. Entrenamiento en las aplicaciones y/o en la base de datos.

En la tarea de arquitectura técnica, con la información suministrada de la fase de análisis, se diseña la arquitectura técnica del sistema propuesto, especificando el hardware, el tamaño de la base de datos, la especificación de la red, y el software para los diversos entornos (desarrollo, pruebas, producción, etc.). También se desarrolla la estrategia de respaldo del sistema en caso de emergencia.

Para la migración de la base de datos y/o la aplicación se emplea la información recolectada en las fases anteriores y se diseña un plan de migración que tenga en cuenta los problemas identificados y sus soluciones.

En la migración de datos, se diseña la estrategia de cómo y cuándo se migrarán los datos, esto incluye el diseño de las estrategias de limpieza de los datos e identificar las transformaciones de datos necesarios.

En la tarea de pruebas, el cliente diseña los scripts de las pruebas de unidad, que incluyen una lista de los objetos a ser probados. En la tarea de transición, se define la estrategia de cut-over para el despliegue al sistema de producción. Esto incluye cronogramas, hardware e impacto al negocio.

Cualquier estrategia o diseño debe ser aprobado por el cliente antes de ser puesto en funcionamiento.

Mientras, la tarea de administración del proyecto continúa haciendo ajustes y llevando controles de los cambios y problemas, y realizando seguimiento del desarrollo del proyecto.

- **Fase de Migración:** El objetivo de esta fase es la generación de scripts y reportes concisos y confiables que permitirán la generación de esquemas, datos y programas almacenados de su base de datos de origen en su base de datos Oracle. Dentro de las tareas que se llevan a cabo en esta fase se encuentran: entrenamiento, migración de la base de datos, migración de la aplicación, migración de datos, pruebas y administración del proyecto.

Para la tarea de entrenamiento, se llevarán a cabo los entrenamientos del personal necesario para que obtengan el conocimiento y las habilidades necesarias para llevar a cabo la migración de la base de datos y/o la aplicación.

El objetivo de la migración de la base de datos es mover exitosamente todos los objetos de los esquemas de la bases de datos de desarrollo (DEV) a la

base de datos de prueba (TEST) de Oracle. La migración de la base de datos se realiza por etapas; primero se debe realizar la migración de los esquemas antes de iniciar la migración de los datos, mientras que la migración de los programas almacenados (vistas, triggers, procedimientos almacenados y funciones) se puede realizar en paralelo con los esquemas. Para la migración del esquema, se valida el esquema en el entorno de desarrollo, se hacen modificaciones y verificaciones manuales del mismo, se cambia la arquitectura del esquema si es necesario, se crea el reporte de la migración del esquema, entre otras actividades que serán más o menos específicas, técnicamente, dependiendo del proyecto. Para la migración de los programas almacenados, se lleva a cabo pasos similares a los realizados para la migración de los esquemas, se diferencian en las pruebas a realizar para validar los programas almacenados.

Para la migración de la aplicación, se implementan soluciones para los problemas identificados en la fase de análisis. Entre las actividades que se realizan para esta migración se incluye la preparación para la migración (herramientas, reportes, etc.), el análisis de la aplicación, la tarea de migración como tal, la investigación y solución de problemas durante la migración (se investiga problema por problema), se revisan los cambios de código y se presenta la aplicación migrada para realizar pruebas.

En esta fase para la migración de datos se crean los scripts de migración de datos, se verifican y modifican según sea necesario, se crean reportes de migración, se crean instrucciones para los scripts de migración de datos y se realizan más pruebas sobre dichos scripts.

En la tarea de pruebas, se prueban completamente los procesos de migración, dentro del entorno de pruebas para los esquemas, los programas

almacenados, la aplicación y la migración de los datos. También se realizan pruebas de unidad para la base de datos. Cada uno de ellos tendrá sus instrucciones respectivas y scripts. Luego de las pruebas se firmarán las aceptaciones de las migraciones para seguir a la siguiente fase.

En lo que respecta a la administración del proyecto, se llevarán acabo actividades como las de las fases anteriores, es decir, actividades de seguimiento de progreso, control de problemas y sus soluciones, reportes de estado del proyecto, etc.

- **Fase de Transición:** La fase de transición incluye la creación del nuevo sistema recientemente migrado, las pruebas del cliente, la conversión de los datos y su puesta en producción. Entre las tareas que se deben realizar en esta fase se encuentran: arquitectura técnica, migración de la base de datos, migración de la aplicación, migración de los datos, pruebas, transición y administración del proyecto.

Dentro de la tarea de arquitectura técnica, se organiza y arma la estructura técnica para los entornos de desarrollo (DEV), pruebas (TEST) y producción (PROD), instalando y configurando hardware, software, redes, estrategias de respaldo y recuperación, herramientas y programas para la administración del sistema, etc.

Para las tareas de migración de la base de datos y/o la aplicación, se prepara y configura el entorno de desarrollo, los pasos necesarios para esta tarea son de nivel más técnico y dependerán del proyecto. Esta configuración se modificará, basado en los resultados de la tarea de pruebas en esta fase.

Para la migración de los datos, el Plan de Migración de Datos es creado describiendo cómo se llevará a cabo la migración de los datos, qué máquinas

y recursos se necesitarán, y cuánto durará la migración. Los scripts de migración son revisados y mejorados considerando las necesidades de transformación o limpieza de los datos. Luego se realizará una migración hacia alguno de los entornos, prueba o desarrollo, con los scripts de migración mejorados y se validarán los datos migrados.

Para la tarea de pruebas, estas se llevan a cabo en el sistema de pruebas (TEST). Este entorno es una imagen del último estado del entorno de desarrollo (DEV), bases de datos y aplicaciones, a esta nueva instancia se migran los datos. El cliente prueba la base de datos al realizar pruebas sobre todos los objetos de los esquemas. Para probar las aplicaciones se recorren los casos de pruebas diseñados en fases anteriores y se prueban los datos, verificando que los scripts de migración de datos mejorados limpiaron, transformaron y consolidaron los datos como se esperaba.

Para la tarea de transición, se migra al sistema de producción, congelando y tomando una imagen del entorno de pruebas (TEST) e instalándolo en producción, incluyendo los procesos de migración de datos, etc. Se hace el corte (o cut-over) al sistema nuevo de producción y se desconecta el sistema antiguo.

Las tareas de administración del proyecto siguen siendo iguales a las fases anteriores, como monitoreo de actividades, control de cambios, seguimientos de problemas y su solución, etc.

- **Fase de Producción:** La fase de producción en realidad se trata más sobre el soporte post-migración a producción, donde se resuelven problemas pendientes y se afina la aplicación y/o la base de datos, dependiendo del caso.

6.2.3.2. METODOLOGÍA DE PRUEBAS

De la misma forma que Oracle diseñó una metodología que puede ser empleada para migraciones hacia su base de datos o sus aplicaciones, ha creado una metodología para probar o testear su base de datos.

Durante la fase de pruebas, se llevan a cabo pruebas sobre la base de datos Oracle y la aplicación (si es necesario) para asegurarse, que:

- ✓ Los datos migrados estén completos y sean correctos.
- ✓ Las aplicaciones funcionen de la misma forma que antes.
- ✓ La base de datos está produciendo los mismos resultados que la base de datos anterior.
- ✓ La base de datos Oracle y la aplicación cumplen con los requerimientos operacionales y de desempeño.

Se podría iniciar las pruebas a la aplicación y a la base de datos Oracle, realizando las pruebas que se llevaban a cabo sobre el sistema antiguo, para determinar qué tan cercanos están los resultados.

Muchas restricciones modelan el tipo y el número de pruebas que se realizan sobre una base de datos. El plan de pruebas puede contener una o todas las siguientes:

- Validación sencilla de datos.
- Ciclo completo de pruebas dirigidas a pruebas de unidad individuales.
- Pruebas de aceptación y del sistema

Pero la estrategia de pruebas debe adaptarse a las necesidades de su organización, aunque la metodología que ellos sugieren y utilizan es un *Modelo V con Migración de Base de Datos*.

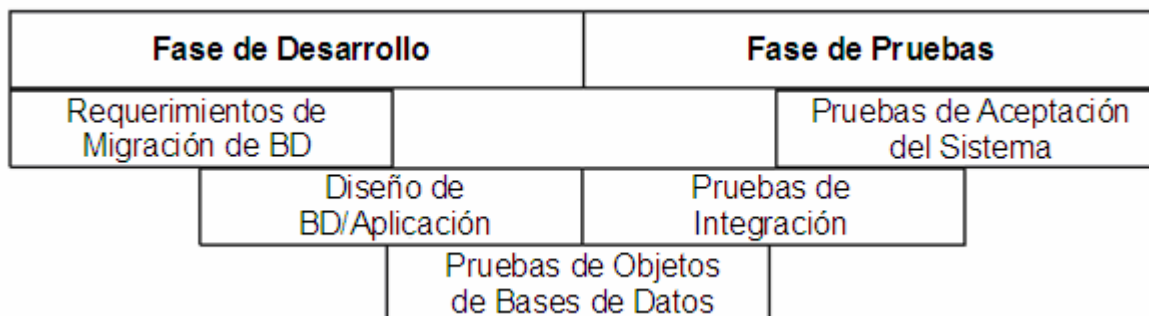


Figura 8: Traducción del Modelo V con Migración a Base de Datos

Fuente: Oracle Corporation - SQL Developer User Guide

El Modelo V fue originalmente desarrollado basándose en el modelo de desarrollo de software en cascada. Los cuatro procesos principales - *requerimientos, especificación, diseño e implementación* - tienen correspondientemente una fase de *pruebas de verificación y validación*. La *implementación* de módulos es probada con *pruebas de unidad*, el *diseño* del sistema es probado con *pruebas de integración*, las *especificaciones* del sistema son probadas con *pruebas del sistema* y finalmente, las *pruebas de aceptación* verifican los *requerimientos*. El modelo V obtiene su nombre por la distribución de sus fases. Comenzando desde los requerimientos, el sistema es desarrollado una fase a la vez hasta que la última fase, la de implementación, es terminada. En esta etapa se inician las pruebas, comenzando con pruebas de unidad e incrementando un nivel las pruebas hasta que se complete la prueba de aceptación²³.

A pesar de que el modelo V fue diseñado originalmente para un desarrollo secuencial, con modificación, dependiendo de las necesidades, puede soportar un

²³ "Software Engineering" – I. Sommerville, Addison Wesley, 5ta Edición. 1999.

desarrollo iterativo, que es lo que Oracle hizo cuando comenzó a utilizar el modelo V. Llevaron en paralelo las pruebas y el desarrollo, teniendo como el nivel más bajo las pruebas de los objetos de la base de datos, continuando con pruebas de integración para el diseño de la aplicación y la base de datos y terminando con pruebas de aceptación del sistema para los requerimientos de migración de la base de datos.

A continuación se explicará un procedimiento para probar las bases de datos Oracle. Hay otros métodos disponibles y pueden ser más adecuados para las necesidades y requerimientos del negocio. Pero este método es el básico recomendado por Oracle para probar las migraciones a su base de datos. Para probar la base de datos Oracle:

1. Crear una versión controlada de la base de datos migrada. Recomiendan mantener los scripts de migración de la base de datos en un sistema de control de archivos.
2. Diseñar un conjunto de casos de prueba que puedan ser utilizados para probar la base de datos Oracle, desde la unidad mínima hasta el nivel del sistema. Los casos de prueba deberían:
 - (a) Garantizar los siguiente:
 - Todos los usuarios en la base de datos de origen han migrado exitosamente.
 - Privilegios y permisos para los usuarios están correctos.
 - Las tablas tienen la estructura correcta, los valores por defecto están funcionando correctamente, y no se presentaron errores durante el mapeo o generación.
 - (b) Validar que los datos migraron exitosamente haciendo lo siguiente:
 - Comparar el número de filas en la base de datos Oracle con el número de filas en la base de datos fuente o de origen.

- Calcular la suma de columnas numéricas en la base de datos Oracle y comparar con la base de datos fuente o de origen.
- (c) Asegurar que lo siguiente aplique a los constraints o restricciones:
- No se pueden ingresar claves primarias duplicadas.
 - Claves foráneas evitan que ingresen datos inconsistentes.
 - Constraints de validación o chequeo evitan que ingresen datos inválidos.
- (d) Verificar que índices y secuencias se hayan creado exitosamente.
- (e) Asegurar que las vistas hayan migrado exitosamente haciendo lo siguiente:
- Comparar el número de filas en la base de datos Oracle con el número de filas en la base de datos fuente o de origen.
 - Calcular la suma de columnas numéricas en la base de datos Oracle y comparar con la base de datos fuente o de origen.
- (f) Asegurar que los triggers, procedimientos y funciones hayan sido migrados exitosamente. Verificar que los valores correctos sean retornados para los triggers y las funciones.
3. Ejecutar los casos de prueba contra la base de datos migrada.
4. Crear un reporte que evalúe los resultados de los casos de prueba. Estos reportes permiten evaluar los datos y cualificar los errores, informar sobre problemas y mantener un control sobre las versiones de la base de datos.
5. Si se aprueban todas las pruebas, se sigue al paso 7.

Si todas las pruebas en los casos de pruebas pasan o contienen errores aceptables, la prueba pasa. Si ocurren errores aceptables, estos deben ser documentados en un reporte de errores para hacer seguimiento y para propósitos de auditoría.

6. Si los casos de prueba fallan:
- (a) Identificar la causa del error.

- (b) Identificar el caso de prueba necesitado para verificar este error.
 - (c) Registrar el error en la versión controlada del código de la base de datos migrada en el reporte de problemas.
 - (d) Agregar el caso de prueba y una descripción del problema al sistema de rastreo de incidentes o problemas de la organización. Además del caso de prueba, en el registro de incidentes, se debería incluir:
 - Una descripción clara y concisa del problema encontrado.
 - Una descripción completa del entorno, como la plataforma y la versión de control de las fuentes.
 - Adjuntar los resultados de la prueba, si es útil.
 - Indicar la frecuencia y la pronosticabilidad del problema o incidente.
 - La secuencia de eventos que llevan al problema o incidente.
 - Descripción de los efectos en la prueba actual, pasos de diagnóstico tomados y los resultados vistos.
 - Descripción de los efectos persistentes, si existen.
 - (e) Intentar resolver el problema.
 - (f) Retornar al paso 1.
7. Identificar las pruebas de aceptación que pueden ser utilizadas para asegurar que la base de datos Oracle esté en un nivel de calidad aceptable.

Estos pasos son los recomendados por Oracle a sus partners y clientes cuando van a realizar una migración de bases de datos y/o aplicaciones hacia Oracle; son específicos para su motor y son básicos. El nivel de detalle de las pruebas para los datos es mínimo y poco desarrollado, como en otras metodologías.

6.2.4. METODOLOGÍA DE MIGRACIÓN DE SOFTEK

Softek al ser una de las empresas líderes en tecnología para la movilidad de los datos, ha desarrollado una metodología que se basa en mejores prácticas que han ido observando, realizando y practicando.

La metodología de Softek consta de tres fases o etapas, que son *Planear*, *Migrar* y *Validar*, donde cada una de ellas cuenta con un conjunto de actividades que la identifican y describen. En la **Figura 9** se ilustra la Metodología para la Migración de Datos de Softek.

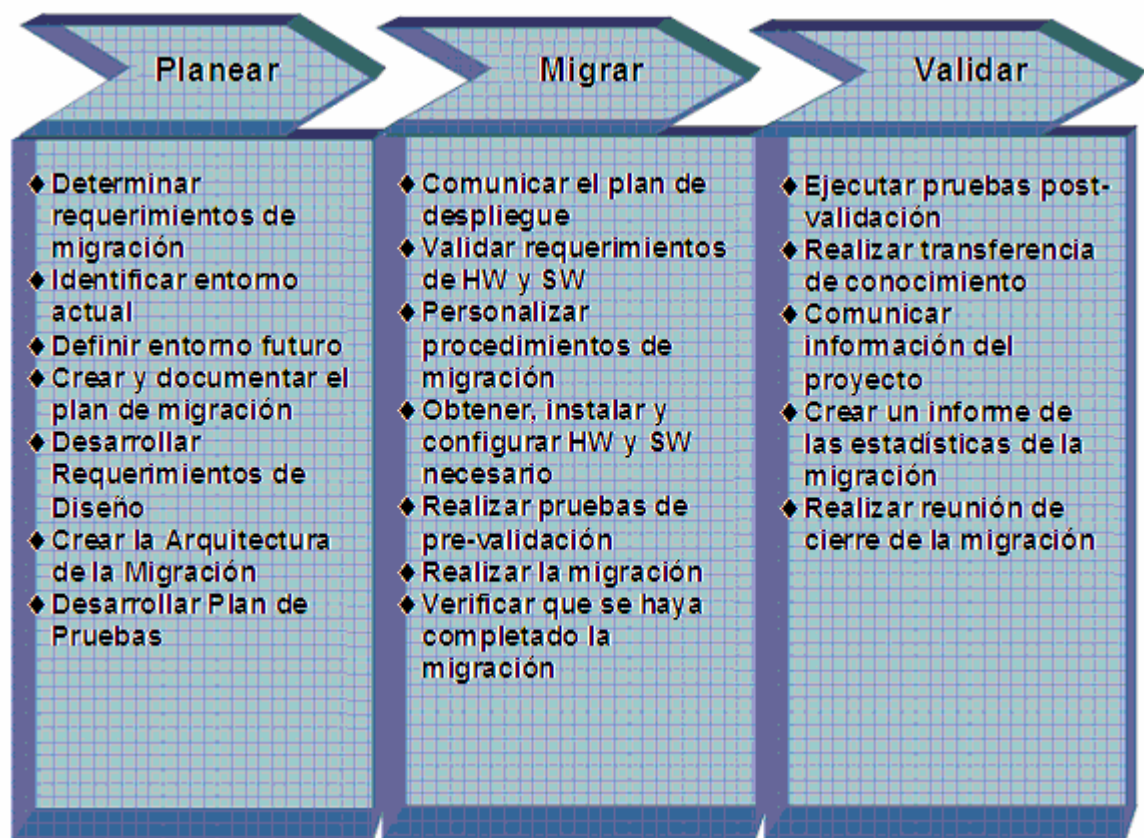


Figura 9: Traducción de la Metodología para la Migración de Datos
Fuente: Softek – “2006 Best Practices for Data Migration”. Pág. 5.

- **Planear:** Basándose en el estudio que la empresa realizó en el 2005²⁴, se establece que el proceso de planeación debería incluir: la definición de los requerimientos de migración, identificación del entorno actual y futuro, creación y documentación de plan de migración. Los requerimientos de diseño deberían incluir la arquitectura de la migración, requerimientos específicos de hardware y software, procedimientos de migración, y planes de despliegue y pruebas.

Una planeación para un proceso de migración apropiado involucra a más personas que solo el equipo de TI, incluye a los dueños de las aplicaciones y los datos que serán migrados, especialmente cuando se determina qué tan importante es una aplicación o un conjunto de datos para el negocio. Adicionalmente, cuando los dueños de las aplicaciones están involucrados en la planeación de la migración, pueden coordinar sus actividades adecuadamente; para evitar incompatibilidades horarias o de utilización de los datos y/o aplicación cuando se vaya a realizar la migración. El tenerlos involucrados puede eliminar o minimizar los posibles conflictos.

Dentro de esta fase también incluye el diseño del plan de pruebas, en el cual se sugiere involucrar a los dueños y usuarios claves de las aplicaciones y de los datos, para así poder construir un conjunto de casos de prueba con mayor grado de conformidad a las necesidades del negocio.

Es importante entender los requerimientos de diseño, tales como requerimientos de migración, cronograma, proveedores y configuración del hardware. Hay que entender el hecho de que hay una variedad de productos de software que pueden ser empleados para llevar a cabo la migración de

²⁴ "Softtek's 2005 Worldwide Data Migration Survey". Fragmentos y traducción de "Data Migration: Everyone is doing it - but are they doing it right?" de Enterprise Strategy Group. Marzo 2005.

datos, cada uno con sus fortalezas y debilidades, que incluyen desempeño, soporte a diferentes sistemas operativos, soporte a los diferentes dispositivos de almacenamiento, y si se requiere tiempo de inactividad de las aplicaciones para llevar a cabo la migración de los datos. Los requerimientos específicos ayudarán a determinar la tecnología de software más adecuada para la migración.

Las tablas, **Tabla 1** y **Tabla 2**, son un ejemplo de un plan de migración y requerimientos de diseño propuestos por Softek. En el plan de migración, se listan las posibles actividades de mayor importancia de manera sencilla, donde se identifica la persona a la cual dicha tarea fue asignada, el estado actual de la tarea y la fecha en la cual se completó la tarea. La **Tabla 1** es un ejemplo de requerimientos de diseño, donde se especifican los requerimientos del entorno de servidores, el entorno de almacenamiento y el entorno de la red; aunque se podrían agregar más especificaciones si es necesario.

Actividad	Asignado A	Estado	Fecha Culminación
Establecer un equipo de gestión de migración.			
Recolectar los horarios de disponibilidad y producción.			
Documentar procedimientos de Control de Cambios para que puedan ser incorporados a los planes y procedimientos de migración.			
Documentar el cronograma para las actividades de cambio de hardware y de migración de datos.			
Anunciar la migración al menos 30 días antes de la fecha programada para la migración.			
Recolectar información sobre el entorno de los			

servidores y las aplicaciones (listas y diagramas).			
Trabajar con el proveedor para entender y conocer la nueva configuración.			
Crear un equipo técnico de migración.			
Seguridad: Informar a los equipos de seguridad sobre la migración.			
Programar una simulación de la migración que incluya a todos los miembros en el equipo de migración y una muestra de los datos, que le permita a los dueños de las aplicaciones realizar procesos de verificación pre- y post-migración.			
Seguir el proceso de control de cambios requeridos.			
Emplear una "Lista de Chequeo del Plan de Migración" para asegurar que todos los pasos pre-migración se lleven a cabo.			

Tabla 1: Ejemplo de un Plan de Migración según Softek

Fuente: Traducción de Softek – “2006 Best Practices for Data Migration”. Pág. 6.

Servidores	
	Fabricante del Servidor (Compaq, Dell, IBM, HP, Sun)
	Número de Procesadores
	Número de particiones lógicas (LPARs) o Dominios
	Tipo de Sistema de Archivos (UFS, VxFs, HFS, JFS, JFS2, NFS, NTFS, FAT, FAT32)
	Sistema Operativo (Windows, AIX 5.1, z/OS 1.4)
	Direccionamiento del Sistema Operativos (32 bits, 64 bits)

	Base de Datos a ser movida (DB2, Informix, Oracle, SQL Server, Sybase)
	Versión de la base de datos
	Tamaño de la base de datos
	Requerimientos de disponibilidad de la base de datos
	Cluster (MSCS, Veritas, Sun, HACMP, etc.)
Almacenamiento	
	Proveedor y modelo (EMC, HDS, IBM, STK, Dell, HP)
	Tipos de Canales (ESCON, FICON, Fibra, iSCSI, SAN)
	Mapeo lógico a físico (RAID-1 vs. RAID-5, etc.)
	Número de fuentes de las cuales se migrará
	Destinos de la migración
Red (si aplica)	
	Topología
	Velocidad de la red
Bases de Datos	
	Fuentes
	Destino
	Número de objetos en la fuente
	Tipos de objetos
	Volumen de datos

Tabla 2: Ejemplo de Requerimientos de Diseño según Softek

Fuente: Traducción de Softek – “2006 Best Practices for Data Migration”. Pág. 7.

- **Migrar:** Entre las actividades que sugieren llevar a cabo para la fase de migración, se encuentran:
 - Comunicar el plan de despliegue a producción.
 - Validar los requerimientos de hardware y software.
 - Obtener, instalar y configurar el hardware y el software.

- Personalizar los procedimientos de migración.
- Llevar a cabo pruebas de pre-validación.
- Realizar la migración.
- Verificar que la migración haya sido completada.

Como se mencionaron en las actividades, en la fase de migración, se deben comunicar todos los planes que estén involucrados y/o relacionados con la migración; obtener, instalar y configurar el hardware y software necesario. Recomiendan una prueba de validación de datos antes de la migración, en adición a las pruebas de validación post-migración. Estas pruebas confirmarán que los datos están en el mismo estado y disponible después de la migración.

Pero la tarea principal de esta fase es la migración como tal de los datos, la cual se debe realizar con todo el cuidado disponible y con las verificaciones necesarias para tratar de reducir al mínimo los problemas durante y después de la migración.

➤ **Validar:** Entre las actividades de post-validación se incluyen:

- Ejecutar pruebas de post-validación.
- Hacer la transferencia de conocimiento (si se requiere).
- Comunicar información del proyecto.
- Generar un informe de las estadísticas de la migración.
- Tener una reunión de cierre de migración.

Luego de completar la migración, sugieren recopilar las estadísticas de la migración y preparar un informe que señale lo que funcionó, lo que no funcionó y las lecciones aprendidas. El informe debería ser compartido con todos los miembros del equipo. Este tipo de informes son críticos para la construcción de procesos repetitivos y consistentes a través de un mejoramiento continuo de procesos, creciendo sobre lo que funcionó o cambiando lo que no funcionó.

Además, documentar el proceso de migración puede servir para el entrenamiento del personal en el futuro, lo cual puede ser beneficioso para las próximas migraciones, para reducir riesgos y costos.

Dentro de ese informe se debería incluir las pruebas realizadas y los resultados de las mismas y la importancia que tiene cada prueba, dependiendo del área o áreas a las cuales estaba dirigida.

Al igual que otras metodologías, está orientada más a ser empleada con sus herramientas, pero es de las metodologías más sencillas y flexibles que pueden ser estudiadas y moldeadas a las necesidades de una organización que está en el proceso de realizar una migración de datos. No hacen mención del tipo de pruebas que se deberían realizar, pero persiste la presencia de las actividades de prueba, como la metodología, las pruebas se generarían con el apoyo de los demás miembros del equipo y de los dueños de los datos.

6.3. COMPARACIÓN DE LOS ENTORNOS

Si se quiere realizar un análisis comparativo del estado en el que se encuentran los procesos y metodologías para las Pruebas para la Migración y Validación de datos en el día de hoy en Medellín y a nivel internacional, primero se definirán los aspectos por medio de los cuales se realizará el análisis, en segundo lugar se hará el análisis como tal de cada aspecto en ambos entornos, en Medellín y a nivel internacional, y por último, se intentará realizar una posible proyección de hacia dónde se podría estar dirigiendo esta área y qué se puede esperar ver en un futuro no muy lejano.

Primero, definamos los aspectos que se emplearán para realizar el análisis. Dado que el énfasis de esta investigación son las Pruebas para la Migración y Validación de Datos, se iniciará el análisis con las metodologías para realizar la Migración y Validación de Datos; el siguiente aspecto serán las pruebas empleadas para los procesos de migración, luego se mirarán las herramientas empleadas, tanto para las migraciones como para las pruebas, y por último, se hablará sobre la Calidad de los Datos y los planes para los mismos.

Como se dijo, el primer aspecto que se analizará son las Metodologías para las Migraciones y Validaciones de Datos.

En Medellín: Muy pocas empresas cuentan con una metodología diseñada específicamente para realizar proyectos de migración. La mayoría simplemente estructuran un proyecto general, sin consideraciones específicas para las migraciones de datos, que más tarde tienen que ser corregidos.

Internacional: La mayoría de las empresas ya tienen una metodología estructurada y probada en muchas ocasiones, para llevar a cabo migraciones de datos. En la actualidad, muchas se encuentran estandarizando y mejorando dichas metodologías. Muchas de esas metodologías han salido de experiencias vividas y recolección de mejores prácticas.

En la actualidad se podría decir que las empresas en Medellín se encuentra un paso detrás en lo que respecta al empleo y desarrollo de metodologías específicas para este tipo de proyectos. El número de proyectos de migración está creciendo y seguirá creciendo en los próximos años, eso empujará a las empresas a tener metodologías más confiables y repetitivas, que les puedan garantizar mayores probabilidades de éxito.

El siguiente aspecto a mirar son las pruebas para la migración y validación de datos.

En Medellín: Las empresas realizan pruebas empleando los conocimientos de los usuarios y pruebas que integran a los scripts o programas de migración para hacer una valoración en tiempo de ejecución de la migración. Se realizan pruebas básicas sobre los datos al momento de la migración.

Internacional: Realizan el mismo tipo de pruebas que en Medellín en la actualidad, pero adicionalmente realizan pruebas sobre los datos mismos dentro de la base de datos (referencias cruzadas con la base de datos de origen, chequeo cruzado de columnas y filas, etc.). Se está comenzando a emplear las técnicas de pruebas empleadas en el testing de software para probar las bases de datos, como pruebas de caja negra y caja blanca, pruebas de unidad orientados a las bases de datos, etc.

Las diferencias son mínimas en este aspecto, dado que las otras técnicas empleadas a nivel internacional son nuevos enfoques que se están presentando en la comunidad tecnológica.

El tercer aspecto que se empleará para analizar los dos entornos, son las herramientas empleadas para realizar las migraciones y para realizar las pruebas.

En Medellín: Hay uso de pocas herramientas que ayuden a automatizar el proceso de migración y realizar las pruebas. Tanto la migración como las pruebas son llevadas a cabo empleando lenguajes de programación para desarrollar programas que las realicen.

Internacional: Hay mayor uso de herramientas para automatizar los procesos de migración, aunque aún se emplean los lenguajes de programación en algunas migraciones. De igual manera, se están comenzando a emplear herramientas para automatizar las pruebas, aunque estas no están enfocadas simplemente a este tipo de pruebas, son empleadas para todo tipo de pruebas sobre las bases de datos.

La falta de conocimiento sobre las herramientas en las empresas encuestadas en Medellín es lo que marca la diferencia principal y en otras ocasiones son los costos para adquirir dichas herramientas. A pesar de que a nivel internacional están empleando herramientas para realizar pruebas sobre las bases de datos, muchas de esas herramientas son nuevas y aún están siendo probadas y mejoradas.

El último aspecto que se empleará para el análisis, es el uso de planes de Calidad de Datos.

En Medellín: Sorprendentemente, solamente una de las empresas encuestadas emplea planes de calidad.

Internacional: Entre todos los encuestados, todas las empresas ya tenían un plan de Calidad de Datos en el cual se podían basar para la migración de datos, al momento de hacer limpiezas y/o transformaciones de los datos.

Los datos se están convirtiendo en un activo de gran valor en todas las empresas en el Mundo, y Colombia no es la excepción, con la gran acogida que están teniendo soluciones tecnológicas como las de Inteligencia de Negocios, las empresas y sus directivas comenzarán a ver la importancia que tiene un plan de

Calidad de Datos y los beneficios que este le puede traer, no solo a nivel de sus sistemas sino para su negocio también.

En los siguientes años podríamos esperar ver una mayor adopción de metodologías y mejores prácticas en Medellín y en Colombia; con la adopción de estas metodologías comienzan las necesidades de herramientas y técnicas para realizar pruebas para proyectos de migración y validación de datos. De igual forma, se comenzarán a ver más y más empresas implementando planes de calidad.

7. HERRAMIENTAS PARA MIGRACIÓN Y VALIDACIÓN DE DATOS

En el pasado y aun hoy en día, las migraciones son programadas para ser realizadas en horarios de bajo uso de los sistemas o en horas fuera de las de trabajo de la organización; pero las organizaciones actuales tienen ventanas de tiempo muy pequeños o en algunos casos inexistentes en los cuales no se utilizan los sistemas. En lugar de posponer la adquisición o las migraciones a nuevas tecnologías para evitar al máximo las complicaciones con la migración de datos, se pueden emplear tecnologías más recientes, que tratan de ayudar a mantener al mínimo las complicaciones y permitir un mayor grado de flexibilidad al momento de planear, diseñar e implementar el proceso de migración de los datos.

Pero antes de elegir una herramienta sobre otra, se deben tener en cuenta algunos factores claves cuando se están evaluando las opciones. En la siguiente tabla se explicarán los factores a considerar, estos fueron recopilados luego de analizar la situación actual y lo que se necesita lograr con las migraciones de datos. Algunos son aspectos empleados por proveedores, otros son simplemente consideraciones que nacieron de esta investigación; no todas las herramientas que se mirarán en esta sección cumplirán con todos estos aspectos.

Factores Claves	Descripción
Desempeño general.	Qué tan rápido los datos pueden ser transferidos entre el sistema de origen y el destino, considerando la sobrecarga que se le aplicarán a las aplicaciones.

Seguridad y Protección de los datos.	Capacidad de recuperación ante fallas. Si algo sale mal, la migración puede ser terminada y el procesamiento puede ser reiniciado o continuado.
Migraciones a múltiples plataformas.	La mayoría de las migraciones se realizan entre productos de distintos proveedores. Capacidad de la herramienta de permitir migraciones entre productos de distintos proveedores. Qué plataformas soporta.
Tiempos de inactividad permitidos.	Las aplicaciones tienen diferentes grados de criticidad dentro del negocio y por ello tienen diferentes tiempos de inactividad aceptables. La herramienta permite migraciones online u offline (cuánto tiempo de inactividad requiere para realizar la migración).
Flexibilidad en diseño del proceso.	Qué tan personalizable es el proceso de migración con la herramienta o posee una metodología implícita en su construcción.
Volumen de datos manejable.	Cuál es el volumen de datos que maneja sin presentar problemas. Tamaño de los proyectos que es capaz de soportar.
Pruebas.	Permite personalizar las pruebas a realizar durante la migración o viene con un conjunto de pruebas preestablecidas para la migración. Qué tipo de pruebas permite realizar. Una de las más importantes que debería estar presente en toda herramienta es la detección automática de violaciones de integridad.
Reportes de estado de migración.	Deberían brindar reportes de manera sencilla y fácil de entender sobre el estado de la migración. De igual manera, presentar los resultados de la migración, es decir, como conteo de registros creados, registros con

	problemas, etc.
--	-----------------

Tabla 3: Factores Claves para la selección de una herramienta de migración.

Fuente: Recopilación de diversas fuentes: ESG, Softek, Oracle, etc. Elaboración propia.

El peso de estos factores en la decisión sobre cuál herramienta emplear dependerá de las necesidades de la organización para la migración de datos, donde para una organización, la capacidad de realizar migraciones entre los productos de diversos proveedores es lo más importante; para otro podría ser la cantidad de tiempo de inactividad que tendrán sus aplicaciones, y en este caso lo más probable es que elijan un sistema que reduzca ese tiempo al mínimo o permita realizar una migración online sin afectar las aplicaciones.

La decisión de si una organización emplea una herramienta de migración o un grupo de desarrolladores para realizar la migración de manera manual, depende en la mayoría de los casos de la comparación de los costos de la herramienta contra los costos de las horas de trabajo del equipo de desarrollo. Organizaciones que conocen de la existencia de dichas herramientas las eligen sin pensarlo dos veces, no siendo la única opción para llevar a cabo la migración. Sin embargo, algunas de las personas que conocen sobre la existencia de estas herramientas y no las emplean porque no conocen los beneficios que les pueden brindar.

Las herramientas para la migración de datos disponibles son generalmente divididas en tres categorías, cada una con sus ventajas y desventajas. Dichas categorías son:

- *Host based:* Son recomendadas para las migraciones en entornos de almacenamiento heterogéneos. Estas herramientas toman la forma de

software independiente del proveedor de almacenamiento y por lo general son de bajo costo. Una empresa que presenta estas herramientas es Softek.

- *Array based*: Son recomendados para las migraciones en entornos de almacenamiento homogéneos. Por lo general son herramientas propietarias del proveedor del sistema de almacenamiento. Un ejemplo de este tipo de herramientas son las suministradas por IBM, Hitachi Data Systems y Oracle.
- *Appliance/Fabric based*: Son herramientas de terceros, que liberan del trabajo de migración a los dos anteriores. Un ejemplo de un proveedor de este tipo es Falconstor.

Las migraciones manuales o desarrolladas por programadores del equipo aun es un método válido, que depende de las necesidades del proyecto, puede ser tan pequeño que no justifica la adquisición de una herramienta de migración.

En la actualidad, las migraciones no sólo involucran de base de datos a base de datos, sino simplemente migración de archivos o data files de un sistema a otro, la complejidad de estas migraciones en lo que respecta el mapeo de datos no es alta, pero de igual manera guarda complejidad el proceso de migración, porque en muchos casos son de aplicaciones que no se pueden tener inactivas. Este tipo de migraciones requiere de herramientas de migración que por lo general son online.

Entre los proveedores de herramientas para migraciones, distintas a migraciones entre bases de datos, se encuentra Softek, Veritas y Legatos. Debido al enfoque de este proyecto de grado, sobre las migraciones en bases de datos específicamente, no se hará referencia ni se explicarán herramientas para migraciones que no involucren bases de datos, pero hay que tener en cuenta que sus principios, metodologías y objetivos son iguales a los empleados para las migraciones en bases de datos.

Algunas de las herramientas que se analizarán a continuación, son:

- ✓ Oracle SQL Developer Migration Workbench 10.1.0.4.0
- ✓ Oracle Database Migration Verifier
- ✓ AdventNet SwisSQL – Data Migration Tool 4.9
- ✓ MySQL Migration Toolkit

Adicional a las herramientas para realizar migraciones y validaciones de datos, a continuación aparece una lista de herramientas empleadas para llevar a cabo pruebas sobre las bases de datos, no específicas para las migraciones, pero que pueden ser empleadas para tal fin. Se describirán las características de una de ellas y se presentarán en una tabla otras herramientas y los enlaces Web a sus páginas. Algunas son herramientas libres que aún están siendo desarrolladas y mejoradas. La herramienta que será descrita es:

- ✓ Microsoft Visual Studio 2005 Team Edition for Database Professionals

Las herramientas que se listarán en la siguiente tabla son para realizar pruebas de unidad sobre las bases de datos.

Herramienta	Sitio Web
DbUnit	http://www.dbunit.org/
NDbUnit	http://www.ndbunit.org/
Quest Code Tester for Oracle	http://www.quest.com/code%2Dtester%2Dfor%2Doracle/
XUnit	http://weblogs.asp.net/roshero/archive/2004/10/05/238201.aspx

Tabla 4: Lista de Herramientas de Prueba para Bases de Datos

Fuente: Elaboración propia.

7.1. ORACLE SQL DEVELOPER MIGRATION WORKBENCH



Oracle SQL Developer Migration Workbench

El Oracle SQL Developer Migration Workbench (10.1.0.4.0) es una herramienta nueva, redesarrollada, que entiende ampliamente las funcionalidades y la usabilidad de la versión anterior. Su fuerte integración con SQL Developer brinda una interfaz de usuario mejorada.

Es una herramienta intuitiva que permite la migración de bases de datos de terceros a bases de datos Oracle 9i y Oracle 10g, incluyendo objetos de los esquemas, triggers y procedimientos almacenados de manera muy sencilla.

La siguiente tabla contiene la lista de las bases de datos de terceros soportadas por Oracle SQL Developer Migration Workbench.

Bases de Datos Soportadas	Versiones
DB2/400	V4R3, V4R5
DB2/UDB	6.x, 7.1, 7.2
Informix Dynamic Server	7.3, 9.1
Microsoft Access	97, 2000, 2002, 2003
Microsoft SQL Server: 6.5	6.5, 7.0, 2000, 2005
MySQL	3.23, 4.0, 5.0
Sybase Adaptive Server	11, 12

Tabla 5: Bases de Datos soportados por Oracle SQL Developer Migration Workbench

Fuente: Oracle SQL Developer Migration Workbench Datasheet

http://www.oracle.com/technology/tech/migration/workbench/pdf/new_omwb_ds.pdf

Entre las principales características del Migration Workbench, se encuentran:

- *Interfaz de Usuario Mejorado:* Ahora está integrado al Oracle SQL Developer, beneficiándose de la GUI mejorada de dicha herramienta.
- *Migración con privilegios mínimos:* Los usuarios pueden migrar objetos desde la base de datos origen a la destino sin necesidad de tener privilegios de DBA. Les permite migrar objetos a los cuales sólo ellos tienen acceso y no requiere privilegios especiales para funcionar.
- *Migración Rápida:* Manejado por un wizard/asistente para facilitar la migración de esquemas y datos a Oracle. Soporta la migración con privilegios mínimos.
- *Generación de scripts DDL:* La generación de DLL's para la creación de los esquemas Oracle le da al usuario control total sobre las salidas de los DDL.
- *Modalidad de Datos Online:* El movimiento de datos configurable, de alta velocidad y paralelo significa que los datos pueden ser migrados lo más rápido posible.
- *Movimiento de Datos offline:* Los usuarios pueden generar scripts específicos para una plataforma para descargar los datos de una base de datos de un tercero, junto con los archivos de import de SQL*Loader para permitir la carga de datos cuando sea necesario.
- *Navegador de Migración:* Representación visual de la base de datos origen y su equivalente en Oracle, están disponibles en una interfaz de usuario intuitiva, suministrando comandos sencillos que guían el proceso de migración.
- *Tecnología de traducción y parsing avanzado:* Puede reconocer T/SQL y traducirlo al equivalente en PL/SQL. Maneja independencias entre objetos de los esquemas, dependencias entre esquemas y es utilizado para transformar procedimientos almacenados, triggers, vistas, reglas y constraints.

- *Control Completo de los Usuarios:* Capacidades personalizables, incluyen la habilidad de elegir el mapeo de todos los tipos de datos, selección de los objetos a migrar y la habilidad de eliminar o renombrar objetos durante el proceso de migración.
- *Log de Migración Interactivo:* Un log o registro de migración mejorado le brinda información sobre el proceso de migración. Los ítems en el log son seleccionables para permitir acceso instantáneo a los objetos problemáticos.

La **Figura 10** muestra el procedimiento que emplea la herramienta para llevar a cabo las migraciones.

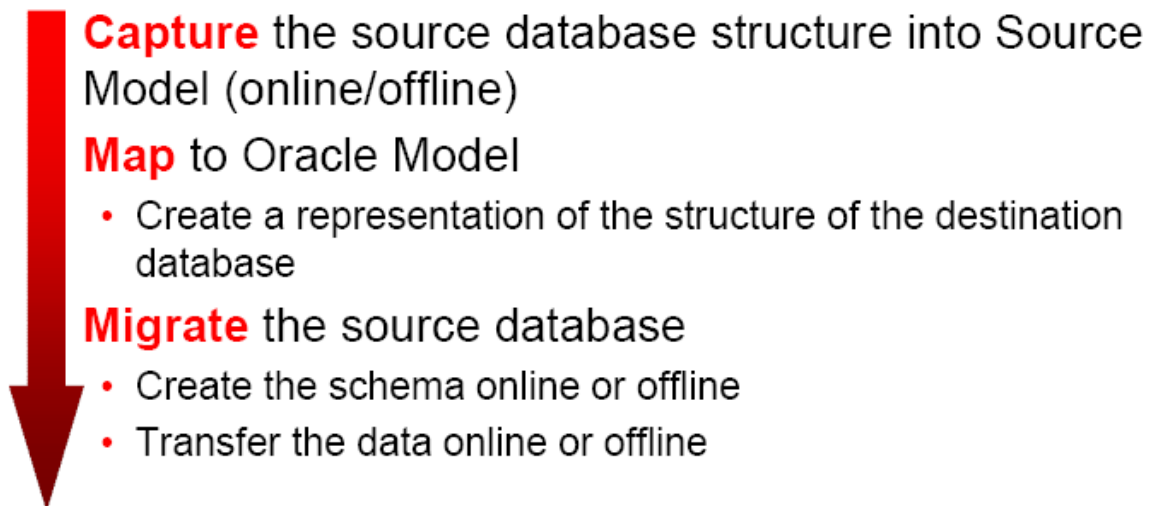


Figura 10: Modelo de Migración de Oracle SQL Developer Migration Workbench.

Fuente: Sitio Web de la herramienta.

<http://www.oracle.com/technology/tech/migration/workbench/>

El procedimiento está compuesto por tres pasos básicos que son:

1. **Captura:** Obtienen la estructura de la base de datos origen al Modelo de Origen, lo pueden obtener online u offline.
2. **Mapeo:** Al modelo Oracle. Crean una representación de la estructura de la base de datos destino.
3. **Migran:** La base de datos origen. Pueden crear los esquemas y transferir los datos online u offline.

VENTAJAS Y DESVENTAJAS

Ventajas:

- Permite migraciones tanto online como offline.
- Verificaciones básicas de validez de los datos.
- Fácil de utilizar.

Desventajas:

- Solo permite migraciones hacia Oracle.

En la **Figura 11** se observa una vista del asistente para migraciones rápidas.

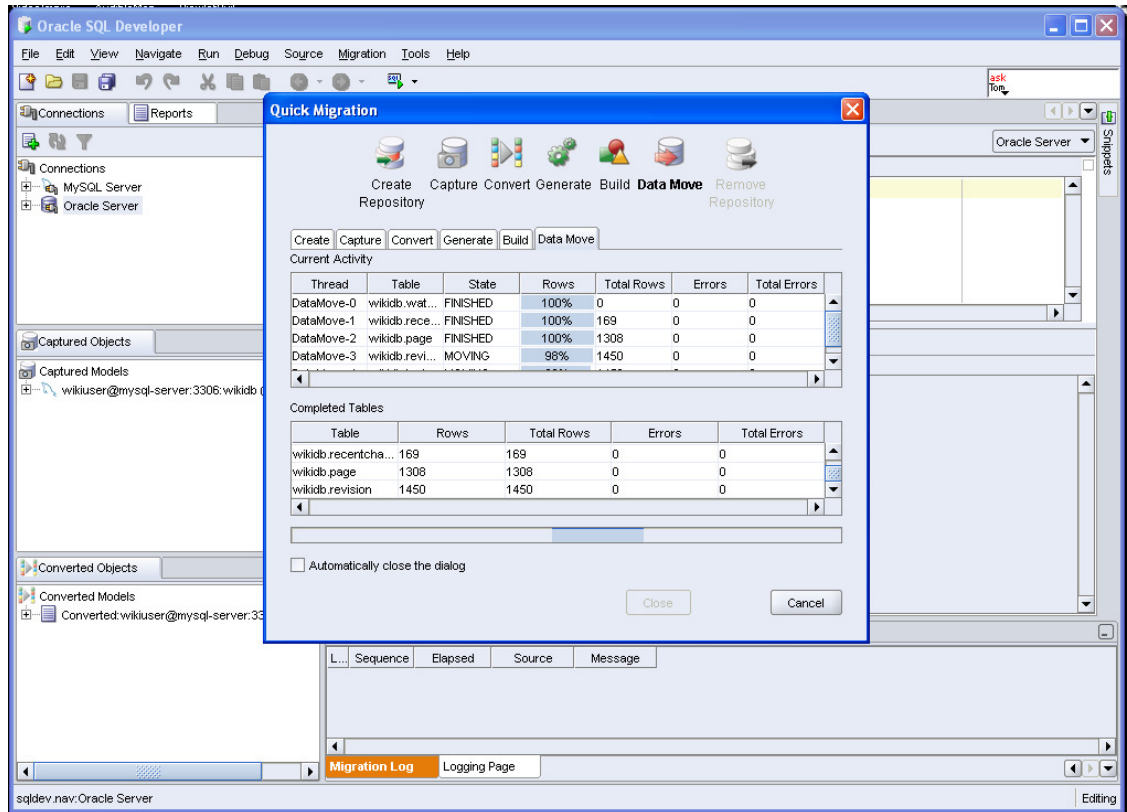


Figura 11: Vista Interfaz de Usuario para la Migración Rápida.

Fuente: Sitio Web de la herramienta.

<http://www.oracle.com/technology/tech/migration/workbench/images/quickmigrate.png>

7.2. ORACLE DATABASE MIGRATION VERIFIER



Oracle Database Migration Verifier

Como complemento al Oracle SQL Developer Migration Workbench, crearon una herramienta separada para realizar la verificación de la migración hacia la nueva base de datos, esta herramienta recibe el nombre de Oracle Database Migration

Verifier (DMV), empleada para asistir en migraciones desde SQL Server y Sybase Adaptive Server hacia Oracle. El trabajo de la herramienta es analizar la base de datos de origen y compararlo con la base de datos Oracle destino, para validar la integridad estructural de los objetos y de los datos dentro de la base de datos.

Las bases de datos entre las cuales se pueden realizar las pruebas de verificación con la herramienta, se encuentran en la siguiente tabla.

Origen		Destino	
Base de Datos	Versión	Base de Datos	Versión
Microsoft SQL Server	2000	Oracle Database	9.2.x
Sybase Adaptive Server	12		10g

Tabla 6: Bases de Datos soportados por Oracle Database Migration Verifier

Fuente: Sitio Web de la herramienta

<http://www.oracle.com/technology/tech/migration/dmv/index.html>

En la actualidad es una herramienta basada en la tecnología java y no forma parte de la suite de Oracle SQL Developer, como Oracle Migration Workbench, pero entre las proyecciones futuras de la herramienta se encuentra el incorporarla a la herramienta y adicionar soporte a más motores de bases de datos.

Entre las características principales de la herramienta se encuentran:

- ↪ Compara esquemas y datos entre las bases de datos de origen y la base de datos Oracle. Confirmando que la base de datos migrada esté completa.
- ↪ Realiza pruebas de conteo sobre las tablas de usuario para verificar la integridad de los datos migrados.
- ↪ Permite la verificación por medio de pruebas sobre las tablas, índices, funciones, procedimientos almacenados, triggers y vistas, para asegurar que hayan migrado exitosa y correctamente.
- ↪ Valida los datos migrados a Oracle con respecto a la base de datos de origen de los datos.
- ↪ Genera un Reporte de Verificación detallado de los resultados de las pruebas de verificación de esquemas y datos, resumido por nombre y tipo de objeto.
- ↪ Su interfaz es por línea de comando. Se configuran las bases de datos a las cuales se conectará a través de los archivos de configuración que acompañan la herramienta.
- ↪ Emplea conectores JDBC como medio de comunicación con las bases de datos.
- ↪ Alta portabilidad por ser desarrollada en java.

La siguiente tabla contiene una lista de los objetos sobre los cuales realiza las pruebas, el tipo de prueba que realiza sobre el objeto y si dicha prueba es aplicable a los dos motores de bases de datos con los cuales puede interactuar la herramienta.

Tipo de Objeto	Prueba de Verificación	SQL Server 2000 (Sí/No)	Sybase 12 (Sí/No)
Tablas	Existencia de la tabla y las columnas de la tabla.	SÍ	SÍ
	Nombre de las columnas de la tabla.	SÍ	SÍ

	Orden de las columnas de la tabla.	SÍ	SÍ
	Tipo de Datos de las columnas de la tabla.	SÍ	SÍ
	Aceptación de valores nulos en las columnas de la tabla.	SÍ	SÍ
	Atributos de las columnas de la tabla.	SÍ	SÍ
	Valores por defecto de las columnas de la tabla.	SÍ	SÍ
Vistas	Existencia de la vista.	SÍ	SÍ
	Nombre de las columnas de la vista.	SÍ	SÍ
	Orden de las columnas de la vista.	SÍ	SÍ
	Atributos de las columnas de la vista.	SÍ	SÍ
Índices	Existencia del índice.	SÍ	SÍ
	Orden de las columnas del índice.	SÍ	SÍ
Funciones	Existencia de la función y de los parámetros de la función.	SÍ	No
	Orden de los parámetros de la función.	SÍ	No
	Tipo de Dato de los parámetros de la función.	SÍ	No
	Atributos de los parámetros de la función.	SÍ	No
Procedimientos Almacenados	Existencia del procedimiento y los parámetros del procedimiento.	SÍ	SÍ
	Orden de los parámetros del procedimiento.	SÍ	SÍ
	Tipo de Dato de los parámetros del procedimiento.	SÍ	SÍ
	Atributos de los parámetros del procedimiento.	SÍ	SÍ
Triggers	Existencia del trigger.	SÍ	SÍ
Datos	Conteo de las filas de la	SÍ	SÍ

	tabla.		
--	--------	--	--

Tabla 7: Tipos de Objetos y pruebas realizables sobre los mismos.

Fuente: Traducción de la información del sitio Web de la herramienta

<http://www.oracle.com/technology/tech/migration/dmv/index.html>

VENTAJAS Y DESVENTAJAS

Ventajas:

- Transportable entre plataformas.
- Realiza pruebas de verificación sobre la mayoría de los objetos de la base de datos.

Desventajas:

- No está integrada al resto de la suite de Oracle SQL Developer.
- Sólo permite hacer verificaciones entre Oracle y otros dos motores (SQL Server y Sybase).
- No permite personalizar las pruebas a realizar.

La **Figura 12** es un ejemplo del reporte que la herramienta les presenta a los usuarios luego de realizar las pruebas de verificación sobre la base de datos.

```

Database Migration Verifier
Source Database :oracle.mtg.schemaverifier.query.SQLServer2KQueryComponent
Oracle Database :oracle.mtg.schemaverifier.query.Oracle9iQueryComponent
Date : June 10, 2004 15:19:22

Schema verification details

Summary
Object Type          Total      Existence Failures      Comparison Failures
TABLE                12         0                        6
VIEW                 1         0                        0
STORED PROCEDURE     4         0                        1
TABLE FUNCTION       0         0                        0
TRIGGER              1         0                        0

Schema Object Existence Failures
Object Type          Object Name

Schema Object Comparison Failures
Object Type          Object Name          Column/Argument Name      Description
TABLE                titleauthor         au_id                      Column Datatype
TABLE                titleauthor         title_id                   Column Datatype
TABLE                sales               titleidind                 Index Existance
TABLE                sales               title_id                   Column Datatype
TABLE                roysched            titleidind                 Index Existance
TABLE                roysched            title_id                   Column Datatype
TABLE                employee            emp_id                     Column Datatype
TABLE                employee            hire_date                  Column Default
TABLE                authors             au_id                      Column Datatype
TABLE                titles              type                       Column Existance
TABLE                titles              title_id                   Column Datatype
TABLE                titles              pubdate                    Column Default
INDEX                UPKCL_taind         UPKCL_taind                Index Column order
INDEX                auidind             auidind                    Index Column order
INDEX                titleidind          titleidind                  Index Column order
INDEX                UPKCL_sales         UPKCL_sales                 Index Column order
INDEX                employee_ind        employee_ind                 Index Column order
INDEX                PK_emp_id           PK_emp_id                   Index Column order
INDEX                UPKCL_auidind       UPKCL_auidind               Index Column order
INDEX                aunmind             aunmind                      Index Column order
INDEX                UPKCL_titleidind   UPKCL_titleidind            Index Column order
INDEX                titleind            titleind                     Index Column order
STORED PROCEDURE     reptq3                 @type                       Argument Existance

Data verification details
Data verification Comparison Failures
Table Name          Source Row Count          Oracle Row Count

```

Figura 12: Muestra del Reporte de Verificación presentado por Oracle DMV

Fuente: Sitio Web de la herramienta

<http://www.oracle.com/technology/tech/migration/dmv/index.html>

7.3. ADVENTNET SWISSQL – DATA MIGRATION TOOL 4.9



AdventNet SwisSQL – Data Migration Tool 4.9 es una suite para la Migración Automática de Bases de Datos que facilita la migración de procedimientos almacenados, esquemas y datos entre bases de datos tales como Oracle, IBM DB2, MS/SQL Server, Sybase, SAP DB, MySQL, PostgreSQL y Access.

Automatiza hasta el 90% de las labores manuales de la migración de una base de datos.

La siguiente tabla contiene la lista de los motores de bases de datos origen que soporta la herramienta y las bases de datos destinos que soporta, junto con sus correspondientes versiones.

Bases de Datos Origen	Bases de Datos Destino
Oracle 8i, 9i y 10g. IBM DB2/UDB para Linux, Unix y Windows 7.1 y 8.x. MS-SQL Server 7, 2000 y 2005. Sybase ASE 11.x, 12.x y 15.0. MySQL 3.23.x, 4.x y 5.0. PostgreSQL 7.3.x, 7.4.x y 8.x. MySQL MaxDB (SAP DB) 7.5.00. MS Access 2000 Archivos de Datos: Excel y CSV	Oracle 9i y 10g. IBM DB2/UDB para Linux, Unix y Windows 7.x y 8.x. MS-SQL Server 2000 y 2005. Sybase ASE 11.x, 12.x y 15.0. MySQL 3.23.x, 4.x y 5.0. PostgreSQL 7.3.x, 7.4.x y 8.x.

Tabla 8: Bases de Datos soportadas por SwissQL Data Migration Tool.

Fuente: Sitio Web de la Herramienta.

<http://www.swissql.com/products/datamigration/data-migration.html>

Entre las características principales que el proveedor menciona sobre la herramienta, se encuentran:

- Adicional a la migración entre diversas bases de datos, permite la migración de datos desde archivos en EXCEL y CSV (Comma Separated Values – Valores Separados por Comas) hacia las bases de datos.

- ↵ Permite la conexión a bases de datos utilizando conectores JDBC, incluye conectores genéricos y permite la conexión a SQL Server a través de Windows Authentication.
- ↵ Migra tablas, índices, vistas y constraints juntos con los datos.
- ↵ Mantiene las relaciones entre las tablas como en la base de datos de origen, identificando las dependencias entre tablas maestras e hijas, de igual manera ordena las tablas con claves primarias y claves foráneas.
- ↵ Soporta la migración de procedimientos almacenados, funciones y triggers si el motor de bases de datos origen y destino son iguales. Actualmente sólo soporta migraciones de este tipo para motores SQL Server, Oracle, DB2 y Sybase.
- ↵ Permite la migración de datos basados en criterios (subgrupos de datos) cuando se está migrando de Oracle, SQL Server, MySQL y Sybase ASE a cualquier base de datos. Definen un criterio de filtración para una tabla (o tablas) que definirá qué filas serán migradas. Si hay una dependencia con otra tabla esa tabla también será migrada.
- ↵ Permite el almacenamiento de los archivos DDL.
- ↵ Crea un respaldo de las tablas en la base de datos origen antes de iniciar la migración.
- ↵ Tiene la opción de migrar datos y esquemas juntos, o por separado.
- ↵ Permite la selección de columnas específicas a ser migradas a una tabla; cambiar el nombre de tablas y columnas en la base de datos destino durante la migración, permite mapear columnas de la base de datos de origen a la columna apropiada en la base de datos destino, y permite el cambio de los tipos de datos de las columnas en una tabla.
- ↵ Interfaz de usuario amigable y fácil de utilizar a través de wizards o asistentes.
- ↵ Suministra reportes detallados y resumidos de la migración. Aunque se pueden ver reportes instantáneos de la migración cuando se presentan problemas.

- ↪ La integridad de los datos y los objetos de la base de datos migrada pueden ser verificados utilizando una herramienta llamada “*Database Migration Verifier*”. Actualmente la verificación solo es soportada para migraciones a SQL Server y Oracle.

Las pruebas que realiza la herramienta como tal son básicas, comenzando por la verificación de la integridad de los datos al migrar y la validación del tipo de datos. Manualmente no permite la configuración de pruebas adicionales. Pero los desarrolladores o encargados del proceso de migración pueden realizar pruebas de estrés y carga de los datos al sistema, contando con los reportes generados por la herramienta para comenzar a determinar un tiempo estimado para la migración, teniendo en cuenta que existe un entorno de pruebas antes de realizar la migración final al entorno de producción.

La herramienta tiene unos pasos específicos para realizar una migración, esos pasos son los siguientes:

1. Creación o selección de un proyecto de migración en el wizard.
2. Selección de la base de datos de la cual se obtendrán los datos.
3. Ingresar la información de conexión con la base de datos de origen y probarla.
4. Selección de la base de datos destino.
5. Ingresar la información de conexión con la base de datos destino y probarla.
6. Selección de los objetos en la base de datos origen que van a ser migrados y especificar, si es necesario, los criterios de filtración de los mismos.
7. Antes de realizar la migración como tal, la herramienta presenta un resumen de la información de la migración, donde detalla las bases de datos origen y destino con su respectiva información de conectividad, de igual manera presenta la información de los objetos a ser migrados y su detalle.

8. Durante el proceso de migración muestra paso a paso la migración de cada objeto y el estado del mismo. Al finalizar la migración se puede revisar el reporte de migración detallado o resumido.

La **Figura 13** es una vista general del wizard al inicio de un proyecto de migración y la **Figura 14** es lo que ven los usuarios durante el proceso actual de migración y el final.

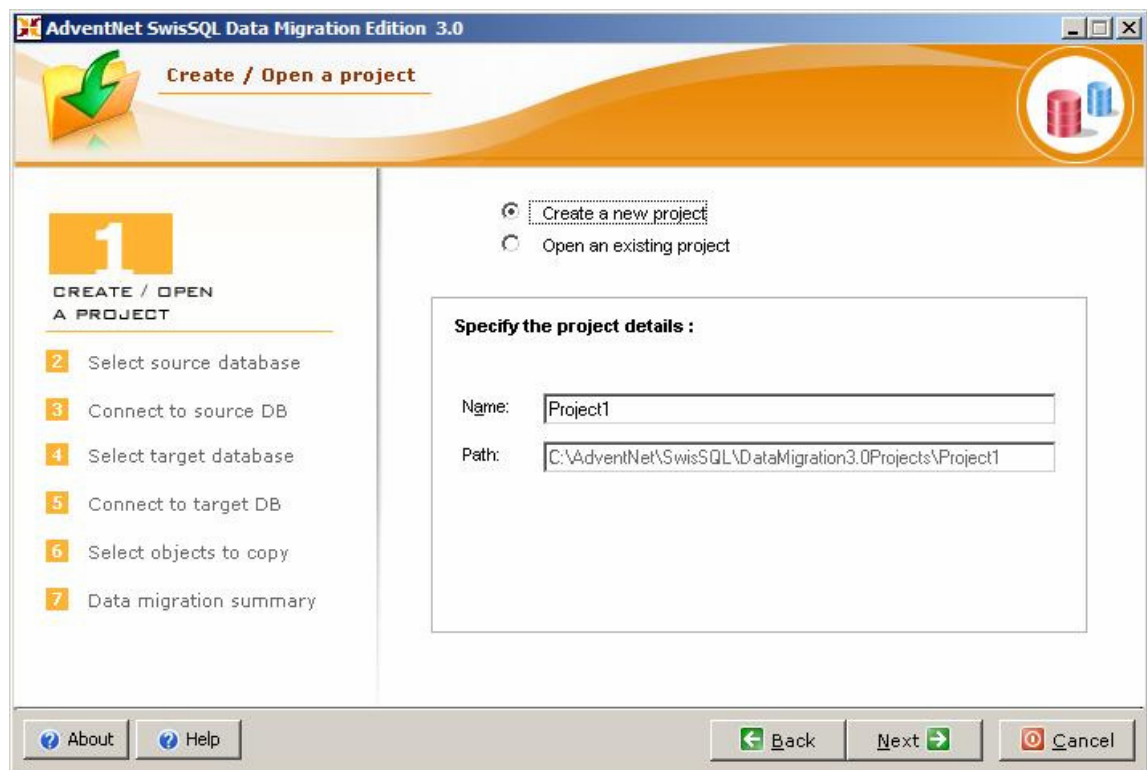


Figura 13: Vista inicial del wizard de SwisSQL – Data Migration Tool 4.9

Fuente: Sitio Web de la herramienta

<http://www.swissql.com/swissql-data-migration-review.html>

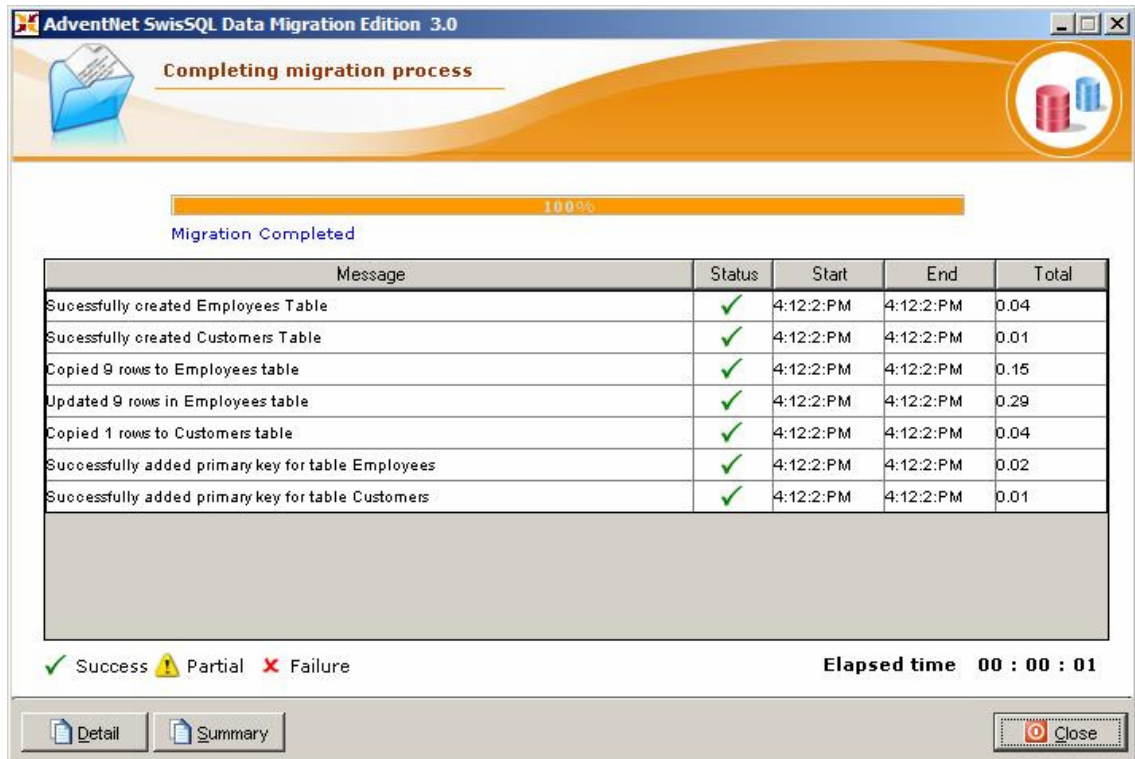


Figura 14: Vista Final del wizard de SwisSQL – Data Migration Tool 4.9

Fuente: Sitio Web de la herramienta

<http://www.swissql.com/swissql-data-migration-review.html>

VENTAJAS Y DESVENTAJAS:

Ventajas:

- ☑ Proceso de migración flexible, abierta y expandible.
- ☑ Rápida migración de datos a través de bases de datos garantizando la integridad de los datos.
- ☑ Re-estructura y transforma esquemas y datos con facilidad durante la migración.
- ☑ Enfoque unificado para la migración entre diferentes bases de datos.

- ☑ Bajo costo.
- ☑ Fácil de utilizar con un wizard.

Desventaja:

- ☒ No traduce códigos de procedimientos almacenados, funciones o triggers entre bases de datos. Requiere de una herramienta adicional para hacer ese tipo de migraciones.
- ☒ Requiere una herramienta adicional para las verificaciones. Y solo permite verificaciones para dos de los motores con los cuales puede realizar migraciones. La herramienta adicional que requiere para hacer las validaciones es el *Database Migration Verifier*, la cual es una herramienta desarrollada por Oracle.

7.4. MYSQL MIGRATION TOOLKIT

MySQL Migration Toolkit es una herramienta gráfica para migrar esquemas y datos desde varias bases de datos relacionales (Oracle, Microsoft SQL Server, Microsoft Access, etc.) a MySQL 5.0 (o posterior). Es un programa Java, por lo que se necesita tener instalado Java Runtime Environment 1.4.2 (o posterior).

En la **Figura 15** se presenta la pantalla inicial de MySQL Migration Toolkit

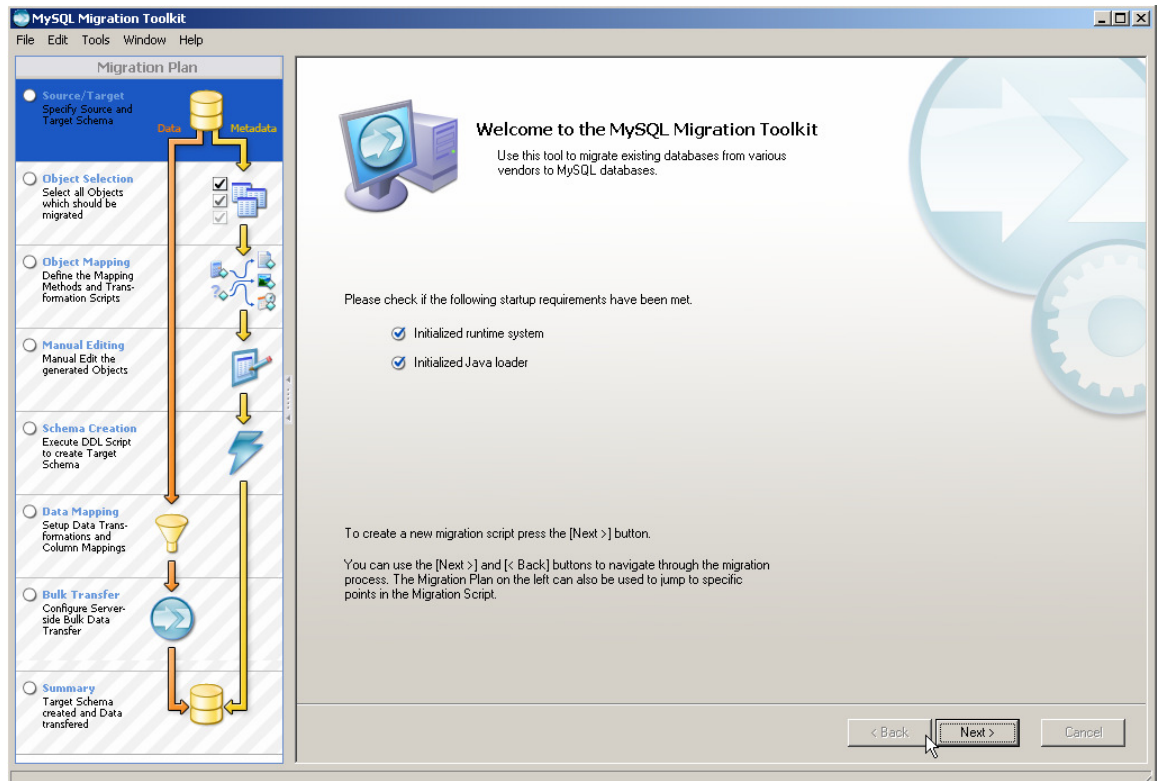


Figura 15: Bienvenido a MySQL Migration Toolkit

Fuente: Sitio Web de la herramienta - <http://dev.mysql.com>

El proceso de migración de datos se lleva a cabo a través de un asistente en ocho pasos.

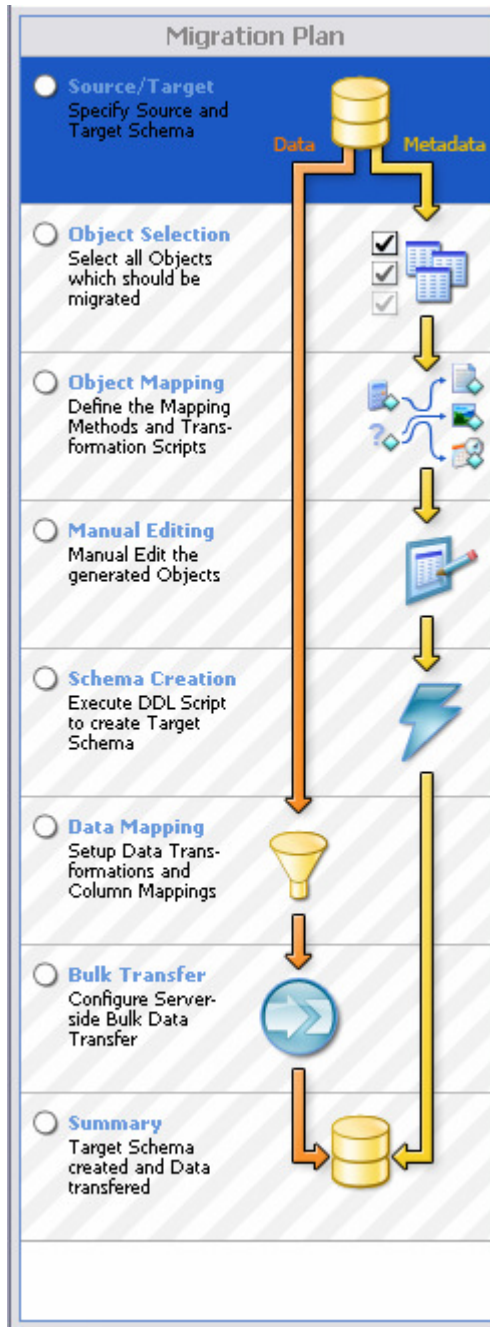


Figura 16: Plan de migración

Fuente: Sitio Web de la herramienta - <http://dev.mysql.com/doc/migration-toolkit/en/mysql-migration-toolkit-overview.html>

1. *Selección de la Fuente/Destino*: En el primer paso, se especifican los parámetros de conexión para los servidores de base de datos de la fuente (MS Access, Generic JDBC, MaxDB Database Server, MS SQL Server, MySQL Server, Oracle Database Server, Sybase Server) y el destino. Se especifica la conexión de la base de datos de origen y de la base de datos de llegada y se realiza una prueba de esta conexión.
2. *Selección de Objetos*: En el segundo paso, se escoge la base de datos que se desea migrar y se seleccionan los objetos, vistas, procedimientos almacenados que serán migrados.
3. *Mapeo de Objetos*: En el tercer paso, se define el método para el mapeo y los scripts de transformación de los objetos.
4. *Edición Manual*: En el cuarto paso, se pueden corregir manualmente los nuevos objetos para la migración y asegurar una transformación apropiada.
5. *Creación de Esquemas*: En el quinto paso, la herramienta, Migración MySQL, ejecuta los scripts DDL para crear los objetos transformados sobre el servidor destino MySQL.
6. *Mapeo de Datos*: En el sexto paso, se especifica cualquier cambio que necesiten los datos antes de ser migrados. Se mapean las columnas y especifican las transformaciones de datos necesarias.
7. *Transferencia de Grandes Volúmenes (Bulk Transfer)*: En el séptimo paso, la herramienta de Migración MySQL transfiere los datos del servidor de la fuente al servidor destino.
8. *Resumen (Sumario)*: En el octavo paso, la herramienta de Migración MySQL crea un informe con el resumen del proceso de migración. Luego se debe actualizar el servidor de MySQL para que aparezca la nueva base de datos, es decir, la base de datos migrada.

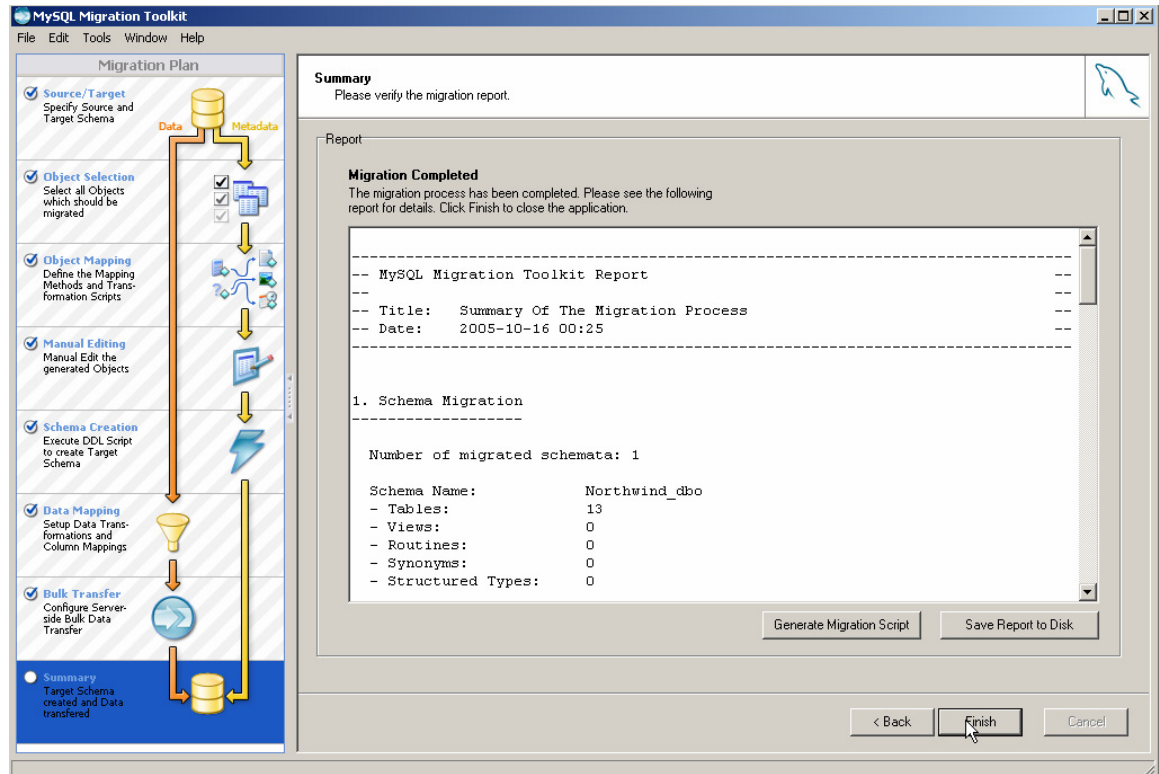


Figura 17: Reporte Final

Fuente: Sitio Web de la herramienta - <http://dev.mysql.com/doc/migration-toolkit/en/mysql-migration-toolkit-overview.html>

VENTAJAS Y DESVENTAJAS

Ventajas:

- Es una herramienta que puede hacer migraciones en línea.
- Sirve para generar scripts de la base de datos.
- Hace validaciones típicas de datos.
- Al terminar la migración muestra un reporte del proceso realizado.

Desventaja:

- ☒ Esta herramienta sólo sirve para hacer migraciones hacia MySQL.
- ☒ Tiene un bug en la interfaz gráfica de Linux, entonces de puede realizar la migración por medio de la interfaz.

7.5. MS VISUAL STUDIO 2005 TEAM EDITION FOR DATABASE PROFESSIONALS

Soporte para el refactoring de renombramiento de objetos de la base de datos

Tiene la capacidad de realizar cambios en el nombre de los esquemas y este cambio se debe realizar en todas las referencias relacionadas con el objeto.

- Reduce el tiempo requerido para la propagación de los cambios en el esquema.
- Reduce los errores causados al realizar los cambios manualmente.
- Simplifica el proceso de hacer cambios en el nombre del esquema, ya que si se realiza el cambio en cualquier parte, este cambio se refleja en los objetos que se relacionan con él en toda la base de datos.

Comparación de esquemas para mantener dos versiones de esquemas sincronizados:

La herramienta de comparación de esquemas de bases de datos proporciona a los administradores una forma de comparar y actualizar esquemas, de manera que se mantengan sincronizadas las bases de datos. Con esto se puede comparar la versión de la base de datos del proyecto con las bases de datos que se encuentran en producción. Con esto se puede crear un script para actualizar las bases de datos y hacerlas equivalentes otra vez.

- Identifica problemas potenciales e inconsistencia de datos antes de liberar el proyecto.
- Reduce el tiempo empleado para localizar las fallas causadas por cambios en el esquema de bases de datos.
- Reduce la cantidad de desarrollo que se tiene que arreglar debido al desarrollo con respecto a la estructura de datos equivocada.

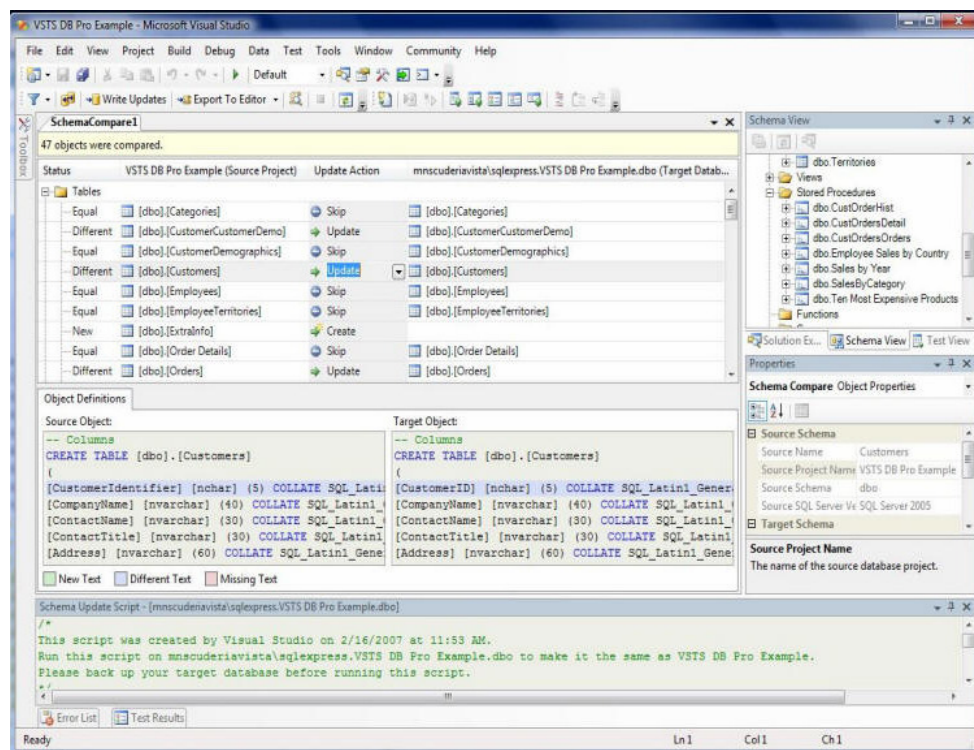


Figura 18: Interfaz gráfica del MS Visual Studio 2005 Team

Fuente: Sitio Web de la herramienta

<http://msdn2.microsoft.com/en-us/teamsystem/bb507197.aspx>

Comparación de datos para tener los datos en dos bases de datos sincronizados:

Es similar al de comparación de esquemas, pero en la comparación de datos se pueden actualizar los datos de la base de datos, en vez de mantenerlos en un esquema sincronizado. Con la comparación de datos se pueden comparar dos bases de datos que tienen el mismo esquema y esto ayuda a identificar problemas potenciales debido a datos contradictorios.

Funcionalidades extendidas para pruebas de unidad

Microsoft Visual Studio 2005 Team Edition for Database Professionals implementa una infraestructura completa para pruebas de unidad para su base de datos. Se pueden escribir las pruebas o se pueden correr las pruebas independientemente de la serie de pruebas que tiene Microsoft Visual Studio 2005. Esta herramienta también proporciona pruebas para la aplicación y para la base de datos.

- Suite de pruebas integrado.
- Personalizable.
- Pruebas de unidad de bases de datos combinados con pruebas de aplicaciones para obtener una suite de pruebas.

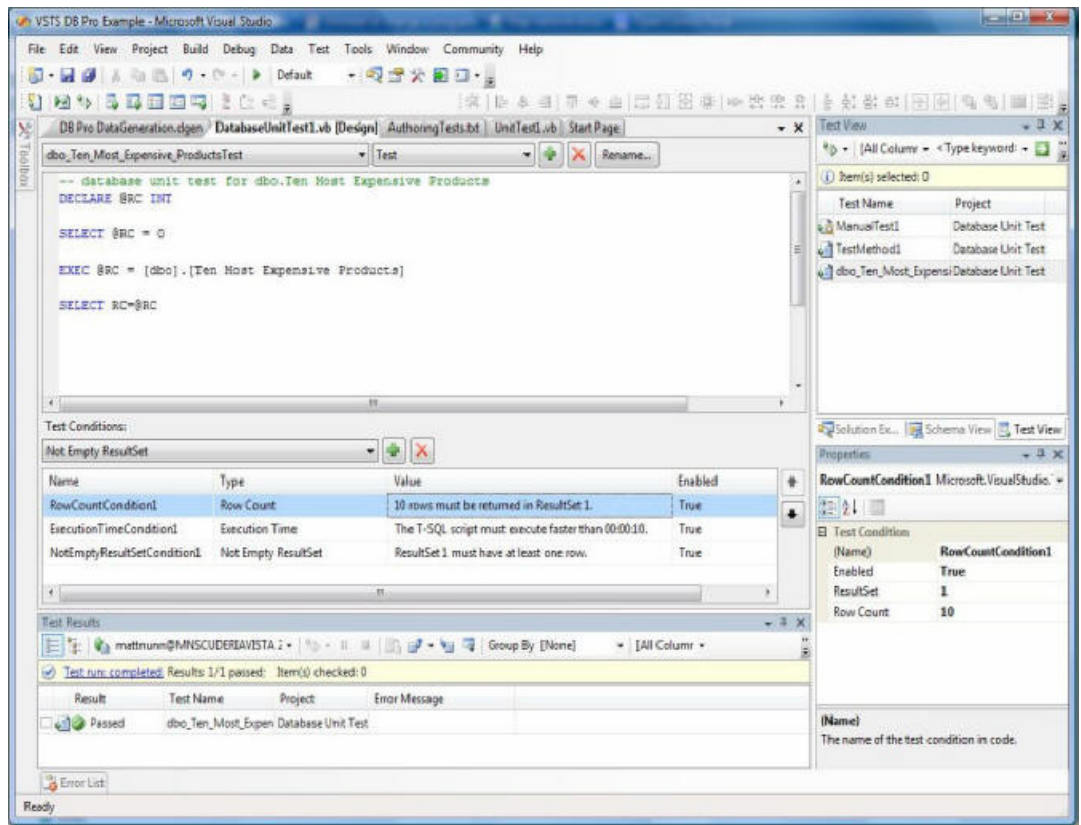


Figura 19: Interfaz gráfica del MS Visual Studio 2005 Team

Fuente: Sitio Web de la herramienta

<http://msdn2.microsoft.com/en-us/teamssystem/bb507197.aspx>

Generador de los datos para definir los sistemas de datos de prueba repetibles

El generador de datos de Microsoft Visual Studio 2005 Team Edition for Database Professionals genera una prueba de datos en su base de datos para realizar un prueba precisa. Se define un plan de generación de datos que produzca el mismo sistema de datos de prueba, de modo que se tenga la confianza que se está corriendo todo el tiempo la misma prueba.

- Generación de datos consistentes para la prueba.
- Datos del mundo real aun así protegidos de problemas legales y de privacidad.

8. CONCLUSIONES

- Los proyectos de migración de datos en principio son todos iguales, todos siguen un mismo objetivo, mover datos de un lugar de almacenamiento a otro, pero todos son diferentes cuando se observan los detalles que lo conforman, no sólo desde el punto de vista tecnológico, sino también el enfoque que cada equipo de trabajo le da a la planeación, implementación y realización de la migración, cada uno tiene sus propios riesgos y necesidades.
- Los proyectos de migración de datos son complejos, no sólo desde el punto de vista técnico, sino también desde el punto de vista del negocio. Pero también es común que las directivas subestimen los riesgos, costos y duración de este tipo de proyectos. Pero por todo lo visto anteriormente, el impacto en el negocio de una migración fallida puede ser grande y desastroso.
- Según International Data Group (IDC), el mercado que ellos han denominado Data Distribution, Migration and Repurposing (DDMR), ha crecido en un 11.3% en el 2001 para llegar a un total de \$1.03 billones de dólares. Para el 2006, investigadores del IDC esperan que el mercado alcance aproximadamente \$2.2 billones de dólares, impulsado por demandas de minería de datos, inteligencia de negocios, recuperación de desastres, reemplazo de sistemas obsoletos e implementación de nuevos sistemas transaccionales. Estas tendencias de crecimiento se están presentando en todo el mundo; en el entorno de sistemas en Colombia, se están presentando cada día más iniciativas tecnológicas que requieren algún tipo de migración de datos.
- El desarrollo de pruebas para la migración y validación de datos, aun está en etapa de desarrollo y crecimiento, las propuestas para desarrollar pruebas de bases de datos a través de pruebas de regresión es una aproximación a nuevas técnicas que no dependen tanto del ojo humano, sino que se vuelven

más programáticas y los desarrolladores tienen mayor visibilidad y hay mayor flexibilidad en lo que se está intentando probar.

- Las herramientas para realizar pruebas para la migración y validación de datos, aún se encuentran un paso detrás de las demás herramientas de prueba, debido en su mayoría al hecho de que las pruebas mismas están en etapa de desarrollo. Estas pruebas están siendo implementadas sobre la base de datos por medio de lenguajes de programación.
- Realizar un buen diseño de Pruebas de Migración y Validación de Bases de Datos es de vital importancia en las compañías, ya que da la posibilidad de obtener datos con mayor calidad, que serán de gran apoyo en la toma de decisiones, y en la creación de estrategias en las compañías.
- El estado del arte tanto de las tecnologías como las metodologías en cualquier lugar del mundo, en la mayoría de las ocasiones, van a depender de las necesidades del entorno en el cual se encuentran, a medida que un entorno demande más de la tecnología, la tecnología demandará más de los desarrolladores y los ingenieros que los implementan. La misma situación aplica con las migraciones y validaciones de datos, entre más demandas tecnológicas se presenten y entre más avancen esas tecnologías que emplean datos y que por ende llevan a la realización de una migración de datos, mejores y más eficientes maneras de realizar las migraciones serán necesitadas, esas necesidades se traducen en metodologías probadas y de alta calidad, y herramientas con mejor desempeño y confiables.
- Las metodologías y tecnologías que giran en torno a las tecnologías de información y el entorno de sistemas están en constante evolución, buscando una mayor calidad en los resultados y productos que pone a disposición de sus usuarios. No es un entorno estático.
- La diferencia entre el estado del arte entre un lugar y otro depende en muchas ocasiones del foco de origen de dicha tecnología o metodología de la cual se esté haciendo referencia. En el caso particular de las metodologías,

paradigmas o enfoques que tienen puntos de origen diferentes y específicos, que también tienen periodos de gestación, desarrollo y auge. Luego de ser aceptadas en la comunidad tecnológica, el conocimiento sobre ellas comienza a expandirse y a ser adoptada en otros lugares tal cual fue creada o con modificaciones que le permitan adaptarse a las necesidades del nuevo entorno en el cual se encuentran. Todo esto se lleva a cabo en diferentes periodos de tiempo.

- Aunque existen algunas herramientas que ayudan a realizar pruebas para la migración y validación de datos, muchas empresas en nuestro medio no las utilizan, ya sea por desconocimiento de ellas o porque tienen costos elevados, llevando a hacer migraciones más manuales que no tienen un nivel de validación tan alto como el que ofrecen las herramientas actuales.

9. RECOMENDACIONES

9.1. CONSIDERACIONES Y RECOMENDACIONES PARA LA MIGRACIÓN Y VALIDACIÓN DE DATOS

- Entre más importantes sean los datos para las operaciones del negocio y entre más complejo el entorno, el proceso de planeación para la migración es más crítico. Un proceso de planeación sólido y bien estructurado puede ayudar a identificar posibles problemas, definir soluciones, estrategias de mitigación de problemas, además de identificar las necesidades técnicas y de negocio para la migración, definir que tipo de tecnología se necesita para la migración, herramientas o solo personal, entre otros aspectos que son críticos durante un proceso de migración y que pueden ser identificados durante la fase de planeación de la migración. Adicionalmente, se deberían establecer puntos de control dentro del plan de migración, para verificar en cada uno de esos puntos cómo va el proyecto.
- Durante la planeación del proceso de migración se debería considerar el orden en el cual los datos serán migrados de un sistema a otro, se tiene que considerar la integridad referencial de los datos. Es decir, se debería trasladar primero los datos de las tablas que sólo contengan claves primarias y luego las tablas que contengan claves foráneas sobre las primeras migradas.
- La limpieza de la base de datos debería ser un proceso continuo, de tal manera que llegada la hora de la migración se pueda migrar la mayor cantidad de datos correctos como sea posible. Los datos deben estar lo más limpios posibles o ser lo más homogéneos posibles si provienen de diversas fuentes. Durante el proceso de migración, los datos de igual forma podrían requerir transformaciones para adaptarse al nuevo sistema, estas deben ser diseñadas

e implementadas con mucho cuidado. En resumen, cuando se realice la migración se debería intentar separar los datos malos de los buenos y separar los procesos de limpieza de los de transformación; los primeros son realizados en la base de datos de origen o fuentes y es realizado por los usuarios del sistema, mientras que la segunda, la transformación, son realizadas al momento de la migración y generalmente lo realiza la herramienta o el programa de carga de los datos en el sistema destino.

- Si se cuenta con un plan de Calidad de Datos o se está comenzado a implementar y se emplea el proceso de migración de datos como plataforma para la misma, lo ideal es comunicar dicho plan a todos los usuarios de los sistemas, explicarles el impacto e importancia que tiene sobre la organización la calidad de los datos, decirles cuál será el aporte de ellos al plan y la importancia de la continuidad y mejoramiento del mismo. Al igual que con el proyecto de migración de datos y demás proyectos de tecnología, éste se debe vender a los usuarios para garantizar un cierto nivel de éxito.
- Uno de los primeros pasos en un proceso de migración de datos es familiarizarse con los datos que se van a migrar: quién es el dueño del dato, qué aplicaciones la utilizan, en qué base de datos se encuentra, etc.
- Al igual que con el plan de calidad, el plan completo de la migración de los datos debería ser comunicado a todos los miembros del equipo y personas claves en la organización, para recibir retroalimentación y observaciones acerca del plan si se encontraron inconsistencias, sugerencias con respecto a los tiempos, diseños, herramientas, etc. Es un trabajo en equipo y como tal la planeación debería ser de las primeras cosas con las cuales el equipo entero se sienta cómodo.

9.2. CONSIDERACIONES PARA LAS PRUEBAS PARA MIGRACIÓN Y VALIDACIÓN DE DATOS

- Se debe elaborar una aproximación metódica a las pruebas de los datos migrados. Dentro de esta metodología se debería desarrollar un protocolo para documentar y reportar problemas del proceso de migración de datos encontrados durante las pruebas.

Entre los errores que se pueden cometer a la hora de realizar las pruebas sobre las migraciones de datos, se pueden mencionar las siguientes:

- *Dejar por fuera a los usuarios del sistema final.* Hay quienes coinciden que no son los más confiables para realizar pruebas a los datos, pero ellos son los dueños de esos datos y probablemente los que mejor los conocen. Otro error que se puede cometer y que está directamente relacionado, es el elegir a los testers equivocados, el elegir a la persona equivocada del negocio para llevar a cabo pruebas sobre los datos puede no dar los resultados esperados y retrasar el descubrimiento de posibles problemas.
- *Llevar a cabo solo pruebas con los reportes y no los datos.* Es un hecho que los reportes han sido empleados como una manera de validar los datos migrados, pero el comportamiento de los datos es distinto cuando se genera un reporte que cuando es empleado en el sistema como tal; es por esta razón que adicional al uso de los reportes como uno de los medios para validar los datos, deben existir pruebas para los datos dentro del sistema.
- Otra prueba que podrían realizar los usuarios con los reportes, es una conciliación de los dos sistemas, es decir, un comparativo entre la información contenida en cada uno de los sistemas, empleando reportes para este fin sería lo más sencillo.

- Cuando se vayan a realizar pruebas sobre el proceso de migración y validación de datos, lo ideal es emplear un subconjunto de datos del sistema antiguo, si el volumen es muy grande para migrar en una prueba, y no emplear datos de prueba inventados o generados por otras herramientas con el propósito de llevar a cabo las pruebas.

Adicional a estos errores que se deben tratar de evitar durante el diseño y la realización de las pruebas a los datos, se deberían tener algunas consideraciones, como:

- Planear y programar suficiente tiempo para realizar las pruebas al proceso de migración y validación de datos, y a los datos mismos.
- Al igual que con los demás documentos del proyecto, los casos de prueba y sus resultados deben ser aprobados por los usuarios dueños del dato, y se recomienda tener implementado un sistema de aprobación y firma de documentos del proyecto, y su posterior almacenamiento.

Cuando se llega a las pruebas de los datos y la base de datos como tal, en el entorno de sistemas actual se pueden ver dos frentes, el de los tradicionalistas y el de la comunidad ágil. Los primeros, los tradicionalistas de la administración de datos, creen que la mejor forma de probar los datos es solamente con la validación de los usuarios. Mientras que en tiempos más recientes la comunidad ágil ha propuesto nuevas opciones y metodologías para las pruebas. Para la comunidad ágil las pruebas a las bases de datos implican validar el contenido, el esquema y la funcionalidad dentro de la base de datos. Desde el punto de vista de

una base de datos relacional esto incluye los datos, la estructura de las tablas, relaciones entre tablas, procedimientos almacenados, funciones, etc.

Pruebas de interfaz de bases de datos validan la base de datos como una caja negra, mientras que pruebas internas de bases de datos la validan como una caja blanca; si se realizan pruebas sobre las bases de datos como tal son más que todo al nivel de la interfaz.

La comunidad ágil no desacredita ni desmerita los métodos actuales de pruebas de las bases de datos y los datos en su efecto, pero proponen nuevas maneras de probar los datos contenidos en la base de datos de manera programática, que pueden ser empleadas en conjunto con las formas tradicionales de pruebas; ellos son concientes de la importancia de los datos dentro de las organizaciones y con estas pruebas tratan de ayudar en el mejoramiento de la calidad de los datos. Entre las pruebas que recomiendan, se encuentran:

- *Verificación de las reglas de dominio de las columnas:* Ciertas columnas sólo deberían permitir ciertos valores.
- *Verificación de las reglas de valores por defecto de las columnas:* Si no se suministra un valor se coloca uno por defecto.
- *Verificación de las reglas de existencia de valores:* Las columnas que no pueden tener valores nulos.
- *Verificación de las reglas de los valores de una fila:* El valor de un dato en una fila puede depender de otro valor en la misma fila.
- *Verificación de las reglas de tamaño.*
- *Pruebas para validar las relaciones entre filas en diferentes tablas:* Es una validación de integridad referencial, por ejemplo, si una fila en una tabla referencia otra fila de otra tabla, esa segunda fila en la segunda tabla debe existir.

Mencionan estas pruebas como un ejemplo, pero creen que las **pruebas de regresión sobre las bases de datos** pueden ayudar a obtener una mayor calidad en los datos y son pruebas que se pueden realizar continuamente sobre la base de datos y los datos, no sólo durante el proceso de migración y validación de datos.

9.3. FACTORES DE ÉXITO PARA UN PROYECTO DE MIGRACIÓN Y VALIDACIÓN DE DATOS

A continuación serán listados algunos factores que se podrían considerar como factores de éxito, que han sido identificados por expertos en el área y que también han sido identificados en el transcurso del desarrollo de este proyecto de grado.

- ✓ Tener el apoyo de los altos directivos.
- ✓ Contar con una metodología sólida, que haya sido probada y ha tenido proyectos que se han completado exitosamente.
- ✓ Liderazgo técnico sólido de alguien que haya completado exitosamente un proyecto similar.
- ✓ Iniciar el proceso de planeación del proyecto de migración tan pronto se decida que se realizará el proyecto. Sea cual sea la metodología empleada, cada fase requiere tiempo y su estimación no se debe realizar a la ligera.
- ✓ Tener los recursos adecuados, tanto técnicos como del negocio. De la misma forma que se necesita un líder con experiencia y/o conocimientos en esta área, la elección de los demás miembros también es de igual importancia. El equipo debe poder integrarse adecuadamente.
- ✓ Garantizar cooperación. Es necesario contar con la cooperación de miembros de la organización por fuera del equipo de trabajo a la hora de realizar las pruebas y validaciones del proceso de migración, y en otros aspectos del

proyecto. Si el proyecto es parte de otro proyecto más grande, deben existir líneas claras de comunicación y un entendimiento conjunto de los requerimientos y necesidades.

- ✓ El saber administrar los riesgos y los cambios, han sido factores de éxito y/o fracaso en todo tipo de proyectos, y un proyecto de migración de datos no es la excepción. El saber manejar y controlar estos aspectos puede definir si un proyecto se extiende más del tiempo estimado o si sus costos se sobrepasan.
- ✓ Tener un plan de pruebas bien estructurado, en el cual se realizan pruebas sobre la base de datos como tal y se le permite a los usuario validar los datos en el sistema, para que puedan ir adquiriendo confianza en el mismo.
- ✓ Otro aspecto importante es el conocer las opciones disponibles para realizar la migración, en términos de metodologías y herramientas disponibles, las decisiones se deben tomar basados en las necesidades de la organización, la importancia de los datos, etc. También es necesario conocer las opciones dentro de la organización que tendrá para realizar la migración, como el tiempo de inactividad disponible, la cantidad de datos a migrar, recursos disponibles, etc.
- ✓ Planear y definir el cronograma para garantizar las mayores probabilidades de lograr una culminación exitosa, no permitir que presiones para adelantar las fechas influyan en la planeación. Debe quedar claro el por qué de esos tiempos y las posibles consecuencias de acelerar el proceso. En muchas ocasiones esos adelantos de fechas implican quitar pasos o procesos dentro del plan, y en la mayoría de los casos el paso que se reduce es el de las pruebas, dejando un mayor margen para errores no detectados durante el periodo de pruebas.
- ✓ El tener medidas comprensivas de la calidad de los datos, la validación y la limpieza de datos es una parte integral del proceso de migración. La forma más clara de obtener esto es contar con un plan de calidad de datos.

Estos solo son algunos aspectos que pueden llevar a un proyecto de migración de datos a una culminación exitosa, pero no siempre es igual para todos.

10. TRABAJOS FUTUROS

El estudio sobre las pruebas para bases de datos en general, y en particular las pruebas para las migraciones y validaciones de datos, es un campo que apenas se está comenzando a desarrollar. Dado el estado en el que se encuentra el mercado actual y el camino que están siguiendo las implementaciones de sistemas (sean tradicionales o para inteligencia de negocios) muestran que se estarán llevando a cabo más migraciones de datos y a gran escala en muchas ocasiones.

Con lo encontrado durante esta investigación se podrían presentar varios trabajos futuros con diversos enfoques, entre los cuales podemos listar los siguientes:

- ✓ Una metodología para pruebas para la migración y validación de datos que no sea particular para una plataforma o proveedor en específico, siendo lo suficientemente general y completo para que se pueda adaptar a cualquier tipo de proceso de migración y validación de datos.
- ✓ Un marco de trabajo completo para el desarrollo de planes de pruebas para bases de datos, integrando las diversas pruebas para bases de datos, desde seguridad hasta el mantenimiento de las mismas.

BIBLIOGRAFÍA

- ➔ **ALAN, ROBERT:** "The Serials Data Migration Dilemma". Technical Services Quarterly 2002.
- ➔ **BECKER, HAL B.:** "Information integrity: a structure for its definition and management." New York: McGraw-Hill, 1983.
- ➔ **CELKO, JOE:** "Joe Celko's Data and Database: Concepts in Practice". Morgan Kaufmann Publishers, 1999.
- ➔ **ECKERSON, WAYNE W.,** "Data Quality and the Bottom Line" Report of the Data Warehousing Institute, January 2002.
- ➔ **ELMASRI, RAMEZ & NAVATHE, SHAMKANT B.:** "Fundamentals of Database Systems". Second Edition. Canadá 1994.
- ➔ **HUDICKA, JOE:** "Why Data Migration Projects Fail".
- ➔ **HURLEY, NANCY:** "Data Migration: Everyone is doing it – but are they doing it right?" Enterprise Strategy Group. Marzo 2005.
- ➔ **LEWIS, WILLIAM E.:** "Software Testing and Continuous Quality Improvement". CRC Press LLC. 2000.
- ➔ **LUQUE RUIZ, IRENE; GÓMEZ, MIGUEL ÁNGEL; LÓPEZ ESPINOSA, ENRIQUE; CERRUELA GARCÍA, GONZALO:** "Bases de datos desde Chen

hasta Codd con ORACLE". Alfaomega Grupo Editor, Primera Edición. México 2002.

- ➔ **MORIARTY, T., HELLWEGE, S.** "Data migration. Database *Programming & Design*". 1998.
- ➔ **MURRAY, CHUCK:** "Oracle Database SQL Developer User's Guide, Release 1.2". Oracle Corporation. 2007
- ➔ **OLSON, JACK E.:** "Data Quality. The Accuracy Dimension". Morgan Kaufmann Publishers - San Francisco, CA. 2003.
- ➔ **PICK, GERSHON:** "Data Migration. Concepts & Challenges". AYM Gael, 2001
- ➔ **SOMMERVILLE, I.:** "Software Engineering" - Addison Wesley, 5ta Edición. 1999.
- ➔ **TOIGO, JON WILLIAM.** "Data Migration Headaches Underscored by Softek Survey". 2005
- ➔ "Matching Data Migration to Availability Requirements" – Softek Technical White Paper. 2005.
- ➔ "The Hidden Cost of Data Migration" – Softek Press Release. 2006
- ➔ "2006 Best Practices for Data Migration" – Softek Press Release, 2006
- ➔ "10 Mistakes to Avoid When Testing a Data Warehouse" por Mark Robinson. DM Direct Newsletter, Diciembre 2004.

- ➔ "Meeting the Data Migration Challenge". Storage Networking Solutions (SNS) Europe, Mayo 2005.
- ➔ "Success Drivers for State-of-the-Art Data Migration" por Arvind Parthasarathi, Senior Director of Solutions at Informatica. Febrero 2006.
- ➔ "Avoiding Data-Migration Pitfalls" por Jim Utsler. IBM systems Magazine, Julio/Agosto 2006.
- ➔ "Data Migration: Plan to Succeed" por Ashish Nadkarni. DM Direct, Enero 2007.
- ➔ "What to consider when testing databases" por Scott W. Ambler. Dr. Dobbs Journal, Junio 2007.
- ➔ Secretaria del Consejo Superior de Administración Electrónica. Ministerio de Administraciones Públicas de España. METRICA Versión 3
<http://www.csi.map.es/csi/metrica3/>
- ➔ <http://www.oracle.com/technology/tech/migration/index.html>,
http://www.oracle.com/technology/tech/migration//workbench/index_sqldev_omwb.html
- ➔ www.agiledata.org
- ➔ <http://www.swissql.com/products/datamigration/data-migration.html>

- <http://www.dulcian.com/papers/IOUG/1999/DataMigrationMethodologyOverview.htm>
- http://www.lti.unico.edu.sv/modules.php?name=Downloads&d_op=getit&lid=10
- <http://www3.uji.es/~mmarques/f47/apun/node6.html>
- <http://dev.mysql.com/doc/migration-toolkit/en/mysql-migration-toolkit-overview.html>
- http://recursostic.javeriana.edu.co/wiki/index.php/Historia_de_las_bases_de_datos_en_Ciencia_de_la_Informaci%C3%B3n#Acerca_de
- http://es.wikipedia.org/wiki/Base_de_datos#Tipos_de_bases_de_datos
- http://es.wikipedia.org/wiki/Bases_de_datos
- http://www.quadernsdigitals.net/datos_web/hemeroteca/r_1/nr_502/a_6850/6850.htm
- <http://www.colorconcepts.com/glossary/d.htm>
- <http://uis.georgetown.edu/departments/eets/dw/GLOSSARY0816.html>
- <http://www.olapreport.com/glossary.htm>

ANEXOS

ANEXO 1. PREGUNTAS DE ENTREVISTAS

Preguntas sobre las Pruebas para la Migración y Validación de Datos

En el entorno tecnológico actual, el refrescamiento e implementación de nuevas tecnologías se están convirtiendo en actividades del día a día de las empresas y el personal de TI, y la mayoría de las empresas se están involucrando en estos proyectos de implementación para mantenerse al día en el entorno de su negocio e intentar obtener una ventaja competitiva, ayudado por la tecnología, con respecto a la competencia.

Todas estas nuevas tecnologías emplean y necesitan datos para su funcionamiento diario; estos refrescamientos e implementaciones tecnológicas están haciendo que los procesos de migración de datos sean más comunes. Las metodologías o prácticas empleadas para llevar a cabo las migraciones de datos pueden variar desde la tecnología implementada, hasta la empresa o la persona que está realizando la migración. Las metodologías y procedimientos empleados en su mayoría son propietarias y la bibliografía disponible es mínima.

Las pruebas de software siempre han sido una temática de gran importancia, desde los comienzos del desarrollo de software y se ha extendido, se puede decir que, a todas las demás áreas involucradas con tecnología. Los planes de prueba son necesitados para ayudar a obtener la mejor calidad, y son necesitados para eliminar o reducir los ciclos de prueba y error de los desarrollos o implementaciones.

El siguiente cuestionario está dirigido a personas que trabajan en el área de tecnología y que han o están involucradas en procesos de migración de datos. El cuestionario forma parte de un proyecto de grado que busca identificar las posibles mejores prácticas o tendencias en el entorno tecnológico para así poder establecer el Estado del Arte de las Pruebas para la Migración y Validación de Datos.

Nombre:	
Organización:	
Fecha:	

Marque con una X todas las respuestas que apliquen y complete como sea necesario.

1. ¿Ha realizado migraciones?

Sí	
No	

2. ¿Que tipo de migración de de datos a realizado?

Para la actualización y/o cambio de software/hardware	
Migración/Implementación de nuevos sistemas.	
Unificación de sistemas.	
Procedimientos de ETL para la implementación de proyectos de BI.	
¿Otros?	

Si realizó otro tipo de migraciones, por favor especifíquelos:

3. ¿Valida los datos migrados?

Sí	
No	

4. ¿Qué tipo de Validaciones son realizados?

5. ¿Ha tenido problemas durante/después de la migración de datos?

Sí	
No	

6. ¿Qué tipo de problemas?

Tipos de datos inconsistentes.	
Longitud de los datos.	
Datos incompletos.	
Datos en los campos incorrectos	
¿Otros?	

Si han ocurridos otros tipos de problemas, por favor lístelos:

7. ¿Tiene una metodología específica para la Migración y Validación de datos?

Sí	
No	

¿La metodología es propia?

Sí	
No	

Si tiene una metodología, por favor explique sobre que se trata y como es aplicada:

8. ¿Durante el proceso/metodología de Migración y Validación de Datos incluye o realiza pruebas (crea planes de prueba para la migración y validación de datos)?

Sí	
No	

Si realiza pruebas, por favor explique los tipos de pruebas que realiza y las razones por las cuales ese tipo de pruebas en particular son realizados:

--

9. ¿Hay un plan de Calidad de Datos para verificar los datos migrados?

Sí	
No	

10. ¿Qué tipo de herramientas son utilizadas para los procesos de Migración y Validación de Datos? Por favor lístelos. (Herramientas es cualquier software y/o lenguaje de programación utilizado para facilitar la migración de datos).

--

11. ¿Las herramientas realizan o permiten realizar pruebas?

Sí	
No	

Si la respuesta es sí, por favor explique como y que tipos de pruebas:

--

12. Las migraciones que realiza son:

Online (los sistemas están disponibles para los usuarios)	
Offline (los sistemas no están disponibles para los usuarios, se consideran en tiempo de inactividad de las aplicaciones)	

Nota del Autor: La información compartida en este documento es con propósito académico y de investigación. Si hay información adicional que puede ser compartida, por favor adjúntelo a este documento.

ANEXO 2. COMPARATIVO DEL ESTADO DEL ARTE DE LAS PRUEBAS PARA LA MIGRACIÓN Y VALIDACIÓN DE DATOS CON RESPECTO A ESTANDARES INTERNACIONALES Y TMM

Si se quiere realizar algún tipo de comparativo del estado de arte de las pruebas para la migración y validación de datos con estándares internacionales como ISO e IEEE, no se encuentran similitudes definidas, ambos casos son muy específicos en el tipo de pruebas para los cuales fueron creados, en este caso pruebas para software y la calidad de las mismas. Pero por el lado de ISO, si se observan los estándares orientados a las definiciones genéricas de pruebas y términos empleados en el área de la tecnología, se pueden observar analogías entre los términos empleados en los distintos tipos de pruebas sea de software o de bases de datos. Puede que en el momento de la creación de dichos estándares no se hayan considerado las pruebas para bases de datos, porque había más interés en el momento en los procesos de estandarización y calidad del desarrollo de software y las pruebas para bases de datos eran una actividad adicional en la mayoría de casos de desarrollo y no eran el enfoque.

Pero en la actualidad siguiendo la línea de desarrollo de frameworks y modelos de madurez, se creó un framework llamado **TMM (Test Maturity Model)**, el cual fue desarrollado por el Instituto de Tecnología de Illinois, entre principios y mediados de 1990. Está ampliamente alineado con CMM (Capability Maturity Model for Software), y su sucesor TMMi estará alineado con CMMi. Cada nivel de madurez de TMM está alineado con uno de CMM o depende de alguno de los niveles previos de CMM.

Sostienen que entre 30% - 40% del costo total de un proyecto está en las pruebas, se le presta poca atención en modelos tales como CMM y CMMi, los cuales se enfocan en los procesos de mejoramiento del desarrollo de software.

TMM también emplea los conceptos de niveles de madurez para los procesos de evaluación y mejoramiento. Identifica áreas de procesos, objetivos de madurez y mejores practicas. Consta de 5 niveles de madurez, que son:

1. **Nivel 1: Inicial.** No hay área de procesos ni objetivos de madurez definidos en este nivel.

Las pruebas son un proceso caótico, son desarrolladas de manera ad hoc luego de que se realice el código, el objetivos de las pruebas es mostrar que el software funciona.

2. **Nivel 2: Definición.**

Áreas de Procesos: Objetivos y Políticas de Pruebas, Planeación de Pruebas, Técnicas y Métodos de Pruebas y Entornos de Pruebas.

En este nivel las pruebas tienen un proceso definido y esta separado del proceso de debugging. En el contexto de la estructuración del proceso de pruebas, se establecen planes de prueba que contienen una estrategia. Se emplean técnicas formales de diseño de pruebas para crear y seleccionar casos de pruebas de los requerimientos. Sin, embargo las pruebas aun se realizan a veces muy tarde en el proceso de la implementación.

3. **Nivel 3: Integración.**

Áreas de Procesos: Organización de las Pruebas, Programa de Entrenamiento para Pruebas, Integrar Pruebas al Ciclo de Vida del Software, y Control y Monitoreo.

En este nivel las pruebas están completamente integradas al ciclo de vida del software. Se reconocen todos los niveles de modelo-V. La planificación de las

pruebas se realiza en etapas muy tempranas del proyecto como un plan de pruebas maestro. La estrategia de pruebas es determinada empleando técnicas de administración de riesgos y se basa en requerimientos documentados. Hay programas de entrenamiento para pruebas y el testing se ve como una profesión.

4. Nivel 4: Medidas y Administración.

Área de Procesos: Revisión de pares, Medición de las pruebas y Evaluación de Calidad del Software.

Las pruebas son un proceso bien definido y medible. Revisiones e inspecciones son realizadas a través del ciclo de vida y son considerados para del proceso de pruebas. Los productos son evaluados empleando criterios de calidad. Los casos de prueba son reunidos, almacenados y administrados en una base de datos central para ser reutilizados y para pruebas de regresión.

5. Nivel 5: Optimización.

Área de Procesos: Prevención de Defectos, Control de Calidad y Optimización de Procesos de Pruebas.

Basados en los resultados que han sido obtenidos al cumplir todos los objetivos de mejoramientos de los niveles anteriores, las pruebas ahora son un proceso bien definido y es capaz de controlar los costos y la efectividad de las pruebas. En este nivel los métodos y técnicas son optimizados y hay un enfoque continuo en el mejoramiento del proceso de pruebas.

Este framework, es aplicable para cualquier tipo de pruebas, sea de software o de bases de datos.

Al mirar las características del modelo de madurez de pruebas anteriormente mencionado y compararlos con el estado actual de las pruebas para migración y validación de datos en Medellín y a nivel internacional; se puede observar que en Medellín, en su gran mayoría se encuentra en el **Nivel 1**, es decir, no hay procesos definidos, es caótico y se confunden las pruebas con debugging y solo una minoría se encuentra en el **Nivel 2** porque están comenzando a definir procesos de pruebas separados del debugging, aunque aun no emplean técnicas formales de diseño de pruebas. Mientras que a nivel Internacional se observa que están estables en el **Nivel 2**, dado que todos tienen una metodología definida y no se confunde el proceso de pruebas con el proceso de debugging, aunque también se observa que algunos están ubicándose en el **Nivel 3** porque están integrando el proceso de pruebas con el ciclo de vida del software o de la implementación, y se comienzan a presentar los planes de pruebas maestros y planes basados en los requerimientos documentados.

La siguiente figura es una muestra grafica de la diferencia en niveles entre Medellín y el entorno Internacional. Si se observa, están a un nivel de distancia. Esto se puede deber en parte a que algunas de las metodologías empleadas en la ciudad son de terceros, que en muchos casos son internacionales, y que no son conocidos en el País, hasta que no se hayan establecido y probado donde fueron desarrollados, para poder luego darse a conocer en otros mercados, como el Colombiano.

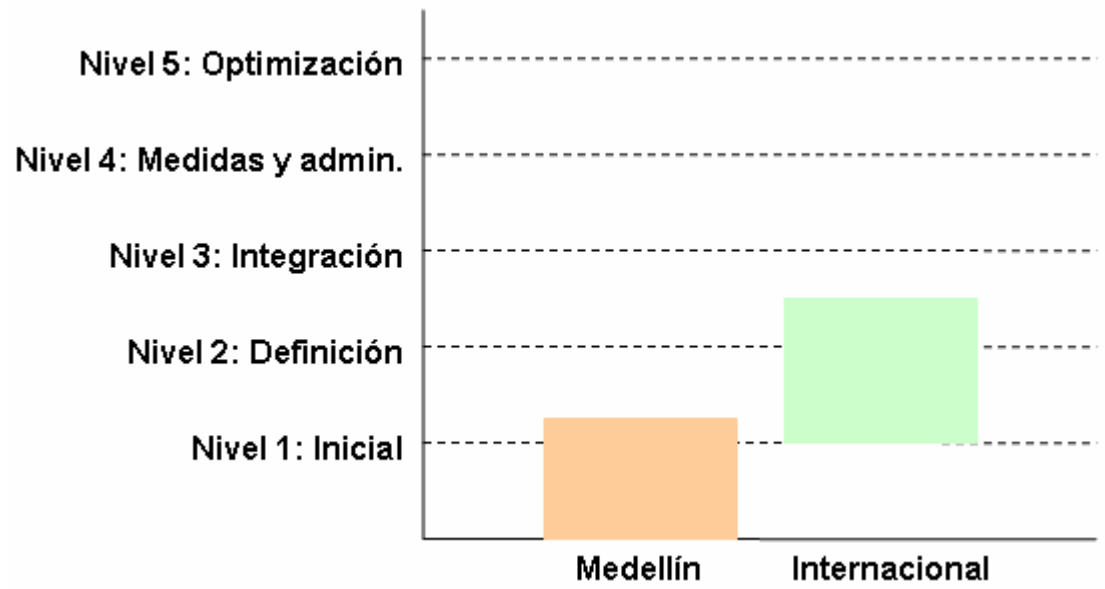


Figura: Estado del Arte de Pruebas para la Migración y Validación de Datos vs. TMM

Fuente: Elaboración Propia