

Universidad de las Ciencias Informáticas

Facultad de Ciencias y Tecnologías Computacionales



“Mercado de Datos para el área de Recursos Humanos”

**Trabajo de Diploma para optar por el título de
Ingeniero en Ciencias Informáticas**

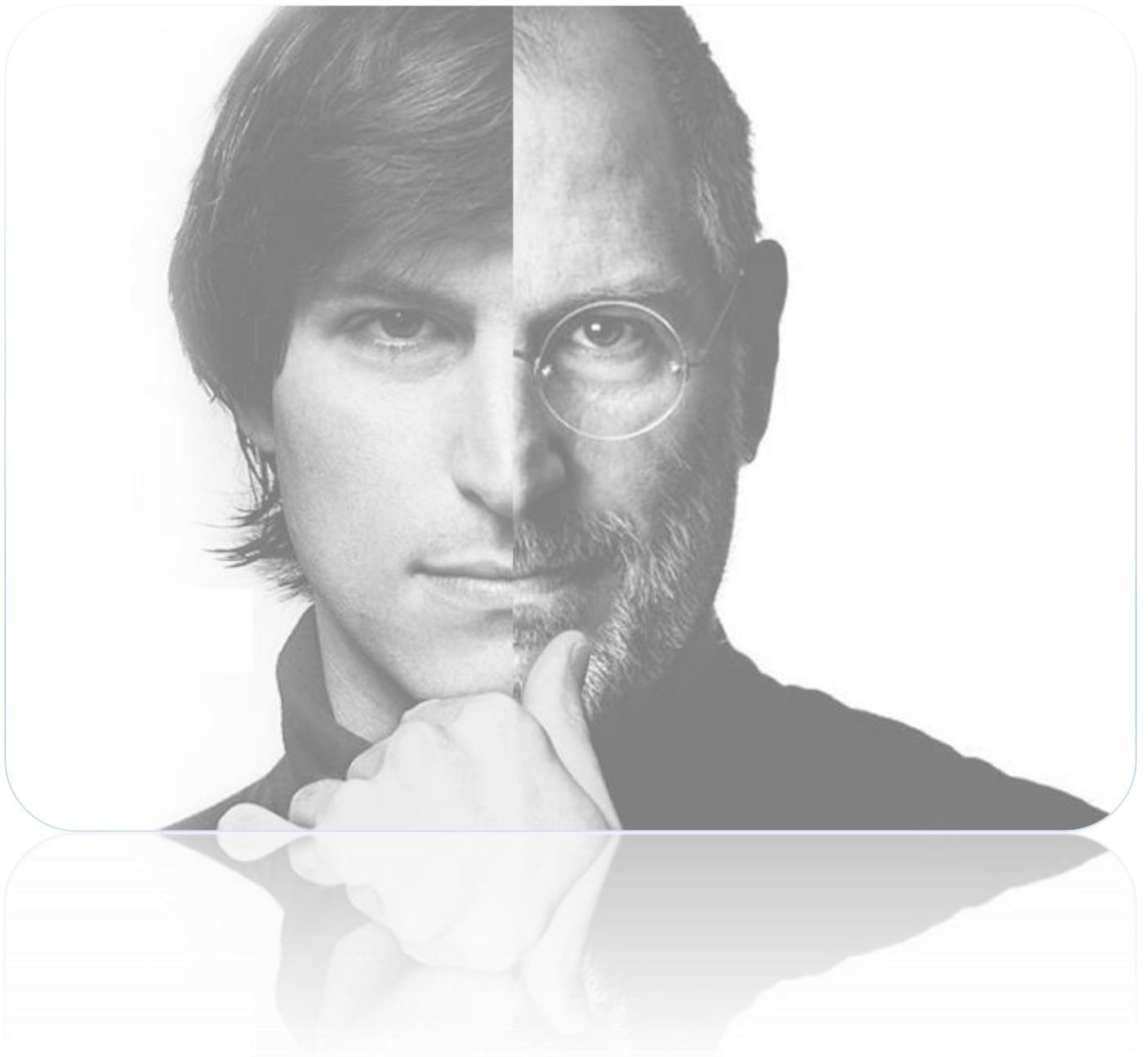
Autores: Suinny Castillo O’Farrill
Yanelis Pérez Hernández

Tutores: MSc. Omar Mar Cornelio
Ing. Irela González Piñera

Co-Tutor: Ing. Juan Pablo Barroso Casanova

La Habana, junio 2017

“Año 59 de la Revolución”



"El único modo de hacer un gran trabajo es amar lo que haces"

Steve Jobs

Declaración de autor

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año 2017.

Suinny Castillo O'Farrill

Yanelis Pérez Hernández

Firma del Autor

Firma del Autor

MSc. Omar Mar Cornelio

Ing. Irela González Piñera

Firma del Tutor

Firma del Tutor

Datos de Autor

Tutor: Msc. Omar Mar Cornelio

Universidad de las Ciencias Informáticas, La Habana, Cuba

Especialidad de graduación: Licenciatura en informática

Correo Electrónico: omarmar@uci.cu

Tutor: Ing. Irela González Piñera

Universidad de las Ciencias Informáticas, La Habana, Cuba

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Correo Electrónico: irela@uci.cu

Co-Tutor: Ing. Jun Pablo Barroso Casanova

Universidad de las Ciencias Informáticas, La Habana, Cuba

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Correo Electrónico: jpbarroso@uci.cu

Autor: Suinny Castillo O’Farrill

Universidad de las Ciencias Informáticas, La Habana, Cuba

Correo Electrónico: sofarrill@estudiantes.uci.cu

Autor: Yanelis Pérez Hernández

Universidad de las Ciencias Informáticas, La Habana, Cuba

Correo Electrónico: yperezh@estudiantes.uci.cu

Agradecimientos de Suinny

Se torna realmente difícil concentrar en solo una página a personas que de una manera u otra me han apoyado tanto. En primer lugar, quisiera agradecer a mi abuela, Consuelo y a mi mamá Niovis, que han sido mi sostén en la vida, mis fuentes de valor y dedicación, por ser quienes se han sacrificado toda su vida por brindarme bienestar, por estar a mi lado en todo momento y apoyarme en todas y cada una de mis decisiones haciéndome creer siempre que puedo lograr lo que me proponga, por todo su amor, mil gracias. A mi papá Richard por siempre estar pendiente de la tesis y brindarme incondicionalmente su apoyo y además sobre todo malcriarme tanto. A mis dos jabaitos Liordy y Liodnys por sus malcriadeces y a la vez por sus cariños los amo.

Quisiera agradecer a mi tía Idalmis por taparme en todas mis locuras y darme al primo que más quiero Alejandro.

A una de las primeras personas que quiero agradecerle mi éxito, que para mí es súper especial en esta escuela es a la May.

A mis dos hermanitas del alma Lisandra y Dayana.

A una persona que todos esperan a que le agradezca Jesús Magdiel.

A mis hermanitos Rafael, Jorge Carlos, Mi gordo bello el Puris y el Negrito Mickey Mouse.

A la Danaysi, a Zarabeth, a Leo, Enier, Juanfri, Erick y Daniel que chicos estos los quiero de corazón y no me arrepiento de compartir con ustedes.

A los chicos de la kizomba, a los de la rueda, la samba, el fútbol.

A mis profesoras de alma que me ayudó mucho Yanelis Benítez y a Hayme Llerena por sus críticas constructivas.

Y a mi dúo de tesis por apoyarme durante todo este proceso tan estresante.

Lo siento si alguien se me queda, a todos ustedes gracias por tenderme la mano cuando la necesité, por ayudarme a encontrar la paciencia y la tranquilidad en los peores momentos; a todos gracias por ser lo que son.

Agradecimientos de Yanelis

La gratitud debería ser un acto constante de cada hora, de cada día, de toda la vida; pero las palabras nunca alcanzan cuando lo que hay que decir desborda el alma y es que el agradecimiento es la memoria del corazón.

Quiero agradecer en primer lugar a las dos personas más importantes en mi vida, esas dos personas que me dieron la vida y que gracias a ellos me he convertido en lo que soy hoy en día; y esas dos personas son mis queridos padres a los cuales les debo mi existencia. No tengo palabras para hacerles llegar cuan agradecida estoy por todos sus consejos, por la educación que me brindaron, por todas sus enseñanzas, por guiar mis pasos a lo largo de todos estos años y no dejar ni por un segundo que cayera, por ayudarme a ser mejor persona cada día, por luchar conmigo frente a las pruebas que la vida me ha deparado, por ser mi guía y ejemplo abnegado a seguir, por su amor incondicional, por su apoyo frente a cualquier circunstancia. Y a pesar de que mi padre ya no se encuentra físicamente conmigo hace algún tiempo, era su sueño verme convertida en toda una profesional y de él aprendí que con esfuerzo, trabajo y constancia todo es posible; y sé que desde arriba sigue guiando mis pasos y es mi guardián que va conmigo a todas partes. A mi mamá por ser la mejor madre de este mundo, gracias por no decaer conmigo nunca y por sobreponerte a las cosas difíciles que nos tocó vivir juntas. Quiero que sepas que estoy muy orgullosa de ti y que toda palabra se queda corta para decirte lo agradecida que estoy por acompañarme incasablemente en mis sueños y por todo lo que en el día a día me has brindado. Gracias a mis bellos padres, si volviera a nacer los elegiría de nuevo como mis padres, son los mejores. Te amo papi... Te amo mamá.

A mis hermanos Ariel, Alexander que a pesar de que no nos vemos mucho de una forma u otra siempre se han preocupado por su única hermanita.

A mi hermano Ulises ese que siempre estuvo a mi lado desde que nací, y que hoy no puede estar aquí por cosas de la vida, quiero que sepas que a pesar de los problemas que hayamos tenido te quiero muchísimo y daría lo que fuera porque volvieras ser ese hermano que me cuidaba siempre, que me daba consejos y que no permitía que nada ni nadie me hiciese

daño; sé que estarás muy orgulloso de que tu hermana finalmente se haya convertido en ingeniera, ahora somos profesionales los dos. Gracias por darme la sobrina más linda de este mundo, a la cualquier quiero con la vida.

A mi tía Hilda mi profesora de toda la vida, Caridad mi peluquera desde pequeña, Teresa, Margarita, gracias por comportarse como mis segundas madres, por apoyarme, quererme, aconsejar, por estar ahí cuando las he necesitado y por estar al pendiente de mí a lo largo de toda mi vida.

A mis primos Miguel y Ebert que son para mí como mis hermanos, a los dos los quiero con la vida y gracias por cuidar de mí siempre.

A mis primas Yenisel, Teima que me dio la posibilidad de ser la madrina de sus dos bellas hijas Elimary y Deymari, a Mayte, a Naomi la futura neurocirujana de la familia y especialmente a Eliani a quien quiero muchísimo y que quisiera que también llegara al final de la meta y se convirtiera también en toda una ingeniera, sería todo un orgullo para la familia.

Durante estos largos 5 años he conocido personas que han dejado en mi corazón una huella imborrable y de cada de una de esas personas me llevo lo más importante y es la amistad, además de todos los recuerdos vividos; especialmente mi súper amiga Clau , que aunque hoy no pudo estar aquí presente porque Brunito no se lo permitió no puedo dejar de agradecerle, gracias por estar ahí cuando más lo he necesitado, gracias por reír y llorar conmigo, por estar a mi lado en las buenas y en las malas , gracias por todos los momentos de locura y diversión vividos en esta etapa universitaria , gracias por ser mi amiga incondicional y espero que dure por siempre.

A mi Midalís mi amiga, la mejor compañera de cuarto que me hubiera podido tocar, mi compañera de estudios y de trabajos, viste como al final lo logramos, gracias a tí te debo parte de este triunfo ya que sabes que mutuamente nos ayudamos a que este sueño se hiciera realidad, sólo quisiera que se cumpliera nuestro deseo de ser también compañeras de trabajo el curso que viene y no olvides invitarme a tu casa cuando esté terminada.

A mis amigas Lisi mi peluquera en la UCI, la mejor que no quepa duda, a Mili por las tertulias que hacíamos hablando de todo un poco y por todas las noches de cafés, a las dos las quiero y las voy a extrañar.

A mi súper amigo del alma Renecillo, el fotógrafo de siempre y autor de las mejores fotos tomadas en estos 5 años gracias por brindarme tu amistad incondicional y por nuestra complicidad en todo momento, te quiero... y espero que sigamos en contacto.

Al negro de Magdiel, no te imaginas lo feliz que estoy de haberte conocido no alcanzaría un libro para enumerar todos los buenos y malos momentos vividos juntos, gracias por todas esas noches-madrugadas de tertulias que teníamos tu y yo, gracias por brindarme tu hombro cuando más lo he necesitado, por aconsejarme en todo momento, en fin, gracias por tu amistad que sé que será para siempre, I Love.

A mi queridísimo amigo Así con el cual tengo más tiempo compartido, a ti decirte que te quiero y admiro muchísimo, al final viste como no importa las piedras que se interpongan en nuestro destino sino la voluntad de seguir adelante, lo conseguimos pese a todo.

A mis compañeros de aula a Yazmin, Karel, Rolyen , Lenia, Cani Osmin, , Ariel ,Karel Rolando el poseído, Idel el despistado del grupo ,Marlon , Jorgito que fue testigo de todo el esfuerzo y dedicación que conllevo la realización de este trabajo, a Raci(everyoin),Leandro que pesar de nuestras discordancias considero como un buen amigo y siempre le voy a agradecer el haber estado en uno de los peores momentos de mi vida, al Yonki que a pesar de haberse incluido en mi círculo de amistad un poco más tarde lo quiero y aprecio muchísimo.

A Mayte una de las integrantes de mi grupo de fiestas, gracias por tu amistad , a Dasnaí Rodríguez , Juanfrí , Leito .

A otras amistades que a las que aprecio mucho como los son Daynelis ,Randy, Yoelvis, Yoísel, Brule,Asley, a Puris, a Yo Pazo el amor de mi vida a Ivoncilla , al negro feo de Michel, al enano del Cuso, a Miguel Angel, a Henry,al flaco malo de Elian, Miguel ,a esas amistades que no llegaron al final de la carrera especialmente a Grendys, mi alma gemela Yoelito , a Saídy y Teus.

A Lijandy por ser una personita súper especial para mí, por ser una de las claves fundamentales en la realización de este trabajo, por estar a mi lado todo este tiempo y por haber llegado a mi vida en el momento preciso.

A mi co-tutor, amigo y excompañero de aula Barra, a ti y solo a ti te debo que este trabajo haya culminado finalmente, gracias por enseñarme que con esfuerzo y dedicación todo es posible, gracias por tu paciencia y dedicación en el día a día, gracias por todas las horas dedicadas en el lab y por transmitirme todo tu conocimiento, te quiero y aprecio mucho.

A mis tutores Irela y Omar, a Omar que a pesar de que su nuevo cargo no le permitió brindarnos toda la atención requerida siempre quiso que la tesis saliera con la calidad que ameritaba y no faltaron nunca sus críticas constructivas y especialmente a Irela, gracias por confiar y haber creído en nosotras, a ti que has sido capaz de moldear nuestra forma de proyectarnos, por guiar cada día el desarrollo de este trabajo, por tus críticas, tus regaños cuando nos veías decaer gracias por tu paciencia y dedicación :Te queremos.

A mi tribunal y oponente que con sus revisiones han logrado que esta investigación sea digna de un profesional, gracias por todas las críticas constructivas realizadas las cuales contribuyeron a que este trabajo saliera con la calidad requerida.

A todos los profesores que contribuyeron a mi formación como profesional durante estos 5 años.

Y por último y no menos importante no pueden faltar mis agradecimientos a mi dúo de tesis, a ti Suinny que hemos compartido tantos y tantos momentos buenos y malos a lo largo de estos 5 años, gracias por ser mi amiga incondicional, gracias por ser tan especial, gracias por no dejarme caer cuando sentía que iba a explotar, gracias por confiar en mí, ya ves como todo sacrificio tiene su premio, finalmente ya podemos decir hoy que somos ingenieras, fue un largo y duro camino pero llegamos a la meta, ojalá y esta amistad dure por siempre negrita, no olvides que tu Perrocolis te quiere con la vida : I love.

En fin, a todas aquellas personas que de una forma u otra han contribuido a que esta investigación tenga éxito y a todos aquellos que he conocido durante estos 5 años.

A todos muchas gracias....

Dedicatoria

Dedico esta tesis a:

Primeramente, a personas que significa mucho para mí y a quienes estaré agradecida siempre que, a pesar de no estar físicamente a mi lado, sé que espiritualmente lo están y me guían en cada paso que doy, mi papito Luisito y mi abuelito Gonzalo, los echo mucho de menos. A mis dos mamás: mi abuelita Consuelo y mi mamá Niovis que las amo con la vida y son mi razón de ser, a mis hermanos que los adoro y a mi papá Richard por siempre estar a mi lado cuando lo necesite y a mi familia en general por amarme y apoyarme siempre en todo y ser el motor inspirador de mis logros.

Suínny

Dedico esta tesis a:

Quiero dedicar esta tesis a mis padres, por quererme, por malcriarme, por ser mi orgullo y ejemplo a seguir en cada paso que doy.

A ustedes que son mi fuente de inspiración va dedicado íntegramente este trabajo.

Yanelis

Mercado de Datos para el área de Recursos Humanos



Resumen

En el Vicedecanato de Administración de la Facultad de Ciencias y Tecnologías Computacionales, se recopilan todos los datos relacionados con las actividades de los Recursos Humanos, donde se incluye lo referente a la nómina para el pago a los trabajadores que laboran en dicha facultad. La información que se maneja en el área de Recursos Humanos es almacenada mediante un modelo de nómina haciendo uso de hojas de cálculo electrónicas, por lo que se necesita de un sistema de almacenamiento de información. La solución del presente trabajo enmarca su desarrollo en los mercados de datos siendo estos de gran utilidad para realizar el análisis de información en el área anteriormente mencionada. En el proceso de desarrollo se utilizó la metodología realizada por el Departamento de Almacenes de Datos que toma como base la metodología de Kimball y se ajusta a las condiciones y características de producción de la Universidad de la Ciencias Informáticas y del Centro de Tecnología de Gestión de Datos. Apoyándose en las herramientas Visual Paradigm, PostgreSQL, Pentaho Data Integration, DataCleaner, Pentaho Report Designer y Pentaho BI Server. Como resultado se obtiene la estructura del modelo de datos dimensional que comprende: las tablas de dimensiones, hechos y las medidas necesarias para la visualización de los datos. Se definen e implementan los mecanismos de extracción, transformación y carga de los datos, además del desarrollo de la capa de Inteligencia de Negocio para la visualización. Se realizaron pruebas a la aplicación para lograr la aceptación del cliente y verificar su calidad.

Palabras Claves: Recursos Humanos, Mercado de Datos, Inteligencia de Negocios.

Mercado de Datos para el área de Recursos Humanos



Abstract

In the Vice-Dean of Administration from the Faculty of Sciences and Computational Technologies, all the data related to the activities of the Human Resources are collected, which includes the reference to the payroll for the payment to the workers laboring in such faculty. The information that is handled in the area of Human Resources is stored by means of a pre-paid model using spreadsheets, therefore is needed an information storing system. The solution of the present work bases its development in the Data Markets being these very useful to carry out the analysis of information in the area mentioned before. In the development process was used the methodology created by the Department of Data Warehouses that is based on the Kimball methodology and conforms to the conditions and production characteristics of the University of Informatics Sciences and the Center for Data Management Technology. Relying on tools such as Visual Paradigm, PostgreSQL, Pentaho Data Integration, DataCleaner, Pentaho Report Designer and Pentaho BI Server. As a result, the structure of the dimensional data model is obtained which comprises: the tables of dimensions, facts and measures necessary for the visualization of the data. The mechanisms of extraction, transformation and loading of data are defined and implemented, in addition to the development of the Business Intelligence layer for visualization. Tests to the application were made to achieve customer acceptance and verify its quality.

Keywords: Human Resources, Data Market, Business Intelligence.

Índice

Introducción.....	1
Capítulo 1: Fundamentos teóricos.	6
Recursos Humanos.....	6
1.1 Almacenes de Datos.....	6
1.1.1 Definición de los Almacenes de Datos.	6
1.1.2 Características, ventajas y desventajas de los almacenes de datos.....	7
1.2 Mercado de Datos.	7
1.2.1 Justificación de la tecnología de almacenamiento seleccionada.	8
1.3 Modelo Multidimensional	8
1.4 Modo de almacenamiento de datos OLAP.....	10
1.5 Metodologías para el desarrollo de Almacenes de Datos.	11
1.5.1 Metodología a utilizar en la solución.....	11
1.6 Extracción, transformación y carga de los datos.....	12
1.7 Herramientas para el desarrollo del Mercado de Datos para el área de Recursos Humanos.....	13
1.7.1 Herramientas de modelado de datos.....	13
1.7.2 Sistema Gestor de Base de Datos.	14
1.7.3 Administrador de Base de Datos.	14
1.7.4 Herramientas para el perfilado de datos.....	15
1.7.5 Herramientas para la Extracción, Transformación y Carga (ETL).....	15
1.7.6 Herramientas de Inteligencia de Negocio (BI).	16
1.8 Conclusiones	17
Capítulo 2: Análisis y diseño del Mercado de Datos.....	18
2.1 Estudio preliminar del negocio. Caracterización del área de Recursos Humanos.	18
2.2 Necesidades de los usuarios.	19

Mercado de Datos para el área de Recursos Humanos



2.3. Requisitos de información.	19
2.4. Requisitos Funcionales.....	21
2.3 Requisitos Funcionales.....	21
2.5. Requisitos no funcionales.	21
2.6. Reglas del Negocio.	23
2.7. Diagrama de Casos de Uso del Sistema.	24
2.8. Arquitectura del Mercado de Datos.	27
2.9. Diseño del subsistema de almacenamiento, integración y visualización.	28
2.10. Esquema de Seguridad.	36
2.11 Conclusiones del capítulo	36
Capítulo 3: Implementación y prueba del Mercado de Datos.	38
Introducción	38
3.1 Implementación del subsistema de almacenamiento.	38
3.1.1 <i>Estándar de codificación</i>	38
3.2 Implementación del Subsistema de Integración de datos.	39
3.2.1 <i>Implementación de las transformaciones</i>	40
3.2.2 <i>Implementación de las transformaciones</i>	41
3.3 Implementación del Subsistema de Visualización.....	42
3.3.1 <i>Implementación de los cubos OLAP</i>	42
3.3.2 <i>Implementación de la Capa de Visualización</i>	43
3.4 Implementación de los reportes candidatos.	44
3.5 Pruebas.	45
3.6 Herramientas para la aplicación de las pruebas.....	46
3.6.1 <i>Casos de pruebas</i>	47
3.6.2 <i>Listas de Chequeo</i>	47

Mercado de Datos para el área de Recursos Humanos



3.7	Resultados de las pruebas.....	49
3.8	Conclusiones del capítulo.	51
	Conclusiones generales.....	53
	Recomendaciones.	54
	Referencias.....	55
	Bibliografía.....	58
	Anexos.....	61

Índice de tablas

Tabla 1: Descripción de los requisitos funcionales.....	21
Tabla 2: Descripción de requisitos no funcionales.....	21
Tabla 3: Descripción de los actores del negocio.	24
Tabla 4: Descripción de los casos de uso del sistema.	25
Tabla 5: Especificación del CUS: “Presentar información sobre los trabajadores.”	26
Tabla 6: Descripción de los hechos.....	29
Tabla 7: Descripción de las dimensiones.	29
Tabla 8: Matriz bus	29
Tabla 9: Reportes candidatos: “Listado de personas evaluado de adecuado” del A.A Trabajador	35
Tabla 10: Reporte candidato: “Obtener las horas trabajadas de los técnicos” del A.ATécnicos	35
Tabla 11: Describe los Roles y Permisos.....	36
Tabla 12: Estándar de codificación para el MD.	39
Tabla 13: Diseño del CP para el CU:”Presentar información sobre las personas”.....	47
Tabla 14: Resultados de las pruebas de sistema.	50

Índice de Figuras

Fig. 1 Esquema estrella.	9
Fig. 2 Esquema copo de nieve.	9
Fig. 3 Esquema constelación.	9
Fig. 4 Diagramas de Casos de Uso.	25
Fig. 5 Arquitectura del Mercado de Datos.	28
Fig. 6 Modelo de datos.	31
Fig. 7 Distribución por tipo de datos.	32
Fig. 8 Distribución de datos nulos para la tabla facultad 6 año 2015.	33
Fig. 9 Diseño para la carga de los hechos.	33
Fig. 10 Diseño para la carga de las dimensiones.	34
Fig. 11 Diseño de la Arquitectura de información.	35
Fig. 12 Trabajo para los hechos.	41
Fig. 13 Trabajo para las dimensiones.	41
Fig. 14 Trabajo general.	41
Fig. 15 Transformación de la dimensión clave.	42
Fig. 16 Transformación para el hecho técnico.	42
Fig. 17 Implementación de los cubos OLAP.	43
Fig. 18 Arquitectura de información.	44
Fig. 19 Implementación del reporte: Listado de Trabajadores con horas extras.	44
Fig. 20 Implementación del reporte: Trabajadores con Misión Internacionalista.	45
Fig. 21 Vista de Análisis: Información de los trabajadores.	45
Fig. 22 Modelo V.	46
Fig. 23 Comportamiento de los indicadores de la lista de chequeo para el artefacto: Especificación de Casos de Uso.	49
Fig. 24 Resultado de las pruebas.	51

Mercado de Datos para el área de Recursos Humanos



Introducción

Los Recursos Humanos (RR.HH) toman vital importancia para las empresas, especialmente si estas constituyen entidades de servicios y productos fundamentalmente, donde el valor global de la organización está directamente ligado a su capital humano y al activo comercial. Dichas empresas tienen como funciones fundamentales; definir qué personal es el que requiere la organización, reclutar y contratar a las personas más adecuadas para la empresa, conseguir que los trabajadores contribuyan a los objetivos de la empresa. Remunerar y retribuir a las personas por sus aportes y lograr una mayor autorrealización del personal en su trabajo.

Es una época en que la información es la clave para obtener una ventaja competitiva en el mundo de los negocios por lo que a nivel mundial la mayoría de las empresas han automatizado sus procesos o se encuentran involucradas en ello. Por esto, las organizaciones buscan cada día soluciones tecnológicas que les brinden mayor rapidez en el manejo de sus datos. Cuba no ha quedado exenta del avance de las nuevas tecnologías de la información por lo que se han creado entidades encargadas de la producción de aplicaciones informáticas para contribuir con la informatización de todas las áreas de la sociedad. Una de estas instituciones es la Universidad de las Ciencias Informáticas (UCI), la que tiene como misión principal producir aplicaciones y servicios informáticos, a partir de la vinculación estudio-trabajo como modelo de formación y sirviendo de soporte a la industria cubana de la informática

Actualmente la UCI se encuentra inmersa en la informatización de la mayoría de sus procesos, sobre todo los relacionados con la gestión y procesamiento de datos, dígame el proceso de gestión académica para el área de secretaria docente y la gestión de la actividad posgraduada para el área de investigación y posgrado. Para la informatización de sus procesos la UCI cuenta con varios centros productivos entre los que se pueden citar el Centro de Tecnologías de Gestión de Datos (DATEC). Este centro tiene como objetivo crear bienes y servicios informáticos relacionados con la gestión de datos, área del conocimiento que agrupa tanto a los sistemas de información, como a los denominados sistemas de inteligencia empresarial o de negocios.

El centro DATEC en conjunto con la Facultad de Ciencias y Tecnologías Computacionales (FCITEC) se encuentra vinculados en la creación de soluciones informáticas para las diferentes áreas de dicha facultad, que permitan la realización de las tareas de los especialistas en un menor tiempo y con una mayor calidad. En la actualidad la FCITEC cuenta con cuatro vicedecanatos: Extensión y Residencia, Investigación y Postgrado, Formación y Administración, este último es el responsable de la gestión de los RR.HH del área.

Mercado de Datos para el área de Recursos Humanos



Su estructura está compuesta por el vicedecano y especialistas los cuales son encargados de seleccionar al personal que labora en la facultad, proporcionar a los trabajadores los medios necesarios para que puedan ejercer su trabajo y lograr que el trabajador satisfaga sus necesidades, llevan además un registro y control de todas las acciones realizadas por los trabajadores para a su vez obtener el impacto en los procesos que intervienen.

Todo ello se realiza de forma manual por lo que no se cuenta con una base de datos para la manipulación y gestión de los datos generados, ya que la información es almacenada en formato duro o en plantillas predefinidas. La información se introduce manualmente en hojas de cálculo electrónicas (Excel) por los especialistas del área, esto trae consigo que se puedan cometer errores en la gestión de los datos estadísticos de los recursos; dígame composición del personal, categoría docente – científica y grado. Además, no existe forma para la obtención de reportes de los certificados médicos, licencias sin sueldo, evaluaciones de desempeño y sanciones administrativas. Se define una poca confidencialidad en los documentos y su información ya que los mismos pueden ser manipulados por personas no autorizadas para ello, esto ocurre por la ausencia de un mecanismo que establezca niveles de accesibilidad a dichos documentos.

Por la situación anteriormente descrita se identifica como **problema de la investigación** ¿Cómo contribuir al proceso de toma de decisiones a partir de los datos del área de Recursos Humanos de la Facultad de Ciencias y Tecnologías Computacionales?

Donde el **objeto de estudio** está dirigido a los de almacenes de datos, enmarcado en el **campo de acción** Mercado de Datos.

Para dar solución al problema de investigación planteado se define como **objetivo general** desarrollar un mercado de datos para la gestión de los Recursos Humanos en la Facultad de Ciencias y Tecnologías Computacionales, que contribuya a la toma de decisiones.

En correspondencia con el objetivo general, se han desglosado los siguientes **objetivos específicos**:

- ❖ Fundamentar la selección de la metodología, herramientas y tecnologías a utilizar en el desarrollo del mercado de datos.
- ❖ Realizar el análisis y diseño del mercado de datos para el área de Recursos Humanos.
- ❖ Implementar el mercado de datos para el área de Recursos Humanos.
- ❖ Validar el mercado de datos para el área de Recursos Humanos.

Mercado de Datos para el área de Recursos Humanos



Para darle cumplimiento al objetivo general planteado fueron formuladas las siguientes **preguntas científicas**:

1. ¿Qué metodología, herramientas y tecnologías utilizar para lograr la implementación del mercado de datos para el área de Recursos Humanos?
2. ¿Qué características y capacidades debe poseer el mercado de datos para el área de Recursos Humanos?
3. ¿Cómo validar el correcto funcionamiento del mercado de datos para el área de Recursos Humanos?

A partir de las preguntas planteadas se definieron las siguientes **tareas de la investigación**:

- ❖ Caracterización de la metodología, herramientas y tecnologías a utilizar en el desarrollo del Mercado de Datos.
- ❖ Levantamiento de requisitos informacionales, funcionales y no funcionales para identificar las necesidades del cliente.
- ❖ Diseño de los casos de uso del mercado de datos para especificar cada una de las funcionalidades del sistema.
- ❖ Definición de los hechos, las medidas y las dimensiones del mercado de datos para el diseño del modelo de datos.
- ❖ Diseño del modelo de datos para identificar los elementos que forman parte del modelo físico de los datos.
- ❖ Definición de la arquitectura del mercado de datos para definir los subsistemas que forman parte de la solución.
- ❖ Diseño del subsistema de integración para guiar los procesos de carga de los hechos y las dimensiones.
- ❖ Diseño del subsistema de visualización que permita la definición de la capa de presentación realizando el diseño de los cubos OLAP (Procesamiento Analítico en Línea).
- ❖ Diseño de los casos de pruebas para guiar el proceso de validación de la solución una vez implementado el mercado de datos.
- ❖ Implementación del modelo de datos para definir la estructura de la base de datos.
- ❖ Implementación del subsistema de integración para poblar el mercado de datos para los Recursos Humanos.

Mercado de Datos para el área de Recursos Humanos



- ❖ Implementación del subsistema de visualización para mostrar las vistas de análisis contenidas dentro de los libros de trabajo según el área de análisis identificada.
- ❖ Aplicación de los casos de prueba para garantizar el correcto funcionamiento del sistema.

Métodos Teóricos

Analítico-Sintético: Se utilizó para el análisis de documentos, materiales y temas relacionados con el desarrollo del Mercado de Datos. Permitted definir los conceptos fundamentales del tema y se usó para especificar la metodología a utilizar, así como los elementos que conforman un Mercado de Datos. La síntesis permitió realizar un análisis de los resultados del sistema desarrollado como parte de la validación y cumplimiento de los objetivos de la presente investigación.

Modelación: Permitted desarrollar los modelos correspondientes al ciclo de vida de desarrollo del Mercado de Datos.

Métodos empíricos

Recolección de información: El método utilizado para la recolección de la información fue mediante la aplicación de la técnica de entrevistas no estructuradas y análisis de documentos. Combinación que permitió la recogida de información durante el proceso de ingeniería de requisitos y la consulta de materiales y normativas para la realización del presente trabajo.

El documento está estructurado de la siguiente manera: introducción, tres capítulos, conclusiones, recomendaciones, referencias bibliográficas, bibliografía y anexos.

Capítulo 1: Fundamentos teóricos.

En este capítulo se abordan las definiciones y los conceptos más importantes sobre los Recursos Humanos, Mercados de Datos, Almacén de Datos, así como las características, ventajas y desventajas de su utilización. Asimismo, se realiza un estudio bibliográfico acerca de las metodologías, herramientas y tecnologías que se utilizan para el desarrollo de un mercado de datos.

Capítulo 2: Análisis y diseño del Mercado de Datos para los Recursos Humanos.

En este capítulo se realiza un análisis del negocio, con el propósito de comprender los principales aspectos de relevancia para la organización, Se definirán las reglas de negocio y los requisitos de información según las necesidades de los clientes, para luego diseñar el diagrama de caso de uso del sistema. Se identificarán los hechos, medidas y dimensiones, quedando conformado el modelo de datos de

Mercado de Datos para el área de Recursos Humanos



la solución. Además, se diseñarán los Subsistemas de Almacenamiento, Integración y Visualización de los datos.

Capítulo 3: Implementación y prueba del Mercado de Datos para los Recursos Humanos.

En este capítulo se realiza la implementación de cada uno de los subsistemas que conforman el Mercado de Datos: almacenamiento, integración y visualización. Además, contendrá las revisiones a la documentación, las cuales se realizan aplicando las listas de chequeo y las pruebas al mercado de datos, se efectúan usando los casos de prueba realizados a partir de los casos de uso.

Capítulo 1: Fundamentos teóricos.

En este capítulo se realizará el análisis del estado del arte de los almacenes de datos (AD) y los mercados de datos (MD), además del concepto de RR.HH y definiciones, características, ventajas y desventajas de los AD y MD. También se realizará el estudio de las metodologías y herramientas a utilizar en la solución.

Recursos Humanos.

En la administración de empresas, se denomina Recursos Humanos al trabajo que aporta el conjunto de los empleados o colaboradores de esa organización. Pero lo más frecuente es llamar así a la función que se ocupa de seleccionar, contratar, formar, emplear y retener a los colaboradores de la organización. Estas tareas las puede desempeñar una persona o departamento en concreto (los profesionales en Recursos Humanos) junto a los directivos de la organización. El objetivo básico que persigue la función de los RR.HH con estas tareas es alinear las políticas de RR.HH con la estrategia de la organización, lo que permitirá implantar la estrategia a través de las personas **(retopyme, 2016)**.

1.1 Almacenes de Datos.

Hoy en día las empresas necesitan herramientas que le permitan actuar de manera correcta en las operaciones que se realizan y las decisiones que se deben tomar, las mismas deben ser rápidas y basadas en buenos cimientos. El poseer información no significa tener respuestas correctas, debido a que la misma puede no provenir de la misma fuente y con la misma estructura.

Existen disímiles tareas como recolectar, procesar, limpiar y transformar la información, las cuales juegan un papel fundamental en el proceso de toma de decisiones, y se pueden considerar engorrosas si se tiene en cuenta que la empresa cuenta con diferentes áreas. Uno de los mecanismos de la inteligencia de negocio que resuelven el problema de análisis y manejo de la información son los almacenes de datos.

1.1.1 Definición de los Almacenes de Datos.

Existen variados conceptos acerca de los AD. En la actualidad los autores más reconocidos son Ralph Kimball y Bill Inmon, este último considerado como el padre de este término.

Según Ralph Kimball un AD es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis, también se determinó que un AD no es más que la unión de todos los MD de una entidad **(Kimball, 2013)**.

Según Inmon, uno de los primeros autores en escribir sobre el tema de los AD, define los mismos como “una recopilación de datos temáticos, integrados, no volátiles y con historial para la toma de decisiones **(Inmon, W. H. 2005)**.”

En la presente investigación se asume que un AD es una colección de datos provenientes de una o varias fuentes almacenadas de forma histórica durante un amplio período de tiempo, la cual puede ser consultada con el objetivo de ayudar en el proceso de toma de decisiones de la empresa.

1.1.2 Características, ventajas y desventajas de los almacenes de datos.

Según definió el propio Bill Inmon (**Inmon, W. H. 2005**), un AD se caracteriza por ser:

Orientado a temas: Los datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.

Variable en el tiempo: Los cambios producidos en los datos a lo largo del tiempo, quedan registrados de manera que los informes que se generan reflejen dichas variaciones.

No volátil: La información almacenada no puede ser modificada ni eliminada, convirtiéndose en información de solo lectura.

Integrado: Como los datos almacenados provienen de fuentes diferentes deben integrarse en una estructura estable que elimine las inconsistencias existentes en los mismos.

Ventajas de los almacenes de datos.

- ❖ Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- ❖ Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios (**sinexus, 2016**).
- ❖ Integración de información procedente de múltiples sistemas externos.

Desventajas de los almacenes de datos.

- ❖ Puede suponer altos gastos, los consumos de mantenimientos son muy elevados.
- ❖ Existe resistencia al cambio por parte de los usuarios (**DataPrix, 2016**).

1.2 Mercado de Datos.

Ralph Kimball define un MD como “un conjunto flexible de datos, idealmente basado en el dato más atómico posible (granular) para ser extraído de las fuentes operacionales y presentado en un modelo simétrico (dimensional), que es más resistente cuando se enfrentan con las más inesperadas consultas de los usuarios (...)” (**Kimball, 2008**).

Tomando como referencia el concepto de Ralph Kimball se puede caracterizar un MD, como una colección de datos que contiene información de un área de negocio específica, se caracteriza por organizar la información de tal forma que se alcance el más alto nivel de detalle. Permite examinar los datos desde diferentes perspectivas, logrando así un mejor control de la información.

1.2.1 Justificación de la tecnología de almacenamiento seleccionada.

Luego de haber realizado un análisis de los sistemas de almacenamiento de datos mencionados anteriormente, para el desarrollo del presente trabajo se decide utilizar un MD por la necesidad de obtener los datos solamente del área de RR.HH, donde los datos están orientados a satisfacer las necesidades particularmente de ese departamento.

1.3 Modelo Multidimensional

Dadas las características de los MD es ideal la utilización en su diseño de un Modelo Multidimensional (MMD), el cual se sustenta del modelo entidad relación. Este tipo de diseño tiene como ventajas sobre el Modelo Entidad-Relación (MER), que es muy flexible, está desnormalizado y orientado a los intereses de un usuario final, aunque esto no significa que existan inconsistencias en los datos (**Comparativa B.I. Open Source, 2010**). El modelo multidimensional se representa a través de la definición de las tablas de hechos y dimensiones.

Tablas de Hechos: Las tablas de hechos son las tablas primarias en el modelo dimensional. Generalmente, almacenan medidas numéricas, las que representan valores de las dimensiones. La llave de la tabla de hecho, es una llave compuesta, debido a que se forma de la composición de las llaves primarias de las tablas dimensionales a las que está unida. Existen tablas de hechos que no contienen medidas, a estas tablas se les denomina tablas de hechos sin hechos, es decir, la relación entre las dimensiones que definen la llave de esta tabla de hecho implica por sí sola la ocurrencia de un evento (**Tabla de Hechos y Dimensiones, 2009**).

Tablas de Dimensiones: Contienen, generalmente, una llave simple y atributos que la describen. Las tablas de dimensiones especifican la organización lógica de los datos y proporcionan el medio para analizar el contexto del negocio. Cada tabla de dimensión tiene una clave principal que se corresponde con exactamente uno de los componentes de la clave compuesta de la tabla de hechos. En dependencia del esquema de diseño que se asuma pueden contener llaves foráneas de otras tablas de dimensión (**Tabla de Hechos y Dimensiones, 2009**).

Para el modelado de los datos en un AD existen tres variantes de modelación, las cuales se detallan a continuación:

Esquema Estrella: posee una sola tabla de hechos que se relaciona con todas las tablas de dimensiones, como se muestra a continuación en la figura 1:



Fig. 1 Esquema estrella.

Esquema Copo de Nieve: es una ampliación del modelo estrella, las dimensiones son normalizadas en dependencia de la jerarquía existente, pero la estructura de una única tabla de hechos se mantiene. A continuación, un ejemplo de este esquema mostrado en la figura 2:

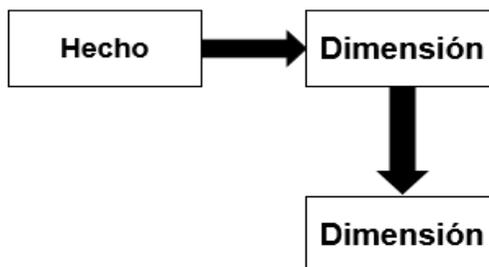


Fig. 2 Esquema copo de nieve.

Esquema constelación de hechos: está compuesto por diversos esquemas de estrella, con la particularidad de que varias tablas de hechos comparten algunas tablas de dimensiones, como se muestra a continuación en la figura 3:

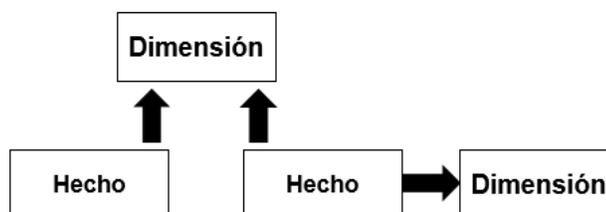


Fig. 3 Esquema constelación.

1.4 Modo de almacenamiento de datos OLAP.

La tecnología de Procesamiento Analítico en Línea (OLAP) permite a los usuarios analizar diferentes dimensiones de datos multidimensionales. Su objetivo es agilizar la consulta de grandes cantidades de datos, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas, utilizadas generalmente para sistemas de ayuda que contribuyan a la toma de decisiones (**BD OLTP y OLAP, 2016**). Existen tres modelos derivados de OLAP: Proceso analítico relacional en línea (ROLAP por sus siglas en inglés), Proceso analítico multidimensional en línea (MOLAP, por sus siglas en inglés) y Proceso analítico híbrido en línea (HOLAP, por sus siglas en inglés); su principal diferencia está dada por la forma de organizar y almacenar los datos.

ROLAP: Los datos son almacenados en filas y columnas de forma relacional. Este modelo presenta la información a los usuarios en forma de dimensiones de negocio. La principal ventaja de esta arquitectura es que permite el análisis de una gran cantidad de datos (**BD OLTP y OLAP, 2016**).

MOLAP: Es una herramienta OLAP que accede a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente (**BD OLTP y OLAP, 2016**).

HOLAP: Permite un análisis híbrido de la información. Lo mejor de las dos tecnologías anteriores. El análisis HOLAP ayuda a reducir costes de hardware ya que se necesita menos espacio en disco que en las Bases de Datos (BD) relacionales. Además, las respuestas de las consultas sobre los BD multidimensionales son más rápidas que sobre las relacionales. Como aspecto negativo, los datos multidimensionales deben ser cargados antes de ser consultados y actualizados cuando se actualizan los datos de la organización (**BD OLTP y OLAP, 2016**).

En la investigación se decide utilizar como modo de almacenamiento ROLAP, debido a las políticas del Centro DATEC y la Universidad en cuanto a la utilización de software libre, pues se utiliza como sistema gestor de BD PostgreSQL, que además de ser libre y poseer una serie de características, las cuales se detallan posteriormente, permite modelar BD relacionales, no así multidimensionales. ROLAP accede directamente a los datos del almacén, soporta técnicas de optimización de accesos, tales como particionado de los datos a nivel de aplicación, soporte a la desnormalización y uniones múltiples, para acelerar las consultas.

1.5 Metodologías para el desarrollo de Almacenes de Datos.

Para guiar el proceso de desarrollo de un software es preciso tener en cuenta la metodología a seguir. La misma se define como la guía que indica qué hacer y cómo actuar cuando se quiere obtener algún tipo de investigación. **(Rodolfo, 2008)** Las metodologías de desarrollo precisan un conjunto de pasos y procesos a seguir que permiten planificar, estructurar y controlar el proceso de desarrollo de software. Desde el surgimiento de las soluciones de AD y BI se han venido desarrollando simultáneamente las metodologías para el desarrollo y la implementación de este tipo de soluciones.

Son muchos los autores que han diseñado o propuesto una metodología para el diseño y construcción de AD, algunos de ellos son Darío Bernabeu con la metodología “HEFESTO”, Inmon, Kimball entre otros. Cada fabricante de *software* de Inteligencia de Negocios busca imponer una metodología con sus productos. Actualmente sobresalen entre la mayoría, los enfoques de Inmon y Kimball.

Inmon en su metodología se basa en un enfoque descendente (top-down) proponiendo primero la creación del AD y a partir de este los MD que se nutrirán de él. Inmon propone la construcción de un repositorio de datos corporativo como fuente de información consistente, consolidada, histórica y de calidad. Como el AD se construye descendentemente los MD se nutren del almacén corporativo, convirtiéndose en un complejo empresarial de BD relacionales **(Hernández, 2013)**.

Por otra parte, Kimball plantea la creación del almacén siguiendo un enfoque ascendente (bottom-up), creando primero los MD independientes por cada área departamental y el Almacén de Datos sería la unión de todos los MD.

1.5.1 Metodología a utilizar en la solución.

Para el desarrollo del MD para el área de RR.HH se define como metodología a utilizar: Propuesta de Metodología para el desarrollo de AD en el DATEC. Esta metodología se ajusta a las condiciones y características de la producción del centro y la universidad, y además toma como base el enfoque de Kimball.

Durante el ciclo de vida de la metodología propuesta de metodología para el desarrollo de AD en DATEC se destacan las siguientes fases de desarrollo:

- ❖ **Estudio Preliminar o Planeación:** En esta fase se realiza un estudio integral de la organización, que se basa en el diagnóstico del negocio, los datos y la infraestructura tecnológica. Además, se definen una serie de aspectos importantes en la gestión de proyecto, tales como el alcance del mismo, los riesgos, la calidad, cronograma, presupuesto, costo y otras actividades.

- ❖ **Requerimientos:** Se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Después se procede a la definición de los requerimientos.
- ❖ **Arquitectura y Diseño:** Se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, definiéndose la arquitectura de información que regirá el desarrollo de la solución.
- ❖ **Implementación:** Se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de Extracción, Transformación y Carga y se configuran e implementan las herramientas de inteligencia de negocios para la obtención de los elementos que se acordaron con el cliente final.
- ❖ **Prueba:** Se realizan las pruebas al sistema desde las Pruebas de Unidad hasta las de Aceptación con el cliente final.
- ❖ **Despliegue:** Se realiza un despliegue piloto en el cual se configuran los servidores y se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona. Posterior a la aceptación del cliente se realiza la carga de los datos así como la Capacitación y Transferencia Tecnológica.
- ❖ **Soporte y Mantenimiento:** Tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.
- ❖ **Gestión y Administración del Proyecto:** A lo largo del ciclo de vida se realizan actividades de control, gestión y chequeo del desarrollo, los gastos, las utilidades, los recursos y demás actividades por parte del Grupo de Dirección del Proyecto **(Hernández, 2013)**.

El MD realizado en la investigación no transitará por todas las fases del ciclo de vida de la metodología a utilizar, no realizará la fase de Despliegue, Soporte y Mantenimiento del proyecto; atendiendo a que estas serán ejecutadas por el personal del Departamento.

1.6 Extracción, transformación y carga de los datos.

Al concluir con la etapa de análisis y diseño, se da paso al proceso de extracción, transformación y carga de los datos (ETL), donde se extraen los datos que proceden de distintas fuentes de información. Estos datos se transforman, pues no siempre se encuentran homogéneos y estandarizados. Por último, se da paso a cargarlos en el depósito de datos o sistema de destino **(Bigatti, 2005)**.

- ❖ **Extracción:** la primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos

distintos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros Excel, pero pueden incluir bases de datos no relacionales u otras estructuras diferentes.

- ❖ **Transformación:** Esta fase es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el AD. Estas acciones se realizan debido a que pueden existir diferentes fuentes de información y es vital conciliar un formato y forma única, definiendo estándares, para que todos los datos que ingresen al AD estén integrados. Puede involucrar la división o la combinación de registros fuentes y la corrección de errores para que los datos queden de la forma deseada.
- ❖ **Carga:** Este proceso es el responsable de cargar la estructura de datos del AD con aquellos datos que han sido transformados y que residen en el almacenamiento intermedio y los datos de los OLTP que tienen correspondencia directa con el depósito de datos. Se debe tener en cuenta que antes de mover los datos al AD es necesario analizarlos con el propósito de asegurar su calidad, ya que es un factor clave que no debe dejarse de lado (**Bernabeu, 2009**).

1.7 Herramientas para el desarrollo del Mercado de Datos para el área de Recursos Humanos.

Para lograr la implementación de los subsistemas de almacenamiento, integración y visualización del MD para el área de RR.HH se utilizan varias herramientas en cada una de las fases que comprenda el proceso. A continuación, se exponen las herramientas definidas por el departamento: Desarrollo de aplicaciones para la implementación de soluciones de AD.

1.7.1 Herramientas de modelado de datos.

El modelado de datos es uno de los elementos más importantes a la hora de iniciar el desarrollo de cualquier sistema. Es la estructura, sobre la que realmente reside la verdadera esencia de la aplicación, incluso determina si el proyecto va a cumplir con su verdadero objetivo. Las herramientas de modelado de datos son de vital importancia para el desarrollo de un software, durante las fases de desarrollo del producto. Las herramientas de ingeniería Asistida por Computadora (*Computer Aided Software Engineering, CASE* por sus siglas en ingles), tiene como objetivo incrementar la productividad y calidad de los productos del software, mejorar la planificación del proyecto, así como reducir el tiempo y costo de su desarrollo la herramienta a utilizar es el Visual Paradigm.

❖ **Visual Paradigm for UML 8.0:**

Es una herramienta CASE profesional que soporta todo el ciclo de vida del desarrollo de software. Además, brinda la posibilidad de modelar todo tipo de diagramas de clases, admite la compatibilidad entre ediciones, la documentación asociada a cada etapa del proceso de desarrollo y generar script para el sistema gestor de base de datos PostgreSQL. Es utilizado por los desarrolladores para facilitar el modelado simultáneo, almacenar los archivos de proyectos y hacer un seguimiento de los cambios (**PARADIGM, 2013**).

1.7.2 Sistema Gestor de Base de Datos.

Un Sistema Gestor de Bases de Datos (SGBD) es un software que permite definir bases de datos, estructurar los datos que serán almacenados y la búsqueda de los mismos. Es una herramienta efectiva que permite el acceso a varios usuarios de manera simultánea. Brindan facilidades y un grupo de funciones con el objetivo de garantizar la confidencialidad, calidad, seguridad e integridad de los datos que contienen, así como el acceso fácil y relativamente rápido de los mismos (**PostgreSQL, 2016**).

❖ **PostgreSQL 9.5:**

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código. Utiliza un modelo cliente/servidor y usa multiprocesos en vez de multihilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando (**OBE y otros, 2015**).

1.7.3 Administrador de Base de Datos.

PgAdmin III es una herramienta de código abierto se encuentra disponible en más de una docena de lenguajes y para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris. Soporta versiones de servidores 7.3 y superiores.

❖ **pgAdmin III v1.20.0:**

Está diseñado para responder a las necesidades de todos los usuarios, ya que facilita la gestión y administración de las bases de datos, ya sea mediante instrucciones SQL o con ayuda de un entorno gráfico. Es una herramienta de código abierto para la administración de bases de datos PostgreSQL y derivados (EnterpriseDB, PostgresPlus, Advanced Server y GreenplumDatabase. Responde a las necesidades de la mayoría de los usuarios, desde escribir simples consultas SQL hasta desarrollar bases

de datos complejas. La interfaz gráfica soporta todas las características de PostgreSQL y hace simple la administración (MICHAUD, 2015).

1.7.4 Herramientas para el perfilado de datos.

DataCleaner 1.5.4:

El perfilado de datos constituye una acción que permite verificar la calidad de los datos, entre otras cosas, permite: suministrar datos precisos y completos que conllevan a mejores resultados a un alto rendimiento aumentando la confianza del usuario, al garantizar que pasen más tiempo analizando los datos y menos tiempo comprobando la calidad de las fuentes de información; aumenta el conocimiento de los datos y permite conocer dónde se están produciendo los errores para corregirlos y construir mejores sistemas de información.

DataCleaner es un motor de procesamiento que fue construido para tareas altamente interactivas, de rendimiento y flexibilidad. Es una herramienta de análisis de calidad de la información que le permite realizar perfiles de datos y validación. El monitoreo es un aspecto central del DataCleaner establece el punto de partida, los objetivos, y para asegurar un proceso de seguimiento de las cuestiones de calidad de datos. Además, encuentra las pautas, valores perdidos, juegos de caracteres y otras características de los valores de la información (Basallo, 2009).

1.7.5 Herramientas para la Extracción, Transformación y Carga (ETL).

Los procesos de ETL son cruciales en la integración de datos, permite a las organizaciones mover datos desde distintas fuentes, reformatearlos, limpiarlos, y cargarlos en otra base de datos. Es una práctica común normalizar los datos antes de combinarlos en el Almacén de Datos mediante herramientas con ese fin.

- ❖ **Pentaho Data Integration v6.0 (PDI):** Permite realizar transformaciones y trabajos de una forma muy sencilla e intuitiva. Es aplicable a diversos tipos de BD como son SQL server, PostgreSQL, MySQL, Microsoft Access. A parte de ser de código abierto y sin costes de licencia, las características básicas de esta herramienta son (Gravitar, 2014):

- ✓ Entorno gráfico de desarrollo.
- ✓ Uso de tecnologías estándar: Java, XML, JavaScript.
- ✓ Fácil de utilizar.
- ✓ Multiplataforma: Windows, Macintosh, Linux.

- ✓ Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones).

1.7.6 Herramientas de Inteligencia de Negocio (BI).

Las herramientas de software de inteligencia de negocios son usadas para acceder a los datos de los negocios y proporcionar reportes, análisis, visualizaciones a los usuarios que lo usarán. La gran mayoría de las herramientas de son usadas inteligencia de negocios por usuarios finales para acceder, analizar y reportar contra los datos que más frecuentemente residen en el Mercado de Datos.

Pentaho Schema Workbench 3.10.0: Herramienta e interfaz de diseño visual que permite crear esquemas de cubos OLAP (en inglés On-Line Analytical Processing). Permite modelar un XML (en inglés Extensible Markup Language) con el diseño del cubo a través de opciones lógicas e intuitivas que no requieren de un manejo avanzado de este formato de archivo y crear, editar, actualizar y publicar esquemas OLAP para que estos sean desplegados por aplicaciones de visualización (**Pentaho, 2015**).

Pentaho Report Designer 6.0: Es una herramienta independiente que forma parte de la unidad de reportes de Pentaho (Pentaho Reporting), que simplifica el proceso de generación de reportes, permitiendo a los diseñadores de reportes crear rápidamente informes sofisticados y ricos visualmente basados en el proyecto de reportes de Pentaho JFreeReport. Posee un diseñador gráfico basado en “arrastrar y soltar” (drag & drop) que provee completo control de acceso a los datos, agrupaciones, cálculos, gráficas, formato para reportes de alta resolución. También está compuesto por un Asistente paso a paso integrado que guía a los diseñadores de reportes durante el proceso de diseño. Plantillas de reportes aceleran el proceso de generación, proporcionando un aspecto consistente y atractivo además de poseer opciones de salida flexibles que incluyen los populares formatos Adobe PDF, HTML, Microsoft Excel, entre otros (**PentahoCorporation, 2005**).

Pentaho BI Server 6.0: Permite elaborar reportes de manera dinámica según las necesidades de usuarios finales. Provee el soporte y la infraestructura necesarios para crear soluciones de BI ante problemas de negocio. Presta servicios de autenticación, registro, auditoría, servicios web y motor de reglas. Además incluye un motor de solución que integra reportes, análisis, tableros de comandos y componentes de minería de datos (**EASY 2005**).

1.8 Conclusiones

En el capítulo que finaliza se trataron los conceptos asociados al MD, estableciéndose las herramientas y tecnologías que posibilitarán un correcto desarrollo del Mercado de Datos para la Facultad de Ciencias y Tecnologías Computacionales seleccionándose como metodología de desarrollo de software Metodología de desarrollo de Almacenes de Datos, la cual guiará todo el proceso de creación del sistema propuesto, ya que cubre las etapas por las que transita el desarrollo de un AD. Para el desarrollo del MD se definieron varias herramientas como Visual Paradigm 8.0 como herramienta CASE. Para el tratamiento de las bases de datos PostgreSQL 9.5 y pgAdmin III v1.20.0. Como herramientas para lograr la calidad de los datos DataCleaner1.5.4 y del sistema Pentaho Data Integration, Pentaho Schema Workbench 3.10.0.1 y Pentaho Report Designer 6.0, cada una de estas herramientas y tecnologías son idóneas para el trabajo con ellas durante el proceso de desarrollo.

Capítulo 2: Análisis y diseño del Mercado de Datos para los Recursos Humanos.

Introducción

En este capítulo se realiza un análisis preliminar del negocio, con el fin de identificar las perspectivas e indicadores a seguir durante el desarrollo del MD para el área RR.HH con el propósito de comprender los principales aspectos de relevancia para la organización. Teniendo en cuenta entre otros elementos el levantamiento de requisitos, las reglas del negocio, los casos de uso y la arquitectura que presenta el sistema. Además, se efectúa el diseño de los subsistemas de integración, almacenamiento y visualización, con las características necesarias para satisfacer las necesidades manifestadas por el cliente.

2.1 Estudio preliminar del negocio. Caracterización del área de Recursos Humanos.

Actualmente la FCITEC cuenta con cuatro vicedecanatos: Extensión y Residencia, Investigación y Postgrado, Formación y Administración, este último es el responsable de la gestión de los Recursos Humanos de la facultad. Su estructura está compuesta por el vicedecano y especialistas los cuales son encargados de proporcionarles a los trabajadores los medios necesarios para que puedan ejercer su trabajo eficientemente. Dichos especialistas llevan un registro y control de todas las acciones realizadas por los trabajadores para a su vez obtener el impacto en los procesos que intervienen. Estos crean una nómina que contiene cincuenta claves por las que los trabajadores se ausentaron, para luego generar un análisis que se realiza todos los meses que trata sobre el pago que reciben todos los trabajadores cada mes. Donde las claves más utilizadas son las siguientes:

- 7 misión internacionalista
- 12 vacaciones
- 26 enfermedad 3 días o menos
- 25 licencia sin sueldo pos natal (hasta 9 meses)
- 21 ausencia justificada
- 20 semana receso docente
- 23 licencia por maternidad
- 22 ausencia injustificada
- 32 licencia sin sueldo (problemas personales)
- 27 enfermedad más de 3 días
- 50 no pagar

- 35 prestación social
- 18 cónyugue acompañante
- 10 prestación de servicios (educ, pcc, ujc, etc)

2.2 Necesidades de los usuarios.

Con la presente investigación se plantea facilitar la toma de decisiones en el área RR.HH. Por tal motivo, es de vital importancia identificar las necesidades de información que poseen los especialistas de dicha área, pues constituyen la base para el correcto diseño del MD.

De esta forma se clasificó la información en dos Áreas de Análisis fundamentales: Área de Análisis Trabajador y Área de Análisis Técnico, las cuales engloban varias temáticas donde cada una contiene su Libro de Trabajo.

2.3 Requisitos de información.

Los requisitos de información representan las necesidades de los usuarios. Dichos requisitos fueron aprobados por el especialista del área RR.HH, donde se detallan brevemente cada uno de ellos:

RI 1. Obtener el listado de técnicos evaluados de deficiente por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 2. Obtener el listado de técnicos evaluados de adecuado por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 3. Obtener el listado de técnicos evaluados de superior por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 4. Obtener el listado de trabajadores evaluados de deficiente por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 5. Obtener el listado de trabajadores evaluados de adecuado por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 6. Obtener el listado de trabajadores evaluados de superior por ubicación, tiempo, sexo, requisito, evaluación cargo.

RI 7. Obtener la cantidad de horas extras de los trabajadores por ubicación, persona, tiempo, cargo, sexo, requisito, evaluación.

RI 8. Obtener la cantidad horas trabajadas de los técnicos por ubicación, persona, tiempo, cargo, sexo, requisito, evaluación, turno.

RI 9. Obtener el listado de personas que tienen afectada la clave ausencias injustificadas por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 10. Obtener el listado de personas que tienen afectada la clave ausencias justificadas por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 11. Obtener el listado de personas que tienen afectada la clave vacaciones por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 12. Obtener el listado de personas que tienen afectada la clave enfermedad 3 días o menos por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 13. Obtener el listado de personas que tienen afectada la clave licencia sin sueldo pos natal (hasta 9 meses) por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 14. Obtener el listado de personas que tienen afectada la clave semana receso docente por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 15. Obtener el listado de personas que tienen afectada la clave licencia por maternidad por ubicación, tiempo, persona, cargo, sexo, requisito, evaluación.

RI 16. Obtener el listado de personas que tienen afectada la clave licencia sin sueldo (problemas personales) por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 17. Obtener el listado de personas que tienen afectada la clave enfermedad más de 3 días por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 18. Obtener el listado de personas que tienen afectada la clave no pagar por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 19. Obtener el listado de personas que tienen afectada la clave prestación social por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 20. Obtener el listado de personas que tienen afectada la clave cónyugue acompañante por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 21. Obtener el listado que tienen afectada la clave prestación de servicios (educ, pcc, ujc, etc) por ubicación, tiempo, cargo, sexo, requisito, evaluación.

2.4 Requisitos Funcionales.

Los requisitos funcionales (RF) son capacidades o condiciones que el sistema debe cumplir. De forma general estos requisitos indican lo que debe hacer el sistema **(29)**.

Los requisitos funcionales describen la función que el sistema debe hacer, de acuerdo con las necesidades y especificaciones del cliente. En la tabla 1 se muestran cada uno de estos requisitos:

Tabla 1: Descripción de los requisitos funcionales.

No	2.3 Requisitos Funcionales	Descripción
1	Extraer datos de la fuente	Permite extraer los datos de la fuente.
2	Realizarla transformación y carga	Permite realizar la transformación de los datos de la fuente.

2.5 Requisitos no funcionales.

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener. Estas propiedades constituyen las características que hacen al producto atractivo, usable, rápido o confiable **(Pressman, 2002)**. Los requisitos no funcionales, representan aquellas características del sistema que le reportan al cliente alguna utilidad, como la usabilidad, la confiabilidad, entre otros aspectos, dándole más confianza y seguridad a la aplicación.

A continuación, en la tabla 2 se describen los requisitos no funcionales definidos para el MD para el área de Recursos Humanos:

Tabla 2: Descripción de requisitos no funcionales.

No	Clasificación	Requisitos no funcionales	Descripción
1	Usabilidad	Facilitar el acceso a los reportes del Almacén de Datos mediante	El usuario podrá encontrar de manera rápida la información que

		la distribución de la información por áreas de análisis.	solicita en el área correspondiente de acuerdo al objetivo de su solicitud.
2	Confiabilidad	Garantizar la persistencia de la información.	Se debe realizar un respaldo total de los datos del MD con una frecuencia mensual.
3	Restricciones de diseño	Lograr la homogeneidad de la estructura de los elementos definidos en el MD.	Las estructuras del MD deben tener un nombre estándar teniendo en cuenta el tipo de estructura que sea.
4	Interfaz	Mostrar todos los textos que aparezcan en la interfaz del sistema en idioma español.	Los mensajes de procesamiento y error, así como los nombres de las vistas y la arquitectura de información deben ser en idioma español y tener un estándar para todo el sistema.
5	Software	Acceso al sistema	El usuario deberá acceder a la aplicación mediante el protocolo HTTP, usando preferiblemente el navegador web Firefox 34.0 o superior.
6	Software	Instalar en las estaciones de trabajo el software necesario para el correcto funcionamiento del sistema.	Las configuraciones de software de las máquinas clientes deben contar con: <ul style="list-style-type: none"> • Firefox 34.0 o superior • Adobe Reader • 2 GB RAM
7	Hardware	Proporcionar características mínimas de hardware a las PC clientes.	Los PC clientes deben contar con los siguientes requerimientos de hardware: <ul style="list-style-type: none"> • 2 GB RAM • Procesador Dual Core 2.0 Hz

8	Hardware	Proporcionar características mínimas de hardware a los servidores.	Los servidores deben contar con los siguientes requerimientos de hardware: <ul style="list-style-type: none"> • 4 GB RAM • Procesador AMD A8 • Almacenamiento 1 TB
---	----------	--	---

2.6 Reglas del Negocio.

Las reglas de negocios (o las directivas empresariales) definen y controlan la estructura, el funcionamiento y la estrategia de una organización, son la acción, práctica o procedimiento dentro de una determinada actividad o esfera (AUQUI, 2013). Son además las condiciones que regulan un proceso de negocio, las cuales se deben seguir para cumplir con las necesidades de información de los clientes. A continuación, se muestran las reglas del negocio identificadas:

Reglas de Variables.

- RN1: El análisis de la prenomina es realizado por el especialista del área de RR.HH.
- RN2: Los campos vacíos que se muestran en las tablas significan que los trabajadores no tuvieron implicación en ninguna clave.
- RN3: cantidad_horas_extras= sumatoria de todas las horas extras de los trabajadores de la facultad.
- RN4: cantidad_horas_trabajadas= sumatoria de todas las horas trabajadas de los técnicos de la facultad.
- RN5: Días de incidencia= Todas las claves afectadas de los trabajadores y de los técnicos.

Reglas de Transformación.

- RN6: El campo sexo se tratará como (F) femenino, (M) como masculino y para el caso en que aparezca un campo con valor (D) como desconocido.
- RN7: El campo evaluación se tratará como (S) superior, (A) como adecuado, (D) como desconocido y (NE) como no evaluado y para el caso en que aparezca un campo vacío tomará valor (D) como desconocido.
- RN8: El campo requisito se tratará como (S) si y (N) como no y para el caso en que aparezca un campo con valor (D) como desconocido.

2.7 Diagrama de Casos de Uso del Sistema.

El diagrama de casos de uso del sistema (CUS) es una forma de representar gráficamente la relación existente entre los actores y casos de uso del sistema, de manera que sirva de guía para reflejar las metas y funcionalidades que persigue el negocio (**Pressman, 2002**).

Los CUS fueron identificados a partir del agrupamiento de los requisitos de información y requisitos funcionales, para un total de dos actores que se relacionan con ellos, según sus responsabilidades. En el mismo se encuentra reflejado el patrón concordancia en adición en el caso de uso extendido: “Realizar operaciones sobre reportes” ya que se utiliza para modelar una parte del caso de uso que puede considerarse como un comportamiento opcional. En la tabla 3, se presenta la descripción de las responsabilidades de cada uno de los actores en el sistema.

Tabla 3: Descripción de los actores del negocio.

Actores	Descripción
Especialista	Es el responsable de analizar la información de los diferentes indicadores del área RR.HH.
Administrador de ETL	Es el responsable de la extracción, transformación y carga de los datos.

Diagrama de Casos de Uso del Sistema.

En la figura 4 se muestra el diagrama de casos de uso del sistema.

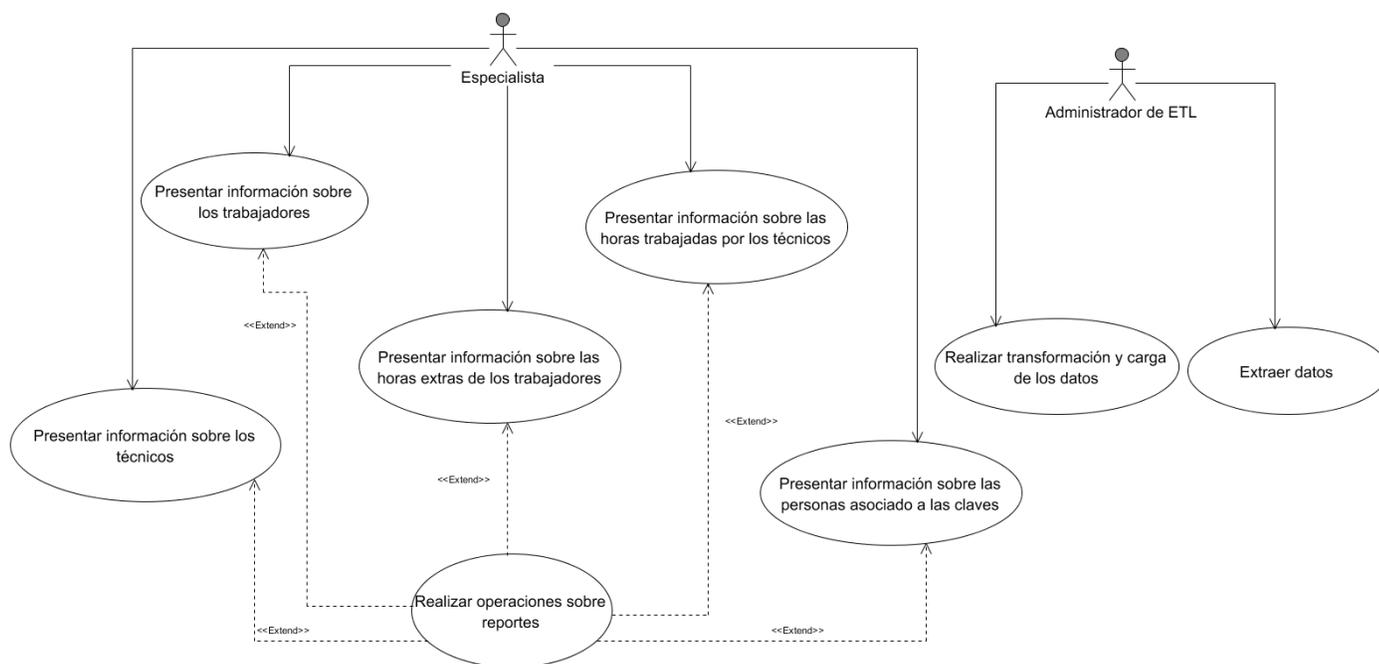


Fig. 4 Diagramas de Casos de Uso.

Descripción de Casos de Uso del Sistema.

En la tabla 4 que se muestra a continuación se describen los CUS, estos son parte del análisis y ayudan a describir qué es lo que el sistema debe hacer y cómo este interactúa con el usuario. Los CUS son fragmentos de funcionalidad que ofrece el sistema para aportar un resultado de valor para sus actores. Recoge una descripción de cada uno de los casos de uso del sistema para una mejor comprensión de ellos.

Tabla 4: Descripción de los casos de uso del sistema.

No	Caso de uso del sistema	Descripción
1	Presentar información sobre los trabajadores	Visualiza los reportes de los indicadores seleccionados sobre información referente a los trabajadores.
2	Presentar información sobre los trabajadores asociados a las claves.	Visualiza los reportes sobre las claves que tienen afectadas los trabajadores.
3	Realizar la transformación y carga de los datos.	Realiza la transformación y carga de los datos necesarios para la construcción del MD.
4	Extraer datos.	Realiza la extracción de los datos necesarios de los ficheros .xls.
5	Presentar información sobre los técnicos.	Visualiza reportes sobre información de los técnicos.

6	Presentar información sobre las horas extras de los trabajadores.	Visualizar reportes sobre la medida horas extras para los trabajadores.
7	Presentar información sobre las horas trabajadas de los técnicos.	Visualiza reportes sobre las horas trabajadas de los técnicos.
8	Realizar operaciones sobre reportes.	Visualiza las operaciones de todos los reportes.

Especificación de casos de uso.

La tabla 5 representa la especificación del CUS “Presentar información sobre los trabajadores”. El resto de los casos de uso se encuentran detallados en el artefacto “Especificación de Casos de Uso: para el área de Recursos Humanos.

Tabla 5: Especificación del CUS: “Presentar información sobre los trabajadores.”

Objetivo	Extraer información sobre los trabajadores.	
Actores	Especialista	
Resumen	El CU inicia cuando el especialista desea consultar la información referente a los trabajadores. El CU finaliza una vez que los datos seleccionados por el especialista son mostrados.	
Complejidad	Media.	
Prioridad	Crítico.	
Precondiciones	<ul style="list-style-type: none"> • El especialista tiene que estar autenticado. • Disponibilidad de las fuentes. • Los reportes relacionados con los trabajadores fueron creados. 	
Post condiciones	Los reportes correspondientes fueron consultados por el especialista.	
Flujo de eventos		
Flujo básico		
	Actor	Sistema
1	Selecciona el A.A Trabajador.	
2		Muestra los libros de trabajo del área seleccionada.
4	Selecciona el libro de trabajo.	
5		Muestra los reportes contenidos en la libro de trabajo.

6	Selecciona el reporte que desea analizar.	
		Muestra en una interfaz el listado correspondiente con la opción marcada.
5		Finalizando el caso de uso.
Prototipo		
Flujos alternos		
1. No existen personas con la información seleccionada		
	Actor	Sistema
6.1		Muestra el mensaje "No existen personas con la información seleccionada".
Prototipo		
Relaciones	CU Incluidos	No aplica.
	CU Extendidos	No aplica.
Requisitos no funcionales	Documento: "Especificación de requisitos de software".	
Asuntos pendientes		

2.8 Arquitectura del Mercado de Datos.

El MD para el área de Recursos Humanos está ordenado mediante una arquitectura compuesta por el subsistema de integración, subsistema de almacenamiento y subsistema de visualización. A continuación, una descripción de cada uno de estos Subsistemas.

- En el subsistema de integración de datos es donde se realizan todos los procesos ETL en los cuales se extrae, se limpia e integra toda la información almacenada en los sistemas fuentes a través de transformaciones.
- En el subsistema de almacenamiento es donde se guarda toda la información que ha sido transformada en el subsistema de integración.

- En el subsistema de visualización de la información es donde se muestra toda la información almacenada al cliente, a través de reportes operacionales y vistas de análisis. Los mismos permiten al cliente realizar un análisis de toda la información procesada.

En la siguiente figura 5 se muestra la arquitectura diseñada para el MD:

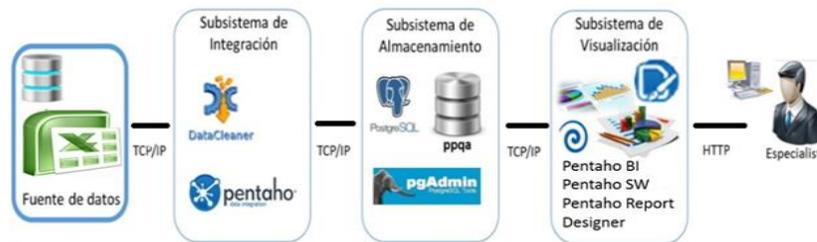


Fig. 5 Arquitectura del Mercado de Datos.

La fuente de datos está compuesta por ficheros Excel que contienen los datos de los RR.HH de toda la facultad. Primeramente, se encuentra el Subsistema de integración que obtiene la información mediante proceso de integración de los datos para su almacenamiento posterior. Luego el subsistema de almacenamiento recibe la información integrada y la almacena en la BD destino, la cual es soportada por el SGBD PostgreSQL 9.5 y es administrada por los usuarios que tienen acceso a la información a través del pgAdminIII 1.20.0. Finalmente, en el subsistema de visualización se muestran los reportes mediante la herramienta Pentaho Report Designer 6.0, el Pentaho Schema Workbench 6.0 donde se crean los cubos OLAP y el Pentaho Business Intelligence 6.0 muestra la información a través de vistas de análisis de tal forma que es accedida por los usuarios autorizados a realizar los reportes.

2.9 Diseño del subsistema de almacenamiento, integración y visualización.

En el diseño del subsistema de almacenamiento se identifican las tablas dimensiones y las tablas hechos, dando lugar a la creación del matriz bus y el modelo de datos. Posteriormente, en el diseño del subsistema de integración se realiza el perfilado de los datos y el diseño de los procesos de integración, sentando las bases para una correcta implementación de dichos procesos. Finalmente, en el diseño del subsistema de visualización se crean los reportes quedando implementada la propuesta de solución.

Diseño del subsistema de almacenamiento.

Luego de un profundo estudio realizado sobre la información que se recoge en el área RR.HH, se identificaron 21 requerimientos de información obtenidos a partir de las necesidades del cliente, fueron

creadas 2 tabla de hechos, 9 tablas dimensiones y 4 medidas. A continuación, en la tabla 6 y 7 se describen los hechos y dimensiones identificadas:

Tabla 6: Descripción de los hechos.

No	Hechos	Descripción
1	hech_trabajador	Contiene la medida correspondiente a la cantidad de horas extras y días de incidencia.
2	hech_técnico	Contiene la medida horas trabajadas y días de incidencia.

Tabla 7: Descripción de las dimensiones.

No	Dimensiones	Descripción
1	dim_persona	Contiene información importante sobre las personas.
2	dim_cargo	Representa el cargo de las personas.
3	dimsexo	Representa el sexo de las personas.
4	dim_clave	Representa y contiene todas las claves de ausencias que se utilizan en la prenomina.
5	dim_requisitos	Representa si cumple con los requisitos si(S) o no(N).
6	dim_ubicacion	Representa y contiene la entidad en la que se labora o se laborara.
7	dim_tiempo	Contiene el día, el mes y el año en el que se realiza la prenomina.
8	dim_evaluacion	Representa todas las evaluaciones: Superior, Adecuado, No evaluado y Deficiente.
9	dim_turno	Representa los turnos de los técnicos, 1er turno, 2do turno.

Matriz bus.

La matriz bus representa la relación existente entre las dimensiones y los hechos del MD para el área de RR.HH. Constituye una validación del análisis realizado evitando que exista solapamiento entre los hechos. Las columnas de la matriz representan los hechos identificados en el MD y las filas las dimensiones utilizadas. Las celdas marcadas con una X indican que la fila de dimensión está relacionada con la columna del hecho. A continuación en la tabla 8, se presenta la matriz correspondiente al MD para el área de RR.HH:

Tabla 8: Matriz bus

Matriz BUS

Dimensiones	Hechos	
	H1	H2
D1	X	X
D2	X	X
D3	X	X
D4	X	X
D5	X	X
D6	X	X
D7	X	X
D8	X	X
D9		X

Luego de obtener la matriz bus y evaluar los resultados arrojados por la misma, se procede a la estructuración del modelo de datos.

Justificación del modelo utilizado en la propuesta de solución.

El modelo propuesto para el desarrollo del MD para el área de RR.HH se basa en un esquema de constelación de hechos, siendo este el más adecuado a las necesidades del área RR.HH, ya que existen dimensiones compartidas para distintas tablas de hechos. En la figura 6 se ilustra el modelo propuesto reducido. En el se encuentra el modelo de datos ampliado.

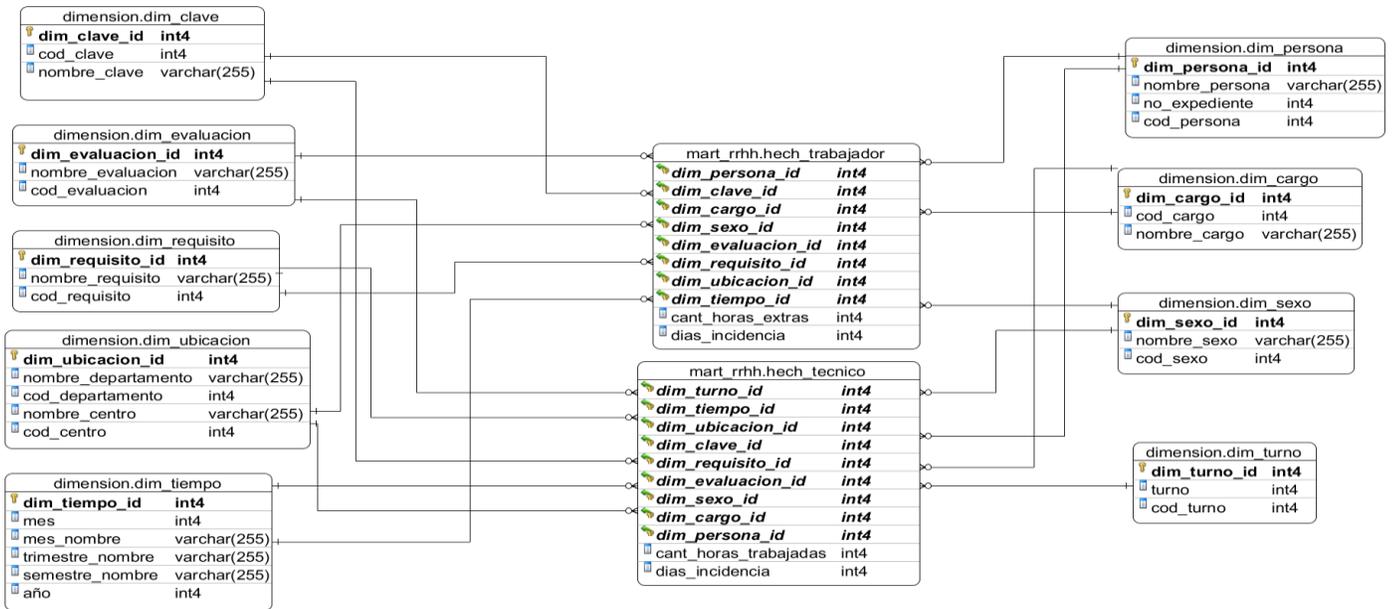


Fig. 6 Modelo de datos.

2.9.2 Diseño del Subsistema de Integración.

Uno de los elementos más importantes del MD para el área de RR.HH es el diseño del subsistema de integración del cual forma parte el **perfilado de datos**, que permite realizar un análisis profundo de los datos provenientes de la fuente para conocer el estado en que se encuentra la información así como su calidad y estructura (Basallo, 2009).

- ❖ **Replicación de datos:** Consiste en crear y mantener copias de las bases de datos, donde generalmente, un servidor contiene la copia primaria de la base de datos y otros mantienen las copias esclavas de la misma.
- ❖ **Integración de información empresarial:** Consiste en crear un intermediario que sirva como canal de consulta y representación de la información recuperada y que contenga los directorios de la base de datos.
- ❖ **Integración de aplicaciones empresariales:** Consiste en la integración de varias aplicaciones con tecnologías incompatibles y gestionadas independientemente. Esta permite que dichas aplicaciones se comuniquen e intercambien información entre sí.
- ❖ **ETL:** Esta consiste en la extracción de datos de sistemas fuentes, transformarlos según las necesidades de los usuarios, y cargarlos en un sistema destino, donde tanto la fuente origen como el destino son bases de datos o archivos.

Para este proceso de integración de datos es utilizada la última estrategia

Perfilado de datos.

El perfilado de datos consiste en analizar las fuentes de datos para identificar los tipos de datos presentes y la calidad de los mismos. Permite verificar los distintos tipos de datos que contiene la fuente y obtener estadísticas sobre valores incorrectos, duplicados y nulos que se encontraron para poder corregirlos. La figura 7 muestra la distribución de los datos y la 8 muestra la distribución de los datos nulos:

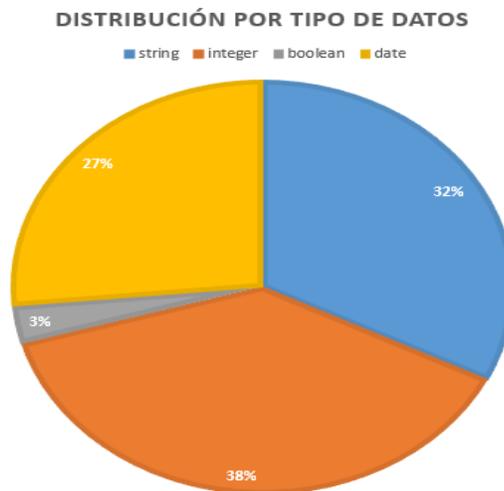


Fig. 7 Distribución por tipo de datos.

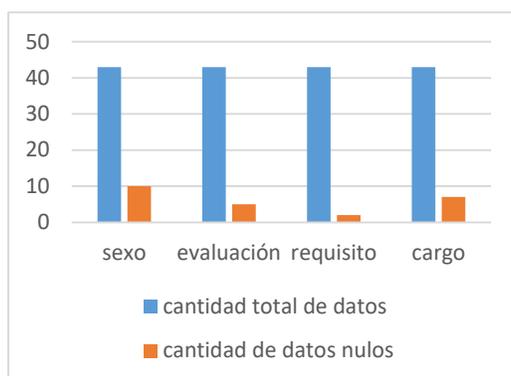


Fig. 8 Distribución de datos nulos para la tabla facultad 6 año 2015.

Diseño de las transformaciones.

Las transformaciones son el elemento fundamental del proceso de ETL, en el diseño de las transformaciones se detalla cada uno de los pasos a seguir para efectuar la carga de las dimensiones y los hechos en el MD para el área de RR.HH. Las figuras 9 y 10 a continuación muestran lo detallado de forma general del diseño de las transformaciones para la carga de las dimensiones y la carga de los hechos del MD para el área de RR.HH.

Diseño para la carga de hecho.

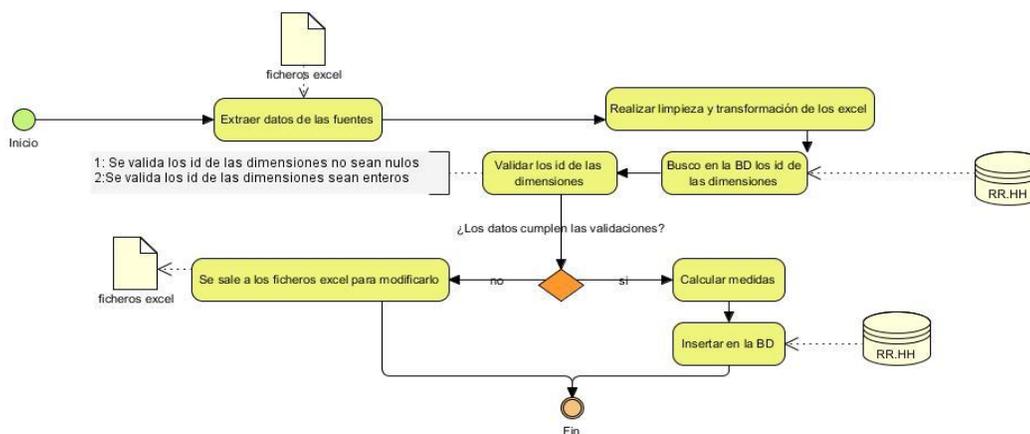


Fig. 9 Diseño para la carga de los hechos.

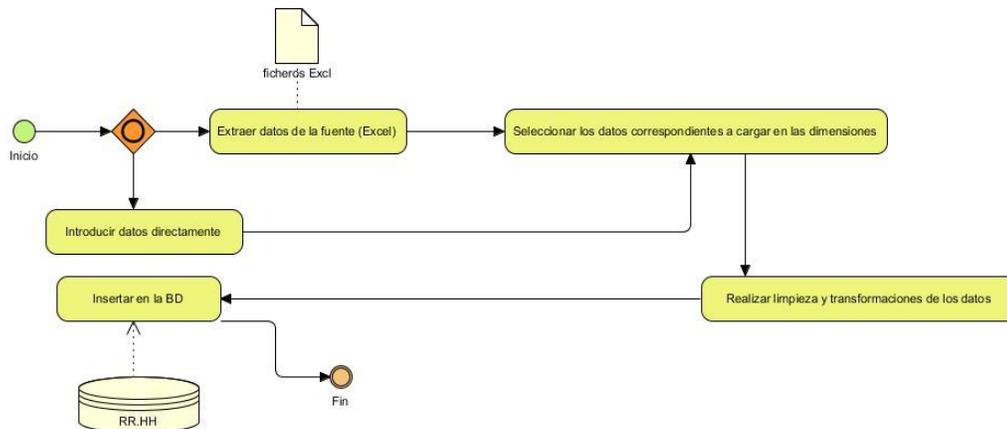


Fig. 10 Diseño para la carga de las dimensiones.

2.9.3 Diseño del Subsistema de Visualización.

El diseño del subsistema de visualización se realiza con el objetivo de organizar las vistas de análisis por áreas de análisis, facilitando al usuario una búsqueda rápida de la información.

Arquitectura de la Información

Con la creación del MD para el área de RR.HH se podrá almacenar la información que pertenecientes a dicha área, para ser analizada por las personas capacitadas y sirviendo de apoyo al proceso de toma de decisiones. En la investigación se identificaron 2 áreas de análisis que se divide en 6 libros de trabajo los que incluyen las tablas de salida definidas por el cliente.

Áreas de Análisis (AA): Son las secciones en las que se divide el MD para el área de RR.HH, las que contienen los diferentes libros de trabajo y las vistas de análisis.

Libro de Trabajo (LT): Son las diferentes categorías en las que se pueden agrupar cada una de las vistas de análisis.

- **Reportes:** Informe que expresa mediante tablas los valores de los indicadores analizados.

A continuación, en la figura 11 se muestran las áreas de análisis definidas:

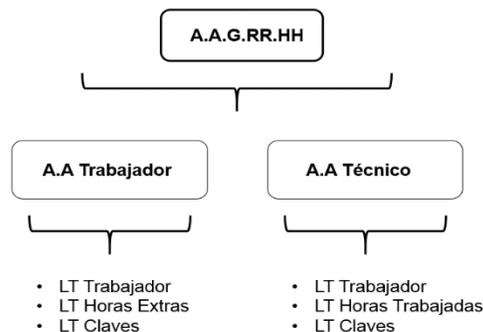


Fig. 11 Diseño de la Arquitectura de información.

Diseño de reportes candidatos.

Los reportes candidatos responden a las necesidades del cliente y a las salidas que debe mostrar el sistema. A continuación, se muestran en la tabla 9 y 10 se describen algunos reportes candidatos de algunas de las áreas definidas:

Tabla 9: Reportes candidatos: “Listado de personas evaluado de adecuado” del A.A Trabajador

Área de análisis (AA)	Técnicos
Libro de Trabajo (LT)	LT Horas trabajadas.
Reporte (Tabla de Salida – TS)	Horas trabajadas de los técnicos.
Descripción	Representa las horas de trabajo de los técnicos de la facultad.
Elementos del reporte	Turno, nombre, número del solapín.
Frecuencia de emisión	Mensual.
Funciones	Listar.

Tabla 10: Reporte candidato: “Obtener las horas trabajadas de los técnicos” del A.ATécnicos

Área de análisis (AA)	Técnicos
Libro de Trabajo (LT)	LT Horas trabajadas.
Reporte (Tabla de Salida – TS)	Horas trabajadas de los técnicos.

Descripción	Representa las horas de trabajo de los técnicos de la facultad.
Elementos del reporte	Turno, nombre, número del solapín.
Frecuencia de emisión	Mensual.
Funciones	Listar.

2.10 Esquema de Seguridad.

En el MD para el área de RR.HH es de gran importancia la seguridad de la información, pues los datos que maneja son de vital importancia para la facultad. Con este objetivo se definieron roles para darle permisos cada uno de los usuarios que interactúan directamente con el sistema. En la tabla 11 se muestra la descripción de los roles y permisos que se darán.

Tabla 11: Describe los Roles y Permisos

Roles	Aplicación		Base Datos	
	Lectura	Escritura	Lectura	Escritura
Administrador ETL			X	X
Especialista	X			

Conclusiones del capítulo

En este capítulo del Mercado de Datos para el área de RR.HH de la facultad FCITEC finalizada la etapa de análisis y diseño se puede concluir que:

- ❖ Se identificaron 21 RI y 2 RF agrupados en siete CUI y CU respectivamente, sirviendo de base para elaborar el diagrama de Casos de Uso del Sistema. De igual manera se identificaron ocho RN permitiendo definir las reglas de transformación que serán utilizadas en los procesos de integración.
- ❖ El modelo dimensional diseñado representa las relaciones entre las dos tablas de hechos, las nueve dimensiones y las cuatros medidas identificadas para el MD.
- ❖ Se diseñaron dos transformaciones, una para la carga de los hechos y otra para la carga de las dimensiones, con el fin de poblar la BD.

Mercado de Datos para el área de Recursos Humanos



- ❖ Se diseñaron los 32 reportes candidatos en correspondencia con las necesidades de información de los clientes.
- ❖ Se definieron los roles y permisos para el MD, lo que contribuye a la seguridad de la aplicación.

Capítulo 3: Implementación y prueba del Mercado de Datos para los Recursos Humanos.

Introducción

En el presente capítulo se describirá la estructura de los datos del almacén, se realizará la implementación de los subsistemas de integración y el subsistema de visualización de datos definiéndose la arquitectura de integración y explicándose los procesos de extracción, transformación y carga de los datos, así como la implementación de los trabajos y la generación de reportes utilizados en el MD. Además, se realizarán pruebas que parten de la aplicación de listas de chequeo y casos de pruebas, con el objetivo de verificar el correcto funcionamiento del MD, así como la validación de la propuesta de solución para contribuir a la toma de decisiones.

3.1 Implementación del subsistema de almacenamiento.

En la implementación del subsistema de almacenamiento se realiza el desarrollo de la estructura física del MD, además se definen todos los estándares de codificación que van a poseer las estructuras del Mercado de Datos, para facilitar la comprensión por parte del cliente.

Estructura de los datos:

Una estructura de datos es una vía factible para organizar una colección de datos. En la BD la información se encuentra organizada en estructuras lógicas que facilitan una correcta manipulación de la misma. Estas estructuras son denominadas esquemas que están conformados por una o varias tablas.

Esquemas: Los esquemas en una BD representan una manera eficiente de tener organizada la información. Estos pueden contener funciones, operadores y tipos de datos. Esta estructura le permite al usuario tener acceso a ellos siempre y cuando posea los permisos adecuados.

Para el desarrollo de la solución se cuenta con los siguientes esquemas:

- ❖ **Dimensión:** Este esquema contiene 9 dimensiones identificadas para el MD de RR.HH de la CITEC.
- ❖ **Hecho:** Este esquema contiene los hechos identificados para el MD de RR.HH.
- ❖ **Metadatos:** Este esquema contiene información relacionada con la fecha de la última carga.

3.1.1 Estándar de codificación.

Con el objetivo de organizar la estructura del almacén de datos, se formaliza un modelo, norma, patrón o estándar de codificación. Esta acción permite a los desarrolladores entender cada una de las estructuras.

En la tabla 12 se muestran como quedaron definidos estos estándares.

Tabla 12: Estándar de codificación para el MD.

Tipo de Objeto	Función	Nomenclatura	Descripción
Esquemas	Dimensiones	dimensión	Esquemas donde se encuentran las tablas de dimensiones
	Hechos	mart_rr.hh	Esquema donde se encuentra la tabla de hechos.
	Metadatos	md_fecha	Esquemas donde se encuentra la tabla con la información del metadato fecha.
Tablas	Dimensiones	dim_[nombre]	Tablas dimensionales utilizadas como perspectivas de análisis.
	Hechos	hech_[nombre]	Tablas de hechos que definen las principales medidas requeridas para calcular.

3.2 Implementación del Subsistema de Integración de datos.

El proceso de integración de los datos consta de tres etapas fundamentales relacionadas entre sí: extracción, transformación y carga de los datos. Para identificar y corregir los problemas se realiza la limpieza de los datos que permite llenar valores ausentes y corregir errores. Una vez que los datos son transformados se cargan, poblando las dimensiones y los hechos que conforman la estructura del MD para el área de Recursos Humanos.

Según **(Kimball, 2004)**, ETL consiste en:

- ❖ **Extracción:** La primera parte del proceso de ETL consiste en extraer los datos desde los sistemas de origen. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.
- ❖ **Transformación:** Cualquier operación realizada sobre los datos para que puedan ser cargados en el Almacén de Datos o se puedan migrar de éste a otra base de datos. La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en

datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos.

- ❖ **Carga:** Consiste en almacenar los datos en la base de datos final. La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino.

Ralph Kimball define 34 sistemas ETL clasificados en cuatro grupos, donde para llevar a cabo el proceso de integración de datos del presente trabajo fueron utilizados los siguientes subsistemas (**Díaz, 2011**):

Extracción

- ❖ Sistema de extracción: permite la extracción de los datos desde la fuente origen hasta la fuente destino.

Limpieza y conformación

- ❖ Limpieza de datos: implementa los procesos de calidad de los datos que permiten detectar las incoherencias de calidad.
- ❖ Rastreo de eventos de errores: captura todos los errores que proporcionan información valiosa sobre la calidad de los datos y permite la mejora de los mismos.

Entrega

- ❖ Claves subrogadas: permite crear claves subrogadas independientes para cada tabla.

Gestión

- ❖ Programador de trabajos: permite gestionar las ETL de la categoría de trabajos.
- ❖ Repositorio de metadatos: captura los metadatos de los procesos de ETL, de los datos del negocio y de los aspectos técnicos.

3.2.1 Implementación de las transformaciones.

Un trabajo o “Job” es un conjunto de tareas con el objetivo de realizar una acción determinada. Estos permiten ejecutar varias transformaciones o trabajos previamente diseñados y organizar una secuencia de ejecución de éstos. Los trabajos se encuentran en un nivel superior de las transformaciones. En la presente investigación se realizaron tres trabajos, uno es destinado a ejecutar las transformaciones para la carga de las 9 dimensiones, otro para ejecutar las transformaciones para la carga de los dos hechos y el general que ejecuta los dos trabajos anteriores. A continuación, será descrito este proceso en las figuras 12,13 y 14:

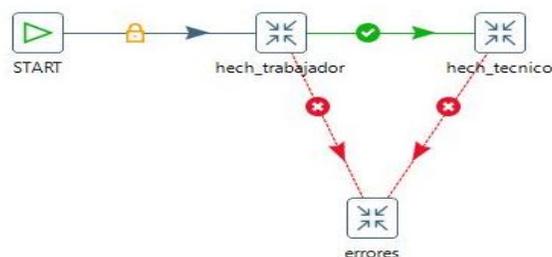


Fig. 12 Trabajo para los hechos.

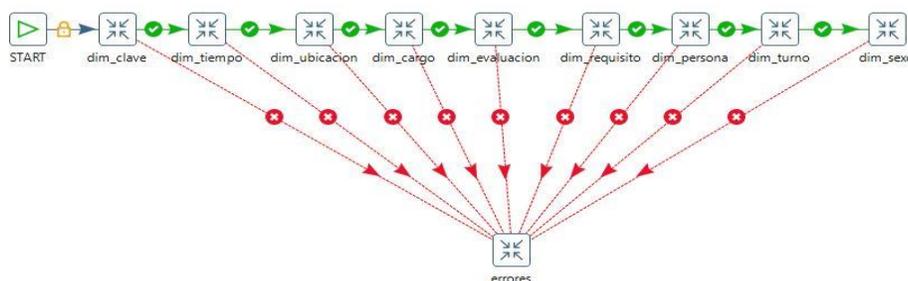


Fig. 13 Trabajo para las dimensiones.



Fig. 14 Trabajo general.

3.2.2 Implementación de las transformaciones.

En la presente investigación se realizó un flujo de transformación para la carga de cada una de las tablas pertenecientes al esquema mart_rr.hh. Para las dimensiones, la transformación se realizó a partir de la carga de los indicadores de cada uno de los modelos de las fuentes de datos, estos contienen los campos que son necesarios para poblar la BD para luego pasar a las transformaciones adecuadas. Las siguientes figuras 15 y 16 muestran las transformaciones de una dimensión y un hecho.

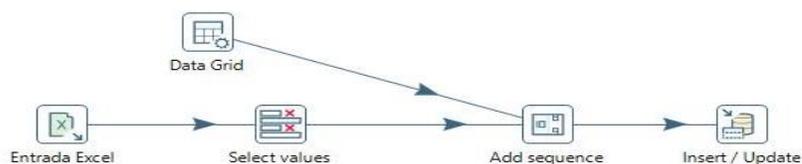


Fig. 15 Transformación de la dimensión clave.

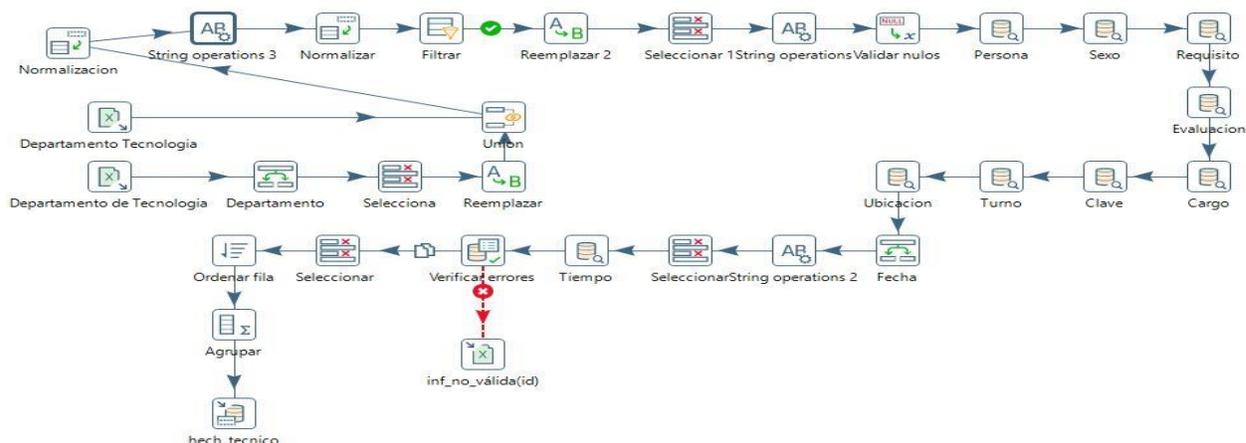


Fig. 16 Transformación para el hecho técnico.

3.3 Implementación del Subsistema de Visualización.

Luego de realizado el proceso de ETL, como la implementación de los subsistemas de visualización, se describen los elementos asociados a la creación de los cubos OLAP y de los reportes para la visualización y análisis de la información.

3.3.1 Implementación de los cubos OLAP.

Para la implementación de los cubos del MD, se definieron las medidas, dimensiones y niveles de jerarquía que componen el esquema para el área de RR.HH. La figura 17 muestra la estructura de los cubos.



Fig. 17 Implementación de los cubos OLAP.

3.3.2 Implementación de la Capa de Visualización.

El mapa de navegación se crea con el objetivo de que el cliente tenga una mejor visualización de cómo está estructurada la información referente al MD sobre los RR.HH, contribuye a definir el entorno de análisis del área de Recursos Humanos. El A.AG_Recursos_Humanos, está dividida en dos Áreas de Análisis. El A.A_Trabajador, contiene tres libros de trabajo y 16 reportes o tablas de salida, mientras que el A.A_Técnico contiene tres LT y 16 reportes, conformando un total de 32 reportes que satisfacen las necesidades del cliente.

- A.A.G.RR.HH: Agrupa la información referente a todos los Recursos Humanos de la CITEC.
- A.A Trabajador: Contiene la información referente a todos los trabajadores de la facultad.
- A.A Técnico: Contiene la información referente a todos los técnicos de la facultad.
- LT Trabajador: Contiene la información referente a los indicadores de los trabajadores.
- LT horas extras: Contiene la información referente a todos los trabajadores con horas extras.
- LT claves: Contiene la información referente a todos los trabajadores con alguna clave afectada.
- LT horas trabajadas: Contiene la información referente a las horas trabajadas por los técnicos.
- LT claves: Contiene la información referente a todos los técnicos con alguna clave afectada.
- LT Técnico: Contiene la información referente a los indicadores de los técnicos.

La figura 18 muestra la arquitectura de información diseñada para el MD para el área de RR.HH:



Fig. 18 Arquitectura de información.

3.4 Implementación de los reportes candidatos.

Una vez realizado el MD se podrán realizar reportes y vistas de análisis con el objetivo de obtener información relevante de los datos que se encuentran almacenados para contribuir a la toma de decisiones. Para la implementación de la solución se pueden generar varios reportes diferentes y vistas de análisis, teniendo en cuenta la información que se maneja en el MD. Los mismos se deben mostrar en formato PDF y Excel. A continuación, se muestran algunos ejemplos de los reportes y algunas vistas que se pueden observar en las figuras 19, 20 y 21:



Facultad de Ciencias y Tecnologías Computacionales

Hora de acceso: 07 junio, 2017, 05:32
Facultad de Ciencias y Tecnologías Computacionales

Listado de Trabajadores con horas extras

Nombre y Apellidos	Cargo	Sexo	Requisito	Evaluación	Departamento	Horas extras	Mes	Año
Eugenio R. Moreno Pantoja	Chofer "D"	M	S	A	Facultad 6	0	Agosto	2015
Eugenio R. Moreno Pantoja	Chofer "D"	F	S	S	Facultad 6	64	Mayo	2015
Eugenio R. Moreno Pantoja	Chofer "D"	F	S	S	Facultad 6	56	Marzo	2015
Eugenio R. Moreno Pantoja	Chofer "D"	F	S	A	Facultad 6	32	Julio	2015
Eugenio R. Moreno Pantoja	Chofer "D"	F	S	A	Facultad 6	32	Noviembre	2015
Eugenio R. Moreno Pantoja	Chofer "D"	F	S	S	Facultad 6	48	Enero	2015
Eugenio R. Moreno Pantoja	Chofer "D"	F	S	A	Facultad 6	48	Abril	2015
Eugenio R. Moreno Pantoja	Chofer "D"	F	S	S	Facultad 6	56	Octubre	2015
Eugenio R. Moreno Pantoja	Chofer "D"	F	S	S	Facultad 6	24	Diciembre	2015
Total	9	9	9	9	9	360	9	9

Fig. 19 Implementación del reporte: Listado de Trabajadores con horas extras.



Año: 2015
 Mes: Abril

Departamento de Programacion

Trabajadores con Misión Internacionalista

Nombre y Apellidos	Cargo	Sexo	Requisito	Evaluación	Cantidad de días
Asnay Guirola González	Profesor	M	S	A	24
Yanet Espinal Martin	Profesor	F	S	A	24
Yaneisi PérezHeredia	Profesor	F	S	A	24

Fig. 20 Implementación del reporte: Trabajadores con Misión Internacionalista.

Persona	Cargo	Ubicación	Sexo	Evaluación	Requisito	Medidas	
Todos	Todos	Todos	Todos	Todos	Todos	• Cantidad de horas extras 61.364	
Adisley Reyes Crespo	Todos	Todos	Todos	Todos	Todos	0	
Adonis Ricardo Rosales García	Todos	Todos	M	S	S	0	
					N	0	
					S	Todos	0
					NE	Todos	0
					M	Todos	Todos
		UCI	Todos	Todos	Todos	0	
		Departamento de Ciencias Basicas	Todos	Todos	Todos	0	
	Profesor	Todos	Todos	Todos	Todos	0	
Aida Perez Lopez	Todos	Todos	Todos	Todos	Todos	0	
Alexis René Rodríguez León	Todos	Todos	Todos	Todos	Todos	0	

Fig. 21 Vista de Análisis: Información de los trabajadores.

3.5 Pruebas.

Una vez que se ha dado por concluida la implementación, se debe dar paso a la etapa de las pruebas. Estas garantizan que se haya cumplido con las especificaciones iniciales que fueron definidas para el MD.

La prueba es el proceso de ejecución de un programa con el fin de encontrar deficiencias. Una prueba tiene éxito si descubre errores que no han sido detectados hasta entonces. Para llevar a cabo las pruebas del MD, se ha empleado el método utilizado por el centro DATEC para garantizar el buen funcionamiento y la calidad de los productos de almacenes de datos.

El siguiente modelo es una representación de dos cascadas relacionadas con un vértice común en la codificación, donde la cascada izquierda muestra las actividades relacionadas con el análisis y diseño del MD y la cascada derecha muestra las actividades relacionadas con el aseguramiento de la calidad mediante los tipos de pruebas aplicadas. La figura 22 muestra el modelo:

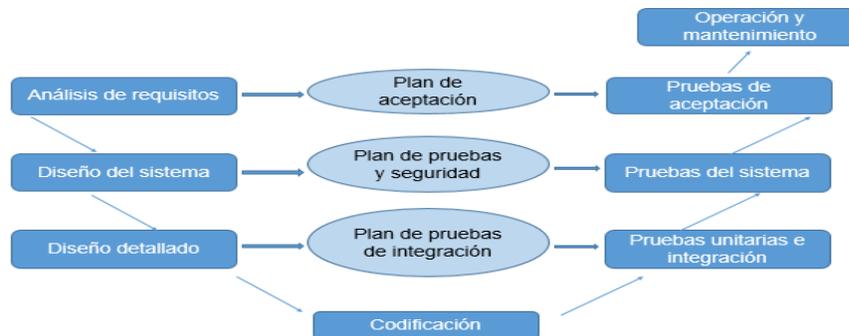


Fig. 22 Modelo V.

Según Lantares (**Lantares, 2003**), estas pruebas consisten en:

Pruebas unitarias: Constituyen la primera fase de las pruebas dinámicas y se realizan a cada componente o módulo del software de manera individual.

Prueba de integración: Estas pruebas consisten en comprobar el sistema como un todo para determinar la integración de los componentes del mismo, evaluar su correcta interfaz, funcionalidad y desempeño.

Pruebas de sistema: Estas pruebas se realizan para comprobar el correcto funcionamiento del sistema y validar el cumplimiento de los requisitos de información y funcionales definidos por los clientes.

Pruebas de aceptación: Estas pruebas son realizadas directamente por los clientes para verificar los requisitos o necesidades planteadas por ellos. Tienen como objetivo demostrarle al cliente que la funcionalidad está terminada correctamente.

3.6 Herramientas para la aplicación de las pruebas.

Para lograr que un producto de software tenga calidad, es necesario realizar un conjunto de evaluaciones durante todo el proceso de desarrollo que implique al cliente y desarrollador. Con esto se garantiza que el sistema desarrollado cumpla con los requisitos, tanto funcionales como no funcionales.

Existen diferentes tipos de pruebas, cada uno aplicable en un entorno diferente de acuerdo a los objetivos que se persigan en su realización. Para verificar el correcto funcionamiento del Mercado de Datos se realizaron las siguientes pruebas.

3.6.1 Casos de pruebas.

Para la validación del MD para el área RR.HH, se diseñaron los siguientes Casos de Pruebas (CP) asociado a los casos de uso de información, con el propósito de comprobar que la documentación del producto se corresponde con lo establecido en la aplicación, ejecutándose sobre ciertas condiciones para descubrir errores antes de la entrega del software al cliente. La tabla 13 muestra el diseño de un caso de prueba:

Tabla 13: Diseño del CP para el CU: "Presentar información sobre las personas".

Escenario	Descripción	Variables de entrada	Variables de salida	Respuesta del sistema	Flujo central
EC 1: Obtener la cantidad de horas extras por ubicación, persona, tiempo, cargo, sexo, clave, requisito, evaluación.	Muestra información sobre las horas extras de las personas por ubicación, persona, tiempo, cargo, sexo, clave, requisito, evaluación	Ubicación	listado de personas	El sistema muestra todas las variables disponibles para el análisis, ubicado en las filas y las columnas que pueden ser visualizadas en el reporte.	El usuario se autentica, selecciona el A.A.G Análisis de Recursos Humanos / A.A Trabajador / L.T Horas extras / Se selecciona el reporte correspondiente.
		Tiempo	listado de personas		
		Persona	listado de personas		
		Cargo	listado de personas		
		Sexo	listado de personas		
		Evaluación	listado de personas		
		Clave	listado de personas		
		Turno	listado de personas		
		Requisito	listado de personas		

3.6.2 Listas de Chequeo.

Las listas de chequeo que tienen como objetivo examinar cuidadosamente áreas importantes, para lograr identificar problemas comunes, encontrar errores y solucionarlos en un espacio breve de tiempo.

En esta investigación se aplicaron las siguientes listas de chequeo a los artefactos de los procesos de ETL con el fin de evaluar y verificar el potencial de cada uno de ellos, midiendo la confiabilidad y calidad de los datos cargados, a continuación, los artefactos los cuales se le aplicaron las listas de chequeo:

- ❖ Lista de chequeo del Mapa Lógico de Datos.
- ❖ Lista de chequeo al Documento de Especificación de Casos de Uso.
- ❖ Lista de chequeo al artefacto Documento de Especificación de Requisitos.

- ❖ Listas de chequeo del Perfilado de Datos.

Las listas de chequeo responden a los siguientes elementos:

- ❖ **Peso:** define si el indicador a evaluar es crítico o no.
- ❖ **Indicadores a evaluar:** son los indicadores a evaluar en las secciones Estructura del documento, semántica del documento e indicadores definidos por las diferentes etapas.
- ❖ **Evaluación:** es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de que exista alguna dificultad sobre el indicador y 0 en caso de que el indicador revisado no presente problemas.
- ❖ **No procede:** se usa para especificar que el indicador no es necesario evaluarlo en ese caso.
- ❖ **Cantidad de elementos afectados:** especifica la cantidad de errores encontrados sobre el mismo indicador.
- ❖ **Comentario:** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo.

Las listas de chequeo contienen diferentes indicadores a evaluar los cuales se encuentran distribuidos en las siguientes secciones:

Estructura del documento: ¿Contiene todos los aspectos definidos por el expediente del proyecto?

Elementos definidos para el modelo de desarrollo: Contiene todos los indicadores a evaluar durante la etapa de análisis de datos, a continuación, algunas preguntas relacionadas con este indicador:

- ¿Se realizó un estudio preliminar de la entidad cliente?
- ¿Se identificaron los requisitos de información?
- ¿Se realizó el modelo de datos correspondientes al mercado de datos?

Semántica del documento: Contiene todos los indicadores a evaluar respecto a la redacción y ortografía, a continuación, algunas preguntas relacionadas con este indicador:

- ¿Se han encontrado errores ortográficos en los entregables?
- ¿Se realizaron los diseños de los casos de pruebas?
- ¿Se realizaron las pruebas de aceptación?

Luego de haber aplicado las listas de chequeo fueron detectadas seis NC. La figura 23 muestra el comportamiento de los indicadores identificados correspondiente a la lista de chequeo para el Mapa Lógico de Datos, en la que fueron definidos, para la Estructura del documento, un total de dos indicadores, de ellos, uno fue crítico y una NC detectada. En cuanto a los Indicadores definidos por la etapa, se identificaron tres indicadores, de ellos, dos fueron críticos y se detectaron dos NC. Para la sección correspondiente a la

Semántica del documento, se identificaron cuatro indicadores, de estos, tres fueron críticos y tres NC detectada.

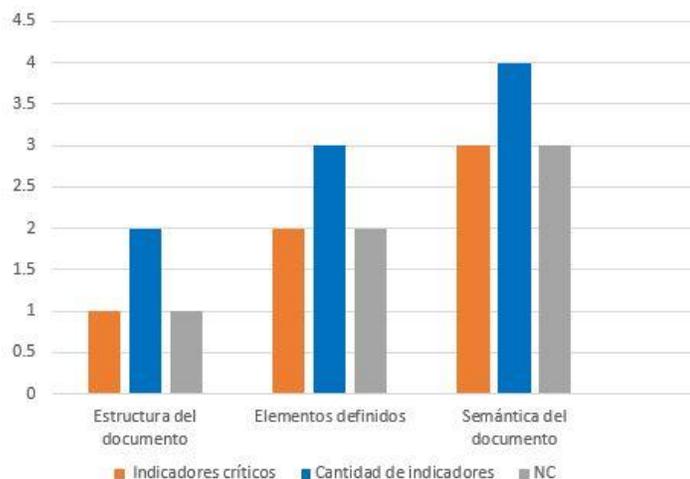


Fig. 23 Comportamiento de los indicadores de la lista de chequeo para el artefacto: Especificación de Casos de Uso.

3.7 Resultados de las pruebas.

Seguidamente se detallará de forma breve cada una de las pruebas aplicadas de acuerdo a la propuesta del modelo antes descrito en el epígrafe 3.5, así como los resultados obtenidos en cada una de ellas:

❖ Pruebas unitarias y de integración

Las pruebas unitarias y de integración se le aplicaron al subsistema de almacenamiento, al subsistema de integración de datos y a los diferentes componentes relacionados con la capa de visualización por los miembros del equipo de desarrollo, detectando cinco NC las cuales fueron resueltas en su totalidad:

- Aplicar tratamiento a los EXCEL por la diferencia estructural de los mismos.
- Redefinir las tablas de hechos de manera que responda a las necesidades de los usuarios.
- Intercambiar el orden en algunos componentes con el objetivo de optimizar el proceso de transformaciones.
- Poner nombres sugerentes a los pasos de las transformaciones.
- Comprobar que los datos cargados coincidan con los de la fuente.

❖ Pruebas del sistema

Las pruebas del sistema se realizaron por parte del equipo de desarrollo. Los siete casos de prueba basados en CU (**Artefacto: Especificación de casos de prueba**) permitieron identificar un total de ocho NC, para una primera iteración se encontraron tres conformidades de complejidad Alta, dos de complejidad Media y una de complejidad Baja, para una segunda iteración se encontraron una de complejidad Baja, una de complejidad Media y para una tercera iteración fueron resueltas todas las no conformidades. A continuación, las no conformidades encontradas:

- **NC1:** Las descripciones de los escenarios del caso de prueba alojamiento está mal, debe haber una descripción de lo que realmente se hace.
- **NC2:** No se corresponden el orden de los escenarios del caso de prueba con el orden de los reportes en la aplicación.
- **NC3:** El flujo central descrito en los CP no se corresponde con la aplicación.
- **NC4:** El nombre de las columnas de Variables de entrada y salida del CP están mal nombradas.
- **NC5:** Las descripciones de las variables no están acorde a lo que muestra la aplicación.
- **NC6:** Error ortográfico en las descripciones de los LT Claves.

La tabla 14 muestra los resultados de las pruebas de sistemas realizadas.

Tabla 14: Resultados de las pruebas de sistema.

Iteraciones	Cantidad de NC	Complejidad			NC Resueltas
		Alta	Media	Baja	
Iteración 1	6	3	2	1	6
Iteración 2	2	0	1	1	2
Iteración 3	0	0	0	0	0

❖ Pruebas de aceptación

En conjunto con el cliente se realizaron las pruebas de aceptación de la solución, donde se detectaron tres no conformidades en la primera iteración, las cuales fueron resueltas satisfactoriamente en la segunda iteración, quedando comprobado que el sistema cumple con sus necesidades y que están satisfechos con el producto elaborado. De esta manera se obtuvo la carta de aceptación por parte del cliente (**Anexo#1**).

La figura 24 a continuación muestra las no conformidades encontradas en las distintas pruebas Unidad, Aceptación y Sistema:

- Mostrar los mensajes en español y sin abreviaturas
- Mostrar la información de todos los trabajadores.
- Mostar las horas trabajadas por los técnicos

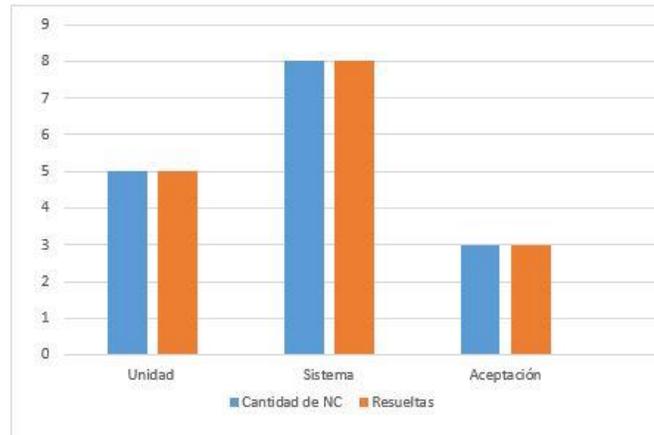


Fig. 24 Resultado de las pruebas.

Siendo avalada por el cliente, donde queda confirmado que el Mercado de Datos está listo para su explotación por parte del especialista del área de RR.HH.

Conclusiones del capítulo.

En este capítulo se diseñó las pruebas tanto a la documentación como al sistema del Mercado de Datos para el área de Recursos Humanos, se ha finalizado con la etapa de prueba por lo que se puede concluir que:

- ❖ Se diseñaron siete casos de pruebas para las pruebas de sistema, con el fin de mejorar la organización y la calidad del MD.
- ❖ Mediante las pruebas unitarias y de integración se detectaron cinco NC de las cuales pertenecen al proceso de ETL con una complejidad Media, dos pertenecen al proceso de visualización con complejidad Media y tres relacionadas con el diseño del modelo de datos con complejidad Alta, todas resueltas satisfactoriamente.
- ❖ Se detectaron con las pruebas de sistema para una primera iteración se encontraron tres conformidades de complejidad Alta, dos de complejidad Media y una de complejidad Baja, para una segunda iteración se encontraron una de complejidad Baja, una de complejidad Media y para una tercera iteración fueron resueltas todas las no conformidades.

Mercado de Datos para el área de Recursos Humanos



- ❖ Mediante las pruebas de aceptación se comprobaron las funcionalidades requeridas con el especialista del área que emitió una carta de aceptación, lo cual permite desplegar el sistema.

Conclusiones generales

Al concluir la investigación, se puede plantear que fueron cumplidos los objetivos trazados y las tareas de la investigación propuesta.

- ❖ El estudio de los principales conceptos relacionados con la investigación, la selección de la Propuesta de Metodología para el Desarrollo de Almacenes de Datos en DATEC y las herramientas a utilizar, permitieron el desarrollo de la solución informática.
- ❖ El análisis y diseño del Mercado de Datos fue realizado teniendo en cuenta los requerimientos acordados con el cliente, quedando definidos los elementos necesarios para la implementación de los subsistemas de almacenamiento, integración y visualización.
- ❖ La implementación de los subsistemas de almacenamiento, integración y visualización, permitió organizar e integrar los datos atinentes al área de RR.HH, apoyando de esta forma la toma de decisiones.
- ❖ Las pruebas unitarias, de integración, del sistema y aceptación realizadas al Mercado de Datos, permitieron validar las funcionalidades definidas del sistema.

Recomendaciones

Al término de esta investigación se recomienda:

- ❖ Agregar al mercado de datos otras técnicas de Inteligencia de negocio como los cuadros de mandos integrales, de manera que se puedan realizar análisis estadísticos más completos en el área Recursos Humanos.
- ❖ Proporcionar cursos de capacitación a los especialistas del área de Recursos Humanos, para que puedan trabajar con la aplicación satisfactoriamente.
- ❖ Extender el Mercado de Datos a las diferentes áreas de Recursos Humanos de la Universidad.

Referencias

Arley, Ricardo Chinchilla. MD: conceptos y metodologías de desarrollo. [En línea] Julio-septiembre de 2011. <http://www.kerwa.ucr.ac.cr/bitstream/handle/10669/619/TM%2024-3%20art%206.pdf?sequence=1>.

AUQUI, D. O. Portal de Tecnologías Aplicadas a Negocios. pis2.wikispaces.com/file/view/TESIS+II++UGEL+04.doc.

Basallo, Y. A. and A. D. Estrada (2009). "Una experiencia en integración de aplicaciones empresariales." Revista Cubana de Ciencias Informáticas 3.

Bernabeu, Dario. Data Warehouse Manager. [En línea] 6 de Mayo de 2009. [Citado el: 30 de Marzo de 2012.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse>.

Bigatti, Ing. Cristian G. Data Warehouse. [En línea] 1 de Diciembre de 2005. . [Citado el: 31 de Marzo de 2012.] <http://www.edutecne.utn.edu.ar/sist-gestion-II>.

Comparativa B.I. Open Source. Open Busines Intelligence. [En línea] 6 de 2010. http://www.stratebi.es/todobi/jun10/Comparativa_OSBI.pdf.

DataPrix. DataPrix. DataPrix. [En línea] DataPrix. [Citado el: 24 de octubre de 2016.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-conceptos-6>.

Díaz, J. C. (2011). Introducción al Business Intelligence.

EASY, P. P. A. M. (2005). "Pentaho BI Platform and Server ", from http://community.pentaho.com/projects/bi_platform/.

Gravitar (2014). "Data Integration.

Hernández, Y. G. (2013). METODOLOGÍA DE DESARROLLO PARA PROYECTOS DE ALMACENES DE DATOS, Universidad de las Ciencias Informáticas.

Imhoff, Claudia, Galemno, Nicholas y G. Geiger, Jonathan. 2003. Mastering Data Warehouse Design Relational . Mastering Data Warehouse Design Relational . 2003.

Inmon, B. 1992. Building the Data Warehouse. Building the Data Warehouse. 1992.

Inmon, W. H. 2005. Building the Data Warehouse. Building the Data Warehouse. 2005.

- INTECO, Laboratorio Nacional de Calidad del Software. 2010.** INGENIERÍA DEL SOFTWARE: METODOLOGÍAS Y CICLOS DE VIDA. Gobierno de España : s.n., 2010.
- Kimball, R. 2006.** The Data warehouse Lifecycle Toolkit. [aut. libro] R Kimball. The Data warehouse Lifecycle Toolkit. 2006.
- Kimball, Ralph, Ross Margy. 2013.** The Data Warehouse Toolkit. Third Edition. 2013.
- Kimball, R. (2004).** The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Wiley Publishing.
- Lantares, S. (2003).** "Data warehouse: la prueba de la calidad." Solutions Lantares <http://www.lantares.com/blog/test-de-calidad-para-data-warehouse-y-bases-de-datos>.
- Martinez, Rafael. 2013.** PostgreSQL-es. PostgreSQL-es. [En línea] 2013. [Citado el: 15 de diciembre de 2015.] http://www.postgresql.org.es/sobre_postgresql.
- MICHAUD, Frederic. 2015.** Identifying Key Attack Surface Resources with Dynamic Analysis. 2015.
- OBE, Regina O. y HSU, Leo S. 2015.** PostGIS in action. s.l. : Manning Publications Co, 2015.
- Paradigm. 2010.** Paradigm, 2010. Paradigm, 2010. [En línea] www.visual-paradigm.com, 2010. [Citado el: 15 de noviembre de 2015.] <http://www.visual-paradigm.com/aboutus/newsreleases/vpum180.jsp>.
- PARADIGM, C. V. (2013).** "Visual Paradigm for UML - UML tool for software application development." from <http://www.visua-paradigm.com/product/vpum1/>.
- Pentaho, C. (2015).** "Schema Workbench." from <http://community.pentaho.com/>.
- PentahoCorporation. (2005).** "Pentaho BI Suite Enterprise Edition." Retrieved 10 de Marzo, 2009, from www.pentaho.com.
- Pérez Niebla, Leonel.** Sistema de Información de Gobierno. Mercado de datos para el área de Construcción. La Habana: Universidad de las Ciencias Informáticas, 2011. Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas.
- PostgreSQL. (2016).** "The work's most advanced open source data base." From <https://www.postgresql.org/>.
- Pressman, R. S. (2002).** Ingeniería del Software, un enfoque práctico.

PRESSMAN, Roger. 2009. Ingeniería del Software. Un enfoque práctico. s.l. : McGraw-Hill/Interamericana, 2009.

prueba, Condiciones y casos de. Condiciones y casos de prueba. Condiciones y casos de prueba. [En línea] <http://www.rmya.com.ar>.

Ramos, N. (2009). "Fundamentos y Casos de Aplicación." Soluciones de Inteligencia de Negocios a su alcance.

retopyme. retopyme. retopyme. [En línea] [Citado el: 29 de octubre de 2016 .] <http://www.retopyme.com.ar/beta/InfodeRRHH.pdf>.

Rodolfo Quispe-Otazu "Crear y usar reglas de Negocio" diciembre 2008. Disponible en: <http://msdn.microsoft.com/es-es/library/aa577691%28v=bts.10%29.aspx>.

sinnexus. 2016. sinnexus. sinnexus. [En línea] 2016. [Citado el: 23 de octubre de 2016.] http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.

teradata. 2016. teradata. teradata. [En línea] 2016.

Tabla de Hechos y Dimensiones. [En línea] 2009. <http://msdn.microsoft.com/es-es/library/ms244679%28v=vs.80%29.aspx>.

Bibliografía

- Arley, Ricardo Chinchilla.** MD: conceptos y metodologías de desarrollo. [En línea] Julio-septiembre de 2011. <http://www.kerwa.ucr.ac.cr/bitstream/handle/10669/619/TM%2024-3%20art%206.pdf?sequence=1>.
- AUQUI, D. O.** Portal de Tecnologías Aplicadas a Negocios. pis2.wikispaces.com/file/view/TESIS+II++UGEL+04.doc.
- Basallo, Y. A. and A. D. Estrada (2009).** "Una experiencia en integración de aplicaciones empresariales." Revista Cubana de Ciencias Informáticas 3.
- Bernabeu, Dario.** Data Warehouse Manager. [En línea] 6 de Mayo de 2009. [Citado el: 30 de Marzo de 2012.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse>.
- Bigatti, Ing. Cristian G.** Data Warehouse. [En línea] 1 de Diciembre de 2005. . [Citado el: 31 de Marzo de 2012.] <http://www.edutecne.utn.edu.ar/sist-gestion-II>.
- Bruegge, B. Y Dutoit, A. 2002.** Ingeniería de Software Orientado a Objetos. 2002.
- DataPrix.** DataPrix. DataPrix. [En línea] DataPrix. [Citado el: 24 de octubre de 2016.] <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-conceptos-6>.
- Díaz, J. C. (2011).** Introducción al Business Intelligence.
- EASY, P. P. A. M. (2005).** "Pentaho BI Platform and Server ", from http://community.pentaho.com/projects/bi_platform/.
- Gravitar (2014).** "Data Integration.
- GAMMA, E., HELM, R. y JOHNSON, R. y VLISSIDES, J. 2009.** Patrones de diseño. Patrones de diseño. [En línea] vico, 2009. <http://www.vico.org/pages/PatronsDisseny.html>.
- Imhoff, Claudia, Galemno, Nicholas y G. Geiger, Jonathan. 2003.** Mastering Data Warehouse Design Relational . Mastering Data Warehouse Design Relational . 2003.
- Inmon, B. 1992.** Building the Data Warehouse. Building the Data Warehouse. 1992.
- Inmon, W. H. 2005.** Building the Data Warehouse. Building the Data Warehouse. 2005.

- INTECO, Laboratorio Nacional de Calidad del Software. 2010.** INGENIERÍA DEL SOFTWARE: METODOLOGÍAS Y CICLOS DE VIDA. Gobierno de España : s.n., 2010.
- Kimball, R. 2006.** The Data warehouse Lifecycle Toolkit. [aut. libro] R Kimball. The Data warehouse Lifecycle Toolkit. 2006.
- Kimball, Ralph, Ross Margy. 2013.** The Data Warehouse Toolkit. Third Edition. 2013.
- Kimball, R. (2004).** The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Wiley Publishing.
- Lantares, S. (2003).** "Data warehouse: la prueba de la calidad." Solutions Lantares <http://www.lantares.com/blog/test-de-calidad-para-data-warehouse-y-bases-de-datos>.
- Larman, C. 2008.** UML y patrones. 2008.
- León, A. y Sotolongo, R. 2010.** Modelo de descripción de arquitectura de almacenes de datos para ensayos clínicos del Centro de Inmunología Molecular. Modelo de descripción de arquitectura de almacenes de datos para ensayos clínicos del Centro de Inmunología Molecular. La Habana : s.n., 2010.
- Martinez, Rafael. 2013.** PostgreSQL-es. PostgreSQL-es. [En línea] 2013. [Citado el: 15 de diciembre de 2015.] http://www.postgresql.org.es/sobre_postgresql.
- MDN. 2009.** MDN. MDN. [En línea] 2009. [Citado el: 26 de octubre de 2015.] <https://developer.mozilla.org/es/docs/Web/JavaScript/Guide/Introducci%C3%B3n>.
- MICHAUD, Frederic. 2015.** Identifying Key Attack Surface Resources with Dynamic Analysis. 2015.
- OBE, Regina O. y HSU, Leo S. 2015.** PostGIS in action. s.l. : Manning Publications Co, 2015.
- ÖVERGAARD, Gunnar y PALMKVIST, Karin,. 2008.** Use Cases: Patterns and Blueprints. 2008.
- Paradigm. 2010.** Paradigm, 2010. Paradigm, 2010. [En línea] www.visual-paradigm.com, 2010. [Citado el: 15 de noviembre de 2015.] <http://www.visual-paradigm.com/aboutus/newsreleases/vpuml80.jsp>.
- PARADIGM, C. V. (2013).** "Visual Paradigm for UML - UML tool for software application development." from <http://www.visua-paradigm.com/product/vpuml/>.
- Pentaho, C. (2015).** "Schema Workbench." from <http://community.pentaho.com/>.

PentahoCorporation. (2005). "Pentaho BI Suite Enterprise Edition." Retrieved 10 de Marzo, 2009, from www.pentaho.com.

PostgreSQL. (2016). "The work's most advanced open source data base." From <https://www.postgresql.org/>.

Pressman, R. S. (2002). Ingeniería del Software, un enfoque práctico.

PRESSMAN, Roger. 2009. Ingeniería del Software. Un enfoque práctico. s.l. : McGraw-Hill/Interamericana, 2009.

prueba, Condiciones y casos de. Condiciones y casos de prueba. Condiciones y casos de prueba. [En línea] <http://www.rmya.com.ar>.

Ramos, N. (2009). "Fundamentos y Casos de Aplicación." Soluciones de Inteligencia de Negocios a su alcance.

retopyme. retopyme. retopyme. [En línea] [Citado el: 29 de octubre de 2016 .] <http://www.retopyme.com.ar/beta/InfodeRRHH.pdf>.

sinnexus. 2016. sinnexus. sinnexus. [En línea] 2016. [Citado el: 23 de octubre de 2016.] http://www.sinnexus.com/business_intelligence/datawarehouse.aspx.

teradata. 2016. teradata. teradata. [En línea] 2016.

uml-diagrams. 2010. uml-diagrams. uml-diagrams. [En línea] 2010. [Citado el: 14 de octubre de 2015.] <http://www.uml-diagrams.org/>.

Wolff, C. (2002). "Implementando un data warehouse." Revista Ingeniería Informática 5.

Anexos

Anexo#1: Carta de Aceptación.

ACTA DE ACEPTACIÓN

De una parte, el Vicedecano de Administración de la facultad CITEC, en lo sucesivo VDA, de la Universidad de las Ciencias Informáticas, representado en este acto por: Lcdo. Carlos Luis Hernández Hernández y por otra parte el Tutor y Director de la Dirección de Servicios Generales, representado en este acto por: MSc. Omar Mar Cornelio y por último los estudiantes: Yanelis Pérez Hernández y Suinny Castillo O'Farrill.

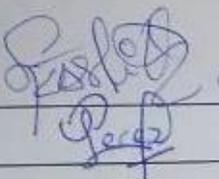
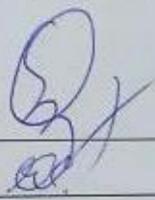
Primero: Que en cumplimiento de los requisitos funcionales han sido efectuadas las implementaciones correspondientes.

CONSIDERANDO: Que los hitos realizados has sido desarrollados con la calidad requerida y bajo las condiciones pactadas y aprobadas por **Las Partes**.

CONSIDERANDO: Que los hitos que se han ejecutado cumplen con los requisitos establecidos.

POR TANTO: **Las Partes** acuerdan formalizar mediante la presente Acta, la aceptación del producto: Mercado de datos para el área de Recursos Humanos.

Y para que así conste, se extiende la presente Acta en tres (3) ejemplares, rubricados por **Las Partes**.

 _____ _____ Entregan	 _____ _____ Reciben
---	---

Anexo# 2: Carta de Aceptación.

ACTA DE ACEPTACIÓN DE LOS REQUISITOS

Se acuerda formalizar mediante la presente Acta la aceptación por parte del Vicedecano de Administración de la facultad CITEC de los requisitos de Información definidos para el desarrollo del Mercado de Datos para el área de Recursos Humanos que se relacionan a continuación:

RI 1. Obtener el listado de técnicos evaluados de deficiente por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 2. Obtener el listado de técnicos evaluados de adecuado por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 3. Obtener el listado de técnicos evaluados de superior por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 4. Obtener el listado de trabajadores evaluados de deficiente por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 5. Obtener el listado de trabajadores evaluados de adecuado por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 6. Obtener el listado de trabajadores evaluados de superior por ubicación, tiempo, sexo, requisito, evaluación, cargo.

RI 7. Obtener la cantidad de horas extras de los trabajadores por ubicación, persona, tiempo, cargo, sexo, requisito, evaluación.

RI 8. Obtener la cantidad horas trabajadas de los técnicos por ubicación, persona, tiempo, cargo, sexo, requisito, evaluación, turno.

RI 9. Obtener el listado de personas que tienen afectada la clave ausencias injustificadas por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 10. Obtener el listado de personas que tienen afectada la clave ausencias justificadas por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 11. Obtener el listado de personas que tienen afectada la clave vacaciones por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 12. Obtener el listado de personas que tienen afectada la clave enfermedad 3 días o menos por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 14. Obtener el listado de personas que tienen afectada la clave semana receso docente por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 15. Obtener el listado de personas que tienen afectada la clave licencia por maternidad por ubicación, tiempo, persona, cargo, sexo, requisito, evaluación.

RI 16. Obtener el listado de personas que tienen afectada la clave licencia sin sueldo (problemas personales) por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 16. Obtener el listado de personas que tienen afectada la clave enfermedad más de 3 días por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 17. Obtener el listado de personas que tienen afectada la clave no pagar por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 18. Obtener el listado de personas que tienen afectada la clave prestación social por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 19. Obtener el listado de personas que tienen afectada la clave cónyuge acompañante por ubicación, tiempo, cargo, sexo, requisito, evaluación.

RI 20. Obtener el listado que tienen afectada la clave prestación de servicios (educ, pcc, ujc, etc) por ubicación, tiempo, cargo, sexo, requisito, evaluación.

MSc. Omar Mar Cornelio

Cliente

Ing. Irela González Piñera

Tutor

Suinny Castillo Ofarrill y Yanelis Pérez Hernández

Tesistas