Recomendaciones Migración de un Data Warehouse a SAP HANA

Sofia Prieto

S. Prieto

Universidad Iberoamericana, Prolongación Paseo de la Reforma 880, Alvaro Obregon, Lomas De santa Fe, 01219 Ciudad de México, D.F., México

M.Ramos, P. Solares.(eds.) Ciencias de la Tecnología de la Información -©ECORFAN, México D.F., 2015.

Abstract

During the last months of 2014 in a tripartite institution have been performed a migration of a Data Warehouse to SAP Hana, the recommendations identified so far are shared through this article, note that only has been migrated a pillar of business lending and without considering the loading data process, simply leaving the previous Data Warehouse transformed data as a collector, concentrator from different sources.

10 Análisis

La institución tripartita cuenta con un Data Warehouse, donde los datos operativos se estructuran específicamente para facilitar Algunos consideran SAP consultas para análisis y con un buen desempeño(Kimball R, 2002), HANA sólo otra base de se encuentra en funcionamiento con la tecnología Microsoft SQL datos, mientras que otros 2008R2 e integration services como software de apoyo para el consideran que es una proceso de carga de datos.

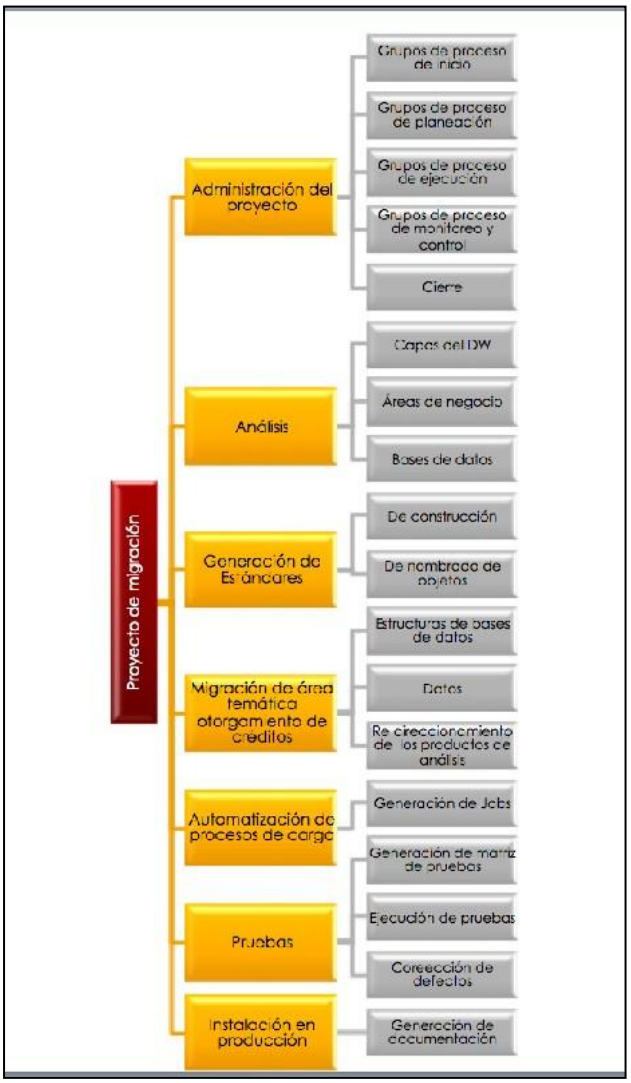
El Data Warehouse es proveído de datos de distintas fuentes de los diferentes pilares de negocio de la institución por medio de un conjunto de procesos en el que los datos de origen operativo son preparados para almacenarse llamado ETL por sus siglas en ingles (Kimball R, 2002) y gestionados por distintas capas hasta llegar a un modelo adecuado para su visualización y análisis La tecnología actual es funcional sin embargo se identifica la necesidad de procesamientos mas rápidos así como una comunicación unificada entre sistemas dado que actualmente la institución cuenta con varias marcas de software lo que hace complicada la comunicación entre ellos.

10.1 Planeación del Proyecto de Migración

La planeación de este proyecto se generó conforme a la mejor práctica de administración de proyectos del PMI®, actualmente no existe una migración del tipo documentada, sólo se encuentra para una migración de SAP BW (solución de tipo Data Warehouse de SAP SE) sin embargo el alcance y los tipos de actividades difieren totalmente del propósito de este proyecto.

La estructura de desglose de trabajo EDT o Work Breakdown Structure WBS por sus siglas en ingles, consiste en el proceso de subdividir los entregables y el trabajo del proyecto en componentes más pequeños y más fáciles de manejar (PMI, 2013). Se puede visualizar en la Figura 1 la realizada para el proyecto.

Figura 1 WBS, fuente: Documentación proyecto de Migraciónel trabajo del proyecto



Antes de comenzar cualquier actividad de desarrollo en SAP HANA es importante considerar una etapa para documentar cuales serán los estándares de construcción de soluciones y objetos a utilizar asi come el de nombrado de objetos conforme a las mejores prácticas de la tecnología y de la organización.

10.2 Estándares de Construcción

Para determinar el est``andar para la generación de soluciones en SAP HANA, se reconstruyó un cubo existente en la nueva tecnología SAP HANA y SAP Data Services (herramienta especializada en ETL), para la ejecución de algunas de las pruebas.

En una clasificación general de desarrollo de soluciones existen dos tipos, la que ejecuta los procesos al vuelo es decir, cada que un usuario entra al reporte o cubo SAP Hana ejecuta todos los objetos de tipo information view, y la que puede ejecutar procesos cuyo resultado sea depositado en una tabla y que proporcionne la información ya procesada a los objetos de tipo infortion views. Las dos opciones son utilizables pero tienen diferente uso y ventajas una sobre la otra, mas adelante explicará esto.

Del estudio realizado por el equipo de proyecto en el que se diseñan seis variables: mediciones de tiempo de ejecución, número de objetos creados en cada escenario, se calificó en base a la experiencia si sería sencillo en mantenimiento si seria reutilizable la trazabilidad del dato y la complejidad del desarrollo. Para obtener el tiempo de ejecución se midió el tiempo en dos momentos del día y realizar un el promedio para la obtención de un dato mas exacto por variables que pudieran. Influir en la mediciónpor ejemplo la concurrencia de la red en la Institución. Los resultados se pueden visualizar en la Tabla 1.

Se desarrollo cuatro tipos de construcciones diferentes, Idenditicadas como Prueba 1, Prueba 2, Prueba 3 y Prueba 4.

10.3 Information Views

Information View son una agrupadción de los tres tipos de objetos en SAP HANA.

Los attribute views que describe una dimensión en el modelado tradicional multimensional, los analytic views cantidades y la relación entre sus atributos y las calculations views un tipo de vista de informaciñon reutilizable, cuyo principal objetivo es caracterizar una transacción (Hickman H, Wells L, 2014)

Tabla 1 Resultado de mediciones de los 4 tipos de construcciones

		Pruebas generadas	Tiempo total en nuevas tecnologías	Tiempo de Ejecución en Microsoft	Número de objetos	El mantenimiento es Fácil?	Es Reutilizable	¿La trazabilidad es Facil?	Complejidad de la construcción	Comentario	Observaciones
Prueba 1		Ejecución de las Calculation View de Hana por Data services (CV Semanal y Diaria)	10:50	30	6	Intermedio	Si	Nio	No	No si es reutilizable, y Si por la cantidad de campos	Alternativa cuando se manda la ejecución del calculo a Hana
Prueba 2	Ver. A)	Ejecución por Data services de un Procedure de Hana (CV Semanal y Diaria)	2:42	30	6	Fácil	No	Si	No	Fácil ya que se invoca un SP	Alternativa para la ejecución de un SP en Hana
	Ver. B)	Ejecución por Data services de un Procedure (8 CV que generan las CV Semanal y Diaria)	1:21	30	12	Fácil	No	Si	No	Fácil ya que se invoca un SP	Alternativa para la ejecución de un SP en Hana, es mas rápida la ejecución de 8 CV's que 2 que unifiquen a las 8
Prueba 3		Ejecución por Data services emulando el proceso de 8 scripts de Integration Services (CRELL002)	14:33	30	12	Difícl	No	No	Si	En su trazabilidad, hay que abir tanto Hana como DS	Alternativa para emular el funcionamiento de un DTS que hoy en día es CRELL002
Prueba 4	Ver. A)	Ejecución por Hana, ejecución de las CV Semanal y Diario (Ejecución en memoria).	2:49	30	28	Fácil	Si	Si	Si	Se construyen muchos objetos	Alternativa para la generación del cubo al vuelo con las CV's Semanal y Diaria
	Ver. B)	Ejecución por Hana, ejecución de las 8 CV que sustituye a las CV Semanal y Diario (Ejecución en memoria).	1:00	30	34	Intermedio	Si	No	No	Trazabilidad complicada por el rastreo en 8 CV's+	Alternativa mas rápido que la anterior debido a las 8 CV's que son mas rápidas
	Ver. C)	Ejecución por Hana, ejecución de las 8 CV se hacen JOINS a las tablas Dimensiones.	0:35	30	34	Fácil	Si	Si	No	Trazabilidad sencilla, ya que se utilizan directamente tablas	Alternativa en la que se utilizan las 8 CV's, una unión de ellas y la estrella se hace directamente con las tablas del modelo

De los resultados obtenidos se observa que las ejecuciones “al vuelo” Prueba 4 son mas rápidas sin embargo el minuto o los treinta segundos que tarda el proceso es el tiempo que el usuario final tiene que esperar para visualizar el resultado dado que se ejecutan todos los pasos necesarios para construir el reporte o análisis requerido en el cubo en el momento. Este tipo de solución es valida cuando la solución pretendida sea de tipo real-time es notorio que se requiere un tipo de solución de ejecución “al vuelo” para obtener los datos mas frescos generados por la operación de negocio.

Para llegar a este tipo de soluciones dependerá mucho de la necesidad del negocio y la estrategia de la solución de Business Intelligence aunado con la arquitectura de los sistemas transaccionales y del Data Warehouse.

Los escenarios Prueba 1, Prueba 2 y Prueba 3 son menos rápidas pero definitivamente más rápidas que la tecnología Microsoft utilizada cuyo tiempo de ejecución es de 30 minutos en un ambiente productivo. El escenario Prueba 3 fue construido con la misma lógica que el proceso de Microsoft simplemente se cambio Integration Services por SAP Data Services y Microsoft SQL2008R2 por SAP HANA obteniendo prácticamente la mitad del tiempo de ejecución.

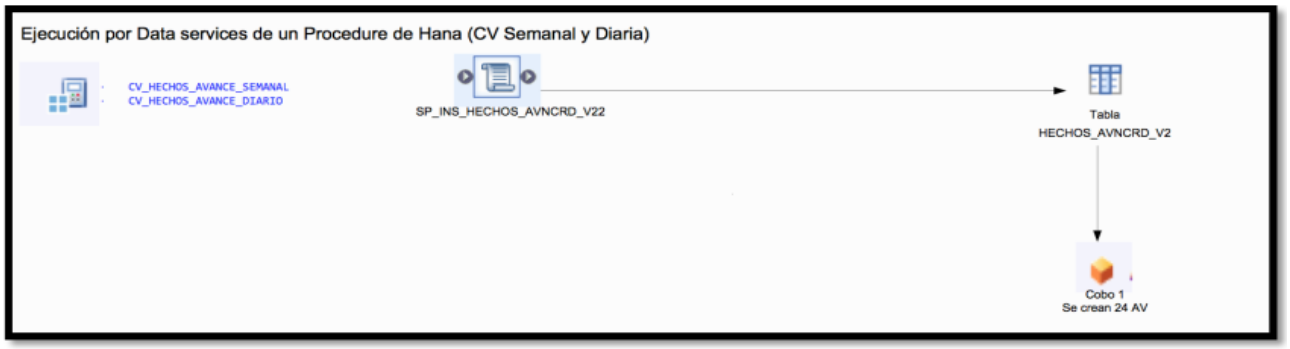
Analizando los diferentes resultados fue elegido por el equipo de proyecto el tipo de solución Prueba 2 Ver.A) por los valores obtenidos en las variables que podemos ver en la Tabla 2

Tabla 2 Variables Prueba 2

Variables	Resultado
Tiempo de ejecución	2 minutos con 42 segundos
Número de objetos creados	6
¿Fácil mantenimiento?	Si, porque solo se realizan 2 *Calculation Views*
¿Reutilizable?	No
Trazabilidad del Dato	Fácil

La recomendación en el caso de que el requerimiento sea un reporte o cubo de formato predefinido, cuyo corte de datos corresponda a una carga batch al Data Warehouse la construcción de la solución debe estar conformada por un prcedimiento alcmacenado que deposite el resultado en una tabla que sea consultada por Attribute views y una Analytic view se (Prueba 2) ilustrada en la Figura 2.

Figura 2 Objetos utilizados en la Prueba 2



10.4 Del estándar de nombrado de objetos

El estandar aquí presentado es el que se generó para la migración del proyecto considerando las mejores prácticas de la tecnología y el estandar de la institución.

Tablas y campos

Se generó la siguiente nomenclatura:

1. Si la palabra es de 4 caracteres, se deja tal cual

2. En caso de ser palabra única, se deja tal cual

3. Si la palabra es mayor a 4 caracteres, considerar lo siguiente:

a. Tomar la primera letra.

b. La primera consonante después de la primera letra, en caso de no existir, la vocal siguiente a la primera letra.

c. La segunda consonante después de la primera letra, en caso de no existir, la vocal siguiente a la primera consonante.

d. Tomar la última letra.

4. Cada palabra debe ser separada por un guión bajo.

5. El nombre siempre será en MAYÚSCULAS.

En la Tabla 3 siguiente tabla se puede ver un ejemplo

Tabla 3

NOMBRE DEL CAMPO	OBSERVACIONES
ID_TIPO_MRCO_AVLO	TIPO=TIPO MRCO=MERCADO AVLO=AVALUO
CV_RCLN_PRDO	RCLN=RECLASIFICACION PRDO=PRODUCTO
ID_UNDD_VLCN	UNDD=UNIDAD VLCN=VALUACION
CV_PAQUETE	EXCEPCION DE NOMBRE DE CAMPO

10.5 Objetos propietarios en la tecnología SAP HANA

Esquemas

1. Considerar la nomeclatura del nombrado para tablas y campo

2. Máximo de 3 letras por área

3. Incorporar el sufijo SA, DW o LL dependiendo la capa correspondiente

4. No incorpora guiones bajos en ningún momento

En la Tabla 4 se puede ver un ejemplo

Tabla 4

CAMPO ORIGEN	CAPA	NOMBRE DEL CAMPO
CREDITO	DW	CRDDW
CARTERA	DW	CRTDW
SUSTENTABILIDAD	DW	SSTDW
ANEC	DW	ANCDW

Ejemplo de nombrado de tablas y campos, fuente: , fuente: Documentación proyecto de Migración

10.6 Infomation views

Attribute views

1. Considerar la nomeclatura del nombrado para tablas y campos
2. Incorporar un Sufico AT_

Analytical Views

1. Considerar la nomeclatura del nombrado para tablas y campos
2. Incorporar un Sufico AV_Calculation Views

1. Considerar la nomeclatura del nombrado para tablas y campos
2. Incorporar un Sufico AV_

10.7 Recomendaciones para la migración de estructuras y datos

En el caso de las estructuras y datos, la recomendación es realizar una ingeniería inversa al modelo de negocio a ser migrado, para ello puede utilizarse SAP Powerdesigner que nos permitirá leer la estructura de bases de datos origen y generar del código para ser ejecutado en SAP HANA con el estándar de nombrado de objetos como campos y tablas elegido.

Una vez que las estructuras están creadas en SAP HANA existen varias formas de incorporar los datos, sin embargo al tratarse de nuestro origen un Data Warehouse no SAP nos quedan las opciones presentadas a continuación en la Tabla 5.

Tabla 5

	Descripción	Ventajas	Desventajas
Smart Data Access	Conexión directa por medio de la replica de la estructura del sistema fuente en SAP HANA	Muy rápido de utilizar No se recomienda para hacer transformaciones	No acepta mas de 86 columnas
Data Services	Herramienta ETL	Útil para hacer limpieza y transformaciones de datos No tiene límite conocido de columnas	Invertir tiempo en la construcción de un job

10.8 Conclusiones

En el caso particular del proyecto de migración mostrado en este artículo una vez migradas las estructuras y los datos del Data Warehouse aún queda trabajo por hacer en el futuro, por ejemplo el rediseño o la migración del proceso ETL, ya sea por SAP Data Services o por SAP LT Replication Server o SAP BW para fuentes SAP. Es muy importante considerar la solución final objetivo dado que es probable que sea necesario replantear la arquitectura e infraestructura empresarial si se tiene como alcance un análisis del tipo "real time".

Para la migración de un Data Warehouse a SAP HANA o implementaciones en SAP HANA la principal recomendación es contar con el apoyo de expertos, como los distintos SAP Partners certificados con la experiencia necesaria en SAP HANA e incluso apoyarse en ellos antes de realizar cualquier inversión para tener una evaluación específica del valor que generará al negocio y qué se requiere para lograrlo. Es común que en las presentaciones de productos de software no se visualicen los escenarios/arquitectura requeridos para la obtención del fin deseado y pudiera variar la implementación de acuerdo a una complejidad del negocio real.

10.9 Referencias

Hickman H, Wells L (2014). Implementing SAP HANA. Boston: Galileo Press.

Kimball R. (2002). The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. New York: Wiley Computer Publishing

Project Management Institute (2013). A guide to the Project management body of Knowledge. Pennsylvania: Project Management Institute, Inc.

Berg B, Sylva P (2013). SAP HANA a Introduction. Boston: Galileo Press

Kimball R. (2002). The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. New York: Wiley Computer Publishing

Project Management Institute (2013). A guide to the Project management body of Knowledge. Pennsylvania: Project Management Institute, Inc.

Newswire, P. R. (2012a). SAP Brings Massive Scale to SAP HANA® Platform NV-SAP-AG-HANA-One: Y.

Newswire, P. R. (2012b). SAP Celebrates One-Year Anniversary of SAP HANA® With Major Milestones CA-SAP-HANA-1-year: Y.

Newswire, P. R. (2012c). SAP Introduces SAP HANA® Cloud, One of the Industry's First In-Memory Cloud Platforms NV-SAP-AG-SAP-HANA: Y.

Newswire, P. R. (2013a). Newsbyte: Deployment Options for SAP® Business Suite Powered by SAP HANA® Deliver Speed and Simplicity SAP-AG-HANA-ERP: Y.

Newswire, P. R. (2013b). Newsbyte: SAP Announces Development Training for New SAP HANA® Platform as "MassiveOpen Online Course" SAP-HANA: Y.

Newswire, P. R. (2013c). SAP Introduces SAP HANA® Enterprise Cloud to Deliver Fast Time-to-Value and Maximum Choice CA-SAP-HANA: Y.

Newswire, P. R. (2014a). SAP Changes the Data Warehousing Landscape With SAP HANA® SAP-BusinessWarehouse: Y.

Newswire, P. R. (2014b). SAP Gives Customers a Simpler Choice and a Clear Path to Adopt SAP HANA® NV-SAP-SE-HANA: Y.

Newswire, P. R. (2014c). SAP® Business Suite powered by SAP HANA® Continues Strong Market Momentum SAP-HANA: Y.

Newswire, P. R. (2014d). SAP® Business Suite powered by SAP HANA® Exceeds 1,000 Customers SAP-HANA-1000customer: Y.

Newswire, P. R. (2014e). State of Indiana Chooses SAP HANA® Platform to Help Fight Infant Mortality SAP-HANA-Indiana: Y.