

Data Warehousing al proceso de capacitación en la Empresa de Telecomunicaciones de Cuba S.A.

Data Warehousing to the training process at the Telecommunication Company of Cuba S.A.

MSc. Alexander Gómez Betancourt¹, Dr.Sc. Lindsay Alonso Gómez Beltrán²

Recibido: 11/2017 | Aceptado: 03/2018

PALABRAS CLAVE

Capacitación
Almacén de datos
HEFESTO
Pentaho
Recursos humanos
Análisis
Datos

KEYWORDS

Training
Data warehouse
HEFESTO
Pentaho
Human resources
Analysis
Data

RESUMEN

Actualmente en la Empresa de Telecomunicaciones de Cuba S.A, se cuenta con una estructura que contempla, dentro de la Dirección Central de Capital Humano, a la Dirección de Formación y Desarrollo, como área rectora del proceso de capacitación. A pesar de tener un sistema de gestión que automatiza parte del proceso de capacitación, se hace difícil el análisis de toda la información generada y necesaria para apoyar la toma de decisiones. En la presente investigación se desarrolló una solución basada en una arquitectura data warehousing para el análisis de datos, utilizando un enfoque de construcción de abajo hacia arriba. Se toman como referencias los indicadores asociados al cumplimiento del plan de capacitación con sus respectivas perspectivas de análisis, creando un mercado de datos, mediante los pasos definidos en la metodología HEFESTO V2.0. Se incorporan procesos automatizados de extracción, transformación y carga de datos transaccionales e históricos, referentes a la gestión de la capacitación, hacia un mercado de datos.

ABSTRACT

Currently in Telecommunications Company of Cuba S.A., has a structure that includes within the Human Capital Central Management to the Directorate of Training and Development, as the guiding area of the training process. Despite having a management system that automates part of the training process, it becomes difficult to analyze all the information that it generates and necessary to support the decisions that are made. In this research, a solution based on a data warehousing architecture for data analysis is developed, using a bottom-up construction approach. The specific indicators associated with the fulfillment of the training plan with their respective perspectives of analysis are taken as reference, creating a data market, using the steps defined in the HEFESTO V2.0 methodology. The solution is presented as support for the decision making of managers and specialists at the different levels involved in the training process. It incorporates automated processes

¹ Empresa de Telecomunicaciones de Cuba S.A. La Habana, Cuba, alexander.gomez@etecsa.cu

² Universidad de Camagüey, Camagüey, Cuba. lindsay.gomez@reduc.edu.cu

of extraction, transformation and loading of transactional and historical data, references to training management, towards a data market. Analytical tools, using the Pentaho Suite version 6.1, allowed to perform deeper analysis, reduce the time spent in recovery tasks and data organization; as well as having the historical information available.

Introducción

En los tiempos actuales, las empresas brindan una mayor atención a su capital humano como base para el buen desempeño de sus procesos. Esto está dado ya que “si una compañía pierde su capital o sus activos, con un buen proyecto empresarial, podrá sobrevivir; pero si aún conservando los activos pierde a su personal y al equipo directivo difícilmente podrá sobrevivir” (Coopers, 2006). Es por ello que existen nuevas estrategias encaminadas a la modernización de las estructuras y de la gestión de los recursos humanos en los entornos empresariales, como actividad esencial en el desempeño de los trabajadores y el cumplimiento de los objetivos (Dolan, Cabrera & Jackson, 2007).

Actualmente la Empresa de Telecomunicaciones de Cuba S.A. (Etecsa), cuenta con una estructura que contempla dentro de la Dirección Central de Capital Humano, a la Dirección de Formación y Desarrollo, como área rectora del proceso de capacitación. Para garantizar esta tarea a todos los niveles, está constituido el Centro de Formación Ramal de Telecomunicaciones, los Centros Regionales (occidente, centro y oriente), los especialistas que gestionan la capacitación en las diferentes unidades organizativas de la empresa y los directivos como principales responsables de la preparación del personal. Debido a la importancia que la empresa le brinda a esta actividad, al alto costo que representa que todos sus trabajadores se encuentren preparados para asumir el reto del quehacer diario, de los constantes cambios en las tecnologías que se implementan, así como de las estrategias y necesidades propias del país; es obligatorio que este proceso cuente con un alto grado de atención a su planeación, gestión, ejecución y control (Guilarte, 2011).

La Dirección de Formación y Desarrollo (DFD) se apoya para el funcionamiento de su proceso del

“Sistema Automatizado de Gestión Estratégica de la Capacitación” (Sagec) (Castillo, 2010), el cual automatiza en gran medida este proceso de capacitación en Etecsa. Sagec es un sistema operacional que centra sus funciones en las etapas de planificación, organización y ejecución de las acciones de capacitación y que registra la información relacionada con los planes anuales de capacitación, de los cursos impartidos, los participantes, los instructores, la logística necesaria, así como de los presupuestos y gastos. Cuenta con una serie de reportes importantes para el desarrollo de la actividad y brinda información de la ejecución y control operativo del proceso. Estos elementos han requerido que la DFD realice análisis del estado de la capacitación y el cumplimiento de la planificación, con el objetivo de tomar decisiones tácticas y operativas para asegurar la ejecución de las acciones de capacitación y además el seguimiento continuo a la evolución del plan de capacitación de cada unidad organizativa. Se han definido un número de indicadores que brindan información necesaria para este fin, los cuales influyen en el cumplimiento de las acciones y del plan de capacitación de cada unidad y de la empresa.

A pesar de tener un sistema automatizado y un proceso definido, se hace difícil el análisis de toda la información generada y necesaria para apoyar las decisiones que son tomadas, así como alinear los procesos claves de la gestión de los recursos humanos con la estrategia de la empresa e integrarlos con el resto de los procesos de gestión. Los datos que son manejados en Sagec son numerosos y detallados, los mismos cuentan con un gran número de relaciones entre sí que complejizan su posterior análisis, teniendo en cuenta que este sistema está diseñado para la ejecución de los planes y otras tareas operativas. Además, mediante el sistema actual no es posible la elaboración de informes dinámicos, flexibles e interactivos, que logren una rapidez en los tiempos de respuesta en el análisis de grandes

volúmenes de datos. Por tales motivos, se han detectado insuficiencias en los sistemas que apoyan la toma de decisiones, que dificultan el buen desempeño de la actividad de capacitación en la empresa.

El objetivo de esta investigación es desarrollar una arquitectura data warehousing que facilite la toma de decisiones en el proceso de capacitación en Etecsa.

Se obtiene una solución basada en una arquitectura data warehousing para el análisis de datos, siendo el aporte más valioso: el apoyo a la toma de decisiones de los directivos y especialistas en los diferentes niveles que intervienen en el proceso de capacitación. Con la propuesta, se pueden incorporar procesos automatizados de extracción, transformación y carga de datos referentes al proceso de gestión de la capacitación.

Materiales y métodos

Se realizó una organización de esta información mediante una arquitectura data warehousing, lo cual permitirá a los especialistas y directivos realizar análisis más profundos, disminuir los tiempos empleados en estas tareas, determinar situaciones de demoras, conflictos o mal funcionamiento del proceso; así como tener la información histórica disponible y poseer informes dinámicos, flexibles e interactivos.

Luego de haber descrito los indicadores que miden el cumplimiento de la planificación de la capacitación en la empresa, se identificaron las necesidades de información asociadas a los mismos. La identificación de estas necesidades se realizó mediante los métodos de observación y entrevistas a los especialistas y directivos que intervienen en el proceso de toma de decisiones en la gestión de la capacitación en Etecsa. Se utilizó la herramienta Pentaho Data Integration (PDI).

Resultados y discusión

A consideración de Moody (Moody, 1983), existen dos factores que afectan el nivel para tomar una decisión: lo competente que sean las personas y el acceso que pueda tener a todos los datos necesarios para tomar una decisión basada en información completa. Numerosos autores como Angeloni (Angeloni, 2003), Ariño (Ariño, 2005), Caixeta (Caixeta & Rodrigues, 2008) han abordado elementos puntuales sobre estos componentes y disímiles investigadores han profun-

dizado sus particularidades. En este sentido se puede enunciar que los componentes fundamentales de la toma de decisiones son: la situación – problema, el individuo (decisor o grupos de decisión), la información y los elementos contextuales. Teniendo en cuenta lo planteado anteriormente, se puede abordar de que la información a partir de datos disponibles, así como su análisis, tiene un rol fundamental en el apoyo de la toma de decisiones.

Los sistemas de soporte a las decisiones (DSS) se centran más en la utilización de las tecnologías de las comunicaciones, datos, documentos, conocimientos y/o modelos para identificar y resolver problemas, tareas del proceso de la toma de decisiones (Power, Sharda y Burstein, 2015). Por otro lado, según Karen y Lares (Karen y Lares, 2005) durante años, la informática dentro de las empresas se ha considerado como una herramienta para el apoyo de las funciones operativas. Aunque Urrutia (Urrutia, 2000) plantea que no solo debe concentrarse en las funciones operativas, sino que deben utilizarse para manejar mejor la información de la que dispone la empresa, con el fin de conseguir ventajas competitivas y generar así nuevos beneficios. Esta información debe ser clara, precisa, dirigida a perfiles, rápida y estar disponible en el momento que se la necesite, de una manera completa y armonizada con otras informaciones.

Teniendo en cuenta la convergencia entre las tecnologías de la información y la administración de empresas, y con el objetivo de generar conocimiento partiendo de datos transaccionales, la inteligencia del negocio, como concepto, atiende la toma de decisiones basándose en herramientas principalmente tecnológicas. Para Aranibar (Aranibar, 2013), estas herramientas están compuestas por: bases de datos gerenciales (almacenes de datos), análisis multi-dimensionales, minería de datos, consultas, reportes complejos y visualización, conformando una arquitectura data warehousing.

DSS en la capacitación empresarial

En un estudio realizado por el Grupo de Consultoría de Boston y la Federación Mundial de Gestión del Personal, para dar respuesta de cómo afrontar los desafíos en recursos humanos (RRHH) en todo el mundo hasta el 2015 (Strack, 2008) se expresó: “Si contemplamos la estrategia, los indicadores de gestión y los RRHH como los tres vértices de un triángulo, veremos que en la mayoría de las empresas los vínculos entre

RRHH y estrategia así como entre RRHH y los indicadores de gestión están rotos o no existen”. Es por ello que Hebrero (Hebrero, 2015) menciona que no es de extrañar que una de las prioridades en los estudios de tendencias de gestión de recursos humanos (GRH) esté relacionada con la analítica de RRHH (Fitz-Enz, 2010), (McAfee y Brynjolfsson, 2012), (Zikopoulos y Eaton, 2011), es decir, con el uso estratégico de la información y la aplicación de técnicas analíticas a datos de RRHH.

Dentro de las etapas de gestión estratégicas mencionadas por Aranibar (Aranibar, 2013) se detallan algunos conceptos que desarrollan la estrategia. Dentro de ellos se encuentra la etapa 5: monitorear y aprender, aquí la empresa comienza a monitorear los resultados de desempeño, actúa para mejorar las operaciones y estrategia basada en nueva información y aprendizaje. Sin embargo, la tecnología actual permite aplicar un enfoque más sistemático. Cada día, el empleado de una empresa emite múltiples señales que miden su rendimiento, grado de satisfacción respecto a su trabajo y compromiso con la organización (Román, Cristóbal, Carlos, Vázquez y Antonio, 2016). Por otra parte, las técnicas de análisis de datos, englobadas bajo la disciplina hoy denominada ciencia de datos (Waller y Fawcett, 2013), están viviendo un nuevo resurgir que puede realizar grandes aportaciones en este campo, por la posibilidad de convertir todos los datos de RRHH en información y conocimiento útil para la empresa (Román, Cristóbal, Carlos, Vázquez y Antonio, 2016).

Arquitectura data warehousing

Kimball y Ross (Kimball y Ross, 2011) hacen referencia a que los sistemas operacionales están optimizados para procesar transacciones de forma rápida. De manera previsible, realizan las mismas tareas operacionales una y otra vez, ejecutando los procesos de negocio de la organización. Dado este enfoque de ejecución, los sistemas operacionales típicamente no mantienen el historial, sino que actualizan los datos para reflejar el estado más actual.

En contraposición a los usuarios de un sistema data warehousing, los cuales, según Kimball y Ross (Kimball y Ross, 2011), necesitan datos detallados para dar soporte a sus preguntas en constante cambio y suelen exigir que se preserve el contexto histórico para evaluar con precisión el rendimiento de la organización en el tiempo; estos sistemas están optimizados

para consultas de alto rendimiento, ya que las preguntas de los usuarios suelen requerir que cientos de miles de transacciones sean buscadas y comprimidas en un conjunto de respuestas.

La concepción de un almacén de datos propuesta por Aranibar (Aranibar, 2013) surge ante la necesidad de usar los datos residentes en sistemas operacionales, requeridos para el planeamiento y la toma de decisiones. Para cumplir estos objetivos, las consultas que preparan y resumen los datos consumen muchos recursos de los sistemas transaccionales, generándose una reducción en el rendimiento de los sistemas que al mismo tiempo está en plena captura de transacciones. En una situación adversa como esta, la alternativa de separar la parte transaccional de la gerencial o estratégica requiere de un modelo de datos y su correspondiente implementación, orientado exclusivamente para reportes y toma de decisiones. Por su parte, Aranibar (Aranibar, 2013) menciona que el aspecto del rendimiento operativo de los sistemas de bases de datos es una de las principales razones por las que fue concebida la arquitectura data warehousing.

Una división en componentes de una arquitectura datawarehousing puede estar conformada de la siguiente manera: (Bernabeu, *HEFESTO* (5a ed.) (Avalos y Figueroa, 2013)

Datos operacionales: conocido también por sistemas de procesamiento de transacciones en línea (OLTP). Representa toda aquella información transaccional que genera la organización diariamente y las fuentes externas.

Gestor de carga: mediante procesos de extracción, transformación y carga (ETL), los que se encargan de extraer los datos desde los OLTP para manipularlos, integrarlos, transformarlos y posteriormente cargar los resultados obtenidos en el almacén de datos.

Gestor de almacén de datos: su finalidad es transformar e integrar los datos fuentes y de almacenamiento intermedio en un modelo adecuado para la toma de decisiones; permitiendo realizar todas las funciones de definición y manipulación del depósito de datos, para poder soportar todos los procesos de gestión del mismo.

Gestor de consultas: este componente realiza las operaciones necesarias para soportar los procesos de gestión y ejecución de consultas relacionales, propias del análisis de datos. Recibe las consultas del usuario, las aplica a la estructura de datos correspondiente y devuelve los resultados obtenidos.

Herramientas de acceso al componente de almacenamiento físico: son conformadas por los sistemas que permiten al usuario realizar la exploración de datos del almacén de datos. Básicamente constituyen el nexo entre el depósito de datos y los usuarios.

Usuarios: son aquellos que se encargan de tomar decisiones y de planificar las actividades del negocio.

Se puede mencionar que básicamente la forma de operar una arquitectura data warehousing se resume de la siguiente manera: (1) los datos son extraídos desde aplicaciones, bases de datos, archivos, etc., esta información generalmente reside en diferentes tipos de sistemas, orígenes y arquitecturas, teniendo formatos muy variados (2) los datos son integrados, transformados y limpiados, para luego ser cargados en el almacén de datos o mercado de datos, principalmente mediante herramientas ETL (Guilarte, 2011). esencialmente la información del almacén de datos se estructura en cubos multidimensionales, ya que estos preparan esta información para responder a consultas dinámicas con un buen rendimiento (Castillo, 2010) los usuarios acceden a los cubos multidimensionales del almacén de datos utilizando diversas herramientas de consulta, exploración, análisis, reportes, etc.

Metodología de desarrollo

Teniendo en cuenta un estudio realizado por Uquillas (Uquillas, 2015) donde expone una valoración de varias metodologías de desarrollo para almacenes de datos por diferentes criterios de evaluación y factores externos, se selecciona HEFESTO como la metodología a utilizar para el desarrollo de la investigación. Dentro de los aspectos fundamentales que se basa esta elección se encuentran que la misma tiene un buen nivel de detalle de las tareas que componen cada fase, es viable para aplicarla en el escenario de la investigación y se podrá definir la arquitectura de los datos mediante el desarrollo de cada una de las fases que la componen.

La última actualización de la metodología de HEFESTO es la versión 2.0 en junio de 2010 y disponible bajo licencia GNU FDL, “se fundamenta en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos” (Uquillas, 2015).

Consta de cuatro fases: análisis de requerimientos, análisis de los OLTP, modelo lógico del almacén de datos y proceso ETL. Puede ser utilizada en cual-

quier ciclo de vida que no requiera fases extensas de requerimientos y análisis, con el fin de entregar una implementación que cumpla con una parte de las necesidades proporcionadas por el usuario (Bernabeu, HEFESTO (5a ed.).

Herramientas para su desarrollo

En Etecsa, la Instrucción 1/2010 de la División de Tecnologías de la Información (Vidal, 2010), plantea promover el uso de estándares para la creación de aplicaciones en código abierto, así como entornos de programación y uso de librerías que facilitan el desarrollo, además de definir estrategias de investigación y desarrollo que permitan lograr la homologación de las aplicaciones más utilizadas en código abierto.

El estudio implementa la arquitectura data warehousing con la Suite Pentaho. A pesar de que el estudio arrojó que no es una herramienta muy utilizada por los usuarios en el mundo, comprende un conjunto de herramientas que funcionan con una buena sinergia, además de que contiene potencialidades similares a las demás de la comparación y que se sitúan en un peldaño alto del cuadrante de Gartner, dentro de las herramientas de código abierto. Otra cuestión a tener en cuenta es el de la existencia en la empresa de una plataforma implementada, instalada, configurada y desplegada, lo que facilita la integración de un repositorio de información con una misma estructura tecnológica, esto beneficia en cierta manera esta implementación.

Diagnóstico del estado actual

Etecsa declara que el proceso de gestión de la capacitación, en las condiciones actuales, garantiza que se alcance en todos los niveles empresariales una capacitación efectiva, definida como el proceso por el cual se le brinda a cada trabajador, capacitación en función de las competencias requeridas para su puesto de trabajo, al menor costo y de acuerdo a las necesidades de su desarrollo personal, y de la calidad de los servicios que debe prestar la empresa (Marín, 2017).

La Resolución 29, del Ministro de Trabajo y Seguridad Social, establece el “Reglamento para la planificación, organización, ejecución y control del trabajo de la capacitación y desarrollo de los recursos humanos en las entidades laborales” (Cattaya, 2006).

En tal sentido, Etecsa, único operador de telecomunicaciones en el país, ha implementado una fuerte

estructura metodológica y administrativa para la gestión de la capacitación de sus trabajadores, teniendo en cuenta el entorno cambiante y altamente competitivo en el que se desenvuelve (Castillo, 2010)

Por tal motivo, la Dirección Central de Capital Humano de Etecsa ha institucionalizado una red de capacitación que tiene como objetivo “asegurar la preparación continua que necesitan los especialistas de la organización, para soportar este constante desarrollo” (Castillo, 2010)

Según la proyección estratégica de la empresa, se define en el área de resultado clave (ARC) #2 (Presidente Ejecutivo, 2016): efectividad en la gestión empresarial, con el objetivo 2.1: elevar la efectividad de la gestión de la capacitación en la empresa con respecto al año anterior y con un criterio de medida 2.1: cumplidas al 100% las acciones previstas en el plan de capacitación, en función del desarrollo de la empresa. Es ese sentido, se han definido varios indicadores que inciden en la evaluación y el cumplimiento del objetivo planteado, los cuales se describen en la Tabla 1.

Identificación de necesidades de información, indicadores, perspectivas

Luego de haber descrito los indicadores que miden el cumplimiento de la planificación de la capacitación en la empresa, se identificaron las necesidades de información asociadas a los mismos. La identificación de estas necesidades se realizó mediante observación y entrevistas a los especialistas y directivos de los diferentes componentes que intervienen en el proceso de toma de decisiones en la gestión de la capacitación en Etecsa. Ya que es de importancia vital que la información que se solicite debe estar soportada en un sistema transaccional, también se realizaron entrevistas a los desarrolladores del Sagec, como principal sistema que automatiza parte del proceso de capacitación en Etecsa.

Partiendo de las preguntas del negocio y realizando un análisis de las mismas se identificaron los indicadores y perspectivas:

Indicadores: Cantidad de acciones cerradas, cantidad de acciones planificadas a finalizar, cantidad de acciones extraplanes cerradas, cantidad de participantes, cantidad de plazas planificadas, total de participantes de las acciones cerradas.

Perspectivas: Modo de formación, modalidad, subsistema, estrategia (de capacitación y desarrollo), entidad, sub estado de acción, proveedor, tipo de pro-

Indicador	Función de cálculo
Cumplimiento del plan de acciones de capacitación. (CPT)	$CPT = AC \div AP * 100$ <p>AC: Total de acciones cerradas AP: Total de acciones planificadas.</p>
Índice de eficacia de planificación (CP)	$CP = \frac{AC - ACE}{AP} * 100$ <p>ACE: Total de acciones extra planes cerradas.</p>
Cumplimiento de participantes (CPP)	$CPP = \frac{TPP}{PP} * 100$ <p>TPP: Participantes que participan. PP: Plazas planificadas.</p>
Indicador de asistencia (CPA)	$CPA = \frac{TPP}{TP} * 100$ <p>TP: Total de participantes.</p>

Tabla 1 Indicadores de capacitación.

grama, ubicación, programa de estudio, estado del programa, plan de capacitación y fecha.

Partiendo de que el área de estudio elegida es la gestión del plan de capacitación, donde se medirán sus respectivos indicadores, se crea el modelo conceptual para mejor entendimiento del alcance del proyecto a desarrollar. El modelo conceptual permite comprender cuáles serán los resultados que se obtendrán, cuáles serán las variables que se utilizarán para analizarlos y cuáles son las relaciones que existen entre estas.

Análisis de los sistemas transaccionales

El objetivo del análisis de los sistemas operacionales o transaccionales es el de examinar los OLTP disponibles que contengan la información requerida, así como sus características, para poder identificar

las correspondencias entre el modelo conceptual y las fuentes de datos. Luego de realizar un análisis de la base de datos del Sagec se determinó solamente realizar el diagrama entidad relación de las tablas que intervienen los indicadores y perspectivas de análisis.

Una vez que se han establecido las relaciones con los OLTP, se deben seleccionar los campos que contendrá cada perspectiva, ya que será a través de estos por los que se examinarán y filtrarán los indicadores. Se analizaron los campos de las tablas a la que se hacía referencia mediante observación del diagrama entidad relación de la base de datos, así como consultas al diseñador de base de datos del Sagec. Con ello se logró un mejor entendimiento de cada campo. Se extraen los campos disponibles en cada una de las tablas de la base de datos que se corresponden con algunas perspectivas de interés, la descripción de lo que representa el campo y si es de interés para los usuarios.

Mediante el análisis del Sagec como principal sistema fuente OLTP se establecieron las correspondencias con el modelo conceptual; realizando una serie de pasos hasta llegar al modelo conceptual ampliado. En este modelo se realiza una descripción de los indicadores identificados con sus respectivos hechos que lo componen y la función de agregación, logrando una comprensión de la forma en que son calculados.

Modelo lógico del almacén de datos

Las uniones son realizadas entre las tablas dimensiones y la tabla de hecho, lo cual permite obtener el diseño del modelo lógico del almacén de datos. Se

crean las uniones partiendo de las tablas dimensiones por su llave subrogada hacia la tabla de hechos con su llave foránea, en una relación de uno a muchos. Se realiza, además, la creación de la llave primaria compuesta por el conjunto de llaves foráneas con las dimensiones de la tabla de hechos “h_plan_accion”; además de la conformación de las tablas dimensiones con sus respectivas llaves primarias compuesta por la llave subrogada.

Integración de datos

Una vez construido el modelo lógico, se procede a poblarlo con datos, utilizando técnicas de limpieza y calidad de datos, procesos de extracción, transformación y carga, etc.; luego se definen las reglas y políticas para su respectiva actualización, así como también los procesos que la llevarán a cabo.

Para realizar el proceso de la carga inicial de los datos desde el sistema fuente se utilizó la herramienta Pentaho Data Integration (PDI) (Meadows, Pulvirenti y Roldán, 2013), la cual está comprendida dentro de la Suite Pentaho. Con esta herramienta se pueden extraer datos desde múltiples fuentes, transformarlos, limpiarlos y conformarlos para cumplimentar los requerimientos de los usuarios; y posteriormente ser cargados hacia el almacén de datos.

Según Bouman y Dongen (Bouman y Van Dongen, 2009) PDI cuenta con dos tipos diferentes de objetos para realizar los procesos ETL que son los trabajos y las transformaciones. El motor del PDI es un componente capaz de interpretar y ejecutar los tra-

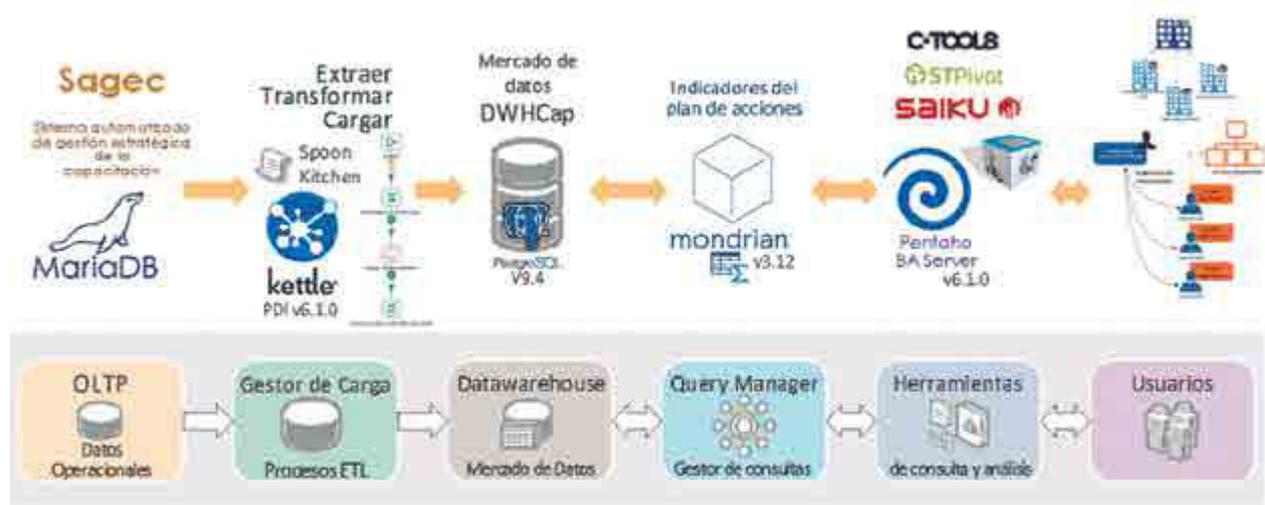


Figura 1. Arquitectura datawarehousing de Capacitación en Etecsa.

bajos y las transformaciones, utilizando para ello, las herramientas Kitchen, Carte y Pan; mientras que para la creación, diseño y pruebas de los trabajos y transformaciones se utiliza la herramienta de integración de datos gráfica Spoon.

Análisis de los datos en el servidor de Pentaho

En esta fase se especificación las necesidades analíticas de los usuarios, partiendo de los informes, reportes, formatos, métodos y mecanismos en que se muestran los indicadores; para llevar a cabo la medición del cumplimiento de la capacitación en Etecsa. Se trabajó en conjunto con los desarrolladores del Sagec, además de con la Dirección de Formación y Desarrollo, los Consejos y Comités de Capacitación para coordinar y validar los reportes y tableros finales, basados en los indicadores de la actividad de capacitación. Posteriormente se trabajó en el desarrollo de estos reportes.

Arquitectura data warehousing

Una arquitectura data warehousing está formada por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema. Por ello es que al diseñar la exposición de cada elemento se hará en forma ordenada y teniendo en cuenta su relación con las demás partes. La figura 1 muestra la propuesta de solución, partiendo del diseño y construcción del mercado de datos, las herramientas utilizadas y demás componentes que intervienen en la misma.

Datos operacionales (OLTP): Comprendido por base de datos Sagec_db en un servidor MariaDB v5.5 y contiene los datos que son generados en el proceso de capacitación de la empresa.

Gestor de Carga: se utiliza la herramienta “Kitchen” del Pentaho Data Integration para la ejecución del trabajo principal J_PLAN_CAPACITACION el cual realiza el proceso ETL desde el OLTP hacia el mercado de datos. Se utilizan parámetros de fecha para la inserción y actualización diaria de los datos.

Datawarehouse: contiene la base de datos CAPDWH en un servidor Postgresql v9.4 el esquema en estrella del mercado de datos diseñado para contener los datos necesarios para la medición de los indicadores. Además, contiene un cubo multidimensional Plan_Capacitacion creados con el fin de poder realizar

consultas de los indicadores a través de los atributos y jerarquías diseñadas.

Gestor de consultas: este componente realiza las operaciones necesarias para el soporte de gestión y ejecución de las consultas de los usuarios, lo comprende el motor Mondrian v3.12, el cual permite la realización de una serie de operaciones sobre los modelos multidimensionales y el cubo Plan_Capacitacion desarrollado.

Herramientas de consultas y análisis: se utiliza la Plataforma de Análisis del Negocio Pentaho v6.1.0 (Pentaho BA Plataform) con una serie de complementos que mejoran las posibilidades de tareas administrativas y de los usuarios para realizar las consultas:

- **Community Dashboard Framework (CDF) v6.1.0.1:** es el núcleo para la ejecución y creación de tableros de control. Es desarrollado por Webdetails.
- **Community Dashboard Editor (CDE): v6.1.0.1:** simplifica la creación, edición y vista previa de los tableros. Desarrollado por Webdetails.
- **Community Graphics Generator (CGG) v6.1.0.1:** permite la exportación de gráficos como imagen, habilita la inclusión de gráficos en los reportes de Pentaho. Es desarrollado por Webdetails.
- **Saiku Analytics v3.10:** Es una suite modular de análisis de código abierto que ofrece análisis OLAP ligero y que permanece fácilmente incrustado, extensible y configurable. Contiene un servidor RESTful que se conecta a los sistemas OLAP existentes. Es de fácil uso, teniendo una analítica intuitiva a través del arrastre y suelta de los componentes. Permite la generación básica de gráficos. Es desarrollado por Meteorite BI.
- **Saiku Chart Plus v2016.05.19:** es un proyecto de código abierto que ayuda a los usuarios de Pentaho BA a crear otros tipos de gráficos y mapas basados en Saiku, Highcharts y Google Maps. Es desarrollado por IT4biz.
- **Pivot4J v1.0:** Pivot4J proporciona una API común para servidores OLAP que puede utilizarse para construir vistas de servicio analítico con interfaz de estilo de pivote. Es una completa aplicación de cliente OLAP, que se puede usar como reemplazo de la aplicación web de JPivot. Es desarrollado por Pivot4J Project.
- **Web Adhoc Query and Reporting (WAQR):** Es una herramienta de generación de informes basada en la tecnología Web 2.0 (AJAX y DHTML)

diseñada para ayudar a los usuarios a generar fácilmente informes dinámicos utilizando la plataforma Pentaho BA. Proporciona una interfaz de usuario que ayuda al usuario a identificar y agrupar rápidamente los datos de interés, aplicar restricciones a los usuarios y generar un informe. Los informes resultantes se pueden ver a través de la función de vista previa de WAQR, o a través de una aplicación web basada en Pentaho. Es desarrollado por Pentaho Legacy.

- Pentaho FusionCharts v4.2.2: permite la creación de visualizaciones de datos Pentaho BA a través de una integración con la biblioteca FusionCharts. Incluye acceso a gráficos en tiempo real, componentes y mapas, todo esto optimizado para dispositivos móviles. El complemento permite integración con CDF y CDE. Es desarrollado por Xpand IT.

Usuarios: responsables de la exploración de los datos y conformado principalmente por los trabajadores del negocio, en este caso por los trabajadores que intervienen directamente en la gestión de la capacitación en la empresa.

Evaluación de los resultados

Con el objetivo de valorar la aceptación de la arquitectura data warehousing como apoyo a la toma de decisiones en el proceso de capacitación en Etecsa, se tomó como criterio para la selección, que los trabajadores y directivos que lo conforman cuenten con 10 años o más de experiencia en la empresa y se desempeñen en funciones afines a la gestión del proceso de capacitación. Esta selección a juicio los declara como especialistas, conformando una muestra no probabilística o dirigida (Battaglia, 2008) de un total de 21 trabajadores.

Para que los especialistas tuvieran una mejor visión de la solución se realizó una presentación de la misma, se configuró la arquitectura data warehousing en un ambiente de pruebas y se crearon cuentas de usuarios para cada uno en el servidor de analítica de negocios de Pentaho; de forma tal que ellos realizaran las pruebas que estimaran con las herramientas presentadas. Posteriormente, se realizó el cuestionario enviado por correo electrónico como adjunto, para ser respondido de forma auto administrada.

De forma general la evaluación arrojó una puntuación de 4,08 siendo favorable la aceptación de la solución de la arquitectura data warehousing como apoyo a la toma de decisiones. En la figura 2 se muestran los porcentajes de respuestas de los especialistas en cada una de las categorías, calculado a partir de la cantidad de veces que cada especialista respondió cada una de



Figura 2. Resultados de la encuesta de satisfacción.

las categorías; donde se puede observar que un elevado por ciento corresponde a las categorías de acuerdo y muy de acuerdo (29% y 48% respectivamente).

Dentro de las sugerencias más significativas emitidas por los usuarios se encuentran, la inclusión de otros objetivos e indicadores de la actividad de capacitación mediante la arquitectura presentada y la necesidad de integrar la interfaz de análisis de los datos del servidor de Pentaho dentro de la interfaz del sistema operacional Sagec.

Conclusiones

Luego de realizar las encomiendas de la presente investigación y de realizar un análisis de los resultados obtenidos en cada una de ellas, se ha llegado a las siguientes conclusiones:

El estudio de los referentes teóricos de los sistemas de soporte a la toma de decisiones arrojó que su aplicación trae consigo beneficios asociados al control y el desempeño administrativo, mediante el análisis rápido, combinado y eficaz de la información.

La caracterización de la arquitectura data warehousing permitió definir las metodologías y herra-

mientas necesarias para la construcción de la solución propuesta.

Se diagnostica el estado actual del proceso de toma de decisiones en la capacitación en Etecsa, determinando los componentes, niveles de dirección y factores que interviene en la misma; además de la fuente de información existente, no suficiente para el análisis de la información.

Se desarrolló una arquitectura data warehousing permitiendo mejorar el acceso a la información; realizar análisis más profundos en menor tiempo y poseer informes dinámicos, flexibles e interactivos.

Se aplicaron cuestionarios que evaluaron satisfactoriamente la aceptación del desarrollo de una arquitectura data warehousing como apoyo a la toma de decisiones en el proceso de capacitación en Etecsa.

Referencia

- Angeloni, M. T. (2003). Elementos intervinientes na tomada de decisão. *Ciência da Informação*, 32(1), 17-22.
- Aranibar, J. C. (2013). *Sistemas de información gerencial para la administración del desempeño empresarial*. La Paz, Bolivia.
- Ariño, M. Á. (Ed.). (2005). *Toma de decisiones y gobierno de organizaciones*: Deusto.
- Arribas Urrutia, A. (2000). Comunicación en la empresa: la importancia de la información interna en la empresa *Revista Latina de Comunicación Social*, (27), 6.
- Avalos, A. M. y Figueroa, Y. C. (2013). Mercado de datos en apoyo a la toma de decisiones sobre el personal docente e investigativo en el departamento de Recursos Humanos de la UCLV. (Trabajo de diploma), Universidad Central “Marta Abreu” de la Villas.
- Battaglia, M. (2008). Convenience sampling *Encyclopedia of Survey Research Methods* (pp. 806-808).
- Bernabeu, R. D. *HEFESTO* (5a ed.). Argentina: Grupo Eglu.
- Bouman, R. y Van Dongen, J. (2009). *Pentaho solutions, business intelligence and data warehousing with Pentaho and MySQL*: Wiley.
- Brizuela Leonard, E. I. y Castro Blanco, Y. (2013). Metodologías para desarrollar Almacén de Datos. *Revista de Arquitectura e Ingeniería*, 7(3), 3-12.
- Caixeta, M. L. y Rodrigues, R. B. (2008). A decisão como resultado de um processo social da informação compartilhada. *Ciência da Informação*, 13(1), 81-104.
- Castillo, O. P. (2010). Lineamientos para el funcionamiento de la red de capacitación.
- Castillo, O. P. (2010). Lineamientos para la actualización de la información de la actividad de capacitación a través del Sistema Automatizado para la Gestión Estratégica de la Capacitación.
- Cattaya, A. M. (2006). *Reglamento para la planificación, organización, ejecución y control del trabajo de la capacitación y desarrollo de los recursos humanos en las entidades laborales*. 10.
- Chun Wei, C. (Ed.). (1999). *La organización inteligente: el empleo de la información para dar significado, crear conocimiento y tomar decisiones*. México DF: Oxford University Press.
- Coopers, P. (2006). *Cuadro de Mando e Indicadores para la Gestión de las Personas*
- Dolan, S., Cabrera, R. y Jackson, S. R. (Eds.). (2007). *La gestión de los recursos humanos*: Mc Graw Hill.

- Fitz-Enz, J. (2010). The new HR analytics: predicting the economic value of your company's human capital investments: AMACOM Division of American Management Association International.
- Guilarte, A. D. (2011). Procedimiento para la gestión de planes estratégicos de capacitación. (Tesis de Maestría), Universidad Central "Marta Abreu" de las Villas.
- Han, J., Pei, J. y Kamber, M. (2011). Data mining: concepts and techniques (3ra ed.): Elsevier.
- Hebrero, C. (2015). La importancia de medir el capital humano. *Revista Capital Humano*, 360(300), 114–116.
- Howar, A. y Ortiz, S. (1971). Decision making and the study of social process. *Acta Sociológica*, 14(4), 213–226.
- Karen, D. C. y Lares, E. A. (2005). *Sistemas de Información para los negocios: Un enfoque para la toma de decisiones* (4 ed.): Mexico: McGraw-Hill.
- Kimball, R. y Ross, M. (2011). *The Data Warehouse toolkit: the complete guide to dimensional modeling* (3ra ed.). Indianapolis: John Wiley & Sons.
- Marín, M. A. (2017). Política para la capacitación en la Empresa de Telecomunicaciones de Cuba S.A. 4.
- McAfee, A. y Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60-68.
- Meadows, A., Pulvirenti, A. S. y Roldán, M. C. (2013). *Pentaho Data Integration Cookbook*: Packt Publishing Ltd.
- Moody, P. E. (1983). *Decision making: Proven methods for better decisions*: McGraw Hill Higher Education.
- Power, D. J., Sharda, R. y Burstein, F. (2015). *Decision support systems*: Wiley Online Library.
- Presidente Ejecutivo. (2016). Carta circular del presidente ejecutivo. (837).
- Román, I. V., Cristóbal, G., Carlos, J., Vázquez, G. y Antonio, J. (2016). TALENT+ Tecnologías avanzadas para la Gestión del Talento. *Revista Procesamiento del Lenguaje Natural*, 57.
- Strack, R. (2008). *Creating people advantage: How to address HR challenges worldwide through 2015*.
- Uquillas Gárces, M. B. (2015). Estudio comparativo de metodologías e implementación de alternativas business intelligence opensource vs. propietarias en entornos tradicionales; caso prototipo en las pymes en el sector agroindustrial. (Tesis de Maestría), Universidad de las Américas, Quito, Ecuador.
- Vidal, W. G. (2010). Lineamientos generales de la migración a código abierto de la empresa de telecomunicaciones de Cuba, S.A., 6.
- Waller, M. A. y Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84.
- Zikopoulos, P. y Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*: McGraw-Hill Osborne Media.

