

**UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS.**



**FACULTAD 6**

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

**Título:**

Mercado de datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI.

**Autores:**

Yanitza de la Caridad Gutiérrez Acea.

Elian Eduardo Castell Legrá.

**Tutores:**

Ing. Alberto Mendoza Garnache.

Ing. Roberto Tellez Ibarra.

**La Habana, julio de 2016**

**“Año 58 de la Revolución”**



*“El futuro de nuestra patria tiene que ser necesariamente, un futuro de hombres de ciencia...”*

*Siichartz*

**Declaración de autoría**

Declaramos ser autores del presente trabajo de diploma “Mercado de datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI”, y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

Autores: \_\_\_\_\_

Yanitza de la C. Gutiérrez Acea

\_\_\_\_\_

Elian E. Castell Legrá

Tutores: \_\_\_\_\_

Ing. Alberto Mendoza Garnache

\_\_\_\_\_

Ing. Roberto Tellez Ibarra

## **Datos de contacto**

**Tutor:** Ing. Alberto Mendoza Garnache

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: Instructor

Años de graduado: 7

Área de investigación: Calidad de Software

Años de experiencia en el tema: 4

Correo electrónico: [agarnache@uci.cu](mailto:agarnache@uci.cu)

**Tutor:** Ing. Roberto Tellez Ibarra

Especialidad de graduación: Ingeniería en Ciencias Informáticas

Categoría docente: Asistente

Años de graduado: 7

Área de investigación: Almacenes de Datos e Inteligencia de Negocios

Años de experiencia en el tema: 6

Correo electrónico: [rtibarra@uci.cu](mailto:rtibarra@uci.cu)

**Autor:** Yanitza de la C. Gutiérrez Acea

Correo electrónico: [ycgutierrez@estudianates.uci.cu](mailto:ycgutierrez@estudianates.uci.cu)

**Autor:** Elian E. Castell Legrá

Correo electrónico: [eecastell@estudiantes.uci.cu](mailto:eecastell@estudiantes.uci.cu)

### Agradecimientos

*Cuando la gratitud es absoluta, las palabras sobran, pero en silencio no sirve a nadie. Es por ello que quiero agradecer:*

*A mi dúo de tesis, por confiar en mí para lograr este sueño ya cumplido, por su dedicación y paciencia. Por ser no solo un compañero de tesis, sino también un amigo incondicional y estar a mi lado durante estos cinco años de carrera en los buenos y malos momentos y ser mi confidente, mi pañuelo de lágrimas. Por ayudarme a retomar mi camino cuando me he sentido perdida, por soportarme cuando estuve molesta o de mal humor. Gracias, eres el mejor.*

*A mis padres Juana y Rolando, por haberme dado educación, un hogar donde crecer, equivocarme, desarrollarme, aprender y donde adquirí los valores que hoy definen mi vida formándome como una profesional. Por apoyarme en mis decisiones, porque me enseñaron a luchar por mis sueños y a enfrentar los retos que la vida me ha puesto por delante. Por desvelarse junto conmigo cuando he tenido pruebas. Por ser mi guía, mi ejemplo a seguir. Los amo.*

*A mi hermana, mi segunda madre, por comportarse como tal y ser mi alma gemela a pesar de la diferencia de edad. Por velar por mí cuando mami no estaba y estar siempre ahí en los momentos más importantes de mi vida y pensar en mí antes de pensar en ella. Por reír conmigo en los momentos felices y llorar juntas en los momentos tristes. Por ser mi hermana, mi amiga, mi todo. A mi hermanito Darian que siempre estuvo pendiente de mí y me brindó su cariño y comprensión. Simplemente, gracias por ser mis hermanos.*

*A mis abuelos, a los que va dedicada esta tesis porque, aunque ya algunos no estén físicamente, han sido mi guía y me han dado fuerzas para seguir adelante. A mi abuela Mayita, que siempre tuvo una respuesta a mis preguntas de niña.*

*A Damián, que ha sido como un padre para mí y me quiere como tal, por complacerme en todos mis caprichos y soportar todos mis berrinches. A Bárbara por su preocupación y siempre estar al pendiente de mí.*

*Al resto de mi familia, porque nunca dudaron de lo que yo podía lograr, mis tíos Toño: por sus consejos, Ida, Mary, Fifi: mi otra abuela, Lázara, Antonia. Mis primos Ray: tierno y adorable y quien se ha convertido en la alegría de la casa, a Dayron: mi compañero de juego en la infancia, mi otro hermano... te quiero, a Yainy, Yandry, Gerson: el pequeñín. A mi madrina Vilma. Mi cuñado Miyoandry que, aunque a veces discutimos en el fondo lo quiero. A Frank: que me ayudó muchísimo durante estos cinco años, luchando por mis comodidades. A Salvador: que estuvo presente siempre que me hizo falta su ayuda y me acogió como su familia.*

*A mis tutores, Tellez: que se convirtió en una persona muy especial para mí y nos atendió siempre, lo mismo por el jabber, que por teléfono o cualquier día de la semana que nos aparecíamos por su laboratorio y fue nuestro guía en el desarrollo de la tesis. A Garnache: que confió en nosotros para asumir la implementación de este trabajo, por su dedicación constante. A ambos, gracias por su ayuda y constancia durante el desarrollo de la tesis. Por estar disponibles siempre que nos surgía alguna duda y no descansar hasta que estuviera resuelta. Por enseñarme a ser responsable y a no descansar mientras haya trabajo por hacer. A Yordanka que, aunque estuvo poquito tiempo, fue con ella con quien dimos los primeros pasos para la realización de este trabajo.*

*A mis profesores que me enseñaron tanto lecciones académicas como lecciones de la vida y me apoyaron y ayudaron en mi formación durante toda la carrera. A Alexis René, que nos apodó como el dúo Pimpinela. A Yamila, Dairys: quienes conocen este trabajo casi de memoria, pues batallaron con nosotros durante todo el desarrollo del mismo. A Yanet que también nos ayudó bastante.*

*Durante estos cinco años he conocido personas que no olvidaré jamás y que han dejado una huella en mi corazón, mis amigos Claudia: la primera persona que conocí cuando entré a la universidad y se*

## *Agradecimientos*

---

*convirtió en mi mejor amiga; Randy: a quien le agradezco los buenos ratos que compartí con él y los consejos que me supo dar siempre que lo necesité, mi compañero de chisme, relajo y chucho; a Rainer: más conocido como Burbuja, Ivonne: una excelente amiga, Yanara: mi compañera de estudio; Alexey: mi gran amigo, mi consejero, mi azabache; Leo: mi compañero de baile, Migue: quien sabe hacerme reír como quiere, Miche: enamorado de mi desde 2do año, te quiero un montón; Ronald, Angélica, los Zecas Company, mis compañeros de aula, de año, apto, de baile, fiestas, los chicos de Cumbiamba, quienes me dieron un espacio en el grupo y confiaron en mí, a los Sensuales. Mis colegas de la FEU, los que me enseñaron a ser una dirigente estudiantil y a los que luego me tocó enseñar y compartieron buenos momentos conmigo, Yaisel, Yoelvis, Lijandy, Susana, Manuel, Yoisbel, Enier, Lio, Osviel, Bernaldo. A Migdalia: la mejor tía de edificio, te quiero mucho.*

*A todos muchas gracias...*

*Yanitza de la C. Gutiérrez Acea*

### Agradecimientos

*A mi dúo te tesis, por confiar en mí para lograr este sueño ya cumplido, por su dedicación, paciencia. Por ser no solo una compañera de tesis, sino también una amiga incondicional y estar a mi lado durante estos cinco años de carrera en los buenos y malos. Gracias, eres la mejor.*

*A mis padres, por haberme dado educación, apoyarme en todas las decisiones difíciles de la vida y guiar este largo camino de 5 años de estudio, gracias por cerrar hoy este ciclo conmigo. Los amo.*

*A mis hermanos, por ser también parte importante de este proceso ya concluido y ser las personas que en todo momento estuvieron junto a mí.*

*A mis abuelos, a los que va dedicada esta tesis, porque aunque ya algunos no estén físicamente, han sido mi guía y me han dado fuerzas para seguir adelante. A mi abuelo Castell que aunque ya no está fue mi segundo padre, que también inició esta etapa conmigo y hoy de una forma o de otra siento que la está terminando también, a mi abuela Zeida, mi segunda madre, quien me crió, siempre me aconsejó y estuvo junto a mí en los momentos más difíciles de mi vida apoyándome en todas mis decisiones y hoy prefirió esperar los resultados en la casa. A mi abuela Marta quien también es un pilar fundamental en mi vida quién también espera los resultados en casa.*

*A Pedro, mi padrastro por su preocupación en todo momento y mostrar interés especial en todos los acontecimientos de estos 5 años de carrera. A Danni, mi madrastra por ser una persona muy especial, preocupada y haber cambiado tantas cosas para bien y estar disponible en cada momento que la necesité. Gracias a los dos por formar parte de mi vida y cerrar también esta etapa hoy junto a mí.*

*Al resto de mi familia, porque nunca dudaron de lo que yo podía lograr, mis tías Maité, Martica y Margarita y a mi primo Damian.*

*A mis amistades Alexis, Jarel, Antonio y Patricia que de una forma u otra siempre han sabido apoyarme en cualquier circunstancia de la vida y siempre los he tenido en el momento que han sido necesarios, además de ser mis mayores “consejeros”.*

## Agradecimientos

---

*A mis otros amigos Yamil, Glennis, Víctor y Randy quienes desde que los conozco hemos compartido momentos impresionantes, de diversión, risas y hasta sustos, los voy a extrañar mucho.*

*A mis tutores, Tellez: quien dio el paso al frente cuando más lo necesitamos, quien estuvo en todo momento que fue necesario junto a nosotros y más que un tutor se comportó como un buen amigo, y hoy podemos decir que todo esto fue posible gracias a su dedicación. A Garnache: que confió en nosotros para asumir la implementación de este trabajo, por su dedicación constante. A ambos, gracias por su ayuda y constancia durante el desarrollo de la tesis.*

*A mis profesores que me enseñaron tanto lecciones académicas como lecciones de la vida y me apoyaron y ayudaron en mi formación durante toda la carrera. A Alexis René, que nos apoyó en todo lo que fue necesario durante la carrera a partir de que lo conocimos en P2.*

*Durante estos cinco años he conocido personas que no olvidaré jamás y que han dejado una huella en mi corazón, mis amigos Rainer y Miguel Ángel: las primeras persona que conocí cuando entré a la universidad y se convirtieron en mis amigos y piquete de aula; Yanara: muy especial, con quien tuve mi primer accidente en la universidad “el 1er día”, y fue mi dúo musical durante la carrera, Daynelis la persona que cuando más lo necesité me compartió las series, Cuso, el enano burlón, a mis compañeros de apartamento Henry y los Juan Pablos, a todos ellos les agradezco los buenos ratos que compartimos; a Leosdany, Julio, Lijandy y a todos aquellos que de una forma u otra compartimos estos 5 años de carrera.*

*También quiero agradecer a Yamila quien a pesar de haber sido mi tribunal me supo aconsejar en momentos de desaciertos míos durante los cortes de tesis, además de dedicarme parte de su tiempo siempre que fue necesario. A Dayris que aunque hoy se convirtió en mi oponente también nos dedicó durante el curso parte de su tiempo y llegó a comportarse como una tutora más en este proceso ya concluido. A Yanet Parra quien también nos ayudó en el último momento a pesar de su estado.*

*A todos muchas gracias...*

*Elían E. Castell Legrá*

### Dedicatoria

*Dedico esta tesis a:*

*Primeramente, a alguien que significa mucho para mí y a quien estaré agradecida siempre, alguien que, a pesar de no estar físicamente a mi lado, sé que espiritualmente lo está y me guía en cada paso que doy, mi abuelo Rolando, que en paz descanse, a quien le juré en su lecho de muerte que terminaría mis estudios como profesional de la informática y le dedicaría mi trabajo de diploma. A ese que cuando me sentía triste, indecisa, nerviosa, me acariciaba la cabeza y dándome un beso en la frente me decía que todo iba a salir bien, que siempre tenía una sonrisa alentadora, una respuesta a todo. Te echo mucho de menos abuelo y te digo que tu India de Guanaroca hoy es una ingeniera como te lo prometió. A mis padres, hermanos, familiares por amarme y apoyarme siempre en todo.*

*Yanitza de la C. Gutiérrez Acea*

*Dedico esta tesis a:*

*Quiero dedicar esta tesis a mis cuatro abuelos, a pesar de que hoy dos de ellos no se encuentran físicamente junto a mí sé que se sienten felices de ver los logros alcanzados y a las que no pudieron venir porque prefirieron por una cuestión de nervios esperar desde casa los resultados que ya son realidad.*

*A mis padres, hermanos, familiares por amarme y apoyarme siempre en todo.*

*Elian E. Castell Legrá*

## **Resumen**

El área de Aseguramiento de la Calidad de Procesos y Productos de trabajo (PPQA) se encarga de la planificación y ejecución de evaluaciones, el monitoreo de las no conformidades detectadas y la realización del Análisis de Tendencias a nivel de centros y de alta gerencia, de manera que apoye al proceso de toma de decisiones estratégicas. El presente trabajo de diploma se enmarca en el desarrollo de un mercado de datos (MD) que contribuya al análisis de la información correspondiente al área PPQA, para la posterior confección del Informe de Tendencias. Para guiar el proceso de desarrollo de la solución se utilizó la Metodología para el desarrollo de proyectos de Almacenes de Datos utilizada por el Centro de Tecnología de Gestión de Datos (DATEC), apoyándose en las herramientas Visual Paradigm, PostgreSQL, PgAdmin III, DataCleaner, Pentaho Data Integration, Schema Workbench y Pentaho BI Server. Una vez concluidos los procesos de análisis, diseño e implementación de la solución se obtuvo como resultado un MD poblado y la capa de visualización donde se muestran los reportes correspondientes a cada indicador del Informe de Tendencias. Para verificar la calidad del producto y lograr la aceptación del cliente se realizaron pruebas al MD, haciendo uso de las herramientas Casos de Prueba y Listas de Chequeo.

**Palabras Claves:** Análisis de Tendencias, Aseguramiento de la Calidad de Procesos y Productos de trabajo, Informe de Tendencias, Mercado de datos.

## **Abstract**

The area of work Products and Processes Quality Assurance (PPQA) is in charge of planning and making assessments, monitoring nonconformities detected and making the Trends Analysis at centers and senior management level, in order to support the strategic decision making process. The present diploma paper is focused on the development of the Data Mart (MD) that contributes to the analysis of the information corresponding to the PPQA area, for further creation of the Trends Report. To lead the development process of the solution it was used the Methodology for the development of Data Warehouses projects used in the Data Management Technology Center (DATEC), supported by the tools Visual Paradigm, PostgreSQL, PgAdmin III, DataCleaner, Pentaho Data Integration, Schema Workbench and Pentaho BI Server. Once concluded the processes of analysis, design and implementation of the solution, it was obtained as a result a populated MD and the visualization layer where the reports corresponding to every indicator in the Trends Report are displayed. To check the quality of the product and achieve the client acceptance, tests were made to the MD, making use of the tools Test Cases and Checklists.

**Key Words:** Trends Analysis, work Products and Processes Quality Assurance, Trends Report, Data Mart.

## Índice de Contenido

Introducción .....	1
Capítulo 1: Fundamentación teórica de los Almacenes de datos. ....	6
1.1 Almacenes de datos. ....	6
1.1.1 Antecedentes de los Almacenes de datos.....	6
1.1.2 Definición de Almacenes de datos. ....	6
1.1.3 Mercado de datos. ....	7
1.1.4 Características de los Almacenes de datos.....	7
1.1.5 Ventajas y desventajas de los Almacenes de datos. ....	7
1.1.6 Componentes de los Almacenes de datos. ....	8
1.2 Metodologías para el desarrollo de Almacenes de datos.....	9
1.2.1 Metodología a utilizar en la solución.....	9
1.3 Herramientas para el desarrollo de Almacenes de datos.....	11
1.4 Conclusiones.....	13
Capítulo 2: Análisis y diseño del Mercado de datos: Análisis de Tendencias. ....	14
2.1 Estudio preliminar del negocio. Caracterización del área PPQA.....	14
2.2 Necesidades de usuario. ....	17
2.3 Requisitos de información. ....	18
2.4 Requisitos funcionales.....	19
2.5 Requisitos no funcionales.....	21
2.6 Reglas de negocio.....	22
2.7 Modelo de casos de uso del sistema.....	24
2.8 Arquitectura del Mercado de datos. ....	27
2.8.1 Subsistema de Almacenamiento.....	27
2.8.2 Subsistema de Integración.....	30
2.8.3 Subsistema de Visualización.....	33

2.9 Políticas de respaldo y recuperación. ....	35
2.10 Conclusiones.....	36
Capítulo 3: Implementación y prueba del Mercado de datos: Análisis de Tendencias.....	37
3.1 Implementación del Subsistema de Almacenamiento.....	37
3.1.1 Estándar de codificación.....	37
3.2 Implementación del Subsistema de Integración de datos.....	38
3.2.1 Transformaciones y trabajos.....	39
3.3 Implementación del Subsistema de Visualización.....	41
3.3.1 Implementación de los cubos OLAP.....	41
3.3.2 Implementación de la capa de visualización.....	42
3.4 Pruebas.....	46
3.4.2 Resultados de las pruebas.....	49
3.5 Conclusiones.....	50
Conclusiones generales.....	51
Recomendaciones.....	52
Referencias bibliográficas.....	53
Bibliografía.....	55
Anexos.....	58

**Índice de Tablas**

Tabla 1: Distribución de las áreas de análisis. ....	18
Tabla 2: Descripción de los RF. ....	19
Tabla 3: Descripción de los actores del sistema. ....	24
Tabla 4: Especificación del CU: "Mostrar información sobre las NC de revisiones".....	25
Tabla 5: Descripción de las tablas de hechos. ....	28
Tabla 6: Descripción de las tablas de dimensiones.....	28
Tabla 7: Matriz bus del MD Análisis de Tendencias. ....	29
Tabla 8: Reporte candidato "Cantidad de revisiones" del A.A. Revisiones.....	34
Tabla 9: Reporte candidato "Porcentaje de respuestas sí según criterios" del A.A Utilidad.....	35
Tabla 10: Reporte candidato "Porcentaje de error de NC según clasificación" del A.A. Liberaciones. ....	35
Tabla 11: Roles y permisos.....	36
Tabla 12: Estándar de codificación para el MD. ....	37
Tabla 13: Diseño de CP para el CU "Mostrar información sobre el plan de revisiones.".....	47
Tabla 14: Resultados de las pruebas de sistema. ....	49

## Índice de Figuras

Fig. 1: Ciclo de vida de la metodología. ....	10
Fig. 2: Diagrama de CU del sistema.....	25
Fig. 3: Arquitectura del MD. ....	27
Fig. 4: Modelo de datos reducido. ....	30
Fig. 5: Distribución por tipo de datos. ....	31
Fig. 6: Distribución de datos nulos para la tabla gespro_ppqa_project_bugs. ....	32
Fig. 7: Diseño de las transformaciones para la carga de dimensiones. ....	32
Fig. 8: Diseño de las transformaciones para la carga de hechos. ....	33
Fig. 9: Diseño del mapa de navegación. ....	34
Fig. 10: Proceso de integración de datos para el hecho plan de revisión y liberación. ....	39
Fig. 11: Proceso de integración de datos para la dimensión estado.....	40
Fig. 12: Trabajo para las dimensiones. ....	40
Fig. 13: Trabajo para los hechos.....	41
Fig. 14: Implementación de los cubos OLAP.....	41
Fig. 15: Arquitectura de información del MD Análisis de Tendencias. ....	43
Fig. 16: Reporte “Adherencia a procesos y productos de trabajo” a nivel de centro. ....	44
Fig. 17: Impacto y Adherencia de las NC usando Dashboard. ....	45
Fig. 18: Modelo V.....	46
Fig. 19: Comportamiento de los indicadores de la Lista de Chequeo para el Mapa Lógico. ....	49
Fig. 20: Resultados de las pruebas. ....	50

### **Introducción**

Actualmente el mundo de la informática ha tenido un avance significativo, el cual puede apreciarse en todas las esferas de la sociedad, siendo el desarrollo de *software* un sector fundamental como centro de todas las grandes transformaciones como la economía digital, la administración del conocimiento y la evolución de las empresas para la automatización de los procesos de gestión y control de la información (Peñaloza 2002).

Dentro del desarrollo de *software* se tienen en cuenta diferentes aspectos como el cronograma, la planificación, el presupuesto, las personas, el alcance y la calidad, definiéndose esta última como “la concordancia con los requisitos funcionales y de desempeño explícitamente establecidos, estándares de desarrollo explícitamente documentados y características implícitas que se espera de todo software desarrollado profesionalmente” (Pressman 1998).

Como parte de las tendencias para el aseguramiento de la calidad en el mundo existe una variedad de modelos entre los que se encuentran: Normas ISO/IEC 12007, Metrical, ISO 15504 y CMMI (Capability Maturity Model Integration), siendo este último un enfoque no radical de mejoras de procesos que proveen a las organizaciones de los elementos esenciales para un proceso efectivo, midiendo así la madurez del desarrollo del *software* en una escala del 1 al 5 (Rosales 2014).

En Cuba existen instituciones y centros investigativos que se dedican al desarrollo de aplicaciones informáticas para contribuir a la informatización de todo el país. Una de estas instituciones es la Universidad de las Ciencias Informáticas (UCI), la cual tiene entre sus proyecciones, desarrollar capacidades productivas que generen productos y servicios informáticos de calidad, sostenibles y con alto valor agregado, según se describe en los objetivos estratégicos de la universidad para el año 2016.

La universidad está inmersa en un proceso de madurez con respecto a los productos que se generan, por lo que se aplica la Actividad de Desarrollo-Producción sobre la base del nivel 2 del modelo de referencia CMMI en su versión 1.3, que consiste en institucionalizar un proceso gestionado bajo el cumplimiento de las políticas y prácticas tanto genéricas como específicas de cada una de las áreas de proceso de dicho nivel (UCI 2014), tales como:

- Planeación de Proyecto (PP).
- Monitoreo y Control de Proyecto (PMC).
- Administración de Requisitos (REQM).
- Administración de Configuración (CM).
- Medición y Análisis (MA).
- Aseguramiento de la Calidad de los Procesos y Productos (PPQA).

En esta última se espera que el aseguramiento de la calidad de los procesos y productos ofrezca reportes que permitan a la organización y a los proyectos, tener una visión objetiva de la calidad de los productos y

servicios y de los procesos ejecutados para desarrollarlos. Adicionalmente se espera que constituya una fuente de información para el apoyo a la toma de decisiones estratégicas y la mejora continua de la actividad de desarrollo-producción, ofreciendo reportes sobre la utilidad de los procesos y productos de trabajo (UCI 2014).

Esta tarea se lleva a cabo a través del Sistema de Gestión de Proyectos (GESPRO), donde se tiene control de todo el ciclo de vida de los proyectos, y el Gestor de Documentos Administrativos (EXCRIBA), donde se almacena la Encuesta de Utilidad, la cual está conformada por un conjunto de preguntas que verifican la utilidad de los procesos establecidos en la universidad a opinión del proyecto, que es el que mayormente implementa los mismos. A partir de lo anterior se realiza el Análisis de Tendencias a nivel de centro y universidad, el cual tiene como objetivo reunir toda la información de las actividades de calidad efectuadas en un período y tras el análisis de dicha información definir tendencias sobre el comportamiento de las no conformidades (NC), para colaborar así en la toma de decisiones. En este proceso se genera un Informe de Tendencias, documento emitido por los asesores de calidad de cada entidad desarrolladora y el coordinador de calidad de la alta gerencia, a partir de los 12 indicadores para medir la calidad en el proceso productivo de la UCI, que son:

1. Estado del plan de revisiones.
2. Tendencia de las NC de revisiones.
3. Impacto de las NC y Adherencia a procesos y productos de trabajo.
4. Análisis de causa.
5. Estado de las NC de las evaluaciones.
6. Apreciación de la utilidad de los procesos y productos de trabajo.
7. Estado de las expectativas de la organización.
8. Estado del plan de liberaciones.
9. No conformidades por tipo de error según su clasificación.
10. Liberaciones por iteraciones.
11. NC por iteraciones por tipo de error según su clasificación.
12. Estado de las NC por iteraciones.

Actualmente estos informes se realizan de forma manual, proceso en el que el asesor de la calidad de la entidad desarrolladora debe acceder al GESPRO y al EXCRIBA, luego contabiliza, de cada proyecto, la información asociada a cada uno de los indicadores antes mencionados. Una vez obtenidos dichos datos, los mismos son introducidos en tablas realizadas en Excel, herramienta privativa del paquete Microsoft Office y como resultado se generan gráficos que conforman el Informe de Tendencia, documento generado en formato PDF. Esta tarea se torna engorrosa debido a que los informes se generan trimestralmente,

trayendo consigo grandes cúmulos de datos, lo que dificulta el proceso de obtención de información; además, dicha información no se encuentra integrada en un mismo sistema, por lo que se hace difícil su análisis, provocando un aumento del costo de tiempo y recursos utilizados para la generación del mismo. Una vez conformado el informe a nivel de centro este es enviado al coordinador de calidad de la Alta Gerencia, el cual recibe un Informe de Tendencia por cada uno de los 14 centros productivos existentes en la universidad, para entonces generar el Informe de Tendencia UCI. Dicho informe es utilizado por la Alta Gerencia para posteriores análisis y como apoyo a la toma de decisiones.

Por todo lo antes descrito se plantea como **problema de la investigación**: ¿Cómo contribuir al proceso de toma de decisiones en el Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI?

Se define como **objeto de estudio**: los almacenes de datos, enmarcado en el **campo de acción**: Mercado de datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI.

Para darle solución al problema planteado, se define como **objetivo general**: Desarrollar el Mercado de datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI, que contribuya al proceso de toma de decisiones, del cual se desglosan los siguientes **objetivos específicos**:

1. Fundamentar la metodología y herramientas a utilizar en el desarrollo del Mercado de datos: Análisis de Tendencias.
2. Realizar el análisis y diseño del Mercado de datos: Análisis de Tendencias.
3. Realizar la implementación del Mercado de datos: Análisis de Tendencias.
4. Realizar las pruebas al Mercado de datos: Análisis de Tendencias.

Con el propósito de cumplir con el objetivo general planteado fueron formuladas las siguientes **preguntas científicas**:

1. ¿Cuáles son los fundamentos teóricos relacionados con el estado del arte de los Almacenes de datos?
2. ¿Qué metodología, herramientas y tecnologías utilizar para llevar a cabo la implementación del Mercado de datos: Análisis de Tendencias?
3. ¿Qué características debe cumplir el Mercado de datos: Análisis de Tendencias para que cumpla con las necesidades de los clientes?
4. ¿Cómo desarrollar el Mercado de datos: Análisis de Tendencias de manera que contribuya al proceso de toma de decisiones?
5. ¿Cómo probar el Mercado de datos: Análisis de Tendencias con el fin de garantizar el correcto funcionamiento del mismo?

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

1. Estudio y análisis de la metodología y herramientas a utilizar para guiar el desarrollo de Almacenes de Datos.
2. Levantamiento de requisitos para identificar las necesidades del cliente.
3. Diseño de los casos de uso del Mercado de datos para identificar la relación entre actores y casos de uso del sistema.
4. Definición de los hechos, dimensiones y medidas del Mercado de datos para el diseño del modelo de datos.
5. Diseño del modelo de datos para identificar la relación entre las tablas de hechos y dimensiones candidatas en la solución.
6. Definición de la arquitectura del Mercado de datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI, para definir los subsistemas que forman parte de la solución.
7. Diseño del Subsistema de Integración para diseñar los procesos de carga de las dimensiones y hechos.
8. Diseño del Subsistema de Visualización para realizar el diseño de los cubos OLAP y de los reportes candidatos.
9. Diseño de los casos de pruebas para aplicarlos posteriormente durante las pruebas del sistema.
10. Implementación del modelo de datos para cumplir con el diseño de la estructura de la base de datos.
11. Implementación del Subsistema de Integración para poblar el Mercado de datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI.
12. Implementación del Subsistema de Visualización para mostrar los reportes contenidos dentro de los libros de trabajo según el área de análisis identificada.
13. Aplicación de las listas de chequeo para comprobar el correcto funcionamiento de los subsistemas de integración y visualización.
14. Aplicación de los casos de prueba para garantizar el correcto funcionamiento del sistema.

### **Métodos teóricos**

- **Histórico-Lógico:** Para el presente trabajo se realizó un estudio del estado del arte de la problemática y de los almacenes de datos en cuanto a ventajas, desventajas, antecedentes y características de los mismos.
- **Analítico-Sintético:** Se utilizó para el análisis de documentos, materiales y temas relacionados con el desarrollo de mercados de datos y de elementos claves de la Actividad de Desarrollo-Producción UCI. 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 del impacto de los resultados del sistema desarrollado como parte de la validación y cumplimiento de los objetivos de la presente investigación.

- **Modelación:** Permitió desarrollar los modelos correspondientes al ciclo de vida de desarrollo del mercado de datos.

### **Métodos empíricos**

- **Recolección de información:** Se utilizó mediante la aplicación de técnicas de recolección de información como **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.

### **Estructura del trabajo de diploma**

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

#### **Capítulo 1: Fundamentación teórica de los Almacenes de datos.**

En este capítulo se realizará el análisis del estado del arte de los almacenes de datos y los mercados de datos, así como definición, características, ventajas y desventajas. También se realizará el estudio de las metodologías y herramientas a utilizar en la solución.

#### **Capítulo 2: Análisis y diseño del Mercado de datos: Análisis de Tendencias.**

En este capítulo se realizará un estudio del negocio, profundizándose en algunos temas relacionados con la gestión de la calidad en la UCI. 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 la solución. Además, se diseñarán los Subsistemas de Almacenamiento, Integración y Visualización de los datos.

#### **Capítulo 3: Implementación y prueba del Mercado de datos: Análisis de Tendencias.**

Este capítulo contendrá la implementación de los Subsistemas de Almacenamiento, Integración y Visualización, teniendo en cuenta los requisitos y necesidades del cliente. Contendrá, además, los resultados de los casos de pruebas y las listas de chequeo aplicadas.

## **Capítulo 1: Fundamentación teórica de los Almacenes de datos.**

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

### **1.1 Almacenes de datos.**

#### **1.1.1 Antecedentes de los Almacenes de datos.**

Durante los años 80 el uso de las tecnologías de la información estuvo enfocado en automatizar los procesos de tipo repetitivo o administrativo, haciendo uso de Sistemas de Información Operacionales o Transaccionales en Línea (OLTP). La función de estos sistemas consiste en recoger y almacenar los datos generados a partir de las aplicaciones informáticas que se utilizan en los distintos departamentos de la organización.

Los sistemas OLTP están diseñados para resolver los problemas de carácter operativo, lo que significa que el tipo de decisiones que soportan están relacionadas con la ejecución y apoyo de las tareas básicas del negocio. Por tal motivo, estos no son útiles para realizar análisis de tipo estratégico.

Para mediados de los años 80 los requerimientos de información en las empresas empezaron a cambiar, ya que se realizaban más consultas que precisaban de más informes. Los departamentos de informática no eran capaces de satisfacer las crecientes demandas de información por lo que surge la necesidad de crear nuevas bases de datos centrales con información resumida y detallada.

A mediados de 1985 aparece el primer AD, su finalidad era la de integrar y organizar los datos que estaban dispersos en múltiples bases de datos, dentro de la compañía. Este primer almacén de datos lo construyó Bill Inmon para un banco en Colorado (Soto 2001).

#### **1.1.2 Definición de Almacenes de datos.**

Con el creciente desarrollo de la informática a nivel mundial se ha evidenciado cómo, con el manejo de grandes volúmenes de información y el análisis de datos que se puede extraer de ellos, ha aumentado aún más la cantidad de información que se genera y procesa, lo que complejiza su almacenamiento. A raíz de esta problemática, el hombre ha sido capaz de desarrollar aplicaciones que dieran solución a lo anteriormente planteado, donde una de estas soluciones son los AD, los cuales sirven de apoyo al proceso de toma de decisiones y contienen información que se puede separar por áreas de análisis o unidades lógicas, llamadas MD.

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

Ralph Kimball, también aportó un concepto de AD, definiéndolo como “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis” (Kimball 2008).

Dadas las definiciones anteriores, en la presente investigación se define el término de AD como una colección de datos estructurados, orientado a temas, integrados, no volátiles y variables en el tiempo, donde la información es almacenada para el apoyo a la toma de decisiones y puede ser agrupada en unidades lógicas llamadas mercado de datos.

### **1.1.3 Mercado de datos.**

Carmen Wolff define el mercado de datos como: “un conjunto de hechos y datos organizados para soporte decisional basados en la necesidad de un área o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un departamento dado, teniendo solo sentido para el personal de ese departamento y sus datos no tienen por qué tener las mismas fuentes que los de otro mercado de datos” (Wolff 2002).

Los MD tienen las siguientes características planteadas por (Ramos 2009):

- Es un subconjunto de un Almacén de Datos (*Datawarehouse* o *DWH*).
- Optimizado para consultas específicas.
- Altamente resumizado.
- Específicas funciones del negocio.
- Orientado a un grupo de usuarios.

### **1.1.4 Características de los Almacenes de datos.**

Según (Inmon 2005) en su definición de AD plantea, cuatro características en término de repositorio de datos:

- **Orientado a temas:** Los datos en la base de 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:** Los datos de todos los sistemas operacionales de la organización se mantienen en las bases de datos de manera consistente.

### **1.1.5 Ventajas y desventajas de los Almacenes de datos.**

La implementación de un AD para solucionar diversos problemas relacionados con el almacenamiento de los datos es muy recomendable ya que según (Orallo 2010), estos:

## *Capítulo 1: Fundamentación teórica de los Almacenes de datos*

---

- Hacen más fácil el acceso a una gran variedad de datos a los usuarios finales.
- Facilitan el funcionamiento de las aplicaciones de los sistemas de apoyo a la decisión tales como informes de tendencia, por ejemplo: obtener los *items* con la mayoría de las ventas en un área en particular dentro de los últimos dos años; informes de excepción, informes que muestran los resultados reales frente a los objetivos planteados a priori.
- Los almacenes de datos pueden trabajar en conjunto y, por lo tanto, aumentar el valor operacional de las aplicaciones empresariales, en especial la gestión de relaciones con clientes.

Así como la implementación de un AD es ventajoso y recomendable, (Orallo 2010) también plantea algunos inconvenientes o desventajas, tales como:

- Los AD no suelen ser estáticos, por lo que su mantenimiento puede tener un costo muy elevado y quedar obsoletos relativamente pronto.
- En ocasiones, devuelven información sub óptima ante una petición de esta, por lo que puede traer consigo pérdida para la organización.
- Debido a la poca relación entre los AD y los sistemas operativos, se hace necesario determinar qué funcionalidades de estos últimos se pueden aprovechar y cuáles implementar en un AD.

### **1.1.6 Componentes de los Almacenes de datos.**

Según (Perdomo 2009), un AD está compuesto por cuatro elementos fundamentales:

**Sistemas fuentes operacionales:** La principal función de estos sistemas es capturar las transacciones del negocio, es por ello que son utilizadas por las compañías o empresas para gestionar las transacciones diarias, operaciones que son almacenadas en fuentes de diversos formatos como base de datos relacional u otros tipos de ficheros según las necesidades del negocio.

Las prioridades principales de este componente son el procesamiento, el rendimiento y la disponibilidad.

**Área de procesamiento:** Este componente consiste en la realización del proceso de Extracción, Transformación y Carga de los datos (ETL), por lo que se invierte la mayor cantidad de tiempo y recurso durante la construcción del AD.

**Área de presentación:** Luego de realizado el proceso de ETL, los datos ya están listos para ser consultados, analizados y hacer reportes sobre ellos. En esta área la información se encuentra diseñada mediante esquemas dimensionales, siendo de gran utilidad para los usuarios durante el proceso de toma de decisiones.

**Herramienta de acceso a datos:** Este componente tiene como misión, consultar la información que se encuentra en el área de presentación del AD. Esta puede ser desde una simple herramienta de consulta de datos hasta una compleja aplicación de modelado o minería de datos.

### **1.2 Metodologías para el desarrollo de Almacenes de datos.**

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 se basa en los conceptos de diseño de bases de datos relacionales, utilizando un enfoque descendente (Top\_Down), donde se implementa el AD y luego los MD que se desarrollen para cada área de la empresa.

Sin embargo, Kimball define una metodología utilizando un enfoque ascendente, en el que se implementan los MD para luego integrarlos en el AD. Este enfoque se centra en el modelo dimensional no normalizado.

#### **1.2.1 Metodología a utilizar en la solución.**

Para el desarrollo de la presente investigación se decidió utilizar la “Metodología para el desarrollo de proyectos de Almacenes de datos”, definida por Yanisbel González Hernández en su tesis de maestría y utilizada por el Centro de Tecnologías de Gestión de Datos (DATEC) para el desarrollo de AD. Dicha metodología toma como base la metodología de Kimball para definir los aspectos específicos del desarrollo de AD, la Guía para los Fundamentos de la Dirección de Proyectos (Guía del PMBOK) para incorporar los principios básicos que permiten una adecuada gestión del proyecto. Además, los temas asociados a CMMI se incorporan a partir del Programa de Mejora, por lo tanto hereda algunos de sus enfoques, artefactos y actividades (Hernández 2013).

Esta metodología define que el diseño y construcción de los AD, consta de siete fases y el flujo de trabajo Gestión de Proyecto, el cual se ejecuta durante todo el ciclo de vida del proyecto. De las siete fases, algunas se ejecutan de forma paralela, tal es el caso de las fases Requisitos y Arquitectura, y la fase Diseño e Implementación, donde también se ejecutan componentes en paralelo, lo que facilita que el desarrollo del AD sea más ágil. La Fig. 1 muestra el ciclo de vida de la metodología.

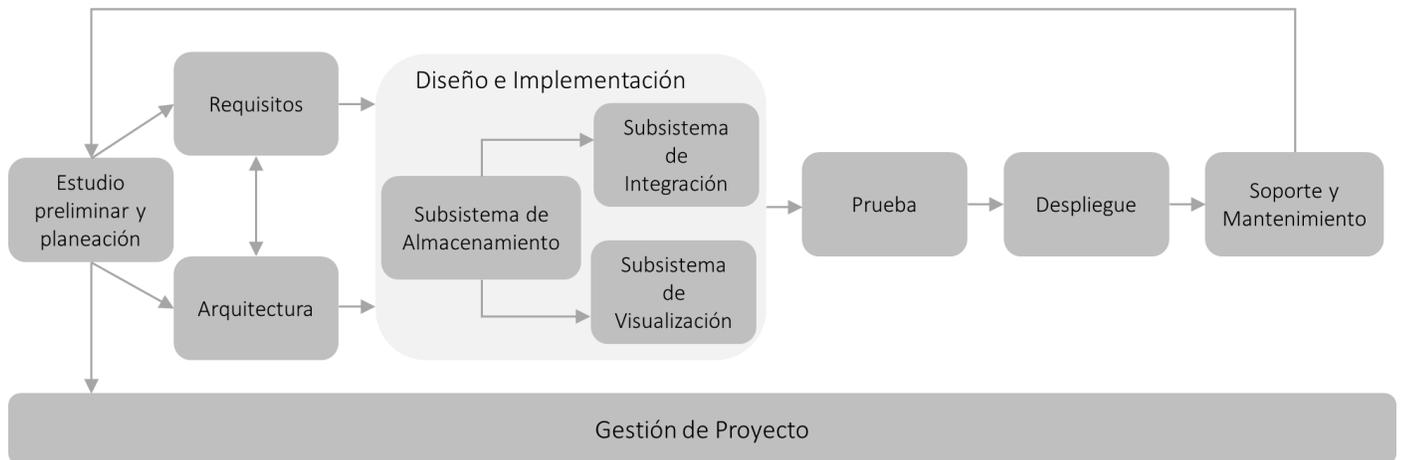


Fig. 1: Ciclo de vida de la metodología.

**Estudio preliminar y 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, etc.

**Requisitos:** En esta fase se realizan entrevistas al cliente para identificar las necesidades del mismo, y a partir de estas realizar el levantamiento de los requisitos tanto de información, como funcionales y no funcionales. También se realiza el análisis de los requisitos definidos, lo que permitirá identificar las estructuras bases del modelo lógico dimensional.

**Arquitectura:** Se definen las vistas arquitectónicas de la solución y aspectos como la seguridad, la tecnología, los subsistemas y la comunicación entre ellos, etc.

**Diseño e implementación:** En esta fase se diseñan e implementan los tres subsistemas que conforman un AD, definiéndose el diseño de las estructuras de almacenamiento de datos, los procesos de integración, como el mapa lógico, los cubos OLAP, y se implementa el repositorio de datos, la integración de datos y la presentación de los mismos.

**Prueba:** En esta fase se realizan pruebas con el objetivo de validar la calidad del producto. Estas pruebas son de tipo Unitarias que permiten comprobar el funcionamiento de cada subsistema o componente, pruebas de integración, para verificar la correcta integración de los componentes y subsistemas, y pruebas de sistema, donde se valida si fueron cumplidos los requisitos de información y funcionales definidos por el cliente.

**Despliegue:** Esta fase se divide en dos etapas. Una primera etapa, que consiste en un despliegue piloto, donde se configuran los servidores, se instalan las herramientas en base a la arquitectura definida y se carga una muestra de los datos en un ambiente controlado que le demostrará al cliente que la solución

funciona, para luego realizar la carga histórica de los datos. La segunda etapa consiste en la capacitación y transferencia tecnológica de la solución al cliente. Esta etapa tiene como finalidad, una solución desplegada en un entorno real y con un correcto funcionamiento.

**Soporte y mantenimiento:** Esta fase comienza cuando la solución está implementada y en explotación. Tiene como objetivo evitar que el sistema quede fuera de servicio debido a fallos en su funcionamiento u obsoleto.

**Gestión de proyecto:** Es un flujo de trabajo que se ejecuta durante todo el ciclo de vida del proyecto, constituyendo la columna vertebral del mismo, pues es aquí donde se controla, gestiona y chequea el desarrollo del AD, el cronograma y otras actividades de gestión y administración de proyecto.

### **1.3 Herramientas para el desarrollo de Almacenes de datos.**

Para llevar a cabo una implementación de la solución que cumpla con las expectativas de los clientes, es de vital importancia la selección de las herramientas para la creación de AD.

#### ➤ **Herramientas para el modelado de los datos.**

En el mundo existe una variedad de herramientas que apoyan a los analistas y desarrolladores de *software* durante todo el ciclo de vida de un proyecto, una de ellas es Visual Paradigm for UML, la cual fue utilizada para el modelado de los datos de la presente investigación.

**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).

#### ➤ **Sistema Gestor de Bases 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 9.2:** Posee características técnicas que la hacen una de las bases de datos más potentes y robustas del mercado. Las características que más se han tenido en cuenta durante su desarrollo son la estabilidad, potencia, robustez, facilidad de administración e implementación de estándares. Utiliza el

## *Capítulo 1: Fundamentación teórica de los Almacenes de datos*

---

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 (Martínez 2013).

### ➤ **Administrador de Bases de Datos.**

PgAdmin III es una herramienta de código abierto que se puede utilizar en Linux, Solaris, FreeBSD, MacOS y las plataformas de Windows para administrar PostgreSQL 7.3 y superior, así como las versiones comerciales y derivados de PostgreSQL.

PgAdmin III 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. Permite acceder a todas las funcionalidades de las bases de datos; consultas, manipulación y gestión de datos, incluso opciones avanzadas como manipulación de motor de replicación Slony-I (PgAdmin 2016). En el presente trabajo se utilizó PgAdmin III en su versión 1.16.1.

### ➤ **Herramienta para el perfilado de datos.**

El perfilado de datos constituye una actividad fundamental para auditar la calidad de los datos, la gestión de datos maestros o programa Data Governance.

En la presente investigación fue utilizada DataCleaner en su versión 3.1. Esta herramienta permite analizar la calidad de la información que se encuentra almacenada en fuentes como: base de datos Oracle, MySQL, PostgreSQL, ficheros Excel, XML, etc. Encuentra pautas, valores perdidos, juegos de caracteres y otras características de los valores de la información. Además, realiza comparaciones de tablas y columnas, lo que permite verificar la consistencia y veracidad de los datos (DataCleaner 2008).

### ➤ **Herramientas para la Extracción, Transformación y Carga (ETL).**

Las herramientas de ETL sirven para facilitar los procesos de extracción, transformación y carga de cualquier sistema, y son especialmente útiles para el AD, cuando se construyen sistemas de AD en proyectos de Inteligencia de Negocios (BI). Una de estas herramientas es Pentaho Data Integration (PDI), la cual fue utilizada para la realización de los procesos de ETL del presente trabajo en su versión 6.0.

**Pentaho Data Integration (PDI):** Herramienta de código abierto y multiplataforma, permite la lectura y escritura de bases de datos, el mapeo de valores y la normalización de los datos en distintas filas y viceversa. Es aplicable a diferentes tipos de bases de datos como PostgreSQL, SQL server, MySQL, entre otras (Gravitar 2014).

### ➤ **Herramientas de Inteligencia de Negocio (BI).**

Las herramientas de Inteligencia de Negocios facilitan un mejor entendimiento y análisis de los datos, apoyando así el proceso de toma de decisiones en las organizaciones (ConexiónEsan 2015). A continuación se presentan las herramientas de BI utilizadas en la presente investigación.

**Pentaho Schema Workbench 3.10.0.1:** 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 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).

### **1.4 Conclusiones.**

En el presente capítulo se analizaron los elementos esenciales de los almacenes de datos (AD) y los mercados de datos (MD), así como sus definiciones, características, ventajas y desventajas. Se describieron las metodologías existentes para la implementación de una solución de AD. Se fundamentó la selección de la “Metodología para el desarrollo de proyectos de Almacenes de datos”, como guía para el desarrollo de Mercado de datos: Análisis de Tendencias. También se realizó la selección de las herramientas que se utilizarán durante el desarrollo del mercado de datos.

### **Capítulo 2: Análisis y diseño del Mercado de datos: Análisis de Tendencias.**

En este capítulo se realizará un estudio del negocio, profundizándose en algunos temas relacionados con la gestión de la calidad en la UCI. Se definirán los requisitos de información según las necesidades de los clientes, para luego diseñar el diagrama de casos de uso del sistema y las reglas del negocio. Se identificarán los hechos, medidas y dimensiones quedando conformado el modelo de datos de la solución. Además, serán definidos y diseñados los Subsistemas de Almacenamiento, Integración y Visualización de los datos.

#### **2.1 Estudio preliminar del negocio. Caracterización del área PPQA.**

El área de Aseguramiento de Calidad de Procesos y Productos de trabajo (PPQA) se encarga de la planificación y ejecución de evaluaciones, así como del monitoreo de las no conformidades (NC), de manera que permiten a los asesores de calidad de cada uno de los centros de la UCI y al coordinador de calidad de la Alta Gerencia, tener una visión objetiva de la calidad de los productos y servicios. Para esto, se realizan, una serie de actividades relacionadas con el seguimiento y control de las NC detectadas en dichas revisiones, para el posterior Análisis de Tendencias.

El Análisis de Tendencias es realizado por los asesores de calidad a nivel de centro y los coordinadores de la alta gerencia y comprende todas las actividades de calidad efectuadas en un período, con el fin definir tendencias sobre el comportamiento de las NC. En el Anexo #1 se muestra una imagen con la descripción más detallada de este proceso.

De este análisis se realiza el Informe de Tendencias, el cual es emitido trimestralmente, donde se recoge toda la información del Análisis de Tendencias y se muestra en base a los 12 indicadores para medir la calidad de los proyectos en la UCI, de manera que sirva de guía para la toma de decisiones en cuanto a comparaciones que se realizan con períodos similares de años anteriores. Estos indicadores son los siguientes:

1. **Estado del plan de revisiones:** Consiste en analizar la cantidad de evaluaciones planificadas, ejecutadas, aplazadas, abortadas, replanificadas y fuera de plan, del Plan de Evaluaciones Gerencial a partir del seguimiento que se le realiza con el fin de controlarlas. Tiene como objetivo analizar la tendencia de estas evaluaciones de tipo RI (Revisiones de inconsistencia), AC (Auditorías a la configuración), RTF (Revisiones técnicas formales) y las PPQA (Aseguramiento de la calidad de productos y procesos) a nivel de proyecto, centro, universidad y en general.
  - **Medida base:** Cantidad de Evaluaciones por estado (e), donde (e) puede ser planificadas, ejecutadas, aplazadas, abortadas, re-planificadas y fuera de plan.

## *Capítulo 2: Análisis y diseño del Mercado de Datos: Análisis de tendencias*

---

2. **Tendencia de las no conformidades de revisiones:** Consiste en analizar la tendencia de las No Conformidades detectadas en las evaluaciones con el fin de caracterizar la adherencia a procesos y productos. De esta manera se muestra la media de revisiones PPQA en los tres niveles y la media RTF de Planificación, Requisitos, Arquitectura y Base de datos.

- **Medida Base:** Cantidad de NC por evaluación (e), donde (e) es PPQA (Proyecto, Centro y UCI) y RTF (Planificación, Requisitos, Arquitectura, Base de Datos).
- **Medida Base:** Cantidad de evaluaciones (e).
- **Medida Derivada:**

$$\text{Media de NC por evaluación (e)} = \frac{\text{Cantidad de NC por evaluación (e)}}{\text{Cantidad de evaluaciones (e)}}$$

3. **Impacto de las NC y Adherencia a procesos y productos de trabajo:** Consiste en analizar por áreas de procesos el impacto de las NC detectadas en las evaluaciones PPQA a nivel de proyecto y centro, con el fin de caracterizar la adherencia de productos y procesos.

- **Medida Base:** Cantidad de NC por área (a) y por tipo (p), **p** es la clasificación Proceso o Producto, **a** es PP, PMC, MA, REQM, PPQA y CM, además de las políticas complementarias (PC) y las políticas genéricas (PG).
- **Medida Base:** Cantidad de NC por área (a) y por impacto (i), **i** es Alto, Medio, Bajo.

4. **Análisis de causa:** Consiste en analizar las NC por procesos y productos de trabajo identificadas durante las evaluaciones con el fin de identificar las causas que producen estos problemas y tomar acciones para mitigarlos.

- **Medida Base:** Cantidad de NC por causa (x)
- **Medida Base:** Cantidad de evaluaciones (e).
- **Medida Derivada:**

$$\text{Media de NC por causa (x)} = \frac{\text{Cantidad de NC por causa (x)}}{\text{Cantidad de evaluaciones (e)}}$$

5. **Estado de las NC de las evaluaciones:** Consiste en analizar el estado de las NC identificadas durante las evaluaciones RTF, PPQA (proyecto, centro, universidad), RI y AC.

- **Medida Base:** Cantidad de NC por estado (n), por evaluación (e), donde el estado puede ser abierta, pendiente, resuelta, atrasada, cerrada.

6. **Apreciación de la utilidad de los procesos y productos de trabajo:** Consiste en analizar la utilidad de los procesos y productos de trabajo de cada área de proceso y los criterios de estas, con el fin de mejorarlos.

- **Medida Base:** Cantidad de respuestas positivas de cada criterio (S).
- **Medida Base:** Cantidad de respuestas negativas de cada criterio (N).

- **Medida Base:** Cantidad de respuestas de cada criterio (T).
  - **Medida Derivada:** Por ciento de respuestas positivas de cada criterio.  $P_s = S * 100/T$
  - **Medida Derivada:** Por ciento de respuestas negativas de cada criterio.  $P_n = N * 100/T$
7. **Estado de las expectativas de la organización:** Consiste en analizar el cumplimiento de las expectativas de la organización con el fin de mejorar los productos y procesos.
- **Medida Base:** Cantidad de respuestas positivas de cada expectativa (S).
  - **Medida Base:** Cantidad de respuestas negativas de cada expectativa (N).
  - **Medida Base:** Cantidad de respuestas de cada expectativa (T).
  - **Medida Derivada:** Por ciento de respuestas positivas de cada Expectativa.  $P_s = S * 100/T$
  - **Medida Derivada:** Por ciento de respuestas negativas de cada Expectativa.  $P_n = N * 100/T$
8. **Estado del plan de liberaciones:** Consiste en analizar la cantidad de liberaciones planificadas, ejecutadas, detenidas, aplazadas, abortadas, replanificadas y fuera de plan del Plan de Liberaciones a partir del seguimiento que se le realiza con el fin de controlarlas.
- **Medida Base:** Cantidad de liberaciones por estado (e) por trimestre (n), donde el estado puede ser, planificadas, ejecutadas, aplazadas, abortadas, detenidas, replanificadas, fuera de plan.
9. **No conformidades por tipo de error según su clasificación:** Consiste en analizar por la clasificación tipo de artefacto, la cantidad de no conformidades por tipo de error detectadas en el proceso de pruebas de liberación, con el fin de minimizar la cantidad de no conformidades.
- **Medida Base:** Cantidad de NC por clasificación.
  - **Medida Base:** Cantidad de NC por clasificación por Tipo de error (n).
  - **Medida Derivada:**

$$\text{Porcentaje de NC}(n) = \frac{\text{Cantidad de NC por clasificación por tipo de error}(n) * 100}{\text{Cantidad de NC por clasificación}(n)}$$

10. **Liberaciones por iteraciones:** Consiste en analizar por la clasificación tipo de artefacto, la cantidad de productos liberados por iteraciones en el proceso de liberación, con el fin de identificar las causas que producen estos problemas y tomar acciones para mitigarlos.

- **Medida Base:** Cantidad de productos liberados por clasificación en la iteración (n).
- **Medida Base:** Cantidad de productos liberados por clasificación.

Nota: los productos liberados pueden ser de documentación o aplicación.

- **Medida Derivada:**

$$\text{Porcentaje de Productos liberados}(n) = \frac{\text{Cantidad de productos liberados en la iteración}(n) * 100}{\text{Cantidad de productos liberados por clasificación}(n)}$$

11. **NC por iteraciones por tipo de error según su clasificación:** Consiste en analizar la cantidad de NC por iteraciones en el proceso de liberación con el fin de identificar las causas que producen estos problemas y tomar acciones para mitigarlos.
- **Medida Base:** Cantidad de NC de aplicación por tipo de error (e) por iteración (n).
  - **Medida Base:** Cantidad de NC de documentación por tipo de error (e) por iteración (n).
12. **Estado de las NC por iteraciones:** Consiste en analizar la cantidad y el estado de las no conformidades por iteraciones en el proceso de liberación con el fin de identificar las causas que producen estos problemas y tomar acciones para mitigarlos.
- **Medida Base:** Cantidad de NC por clasificación resuelta en la iteración (n).
  - **Medida Base:** Cantidad de NC por clasificación pendiente en la iteración (n).
  - **Medida Base:** Cantidad de NC por clasificación no procede en la iteración (n).

Los indicadores 6 y 7 recogen las estadísticas resultantes de las encuestas que se realizan sobre la utilidad de los procesos y las expectativas de la organización en cuanto a los criterios de cada área de proceso. Estas encuestas se realizan mediante la ayuda de ficheros Excel. Para integrar la información que generan las encuestas en una misma fuente y para facilitar la realización de dichas encuestas, se hizo necesario realizar, como parte del presente trabajo, una encuesta web que gestionara las encuestas de manera que sean guardadas en una base de datos para un mejor procesamiento. Dicha aplicación fue realizada a través de la herramienta CMS (en inglés Content Management System) LimeSurvey en su versión 2.0.

El propósito de este Informe de Tendencias es proporcionar a cada centro y a la dirección de la actividad de desarrollo-producción, una perspectiva objetiva del estado de los procesos y productos de trabajo, así como tener una estadística del comportamiento de los proyectos de la universidad, para emitir criterios y tomar decisiones estratégicas en cuanto al comportamiento de los proyectos a partir de lo que genere cada indicador.

### **2.2 Necesidades de usuario.**

Para facilitar el proceso de toma de decisiones por parte del área de gestión de la calidad se requiere de la identificación de las necesidades de información que poseen los asesores de dicha área, ya que estas constituyen la base para el diseño del MD.

De esta manera se clasificó la información en tres grupos o áreas de análisis fundamentales: Revisiones, Utilidad de procesos y Liberaciones, las cuales engloban varias temáticas por las que se agrupan los indicadores antes mencionados. La Tabla 1 muestra la distribución de cada temática en las áreas de análisis y el indicador de calidad al que hace referencia.

Tabla 1: Distribución de las áreas de análisis.

Áreas de análisis	Temáticas	Indicador referente
Revisiones	Estado del plan de revisiones	1
	NC de revisiones	2, 3, 4 y 5
Utilidad de procesos	Apreciación de la utilidad de procesos	6
	Expectativas de la organización	7
Liberaciones	Estado del plan de liberaciones	8
	NC de liberaciones	9, 10, 11 y 12

### 2.3 Requisitos de información.

Los requisitos de información (RI) constituyen el eslabón fundamental para el proceso de Inteligencia de Negocio y la identificación de reportes candidatos, ya que responden a las necesidades de los clientes. En el Anexo #2 se muestran los RI firmados por el cliente.

A continuación se presentan los RI divididos en grupos o áreas de análisis identificadas.

#### RI de Revisiones.

1. Obtener la cantidad de revisiones por estado, trimestre, centro y nivel de revisión.
2. Obtener la media de NC de revisiones por trimestre, centro y nivel de revisión.
3. Obtener la cantidad de NC por impacto de la NC, áreas de proceso, trimestre, centro y nivel de revisión.
4. Obtener la cantidad de NC por el tipo de adherencia, áreas de proceso, trimestre, centro y nivel de revisión.
5. Obtener la frecuencia media de NC por tipo de causa, trimestre, centro y nivel de revisión.
6. Obtener la cantidad de NC de revisiones por estado, trimestre, centro y nivel de revisión.

#### RI de Utilidad de procesos.

7. Obtener el porcentaje de respuestas (sí/no) por área de proceso, tipo de criterio, trimestre, centro y nivel de revisión.
8. Obtener el porcentaje de respuestas (sí/no) por áreas de proceso, trimestre, centro y nivel de revisión.

#### RI de Liberaciones.

9. Obtener la cantidad de liberaciones por estado, trimestre, centro y nivel de revisión.

10. Obtener porcentaje de error de las NC por tipo de error, clasificación de las NC de liberación, trimestre, centro y nivel de revisión.
11. Obtener porcentaje de liberaciones por iteraciones, clasificación de las NC de liberación, trimestre, centro y nivel de revisión.
12. Obtener la cantidad de NC por iteraciones, tipo de error, clasificación de las NC de liberación, nivel de revisión, centro y tiempo.
13. Obtener la cantidad de NC por iteraciones, estado, clasificación de las NC de liberación, trimestre, centro y nivel de revisión.

### **2.4 Requisitos funcionales.**

Los requerimientos funcionales (RF) son capacidades o condiciones que el sistema debe cumplir. De forma general estos requisitos indican lo que debe hacer el sistema (Pressman 2002).

Para el MD Análisis de Tendencias se identificaron 30 RF, los cuales se describen en la Tabla 2.

Tabla 2: Descripción de los RF.

No.	Requisito Funcional	Descripción
RF-1	Autenticar usuario.	Consiste en introducir un usuario y contraseña válidos en el sistema e ingresar al mismo con los permisos concedidos.
RF-2	Adicionar rol.	Adiciona distintos roles con diferentes niveles de permisos a los usuarios.
RF-3	Eliminar rol.	Elimina roles pertenecientes a usuarios del sistema.
RF-4	Visualizar rol.	Visualiza la información del rol.
RF-5	Modificar rol.	Modifica roles pertenecientes a usuarios del sistema.
RF-6	Adicionar usuario.	Adiciona usuarios al sistema.
RF-7	Eliminar usuario.	Elimina usuarios del sistema.
RF-8	Visualizar usuario.	Visualiza la información del usuario.
RF-9	Modificar usuario.	Modifica usuarios del sistema.
RF-10	Adicionar reporte.	Adiciona los diferentes reportes que estarán disponibles en el sistema.
RF-11	Eliminar reporte.	Elimina reportes que están disponibles en el sistema.
RF-12	Visualizar reporte.	Visualizar los diferentes reportes que estarán disponibles en el sistema.

## *Capítulo 2: Análisis y diseño del Mercado de Datos: Análisis de tendencias*

---

<b>RF-13</b>	Modificar reporte.	Modifica los diferentes reportes que estarán disponibles en el sistema.
<b>RF-14</b>	Extraer datos.	Realiza la extracción de los datos de los ficheros o sistemas fuente.
<b>RF-15</b>	Realizar transformación y carga.	Realiza los procesos de transformación y carga de los datos extraídos.
<b>RF-16</b>	Abrir navegador OLAP.	Consiste en determinar las dimensiones que aparecen en las filas y/o columnas, así como los filtros aplicados y las métricas por aparecer.
<b>RF-17</b>	Mostrar editor MDX.	Permite visualizar y editar las consultas MDX que genera el reporte OLAP.
<b>RF-18</b>	Configurar tablas OLAP.	Permite configurar aspectos por defecto de la tabla OLAP como el orden ascendente o descendente de los elementos.
<b>RF-19</b>	Mostrar padres.	Permite mostrar y ocultar el padre del miembro de una jerarquía.
<b>RF-20</b>	Ocultar repeticiones.	Permite ocultar los datos repetidos de las filas y/o columnas.
<b>RF-21</b>	Suprimir filas.	Permite suprimir las filas que contengan valor nulo del reporte sobre el cual se quiere realizar los cambios.
<b>RF-22</b>	Intercambiar ejes.	Permite invertir los ejes del reporte sobre el cual quiere realizar los cambios.
<b>RF-23</b>	Detallar miembros.	Permite detallar los miembros del reporte sobre el cual quiere realizar los cambios.
<b>RF-24</b>	Entrar en detalles.	Permite entrar en los detalles del reporte sobre el cual quiere realizar los cambios.
<b>RF-25</b>	Mostrar datos de origen.	Permite mostrar los datos de origen del reporte sobre el cual quiere realizar los cambios.
<b>RF-26</b>	Mostrar gráfico.	Permite mostrar el gráfico del reporte sobre el cual quiere realizar los cambios.
<b>RF-27</b>	Configurar gráfico.	Permite configurar las propiedades del gráfico. Desde el tipo del mismo hasta propiedades de estilo como tipo de letra, tamaño o color.
<b>RF-28</b>	Configurar impresión.	Permite configurar los parámetros de impresión como el título, disposición de papel, tamaño, etc.
<b>RF-29</b>	Exportar a PDF.	Permite generar un PDF con el contenido del reporte.
<b>RF-30</b>	Exportar a Excel.	Permite generar un Excel con el contenido del reporte.

### **2.5 Requisitos no funcionales.**

Los requerimientos no funcionales (RNF) 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).

En el presente trabajo se identificaron 10 RNF distribuidos entre usabilidad, disponibilidad, restricciones de diseño e interfaz, como se muestra a continuación.

#### **Usabilidad**

**RNF-1** Agilizar el acceso a los reportes del mercado de datos mediante la distribución de la información por áreas de análisis.

El usuario podrá acceder de manera rápida a la información que solicita en el área correspondiente de acuerdo al objetivo de su solicitud, mediante el uso del mapa de navegación diseñado.

#### **Disponibilidad**

**RNF-2** Garantizar el cumplimiento de actualización de los datos en el mercado de datos.

La información contenida en el MD tendrá una precisión y exactitud diaria, ya que se realizará la carga de los datos todos los días a través de tareas programadas en el sistema operativo.

#### **Restricciones de diseño**

**RNF-3** Lograr la homogeneidad de la estructura de los elementos definidos en el mercado de datos.

La estructura del mercado de datos debe tener un nombre estándar teniendo en cuenta el tipo de estructura que sea: **dim\_[nombre]** para las dimensiones y **hech\_[nombre]** para los hechos.

**RNF-4** Utilizar los lenguajes de programación definidos durante la investigación

Para la realización de MD se utilizará como SGBD el PostgreSQL 9.2, a través de la herramienta para la administración pgAdmin III, utilizando el lenguaje PL/pgSQL. Para la realización de las consultas se utilizará el lenguaje MDX.

#### **Interfaz**

**RNF-5** Acceso al sistema.

El usuario deberá acceder a la aplicación mediante el protocolo HTTP, usando preferiblemente el navegador web Firefox 32.0 en adelante.

#### **Interfaces de usuario**

**RNF-6** 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 una estándar en todo el sistema.

### **Interfaces de hardware**

**RNF-7** Definir las interfaces de hardware que soportará el sistema.

El sistema interactuará con la impresora como interfaz de hardware. Esta interacción se ocasionará cuando se necesite imprimir un reporte en formato físico. El acceso a la impresora será mediante el protocolo TCP/IP a través de la interfaz que ofrece el hardware.

**RNF-8** Proporcionar características mínimas de hardware a las estaciones de trabajo.

Para lograr una explotación aceptable del sistema el hardware debe contar con los siguientes requerimientos:

- 1 GB RAM.

**RNF-9** Proporcionar características mínimas de hardware a los servidores.

Para lograr una explotación aceptable del sistema los servidores deben contar con los siguientes requerimientos de hardware:

- 4 GB RAM.
- 1 TB de disco duro.

### **Interfaces de software**

**RNF-10** 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 al menos con:

- Firefox 32.0 o superior.
- Java Virtual Machine 7.0 y Schema Workbench 3.10 en caso de que sea necesario el modelado de nuevos cubos multidimensionales.

## **2.6 Reglas de negocio.**

Las reglas de negocio (RN) especifican restricciones, políticas y normas que se deben cumplir durante el proceso de desarrollo de la solución del MD, por tal motivo fueron identificadas 34 RN, asociadas al comportamiento de las variables correspondientes a cada uno de los indicadores de calidad que se tienen en cuenta para la elaboración del Informe de Tendencias y a las transformaciones durante el proceso de carga de los datos. A continuación se muestran las RN identificadas.

### **Reglas de variables**

- RN1: El Análisis de Tendencias es realizado por el asesor de calidad de cada centro y el coordinador de la calidad UCI.

## *Capítulo 2: Análisis y diseño del Mercado de Datos: Análisis de tendencias*

---

- RN2: Los campos vacíos en las tablas que se muestran en el Informe de Tendencias, significan que no se realizó ninguna actividad de calidad para esa intersección.
- RN3: La información que se introduce referente a las NC debe cumplir con la nomenclatura establecida para las actividades de calidad. (Anexo #3)
- RN4: cantidad\_revisiones= sumatoria de las revisiones de Plan de Evaluaciones por estado.
- RN5: media\_nc\_revisión = cantidad\_nc\_revisiones/cantidad\_revisiones.
- RN6: cantidad\_nc\_impacto = sumatoria de las NC por impacto.
- RN7: cantidad\_nc\_tipo\_adherencia = sumatoria de las NC por tipo de adherencia.
- RN8: cantidad\_nc\_causa = sumatoria de NC por causa.
- RN9: media\_nc\_causa = cantidad\_nc\_causa/cantidad\_revisiones.
- RN10: cantidad\_nc\_revisión = sumatoria de NC de revisión por estado.
- RN11: cantidad\_respuestas\_positivas\_criterio (S)= sumatoria de respuestas positivas por áreas.
- RN12: cantidad\_respuestas\_negativas\_criterio (N) = sumatoria de respuestas negativas por áreas.
- RN13: cantidad\_respuestas\_criterio (T) = sumatoria de respuestas por criterios de cada área.
- RN14: porcentaje\_respuestas\_positivas\_criterio (Ps) =  $S*100/T$ .
- RN15: porcentaje\_respuestas\_negativas\_criterio (Pn) =  $N*100/T$ .
- RN16: cantidad\_respuestas\_positivas\_expectativa (S) = sumatoria de respuestas positivas por áreas.
- RN17: cantidad\_respuestas\_negativas\_expectativa (N) = sumatoria de respuestas negativas por áreas.
- RN18: cantidad\_respuestas\_expectativa (T) = sumatoria de respuestas por criterios de cada área.
- RN19: porcentaje\_respuestas\_positivas\_expectativa (Ps) =  $S*100/T$ .
- RN20: porcentaje\_respuestas\_negativas\_expectativa (Pn) =  $N*100/T$
- RN21: cantidad\_liberaciones = sumatoria de las liberaciones de Plan de Liberación por estado.
- RN22: cantidad\_nc\_clasificación = sumatoria de NC por clasificación.
- RN23: cantidad\_nc\_clasificación\_error = sumatoria de NC por clasificación y tipo de error.
- RN24: porcentaje\_nc = Cantidad de NC por clasificación por Tipo de error \* 100 / Cantidad de NC por clasificación.
- RN25: cantidad\_PLCI (Productos liberados por clasificación en cada iteración) = sumatoria de los productos liberados por clasificación e iteración.
- RN26: cantidad\_PLC (Productos liberados por clasificación) = sumatoria de los productos liberados por clasificación.
- RN27: porcentaje\_PLCI = cantidad\_PLC + 100/ cantidad\_PLC.

- RN28: cantidad\_nc\_aplicación = sumatoria de NC de Aplicación.
- RN29: cantidad\_nc\_documentación = sumatoria de NC de Documentación.
- RN30: cantidad\_nc\_iteración = sumatoria de NC por iteración y estado.

### **Reglas de transformación**

- RN31: Cuando la revisión no pertenece a ningún subproceso se le debe especificar con “Sin proceso”.
- RN32: Cuando las NC no tienen causas se le debe especificar con “Sin causa”.
- RN33: Las NC de liberaciones que no tienen estado, se debe especificar con “Sin estado”.
- RN34: Los productos que no son liberados en alguna iteración, se debe especificar con “No liberada”.

### **2.7 Modelo de casos de uso del sistema.**

El modelo de casos de uso del sistema es una forma de representar gráficamente la relación existente entre los actores y casos de uso del sistema (CUS), de manera que sirva de guía para reflejar las metas y funcionalidades que persigue el negocio (Pressman 2005).

Los CUS fueron identificados a partir del agrupamiento de los RI y RF, para un total de 12 CUS y tres actores que se relacionan con ellos, según sus responsabilidades.

A continuación se presenta la descripción de las responsabilidades de cada uno de los actores en el sistema (Tabla 3) y el diagrama de casos de uso del sistema (Fig. 2), donde para la confección del mismo se utilizaron los patrones de CU CRUD Total, Múltiples actores-Rol común y Concordancia de Adición.

### **Actores del sistema**

Tabla 3: Descripción de los actores del sistema.

Actor	Descripción
Especialista	Se encarga del análisis y la realización de los pedidos de información, así como de la visualización de los mismos.
Administrador	Se encarga de administrar las BD, los usuarios y los roles.
Administrador de ETL	Se encarga de la realización de los procesos de extracción, transformación y carga de los datos.

**Diagrama de CUS**

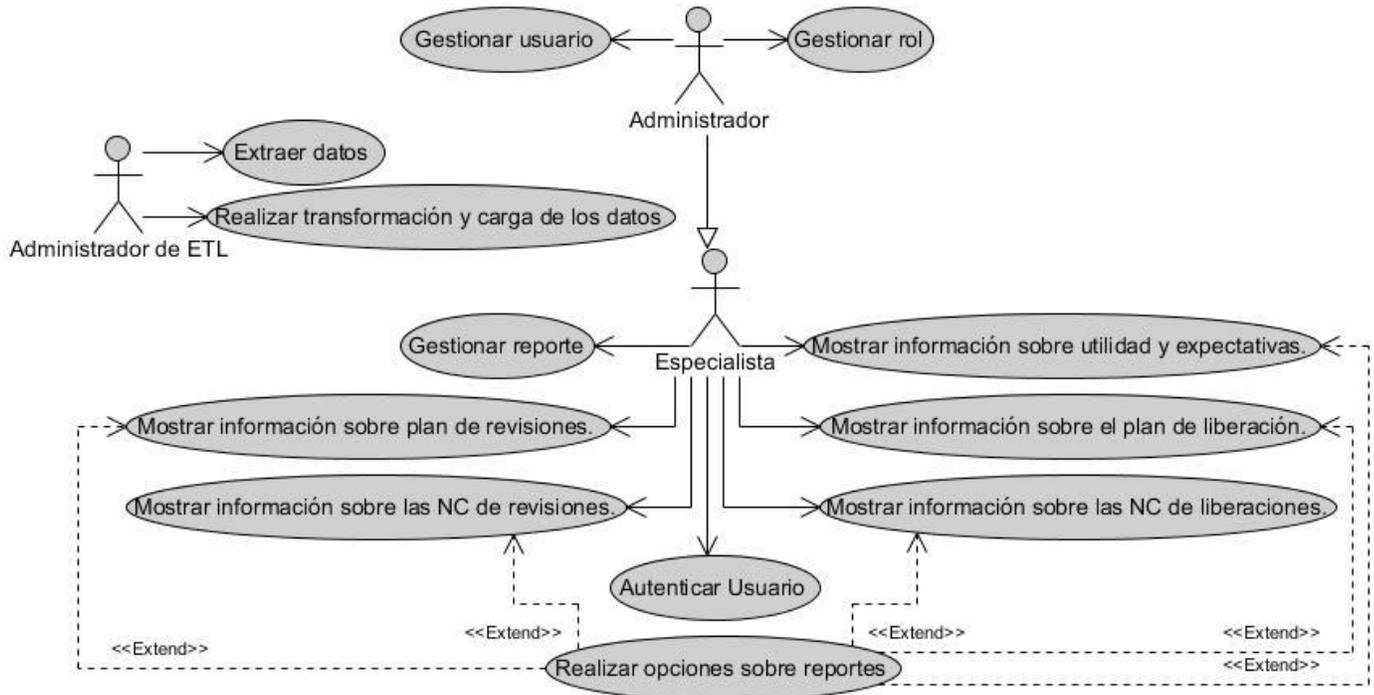


Fig. 2: Diagrama de CU del sistema.

**Especificación de casos de uso**

La Tabla 4 representa la especificación del CU “Mostrar información sobre las NC de revisiones”. El resto de los casos de uso se encuentran detallados en el artefacto “Especificación de Casos de Uso para el proyecto Mercado de datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI”.

Tabla 4: Especificación del CU: "Mostrar información sobre las NC de revisiones".

<b>Objetivo</b>	Mostrar información sobre las NC de revisiones.
<b>Actores</b>	Especialista.
<b>Resumen</b>	El caso de uso inicia cuando el especialista desea consultar la información referente a las NC de revisiones. El especialista selecciona el reporte que desea ver y el sistema muestra la información contenida en él. El caso de uso finaliza cuando el especialista termina el análisis de la información relacionada con las NC de revisiones.
<b>Complejidad</b>	Media.
<b>Prioridad</b>	Alta.
<b>Precondiciones</b>	<ul style="list-style-type: none"> <li>➤ El especialista tiene que estar autenticado.</li> <li>➤ El MD tiene que estar poblado.</li> <li>➤ Los reportes relacionados con las NC de revisiones fueron creados.</li> </ul>

## *Capítulo 2: Análisis y diseño del Mercado de Datos: Análisis de tendencias*

<b>Post condiciones</b>	Los reportes correspondientes fueron consultados por el especialista.	
<b>Flujo de eventos</b>		
<b>Flujo básico Mostrar información sobre tendencias de las NC de revisiones.</b>		
	<b>Actor</b>	<b>Sistema</b>
<b>1</b>	Selecciona el A.A Revisiones.	
<b>2</b>		Muestra los Libros de Trabajos del área seleccionada.
<b>3</b>	Selecciona el Libro de Trabajo.	
<b>4</b>		Muestra los reportes contenidos en el Libro de Trabajo.
<b>5</b>	Selecciona los reportes que desea analizar.	
<b>6</b>		Muestra la información contenida en el reporte seleccionado y brinda opciones al especialista para visualizar los reportes durante su análisis. Ir al CU Gestionar reporte. Finaliza el CU.
<b>7</b>		Brinda opciones para realizar sobre los reportes. Ir al CU Realizar opciones sobre reportes.
<b>Flujos alternos</b>		
No aplica.		
	<b>Actor</b>	<b>Sistema</b>
<b>Opciones de Reportes</b>		
<b>Perspectivas de análisis</b>	<b>Posibles resultados</b>	
	<b>Medidas</b>	<b>Periodicidad</b>
Variables de entrada relacionadas con el CU: <ul style="list-style-type: none"> <li>➤ tiempo</li> <li>➤ nivel de revisión</li> <li>➤ centro</li> <li>➤ estado</li> <li>➤ tipo de adherencia</li> <li>➤ impacto</li> <li>➤ área de proceso</li> </ul>	Variables de salida disponibles: <ul style="list-style-type: none"> <li>➤ media de NC de revisiones</li> <li>➤ cantidad de NC (impacto)</li> <li>➤ cantidad de NC (adherencia)</li> <li>➤ cantidad de NC (estado)</li> </ul>	Rango de tiempo en que se solicitan las variables de salida: <ul style="list-style-type: none"> <li>➤ Trimestral</li> </ul>
<b>Relaciones</b>	<b>CU incluidos</b>	No aplica.
	<b>CU extendidos</b>	Realizar opciones sobre reportes.

<b>Requisitos funcionales</b>	<b>no</b>	Sección: “4.1 Requisitos no funcionales” del documento: “Especificación de requisitos de software”.
<b>Asuntos pendientes</b>		

### 2.8 Arquitectura del Mercado de datos.

Definir una arquitectura adecuada constituye un eslabón fundamental en el proceso de desarrollo de *software*. La arquitectura empleada para el desarrollo del MD consta con tres subsistemas y tres niveles, donde cada subsistema es ubicado en un nivel (Fig. 3).

Las fuentes de datos están compuestas por las BD de todos los centros productivos de la universidad y la BD que se genera a partir de las encuestas de utilidad. Dichas fuentes se encuentran en el sistema gestor de BD PostgreSQL 9.2 y contendrán la información que luego será almacenada en los esquemas definidos en el subsistema de almacenamiento, con apoyo del gestor de BD PostgreSQL y el administrador de BD PgAdmin III. Donde para ello será necesario la realización de los procesos de Extracción, Transformación y Carga de los datos (ETL) definidos en el subsistema de integración, mediante el uso de las herramientas Pentaho Data Integration y DataCleaner. Finalmente, los datos almacenados en el MD, serán mostrados en el subsistema de visualización, el cual contendrá los pedidos de información a los usuarios finales, mediante la herramienta Pentaho BI Server.

Por otra parte, se encuentran los clientes finales o especialistas, los cuales son los encargados de realizar las operaciones sobre los reportes del subsistema de visualización.

En la siguiente figura se muestra la arquitectura diseñada para el MD Análisis de Tendencias:

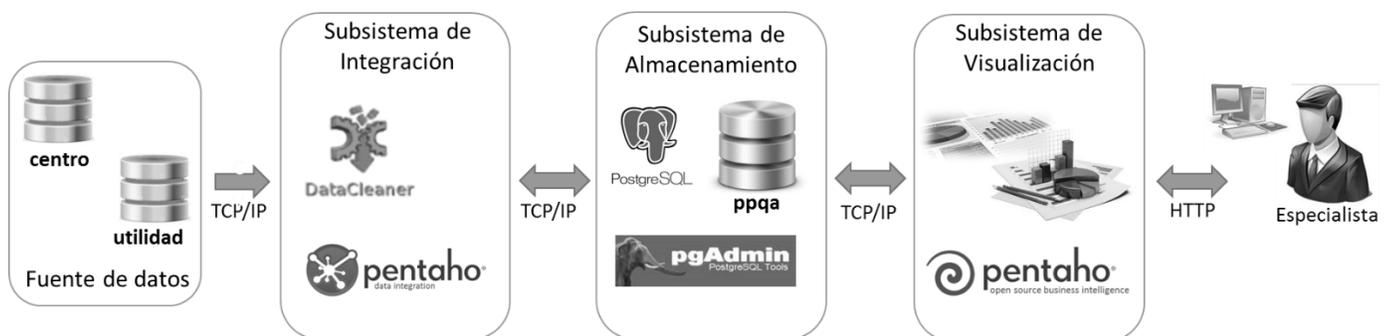


Fig. 3: Arquitectura del MD.

#### 2.8.1 Subsistema de Almacenamiento.

El subsistema de almacenamiento contiene toda la información correspondiente al MD, es por ello que, luego del estudio realizado sobre la información que se recoge en el Análisis de Tendencias y la identificación de los RI a través de las necesidades de los clientes, fueron creadas 6 tablas de hechos, 9 medidas y 12 tablas de dimensiones.

## *Capítulo 2: Análisis y diseño del Mercado de Datos: Análisis de tendencias*

Según (Díaz 2011) se define:

- **Tabla de hecho:** Representa los procesos de negocio de la organización.
- **Medidas:** Representan los indicadores de negocio de los procesos de negocio. Conceptos cuantificables que permiten medir el proceso de negocio.
- **Tabla de dimensión:** Representa una vista para cierto proceso de negocio.

A continuación se presenta la descripción de las tablas de hechos (Tabla 5) y dimensiones (Tabla 6).

Tabla 5: Descripción de las tablas de hechos.

No.	Hechos	Descripción
H1	hech_plan_revision_liberacion	Contiene la medida correspondiente a la cantidad de revisiones y liberaciones según el estado del plan de ambas etapas.
H2	hech_nc_revision	Contiene la medida correspondiente a la cantidad de NC que se detectan en las revisiones.
H3	hech_analisis_causa	Contiene las medidas correspondientes a la cantidad de revisiones y la cantidad de NC por causa.
H4	hech_tendencia	Contiene las medidas correspondientes a la cantidad de revisiones y la cantidad de NC.
H5	hech_utilidad	Contiene las medidas correspondientes a la cantidad de respuestas positivas y negativas dadas en las encuestas de utilidad de cada una de las áreas de proceso y la cantidad de respuestas según las expectativas de cada área.
H6	hech_nc_liberacion	Contiene la medida correspondiente a la cantidad de NC detectadas en la etapa de liberación según la iteración, la clasificación, el tipo de error o el estado de las NC.

Tabla 6: Descripción de las tablas de dimensiones.

No.	Dimensiones	Descripción
D1	dim_tiempo	Define una línea de tiempo específica donde el nivel más bajo es el mes.
D2	dim_centro	Representa y posee información de todos los centros productos de la UCI.
D3	dim_estado	Representa los estados del plan de revisiones y liberaciones y de las NC detectadas en ambas etapas.
D4	dim_nivel_revision	Representa los diferentes niveles de las revisiones (proyecto, centro y UCI) y las revisiones que se realizan en cada uno de ellos.

## *Capítulo 2: Análisis y diseño del Mercado de Datos: Análisis de tendencias*

<b>D5</b>	dim_impacto_nc	Representa el impacto de las NC.
<b>D6</b>	dim_tipo_adherencia_nc	Representa el tipo de adherencia de las NC.
<b>D7</b>	dim_area_proceso	Representa las seis áreas de proceso de CMMI.
<b>D8</b>	dim_causa_nc	Representa las causas de ocurrencia de las NC.
<b>D9</b>	dim_tipo_criterio	Representa el criterio de cada área de proceso.
<b>D10</b>	dim_iteracion	Representa el número de la iteración donde se libera un producto.
<b>D11</b>	dim_tipo_error	Representa el tipo de error de las NC detectadas en la etapa de liberación.
<b>D12</b>	dim_clasificacion_nc	Representa la clasificación de las NC detectadas en la etapa de liberación.

### Matriz bus

La matriz bus representa la relación que existe entre las tablas de hechos y las tablas de dimensiones, permite describir la trazabilidad de los mismos, evitando que ocurra el solapamiento entre las tablas de hechos.

La matriz bus confeccionada permitió identificar en el MD Análisis de Tendencia las dimensiones compartidas como son: dim\_tiempo, dim\_centro, dim\_area\_proceso, dim\_estado y dim\_nivel\_revision (Tabla 7).

A continuación se muestra la matriz bus del MD, donde las filas contienen los hechos y en las columnas se encuentran las dimensiones.

Tabla 7: Matriz bus del MD Análisis de Tendencias.

Matriz Bus												
Hechos	Dimensiones											
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
<b>H1</b>	X	X	X	X								
<b>H2</b>	X	X	X	X	X	X	X					
<b>H3</b>	X	X						X				
<b>H4</b>	X	X		X								
<b>H5</b>	X	X		X			X		X			
<b>H6</b>	X	X	X	X						X	X	X

### Modelo de datos

El modelo de datos representa la relación entre las tablas de hechos y dimensiones identificadas mediante el proceso de análisis. De esta manera los datos del negocio queden reflejados en forma de cubos de datos. El modelo propuesto para la solución del MD Análisis de Tendencias se basa en una topología Constelación de hechos, ya que existen dimensiones compartidas para más de una tabla de hechos. A continuación se muestra un fragmento del modelo diseñado para la solución (Fig. 4), la cual comprende los hechos plan de revisión-liberación y NC de revisión. En el [Anexo #4](#) se encuentra el modelo ampliado.

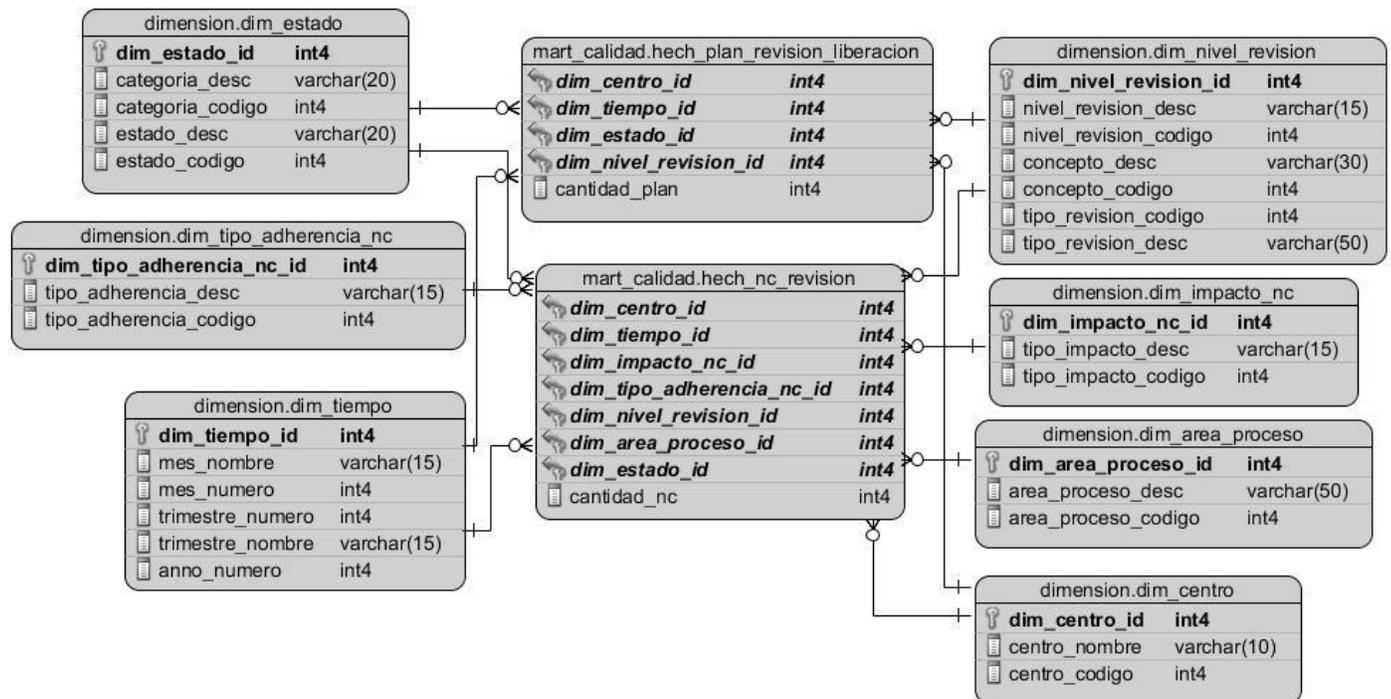


Fig. 4: Modelo de datos reducido.

### 2.8.2 Subsistema de Integración.

En el subsistema de integración se realiza el diseño de las transformaciones como parte del proceso de Extracción, Transformación y Carga de los datos (ETL). Para este proceso de integración de información procedente de varias fuentes son utilizadas diferentes estrategias de integración de datos, según (Basallo and Estrada 2009) estas son:

- **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:** 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.

Esta última estrategia será utilizada en la solución del MD, ya que la fuente de datos que se utilizará para poblar el mismo serán las bases de datos de los 14 centros productivos de la universidad. Además, esta estrategia permitirá extraer dichos datos, transformarlos según las necesidades de los usuarios y las reglas de negocio para luego cargarlos en la base de datos del MD.

### 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. Además, permite verificar los distintos tipos de datos que contiene la fuente y obtener estadísticas sobre valores incorrectos, duplicados o nulos para poder corregirlos.

En el perfilado de datos realizado a la fuente del MD Análisis de Tendencias fueron detectados 59,52 % de datos de tipo textos (String), 32,14 % de números enteros (Integer), 8,33 % de tipo fecha (Date) y 0 % atributos booleanos (Boolean), como se evidencia en la Fig. 5. También fueron analizadas las tablas de la fuente para detectar la cantidad de filas nulas en cada una de ellas. La Fig. 6 muestra la cantidad de datos nulos en la tabla “gespro\_ppqa\_project\_bugs” para cada atributo de dicha tabla.

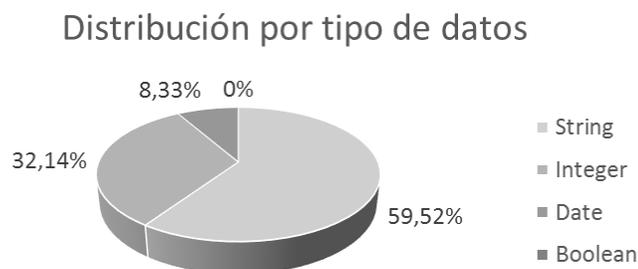


Fig. 5: Distribución por tipo de datos.

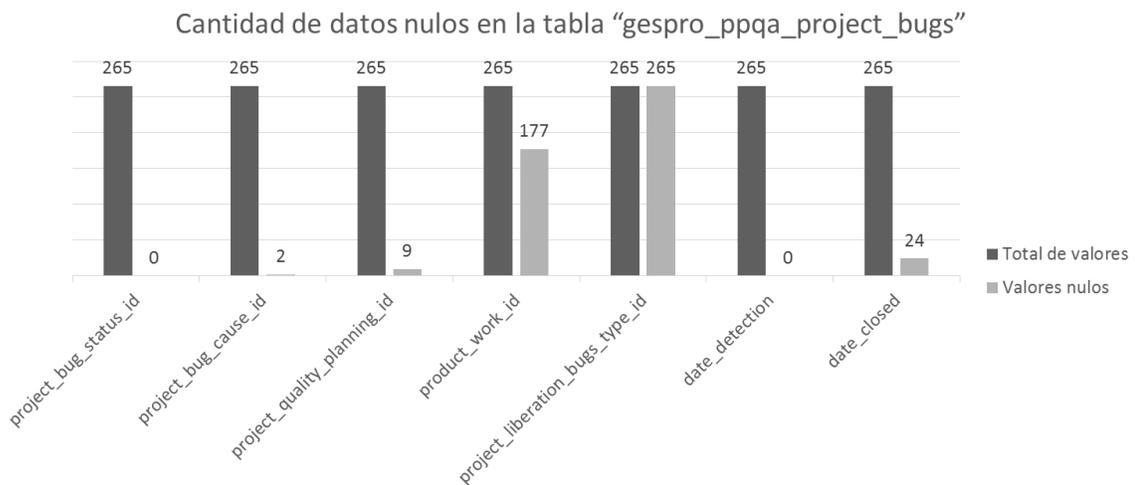


Fig. 6: Distribución de datos nulos para la tabla gespro\_ppqa\_project\_bugs.

### Diseño de las transformaciones

**Diseño para la carga de dimensiones:** Primeramente se extraen los datos de fuentes como las bases de datos "centro" y "utilidad", y/o en algunos casos se introducen directamente. Se seleccionan los datos correspondientes a las dimensiones a cargar y se les realiza la limpieza y transformación, para luego insertarlos en la base de datos "ppqa". La Fig. 7 muestra el diseño de las transformaciones para la carga de las dimensiones.

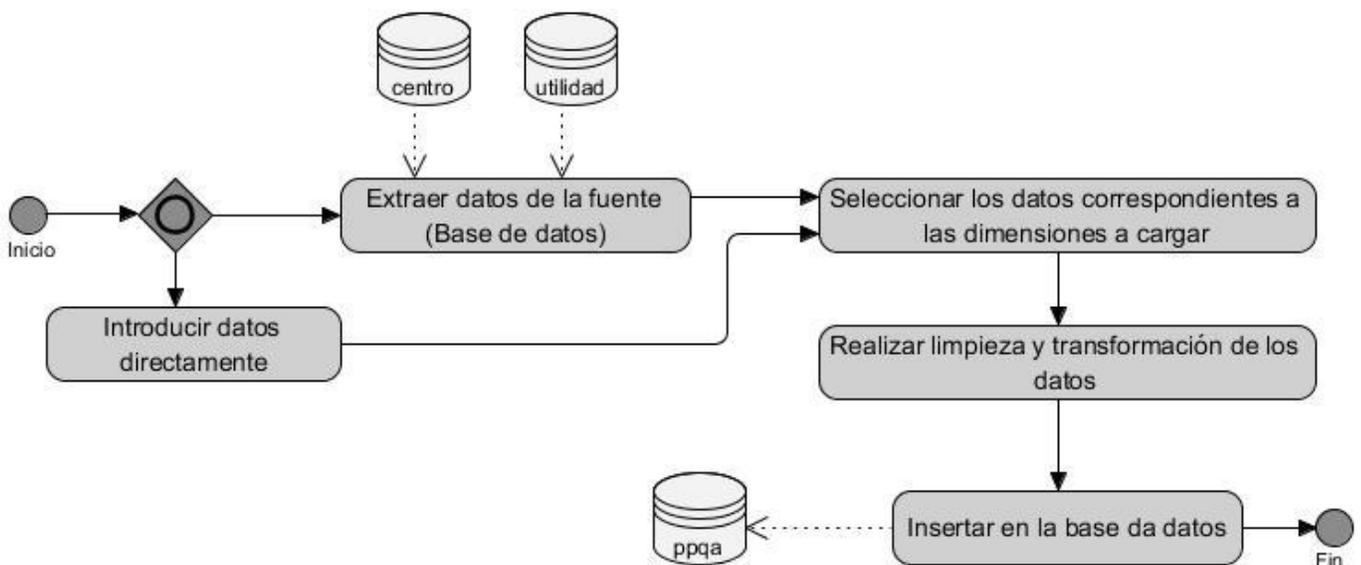


Fig. 7: Diseño de las transformaciones para la carga de dimensiones.

**Diseño para la carga de hechos:** Después de haber cargado todas las dimensiones, primeramente se obtiene y se inserta la información CDC (en inglés Change Data Capture), se extraen los datos de las fuentes "centro" y "utilidad", y se les realiza limpieza y transformación. Se buscan las llaves primarias de cada una

de ellas en la fuente para obtener el nombre del campo y en la base de datos destino, para luego validarlos de manera que cumplan con las restricciones definidas por el especialista. Si las llaves primarias no coinciden con las llaves de las base de datos destino, se almacenan en ficheros Excel para su posterior análisis. En caso correcto, se calculan las medidas y se insertan en la base de datos destino "ppqa". La Fig. 8 muestra el diseño de las transformaciones para la carga de hechos.

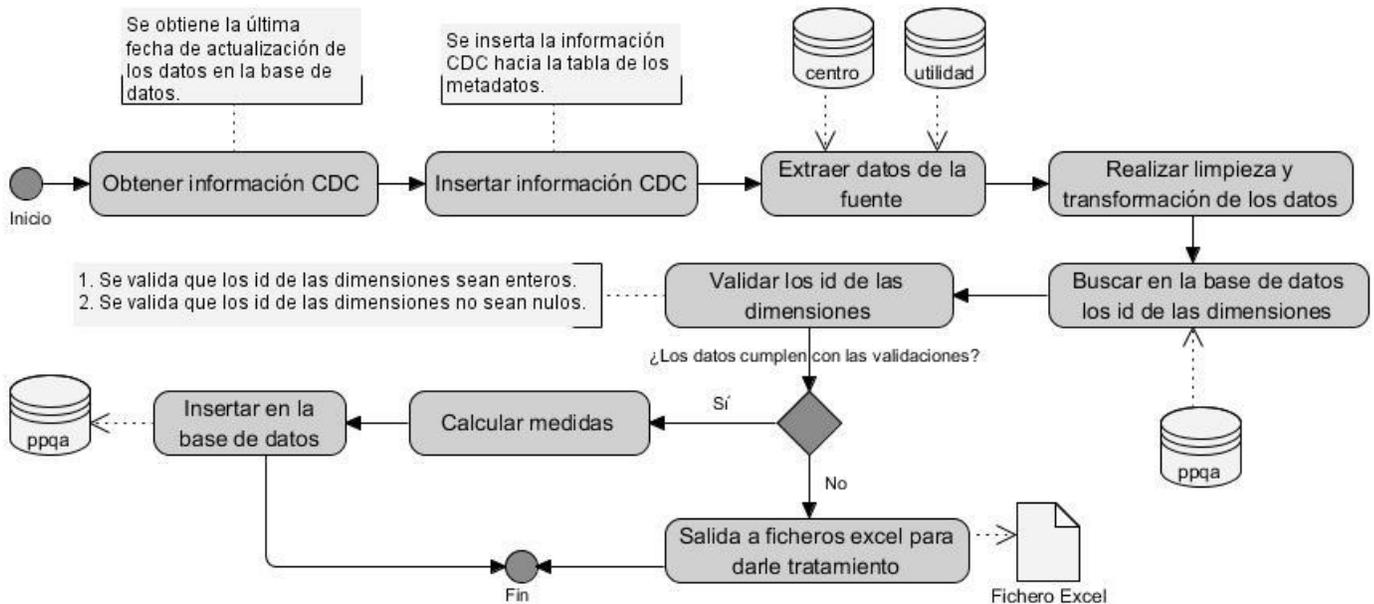


Fig. 8: Diseño de las transformaciones para la carga de hechos.

### 2.8.3 Subsistema de Visualización.

El subsistema de visualización tiene como objetivo definir la arquitectura de información presente en el MD, de manera que sirva como perspectiva de cómo se mostrarán los datos al cliente final.

#### Arquitectura de información

Para el MD Análisis de Tendencias se definió como Área de Análisis General (A.A.G.) Análisis de Tendencias. Además, se identificaron tres áreas de análisis (A.A), Revisiones, Utilidad y Liberaciones, cada uno con sus respectivos Libros de Trabajo (L.T), los cuales suman un total de 12 L.T en correspondencia con la cantidad de indicadores que recoge el Informe de Tendencias. Cada L.T incluye las Tablas de Salida con los reportes candidatos asociados a cada uno. A continuación se muestra la estructura de navegación definida para el MD (Fig. 9).

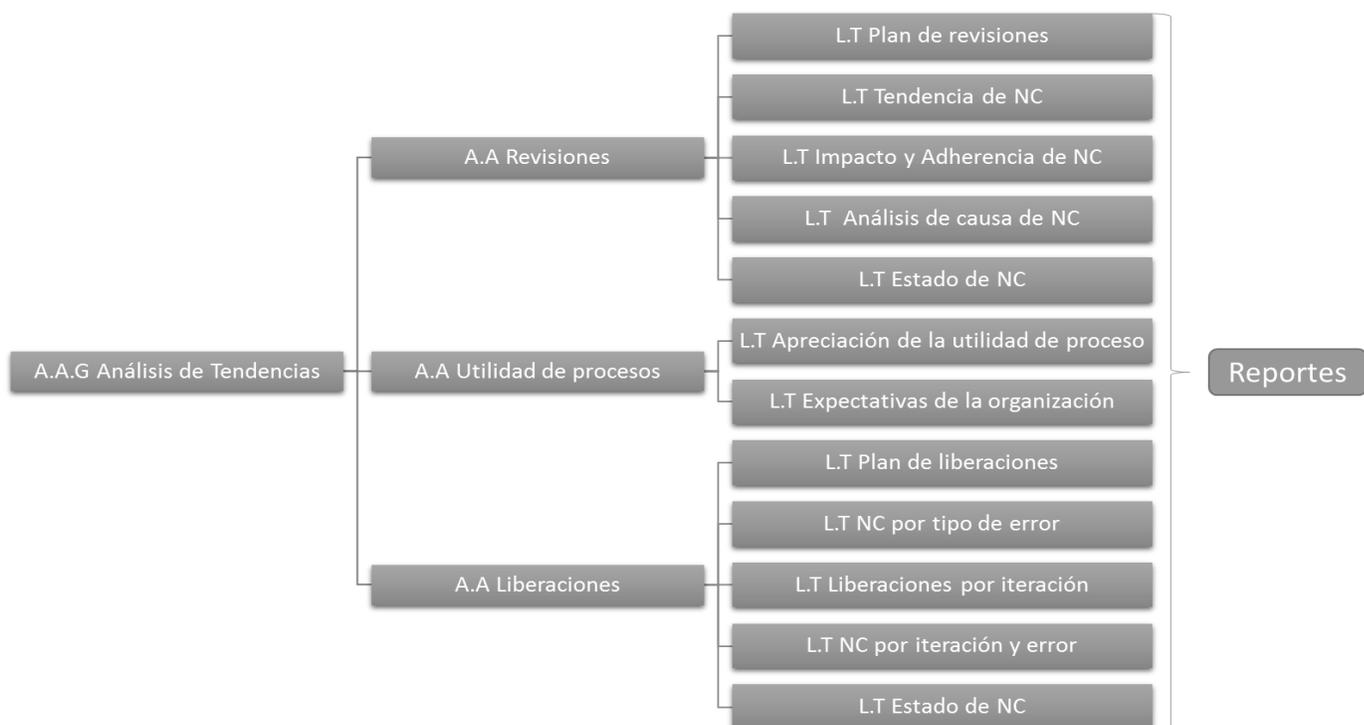


Fig. 9: Diseño del mapa de navegación.

### Diseño de los reportes candidatos

Los reportes candidatos responden a las necesidades de los usuarios y se corresponden con las salidas que debe mostrar el sistema. A continuación se muestran ejemplo de reportes candidatos de cada una de las áreas de análisis definidas, tales como la cantidad de revisiones por tiempo, centro, estado y nivel de revisión, el porcentaje de respuestas positivas por tiempo, centro, nivel de revisión, tipo de criterio y área de proceso y el porcentaje de error de NC según su clasificación por tiempo, centro, nivel de revisión, tipo de error y clasificación de las NC (Tablas 8, 9 y 10).

Tabla 8: Reporte candidato "Cantidad de revisiones" del A.A. Revisiones.

<b>Área de análisis (AA)</b>	Revisiones.
<b>Libro de Trabajo (LT)</b>	Plan de revisiones.
<b>Reporte (Tabla de Salida – TS)</b>	Cantidad de revisiones.
<b>Descripción</b>	Representa la cantidad de revisiones por estado.
<b>Elementos del reporte</b>	Estado del plan de revisiones, Nivel de revisión, Tiempo, Centro.
<b>Frecuencia de emisión</b>	Trimestral.
<b>Funciones</b>	Suma.
<b>Gráfico</b>	Línea.

## *Capítulo 2: Análisis y diseño del Mercado de Datos: Análisis de tendencias*

Tabla 9: Reporte candidato "Porcentaje de respuestas sí según criterios" del A.A Utilidad.

<b>Área de análisis (AA)</b>	Utilidad.
<b>Libro de Trabajo (LT)</b>	Apreciación de la utilidad de procesos.
<b>Reporte (Tabla de Salida – TS)</b>	Porcentaje de respuestas sí.
<b>Descripción</b>	Representa el porcentaje de respuestas positivas y negativas de cada área según el criterio de cada una.
<b>Elementos del reporte</b>	Áreas de proceso, Tipo de criterio, Nivel de revisión, Tiempo, Centro.
<b>Frecuencia de emisión</b>	Trimestral.
<b>Funciones</b>	Porcentaje.
<b>Gráfico</b>	Barra horizontal.

Tabla 10: Reporte candidato "Porcentaje de error de NC según clasificación" del A.A. Liberaciones.

<b>Área de análisis (AA)</b>	Liberaciones.
<b>Libro de Trabajo (LT)</b>	NC por tipo de error según su clasificación.
<b>Reporte (Tabla de Salida – TS)</b>	Porcentaje de error de NC según clasificación.
<b>Descripción</b>	Representa el porcentaje de error de las NC de Aplicación y Documentación.
<b>Elementos del reporte</b>	Clasificación de NC, Tipo de error, Nivel de revisión, Tiempo, Centro.
<b>Frecuencia de emisión</b>	Trimestral.
<b>Funciones</b>	Porcentaje.
<b>Gráfico</b>	Pastel.

### **2.9 Políticas de respaldo y recuperación.**

Las políticas de respaldo y recuperación de datos garantizan la integridad de la información, de manera que se encuentre almacenada en un lugar seguro en caso de fallos en el sistema. Para el MD Análisis de Tendencias se realizarán salvadas a la información contenida en la BD y las copias de seguridad serán realizadas en dos momentos; diariamente a través de tareas programadas en el sistema operativo y trimestralmente, ya que ese es el periodo en el que se generan los Informes de Tendencias. En cuanto a las tablas involucradas, se le realizarán las salvadas a las seis tablas de hechos identificadas y sus dimensiones asociadas.

### **Roles y permisos**

Para el acceso al MD se definieron usuarios por cada rol existente en el sistema para asignar permisos a los usuarios que interactúan con el sistema.

La Tabla 11 muestra los roles con sus permisos asociados.

Tabla 11: Roles y permisos.

Roles	Aplicación		Bases de datos	
	Lectura	Escritura	Lectura	Escritura
<b>Administrador</b>	X	X	X	X
<b>Administrador de ETL</b>			X	X
<b>Especialista</b>	X	X		

### **2.10 Conclusiones.**

En este capítulo se realizó el análisis del negocio del MD Análisis de Tendencias, lo que permitió, a través de las necesidades de los usuarios identificar los requisitos de información, requisitos no funcionales y funcionales; estos últimos garantizaron la confección del diagrama de casos de uso de sistema. La arquitectura propuesta para la construcción del MD, consta de tres subsistemas: Almacenamiento, Integración y Visualización, la cual permitió determinar la estructura, funcionamiento e interacción de cada uno de ellos en el mercado. Además se especificó las tablas de dimensiones y hechos, la matriz buz para evitar el solapamiento los hechos; el diseño de las transformaciones, el perfilado de datos y la arquitectura de información, la cual será una guía para la visualización de los reportes candidatos.

## Capítulo 3: Implementación y prueba del Mercado de datos: Análisis de Tendencias.

En el presente capítulo se realizará la implementación de los Subsistemas de Almacenamiento, Integración y Visualización, teniendo en cuenta los requisitos y necesidades del cliente, de esta forma se implementa el modelo físico a través del *script* del mercado de datos y se realizan los procesos de extracción, transformación y carga de los datos. Contendrá, además, los resultados de los casos de pruebas y las listas de chequeo.

### 3.1 Implementación del Subsistema de Almacenamiento.

Durante la etapa de análisis y diseño del MD fueron definidas las tablas de hechos y dimensiones del mercado de datos, las cuales estarán contenidas dentro de los esquemas definidos para dicho mercado. Estos esquemas permitirán tener organizada la información de manera eficiente. Además, los usuarios podrán tener acceso a estos esquemas siempre y cuando posean los permisos adecuados.

#### Esquemas

- *dimension*: Contiene las 12 dimensiones identificadas para el MD Análisis de Tendencias.
- *mart\_calidad*: Contiene las 6 tablas de hechos identificadas para el MD Análisis de Tendencias.
- *md\_cdc*: Contiene la tabla con la información CDC (Change Data Capture).

#### 3.1.1 Estándar de codificación.

En aras de lograr un mejor entendimiento entre las partes implicadas en el MD fueron utilizados estándares de codificación, los cuales se exponen a continuación en la Tabla 12.

Tabla 12: Estándar de codificación para el MD.

Tipo de Objeto	Función	Nomenclatura	Descripción
<b>Esquemas</b>	Dimensiones	<i>dimension</i>	Esquema donde se encuentran las tablas de dimensiones.
	Hechos	<i>mart_calidad</i>	Esquemas donde se encuentran las tablas de hechos.
	Metadatos	<i>md_cdc</i>	Esquema donde se encuentra la tabla con la información cdc.
<b>Tablas</b>	Dimensiones	<i>dim_[nombre]</i>	Tablas dimensionales utilizadas como perspectivas de análisis.

	Hechos	hech_[nombre]	Tablas de hechos que definen las principales medidas requeridas para calcular indicadores y otras medidas derivadas.
--	--------	---------------	--

### 3.2 Implementación del Subsistema de Integración de datos.

El proceso de ETL es la base fundamental en un AD. Mediante este, se extraen los datos de los sistemas fuentes, se transforman de manera que cumplan con las normas de calidad de los mismos y finalmente se cargan en la base de datos para que luego puedan ser presentados a los usuarios finales.

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 (subsistema 3): permite la extracción de los datos desde la fuente origen hasta la fuente destino.

#### Limpieza y conformación

- Limpieza de datos (subsistema 4): implementa los procesos de calidad de los datos que permiten detectar las incoherencias de calidad.
- Rastreo de eventos de errores (subsistema 5): captura todos los errores que proporcionan información valiosa sobre la calidad de los datos y permite la mejora de los mismos.

#### Entrega

- Claves subrogadas (subsistema 10): permite crear claves subrogadas independientes para cada tabla.

### Gestión

- Programador de trabajos (subsistema 22): permite gestionar ETL de la categoría de trabajos.
- Repositorio de metadatos (subsistema 34): captura los metadatos de los procesos de ETL, de los datos del negocio y de los aspectos técnicos.

#### 3.2.1 Transformaciones y trabajos.

Como parte del proceso de ETL, se realizó la carga en la base de datos de las dimensiones y hechos identificados. A continuación se muestra un ejemplo de transformación para el hecho plan de revisión y liberación (Fig. 10), la cual, antes de cargar los datos de la fuente se obtiene la información CDC asociada a la fecha de la última actualización de la base de datos y se inserta dicha información en la tabla "cdc\_plan\_revision\_liberacion". Luego se comienza con la carga de la fuente de datos (BD), se calcula el mes y el año de las revisiones, se eliminan todos los dígitos que aparezcan, se divide una de las filas a partir del símbolo "-", se filtran las filas para que solo se muestren los tipos de revisiones y en caso de encontrar datos erróneos, estos tendrán salida a un archivo Excel. Luego se selecciona y especifica el centro a cargar, se obtienen y agrupan los identificadores (id) de las dimensiones en las bases de datos y se realiza el tratamiento de errores en caso de que ocurra alguno durante la obtención de los id. Por último se ordena por el año, se agrupa por el tipo de revisión y se cargan los datos en la BD destino (hech\_plan\_revision\_liberacion).

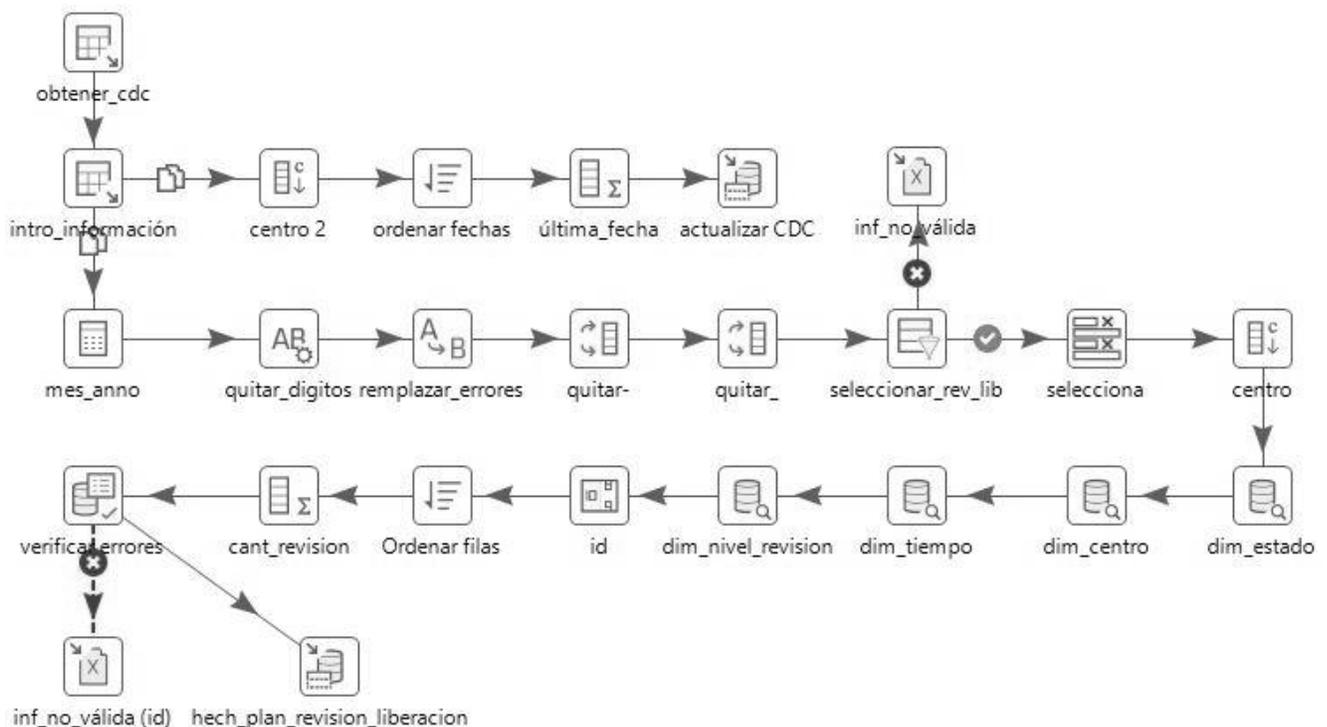


Fig. 10: Proceso de integración de datos para el hecho plan de revisión y liberación.

La dimensión estado es común para los hechos “hech\_plan\_revision\_liberacion”, “hech\_nc\_revision” y “hech\_nc\_liberacion”. La carga de la misma comienza con la extracción de la fuente de datos (BD), una para la entrada del estado de las NC y otra para el estado del Plan tanto de las liberaciones como de las revisiones. Luego se asigna una categoría a cada estado, y un id para dichas categorías, el estado y la dimensión en general, se convierten los id a datos enteros y por último se insertan en la base de datos destino (dim\_estado). La Fig. 11 muestra la transformación descrita.

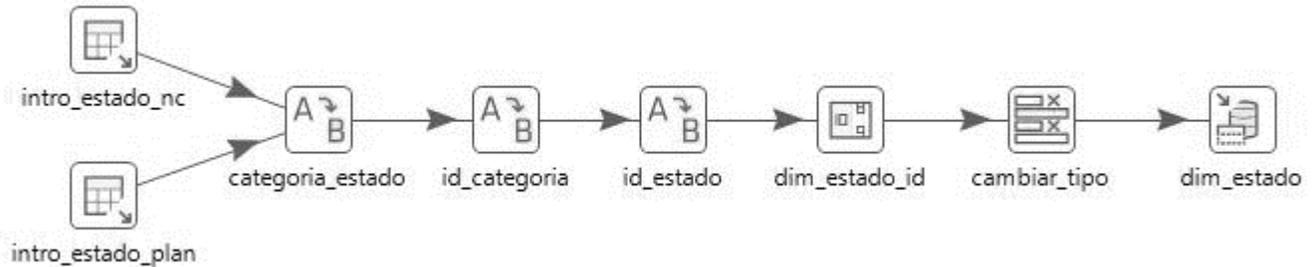


Fig. 11: Proceso de integración de datos para la dimensión estado.

El trabajo o *job* es utilizado para realizar las transformaciones de manera organizada. Para el MD Análisis de Tendencias se realizaron dos trabajos, uno para los hechos (Fig. 13) y otro para las dimensiones (Fig. 12). Para ello, primeramente se realizó una transformación “error” que recoge y muestra mediante un fichero Excel los errores detectados durante la ejecución de los trabajos.

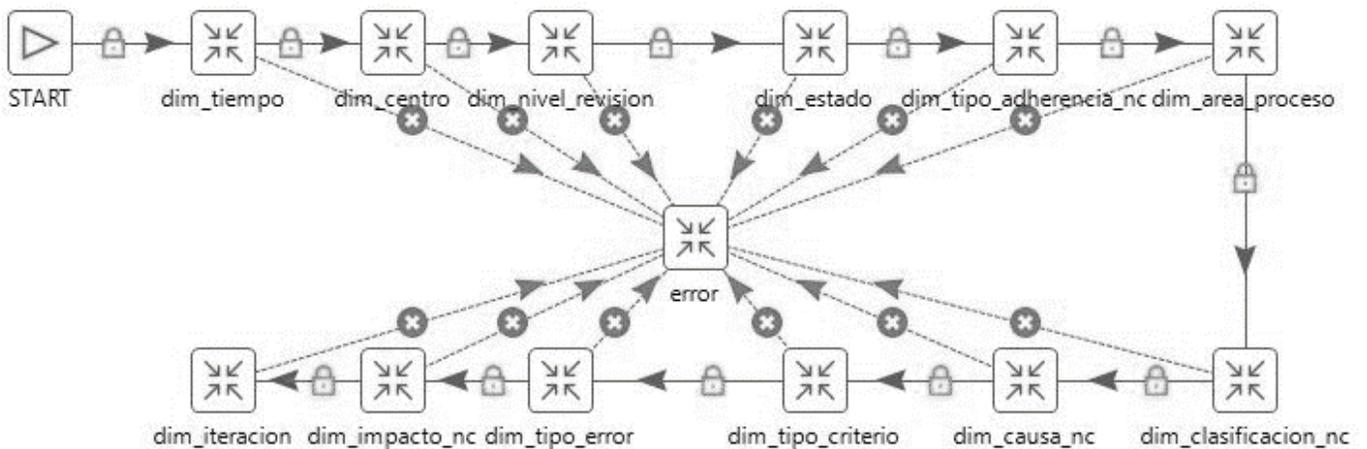


Fig. 12: Trabajo para las dimensiones.

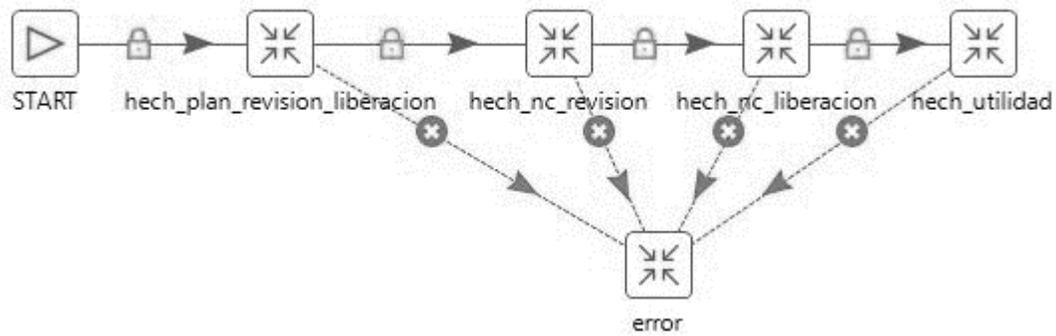


Fig. 13: Trabajo para los hechos.

### 3.3 Implementación del Subsistema de Visualización.

Luego de realizar el proceso de ETL, como parte de la implementación del subsistema de visualización, se procede a la realización de los cubos OLAP (en inglés On-Line Analytical Processing) y la creación de los reportes candidatos.

#### 3.3.1 Implementación de los cubos OLAP.

Para la implementación de los cubos OLAP del MD, se definieron las dimensiones, medidas y niveles de jerarquía que conforma el esquema Análisis de Tendencias y se creó un cubo para cada temática que conforma el Informe de Tendencias (Fig. 14).



Fig. 14: Implementación de los cubos OLAP.

### 3.3.2 Implementación de la capa de visualización.

El mapa de navegación o arquitectura de información está diseñada de manera que permita tener una mejor visualización de cómo estará estructurada la información en el MD Análisis de Tendencias, la cual cuenta con un Área de Análisis General (A.A.G), tres Áreas de Análisis (A.A) y doce Libros de Trabajo (L.T) con sus reportes asociados (Fig. 15).

- **A.A.G Análisis de Tendencias:** Agrupa la información referente a los MD de los 14 centros de la UCI.
- **A.A Revisiones:** Contiene la información referente a las actividades de revisiones que se realizan de cada proyecto.
- **A.A Utilidad de procesos:** Contiene la información referente a las encuestas de utilidad que se realizan sobre cada área de proceso.
- **A.A Liberaciones:** Contiene la información referente a las actividades que se realizan en la etapa de liberación de cada proyecto.
- **L.T Plan de Revisiones:** Contiene la información relacionada a la cantidad de revisiones según el estado del plan de revisiones.
- **L.T Tendencias de NC de revisiones:** Contiene la información relacionada con la tendencia de las NC.
- **L.T Impacto y Adherencia de NC:** Contiene la información relacionada con el impacto y la adherencia de las NC.
- **L.T Análisis de causa de NC:** Contiene la información relacionada a las NC detectadas por causa.
- **L.T Estado de NC:** Contiene información relacionada con las NC de revisión según el estado de las mismas.
- **L.T Utilidad de procesos:** Contiene la información estadística de las respuestas de la encuesta de utilidad de los procesos según el criterio de cada área.
- **L.T Expectativas de la Organización:** Contiene la información estadística de las respuestas de utilidad de todas las áreas de proceso.
- **L.T Plan de Liberaciones:** Contiene la información relacionada al estado del plan de liberaciones.
- **L.T NC por tipo de error:** Contiene la información relacionada con las NC detectadas según el tipo de error.
- **L.T Liberaciones por iteración:** Contiene la información relacionada con los productos liberados en cada iteración.
- **L.T NC por iteración y error:** Contiene la información relacionada con las NC detectadas según la iteración y el tipo de error.

## Capítulo 3: Implementación y prueba del Mercado de Datos: Análisis de tendencias

- **L.T Estado de NC:** Contiene información relacionada con las NC de liberación según el estado de las mismas.

La Fig. 16 muestra la arquitectura de información diseñada para el MD Análisis de Tendencias.

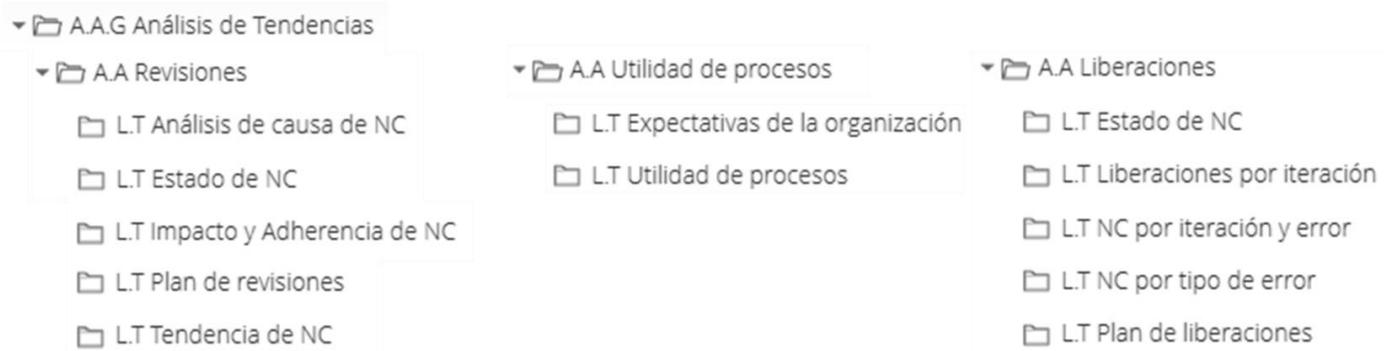


Fig. 15: Arquitectura de información del MD Análisis de Tendencias.

### **3.3.3 Implementación de los reportes candidatos.**

Luego de analizar la fuente de datos que recoge toda la información correspondiente al MD Análisis de Tendencias, fueron confeccionados los reportes candidatos, los cuales contienen la información referente a los pedidos de información establecidos por el cliente. Estos reportes fueron elaborados mediante consultas MDX, las cuales realizan las consultas utilizando los términos hechos, dimensiones y medidas.

La Fig. 15 muestra las vistas de análisis para el indicador “Impacto de las NC y Adherencia a procesos y productos de trabajo” a nivel de centro, con su gráfica correspondiente, donde en la parte superior se visualiza la cantidad de NC detectadas en cada área de procesos según el impacto de la NC (alto, medio y bajo) y en la parte inferior se muestran las NC por el tipo de adherencia que presentan (proceso o producto).

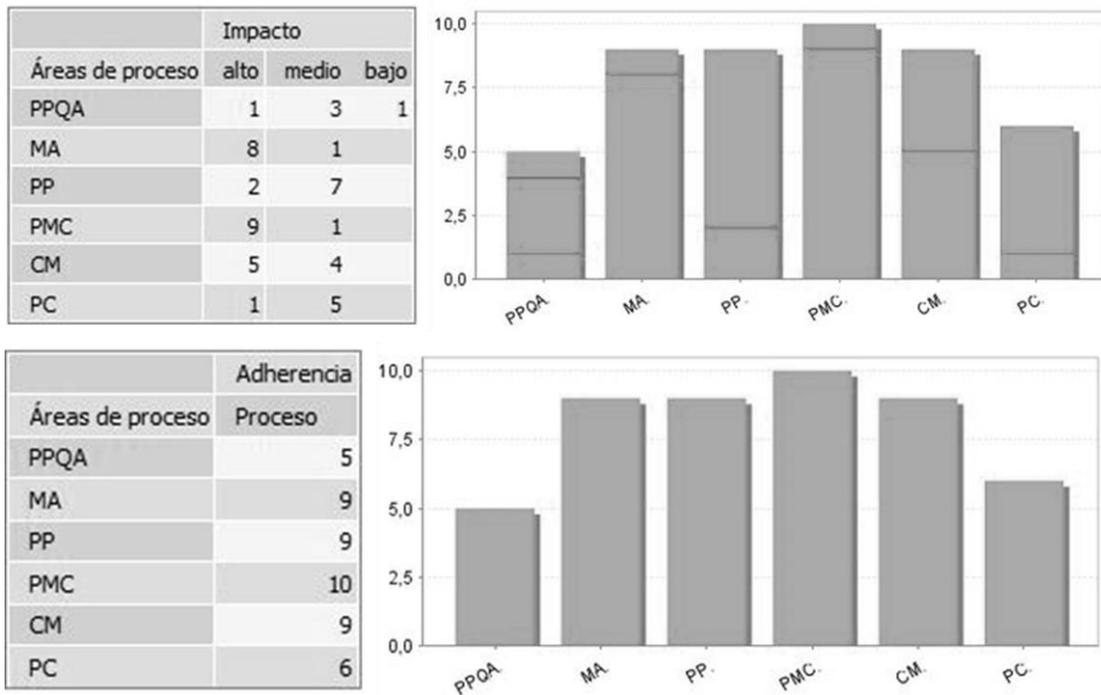


Fig. 16: Reporte “Adherencia a procesos y productos de trabajo” a nivel de centro.

### Dashboard

Un Dashboard (tablero de instrumentos) es una representación gráfica de los principales indicadores (KPI en inglés Key Performance Indicator) que intervienen en la consecución de los objetivos de negocio, y que está orientada a la toma de decisiones para optimizar la estrategia de la empresa. Un Dashboard debe transformar los datos en información y ésta, en conocimiento para el negocio (Elósegui 2014).

Un Dashboard es una aplicación que se utiliza para presentar el contenido de BI de alto nivel a los usuarios finales. Es por ello que para visualizar todos los indicadores del Informe de Tendencias, se realizaron los reportes a través de un Dashboard. La Fig. 17 muestra el Dashboard correspondiente al indicador “Impacto de las NC y Adherencia a procesos y productos de trabajo”, en el que gráficamente se muestra la cantidad de NC detectadas en cada área por impacto y tipo de adherencia de las NC, a nivel de proyecto, centro y universidad.

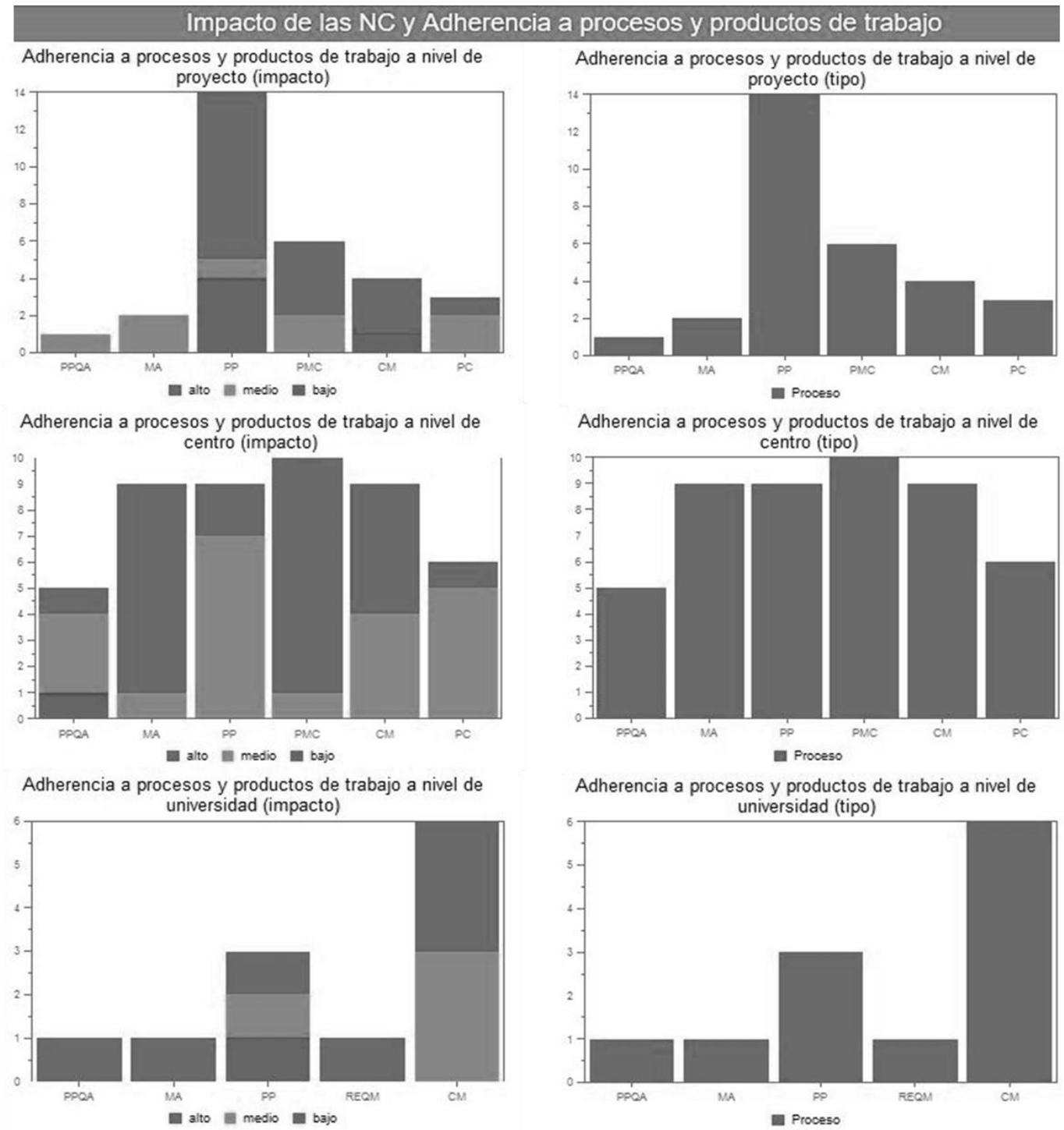


Fig. 17: Impacto y Adherencia de las NC usando Dashboard.

### 3.4 Pruebas.

Durante el desarrollo de cualquier producto de *software* es importante la realización de pruebas al mismo en aras de determinar su calidad. Estas pruebas se realizan desde el inicio del *software* hasta que llegue a manos del cliente, esperando que cumpla con sus expectativas.

Para llevar a cabo las pruebas del MD Análisis de Tendencias fue utilizado el Modelo V, método empleado por el centro DATEC para garantizar el buen funcionamiento y la calidad de los productos de almacenes de datos (Fig. 18). El modelo V 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 siguiente figura muestra la relación entre los componentes del modelo.

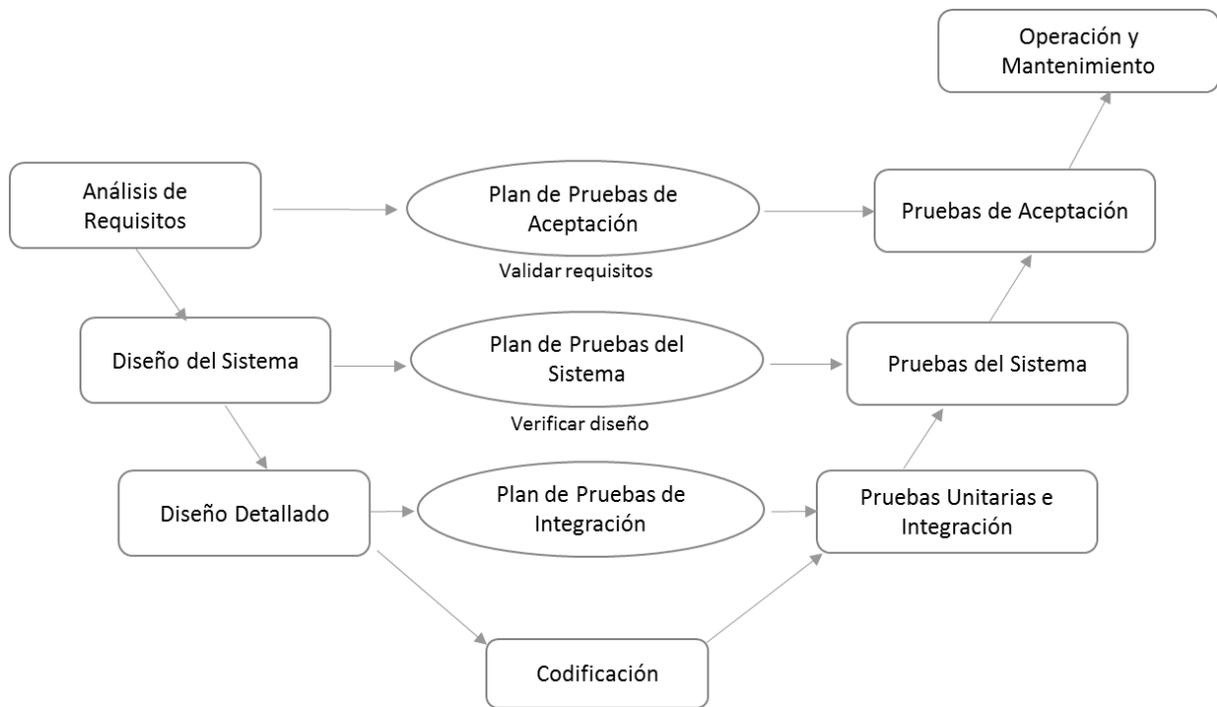


Fig. 18: Modelo V.

Según (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.

## Capítulo 3: Implementación y prueba del Mercado de Datos: Análisis de tendencias

**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.

### 3.4.1 Herramientas para la aplicación de las pruebas.

Para la aplicación de los tipos de pruebas definidos anteriormente se utilizarán las herramientas siguientes:

#### Casos de Prueba

Los casos de pruebas constituyen un conjunto de guías que incluye pasos y resultados esperados durante la ejecución de una prueba funcional del software. Además, describe los pasos detallados que serán seguidos para verificar el *software* (Palacio 2009).

Para el MD Análisis de Tendencias fueron diseñados cinco Casos de Prueba (CP), uno para cada caso de uso de información.

En la Tabla 13 se muestran los escenarios 2 y 3 del CP para el CU "Mostrar información sobre las NC de revisiones" correspondiente al indicador "Impacto de las NC y Adherencia a procesos y productos de trabajo".

Tabla 13: Diseño de CP para el CU "Mostrar información sobre el plan de revisiones."

Escenario	Descripción	Variables de entrada	Variables de salida	Respuesta del sistema	Flujo central
EC 2: Muestra información sobre el impacto de las NC.	Muestra la cantidad de NC de revisiones según su impacto.	Tiempo Centro Nivel de revisión Impacto de NC Áreas de proceso	Cantidad de NC	El sistema muestra todas las variables disponibles para el análisis, ubicadas en las filas y las columnas que pueden ser visualizadas en el reporte.	El usuario se autentica, selecciona el A.A.G Análisis de tendencias / A.A Revisiones / L.T Impacto y Adherencia de NC /Se selecciona el reporte correspondiente.
EC 3: Muestra información sobre la adherencia de	Muestra la cantidad de NC de revisiones por	Tiempo Centro Nivel de revisión	Cantidad de NC	El sistema muestra todas las variables disponibles para el análisis, ubicadas en	El usuario se autentica, selecciona el A.A.G Análisis de

## Capítulo 3: Implementación y prueba del Mercado de Datos: Análisis de tendencias

procesos y productos de trabajo.	el tipo de adherencia.	Impacto de NC	las filas y las columnas que pueden ser visualizadas en el reporte.	tendencias / A.A
		Áreas de proceso		Revisiones / L.T
				Impacto y Adherencia de NC /Se selecciona el reporte correspondiente.

### Listas de Chequeo

Las listas de chequeo consisten en un conjunto de preguntas que verifican el cumplimiento de los objetivos en aras de proporcionar el buen desarrollo del trabajo.

En el presente trabajo se aplicarán las listas de chequeo a los artefactos del proceso de ETL, tales como el Mapa Lógico de Datos, Diccionario de Datos y Perfilado de Datos y al artefacto Salidas del Sistemas del proceso de Inteligencia de Negocio (BI), todas con el objetivo de medir la seguridad de los datos cargados. Las listas de chequeo responderán a los siguientes elementos:

- Indicador a evaluar: Es el indicador a evaluar en las diferentes secciones.
- Peso: Describe la criticidad del indicador a evaluar.
- Evaluación: Forma de evaluar el indicador (1 si existe alguna dificultad, 0 en caso contrario).
- No procede: Especifica la necesidad de evaluar el indicador en ese caso.
- Cantidad de elementos afectados: Especifica la cantidad de errores detectados en el mismo indicador.
- Comentario: Especifica las sugerencias o señalamientos que desee incluir la persona que aplica las listas de chequeo.

Los indicadores a evaluar en las listas de chequeo estarán distribuidos en las siguientes secciones:

- Estructura del documento: Contiene los aspectos definidos por el Expediente de proyecto.
- Indicadores definidos por la etapa: Contiene los indicadores a evaluar durante la etapa de ETL y BI.
- Semántica del documento: Contiene los indicadores a evaluar respecto a la ortografía y redacción.

Luego de haber aplicado las listas de chequeo fueron detectadas 7 NC. La Fig. 19 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 3 indicadores, de ellos, 1 fue crítico y 1 NC detectada. En cuanto a los Indicadores definidos por la etapa, se identificaron 6 indicadores, de ellos, 4 fueron críticos y se detectaron 2 NC. Para la sección correspondiente a la Semántica del documento, se identificaron 4 indicadores, de estos, 2 fueron críticos y 1 NC detectada.

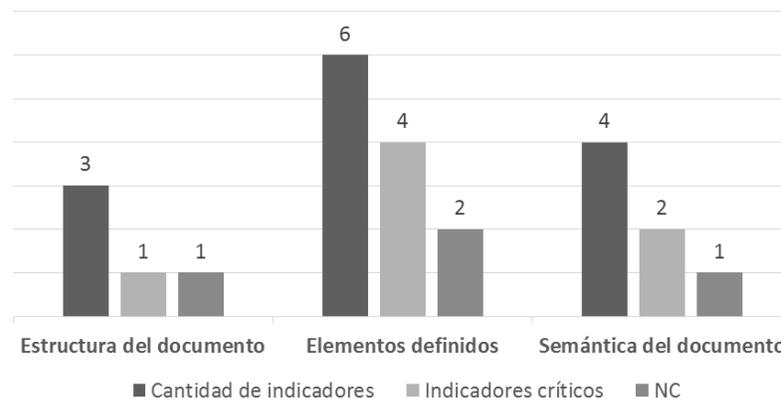


Fig. 19: Comportamiento de los indicadores de la Lista de Chequeo para el Mapa Lógico.

### 3.4.2 Resultados de las pruebas.

Las pruebas realizadas al MD Análisis de Tendencias arrojaron los siguientes resultados:

#### Pruebas unitarias y de integración

Las pruebas unitarias y de integración fueron aplicadas por miembros del equipo de desarrollo y especialistas del centro DATEC. Dichas pruebas arrojaron las siguientes NC, las cuales ya fueron resueltas:

- Redefinir las tablas de hechos de manera que responda a las necesidades de los usuarios.
- Poner nombres sugerentes a los pasos de las transformaciones.
- Optimizar la utilización de componentes en la implementación de las transformaciones.
- Comprobar que los datos cargados coincidan con los de la fuente.

Esta última NC fue resuelta mediante la aplicación de los Casos de Prueba definidos.

#### Pruebas del sistema

Las pruebas del sistema fueron realizadas por el Administrador de la Calidad del centro DATEC. La Tabla 14 muestra la relación con la cantidad de NC detectadas en cada iteración, las cuales ya fueron resueltas, y la complejidad de las mismas.

Tabla 14: Resultados de las pruebas de sistema.

Iteraciones	Cantidad de NC	Complejidad			NC Resueltas
		Alta	Media	Baja	
Iteración 1	6	5	1	0	6
Iteración 2	4	2	1	1	4

### Pruebas de aceptación

Las pruebas de aceptación fueron realizadas por el asesor de calidad de centros DATEC y GEYSED (Geoinformática y Señales Digitales) y a uno de los coordinadores de la alta gerencia, donde en una primera iteración fueron detectadas un total de 2 NC, las cuales fueron resueltas, y en la segunda iteración no se detectaron NC. De esta manera se obtuvo la carta de aceptación por parte del cliente (Anexo #5).

La gráfica representada en la Fig. 20 muestra un resumen de las NC detectadas en los niveles de pruebas: Unidad e Integración, Sistema y Aceptación, en las que se detectaron, 4, 10 y 2 NC respectivamente. Estas NC fueron resueltas satisfactoriamente.

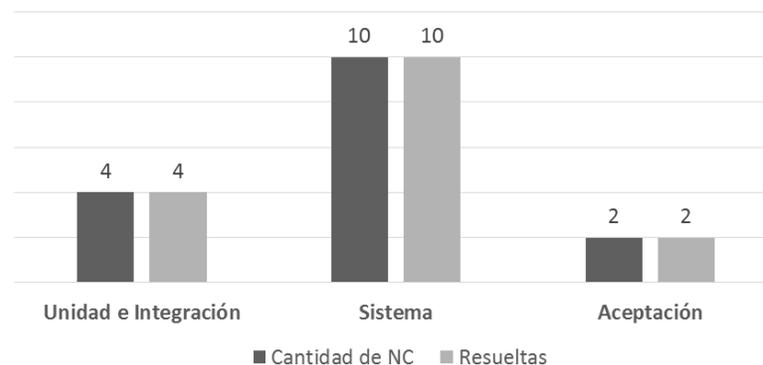


Fig. 20: Resultados de las pruebas.

### 3.5 Conclusiones.

En este capítulo se llevó a cabo la implementación del MD Análisis de Tendencias en base a las necesidades de los usuarios. Para ello se definieron los diferentes esquemas que conforman el MD, uno para las tablas de hechos, otro para las dimensiones y un último esquema que recoge las tablas con la información CDC. Fueron implementados, además, las transformaciones y trabajos que garantizaron la limpieza y estandarización de los datos y los cubos OLAP que permitieron el acceso a los datos correspondientes a las tablas de hechos.

Por otra parte se realizó la implementación del mapa de navegación o arquitectura de información del MD, permitiendo tener una mejor organización en cuanto a la visualización de la estructura de las Áreas de Análisis, los Libros de Trabajo y los reportes. También, con el objetivo de identificar posibles errores en la solución, se aplicaron los casos de pruebas y las listas de chequeo al MD Análisis de Tendencias.

## **Conclusiones generales**

El estudio de diversos temas relacionados con los Almacenes de datos permitió la elaboración del presente trabajo, el cual arrojó como resultado, el MD Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI, dándole cumplimiento, de esta forma, al objetivo por el que fue desarrollado y a las tareas propuestas para su implementación. Es por ello que se puede llegar a las siguientes conclusiones:

- Se confeccionó el marco teórico de la investigación sobre las principales metodologías utilizadas para el desarrollo de Almacenes de datos, donde se seleccionó la “Metodología para el desarrollo de proyectos de AD” como metodología a utilizar para la solución. De esta forma también se definieron y utilizaron las herramientas necesarias para la implementación de MD.
- El análisis del negocio permitió definir 13 requisitos de información, 30 requisitos funcionales, agrupados en cinco y siete casos de uso respectivamente y 10 requisitos no funcionales. El diseño de los Subsistemas de Almacenamiento, Integración y Visualización permitió identificar 6 tablas de hechos, 9 medidas y 12 tablas de dimensiones, dando lugar al modelo lógico y físico del MD, así mismo se diseñó las transformaciones, la arquitectura de información y los cubos OLAP.
- La implementación de los Subsistemas de Integración y Visualización permitió realizar la integración de los datos del área Análisis de Tendencias, mediante la implementación de 19 transformaciones y 2 trabajos. Se obtuvo la capa de visualización, la implementación de 7 cubos OLAP y 13 reportes candidatos.
- Las pruebas aplicadas al MD haciendo uso del Modelo V y mediante la utilización de las herramientas Casos de pruebas y las Listas de chequeo, arrojaron un total de 23 NC distribuidas en los cuatro niveles de pruebas que plantea el modelo utilizado. Estas NC ya fueron resueltas, garantizando así, el correcto funcionamiento del sistema, la calidad de producto final y la obtención de la carta de aceptación por parte del cliente.

## **Recomendaciones**

- Garantizar la escalabilidad del Mercado de datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI, de manera que cumpla con los objetivos estratégicos de próximos años en caso que cambien, aumenten o disminuyan los indicadores que se analizan el en Informe de Tendencias.
- Aplicar técnicas de Minería de datos como árboles de decisión y agrupamiento (clustering) para obtener patrones de predicción y comportamiento de las NC durante el Análisis de Tendencias.

### Referencias bibliográficas

- Basallo, Y. A. and A. D. Estrada (2009). "Una experiencia en integración de aplicaciones empresariales." *Revista Cubana de Ciencias Informáticas* 3.
- Casters, M., R. Bouman and J. Van Donged (2010). *Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration*.
- ConexiónEsan. (2015). "Las 20 herramientas de inteligencia de negocios que debes conocer." from <http://www.esan.edu.pe/apuntes-empresariales/2015/07/20-herramientas-inteligencia-negocios-debes-conocer/>.
- DataCleaner. (2008). "DataCleaner The premier data quality Solution." from <http://datacleaner.org>.
- 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/](http://community.pentaho.com/projects/bi_platform/).
- Elósegui, T. (2014). "¿Qué es un dashboard y para qué sirve?" *Marketing Online*.
- Gravitar (2014). "Data Integration."
- Hernández, Y. G. (2013). *METODOLOGÍA DE DESARROLLO PARA PROYECTOS DE ALMACENES DE DATOS*, Universidad de las Ciencias Informáticas.
- Inmon, W. H. (2005). *Building the Data Warehouse*, John Wiley.
- Kimball, R. (2004). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*, Wiley Publishing.
- Kimball, R. (2008). *The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems*.
- 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>.
- Martínez, R. (2013). "PostgreSQL-es." from [http://www.postgresql.org.es/sobre\\_postgresql](http://www.postgresql.org.es/sobre_postgresql).
- Orallo, J. H. (2010). "Data Warehouse, o Almacén de Datos, y Cubo OLAP." *El Mejor Proyecto* <http://elmejorproyecto2010.blogspot.com/2010/04/data-warehouse-o-almacen-de-datos-y.html>.
- Palacio, L. G. (2009). "Método para generar casos de prueba funcional en el desarrollo de software." *Revista Ingenierías Universidad de Medellín* 8.
- PARADIGM, C. V. (2013). "Visual Paradigm for UML - UML tool for software application development." from <http://www.visua-paradigm.com/product/vpum/>.
- Pentaho, C. (2015). "Schema Workbench."
- Peñaloza, M. (2002). *La industria del software, una oportunidad para México* Entérate.

- Perdomo, C. R. B. (2009). Estudio de la interacción en la consulta y visualización de Almacenes de Datos. Trabajo de Maestría, UNIVERSIDAD AUTÓNOMA DE MADRID.
- PgAdmin. (2016). "PgAdmin." from <https://www.pgadmin.org/index.php>.
- PostgreSQL. (2016). "The work's most advanced open source data base." from <https://www.postgresql.org/>.
- Pressman, R. S. (1998). Ingeniería de Software, un enfoque práctico.
- Pressman, R. S. (2002). Ingeniería del Software, un enfoque práctico.
- Pressman, R. S. (2005). Software Engineering: A practitioner's approach. S.I.
- Ramos, N. (2009). "Fundamentos y Casos de Aplicación." Soluciones de Inteligencia de Negocios a su alcance.
- Rosales, M. (2014). "Qué es CMMI." Software Development <http://software2014calidad1.blogspot.com/2014/04/que-es-el-cmmi.html>.
- Soto, E. G. (2001). Data Warehouse Antecedentes, Situación Actual y Tendencias.
- UCI. (2014). "Aseguramiento de la calidad de procesos y productos de trabajo." Mejora de Procesos de Software Retrieved octubre 24, 2015, from [mejoras.prod.uci.cu](http://mejoras.prod.uci.cu).
- Wolff, C. (2002). "Implementando un data warehouse." Revista Ingeniería Informática **5**.

## **Bibliografía**

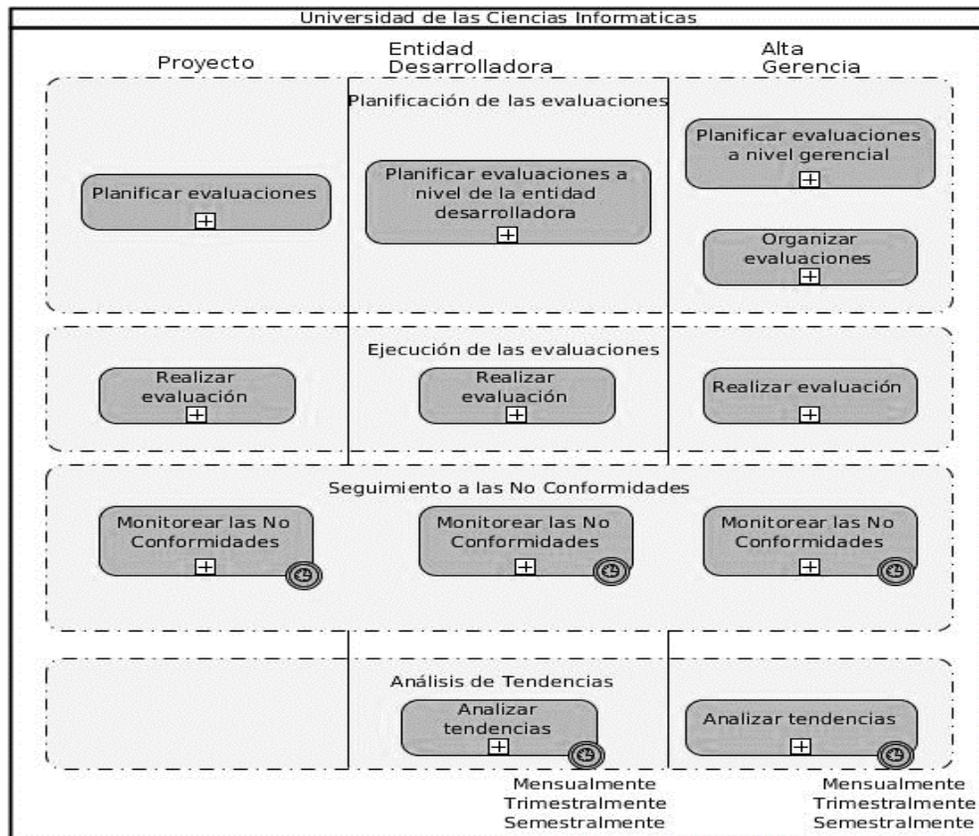
- Basallo, Y. A. and A. D. Estrada (2009). "Una experiencia en integración de aplicaciones empresariales." *Revista Cubana de Ciencias Informáticas* 3.
- Bernabeu, D. R. (2010). *Hefesto: Metodología para la construcción de un Data Warehouse*.
- Brizuela, E. I. L. (2013). "Metodologías para desarrollar Almacén de Datos." *Revista de Arquitectura e Ingeniería* 7.
- Cameron, N. (2003). "PostgreSQL affiliates." *Computerworld*, 2015, from [http://www.computerworld.com.au/article/62894/postgresql\\_affiliates\\_org\\_domain/](http://www.computerworld.com.au/article/62894/postgresql_affiliates_org_domain/).
- Casters, M., R. Bouman and J. Van Donged (2010). *Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration*.
- Company, P. A. H. G. (2009). "Pentaho Analysis Services." from <http://www.dataprix.com/empresa/productos/pentaho-bi-suite>.
- Company, P. A. H. G. (2015). "Data Integration." 2015, from <http://www.pentaho.com/product/data-integration#data-integration->.
- ConexiónEsan. (2015). "Las 20 herramientas de inteligencia de negocios que debes conocer." from <http://www.esan.edu.pe/apuntes-empresariales/2015/07/20-herramientas-inteligencia-negocios-debes-conocer/>.
- DataCleaner. (2008). "DataCleaner The premier data quality Solution." from <http://datacleaner.org>.
- Díaz, J. C. (2011). *Introducción al Business Intelligence*.
- Doménico, T. D. (2011). "Documentación para usuarios y desarrolladores." from <http://wiki.pentaho.com/display/EALes/Inicio>.
- EASY, P. P. A. M. (2005). "Pentaho BI Plataform and Server ", from [http://community.pentaho.com/projects/bi\\_platform/](http://community.pentaho.com/projects/bi_platform/).
- Ellingsworth, J. (2008). "CMS Integrations: LimeSurvey Integration." from <https://moodle.org/mod/forum/discuss.php?d=100630>.
- Elósegui, T. (2014). "¿Qué es un dashboard y para qué sirve?" *Marketing Online*.
- Flores, E. "Metodología de desarrollo de software. El Modelo en V o de Cuatro Niveles. ." from <http://ingenieriadesoftware.mex.tl>.
- García Valón, E., A. E. Góngora Rodríguez and H. Díaz Pérez (2010). *Herramientas para generar las estadísticas cuantitativas de las no conformidades. La Habana*.
- GENBETA. (2012). "DataCleaner, una sencilla opción para mantener las BD ordenadas." Retrieved 19 de marzo de 2016, 2016, from <http://www.genbeta.com/herramientas/datacleaner-una-sencilla-opcion-para-mantener-las-basesde-datos-ordenadas>.

- Gravatar. (2014). "Data Integration." from <http://gravitar.biz/pentaho/>.
- Hernández, Y. G. (2013). METODOLOGÍA DE DESARROLLO PARA PROYECTOS DE ALMACENES DE DATOS, Universidad de las Ciencias Informáticas.
- Inmon, W. H. (2005). Building the Data Warehouse, John Wiley.
- JMeter. (2016). "Apache JMeter." from <http://jmeter.apache.org/>.
- Kimball, R. (2004). The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Wiley Publishing.
- Kimball, R. (2008). The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems.
- Kimball, R. and M. Ross (2002). The Data Warehouse Toolkit. New York.
- 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>.
- LimeSurvey. (2015). "LimeSurvey Manual." from [https://manual.limesurvey.org/LimeSurvey\\_Manual](https://manual.limesurvey.org/LimeSurvey_Manual).
- Martínez, R. (2013). "PostgreSQL-es." from [http://www.postgresql.org.es/sobre\\_postgresql](http://www.postgresql.org.es/sobre_postgresql).
- Nieves Borrero, M. and A. Góngora Rodríguez (2007). Herramienta Informática para automatizar los procesos en el Laboratorio de Calidad de Software: Módulo Gestión de las No Conformidades. La Habana.
- Orallo, J. H. (2010). "Data Warehouse, o Almacén de Datos, y Cubo OLAP." El Mejor Proyecto <http://elmejorproyecto2010.blogspot.com/2010/04/data-warehouse-o-almacen-de-datos-y.html>.
- Palacio, L. G. (2009). "Método para generar casos de prueba funcional en el desarrollo de software." Revista Ingenierías Universidad de Medellín 8.
- 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/>.
- Peñaloza, M. (2002). La industria del software, una oportunidad para México Entérate.
- Perdomo, C. R. B. (2009). Estudio de la interacción en la consulta y visualización de Almacenes de Datos. Trabajo de Maestría, UNIVERSIDAD AUTÓNOMA DE MADRID.
- PgAdmin. (2016). "PgAdmin." from <https://www.pgadmin.org/index.php>.
- Ponniah, P. (2010). Data warehousing fundamentals for IT professionals, John Wiley & Sons.
- PostgreSQL. (2016). "The work's most advanced open source data base." from <https://www.postgresql.org/>.
- Pressman, R. S. (1998). Ingeniería de Software, un enfoque práctico.
- Pressman, R. S. (2002). Ingeniería del Software, un enfoque práctico.

- Pressman, R. S. (2005). *Software Engineering: A practitioner's approach*. S.I.
- Ramos, N. (2009). "Fundamentos y Casos de Aplicación." *Soluciones de Inteligencia de Negocios a su alcance*.
- Roche, R. (2008-2009). "Revista Bioanálisis de Grupo Bio SRL." from [http://www.revistabioanalisis.com/ejemplares/ejemplar/notas/index.php?id\\_edicion=25&id\\_nota=313](http://www.revistabioanalisis.com/ejemplares/ejemplar/notas/index.php?id_edicion=25&id_nota=313).
- Rosales, M. (2014). "Qué es CMMI." *Software Development* <http://software2014calidad1.blogspot.com/2014/04/que-es-el-cmmi.html>.
- Sommerville, I. (2006). *Software Engineering*.
- Soto, E. G. (2001). *Data Warehouse Antecedentes, Situación Actual y Tendencias*.
- UCI. (2014). "Aseguramiento de la calidad de procesos y productos de trabajo." *Mejora de Procesos de Software* Retrieved octubre 24, 2015, from [mejoras.prod.uci.cu](http://mejoras.prod.uci.cu).
- Wolff, C. (2002). "Implementando un data warehouse." *Revista Ingeniería Informática* **5**.

## Anexos

### Anexo #1: Descripción del proceso de PPQA.



**Anexo #2:** Requisitos de información.

**Área de Aseguramiento de la Calidad de Procesos y Productos de trabajo (PPQA)**

**Mercado de Datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI**

1. Obtener la cantidad de revisiones por estado, trimestre, centro y nivel de revisión.
2. Obtener la media de NC de revisiones por trimestre, centro y nivel de revisión.
3. Obtener la cantidad de NC por impacto de la NC, áreas de proceso, trimestre, centro y nivel de revisión.
4. Obtener la cantidad de NC por el tipo de adherencia, áreas de proceso, trimestre, centro y nivel de revisión
5. Obtener la frecuencia media de NC por tipo de causa, trimestre, centro y nivel de revisión.
6. Obtener la cantidad de NC de revisiones por estado, trimestre, centro y nivel de revisión.
7. Obtener el porciento de respuestas (sí/no) por área de proceso, tipo de criterio, trimestre, centro y nivel de revisión.
8. Obtener el porciento de respuestas (s/no) por áreas de proceso, trimestre, centro y nivel de revisión.
9. Obtener la cantidad de liberaciones por estado, trimestre, centro y nivel de revisión.
10. Obtener porciento de error de las NC por tipo de error, clasificación de las NC de liberación, trimestre, centro y nivel de revisión.
11. Obtener porciento de liberaciones por iteraciones, clasificación de las NC de liberación, trimestre, centro y nivel de revisión.
12. Obtener la cantidad de NC por iteraciones, tipo de error, clasificación de las NC de liberación, nivel de revisión, centro y tiempo.
13. Obtener la cantidad de NC por iteraciones, estado, clasificación de las NC de liberación, trimestre, centro y nivel de revisión.



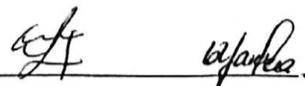
Ing. Dairys Febles Pérez

Cliente



Ing. Alberto Mendoza Garnache

Tutor



Elian E. Castell y Yanitza de la C.

Tesistas

**Anexo #3:** Nomenclatura establecida para las actividades de calidad.

**Revisiones de Adherencia de Proceso y Producto (PPQA)**

**Para las solicitadas a la Dirección de Calidad**

RCU-XX-XXX (Donde RCU: Revisiones de Calidad de la Universidad, XX: año de realización, XXX: número de control consecutivo de revisiones)

**Para las Internas del proyecto:**

RCI-XX-XXX (Donde RCI: Revisiones de Calidad Internas del proyecto y los demás elementos como están descritos anteriormente)

**Para las Internas al proyecto por parte del Centro:**

RCC-XX-XXX (Donde RCC: Revisiones de Calidad Internas del Asesor de Calidad de la Entidad desarrolladora y los demás elementos como están descritos anteriormente)

**Revisiones Técnicas Formales (RTF)**

RTF-XX-Tipo (Donde Tipo de RTF: Planificación, Requisitos, Arquitectura, Base de Datos)

**Auditorías a la Configuración**

AC-XX-XXX (Donde AC: Auditorías a la configuración y los demás elementos como están descritos anteriormente)

**Revisiones de Inconsistencia**

RI-XX-XXX (Donde RI: Revisiones de inconsistencias y los demás elementos como están descritos anteriormente)

**Pruebas de liberación**

PL-XX-XXX (Donde PL: Prueba de Liberación, XX: año de realización, XXX: número consecutivo de liberaciones)

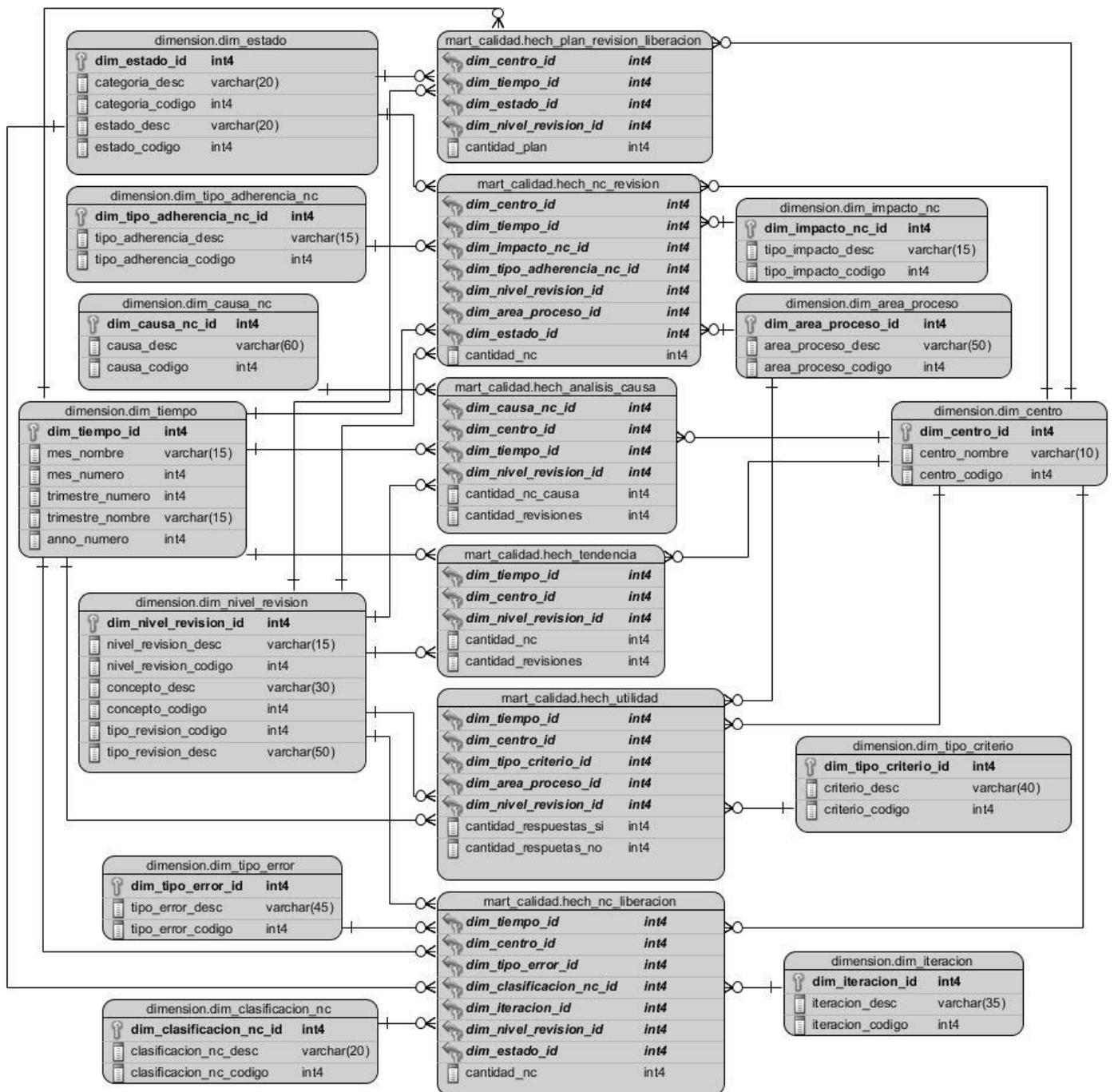
**Pruebas Internas**

PI-XX-XXX (Donde PI: Prueba Interna, XX: año de realización, XXX: número consecutivo de prueba)

**Pruebas de Aceptación**

PA-XX-XXX (Donde PA: Prueba de Aceptación, XX: año de realización, XXX: número consecutivo de prueba)

Anexo #4: Modelo de datos ampliado.



Anexo #5: Acta de aceptación del producto final.



## Mercado de Datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI

### ACTA DE ACEPTACIÓN

Entre el Centro de Tecnologías de Gestión de Datos (DATEC), representado por Dairys Febles Pérez, portador de carné de identidad 87100912899, en su condición de Asesora de Calidad del centro DATEC, facultado para este acto, que en lo sucesivo se denominará la "**PARTE CLIENTE**", por una parte; y por la otra, el equipo desarrollador del **Mercado de Datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI**, representado en este acto por los estudiantes Elian Eduardo Castell Legrá y Yanitza de la Caridad Gutiérrez Acea, portadores de carné de identidad 92051929443 y 93071537613 respectivamente, quienes actúan en su condición de TESISISTAS, suficientemente facultados para este acto, que en lo sucesivo se denominará la "**PARTE PROVEEDORA**"; acuerdan expresamente que:

La **Parte Cliente**, luego de concluir el proyecto **Mercado de Datos: Análisis de Tendencias para la Actividad de Desarrollo-Producción UCI**, determina que el mismo se efectuó satisfactoriamente.

Comentarios:

Y para que así conste se suscribe la presente **Acta** en La Habana a los 2 días del mes de junio de **2016**.

Entrega: Elian E. Castell Legrá

Cargo: Estudiante

Firma: [Firma]

Entrega: Yanitza de la Caridad Gutiérrez Acea

Cargo: Estudiante

Firma: [Firma]

Recibe: Ing. Aldis Joan Abreu Medina

Cargo: Director del Centro DATEC

Firma: [Firma]