

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

**ANÁLISIS, DISEÑO E IMPLEMENTACIÓN UNA SOLUCIÓN DE
INTELIGENCIA DE NEGOCIOS ORIENTADA A CONTROLAR LOS
PROCESOS DE GENERACIÓN Y EMISIÓN DEL DNI (DOCUMENTO
NACIONAL DE IDENTIDAD) EN EL RENIEC (REGISTRO NACIONAL DE
IDENTIFICACIÓN Y ESTADO CIVIL).**

Tesis para optar por el Título de Ingeniero Informático, que presenta el bachiller

Joao Danilo Chavez Yrigoyen

ASESORA: Ing. Sofia Dapello Sarango

Resumen

El presente proyecto tiene como objetivo principal analizar, diseñar e implementar una solución de Inteligencia de Negocios orientada a controlar los procesos de generación y emisión del DNI (Documento Nacional de Identidad) en el RENIEC (Registro Nacional de Identificación y Estado Civil) para poder brindarle una herramienta que a través de indicadores muestre el desempeño del proceso.

La propuesta de una solución de Inteligencia de Negocios surge como alternativa para manejar la gran cantidad de información que posee RENIEC. Mediante el uso de la herramienta se organizará adecuadamente la información generada en cada una de las unidades orgánicas (Gerencias, Jefaturas Regionales y otras) que gestionan el proceso de generación y emisión del DNI, y de ésta forma se brindarán indicadores que son de mucha utilidad a las diferentes unidades orgánicas. De esta manera las personas que deseen acceder a estos indicadores solo deberán utilizar la herramienta.

En el capítulo 1 se presenta la identificación del problema que originó la propuesta de la presente solución, los objetivos y resultados esperados. También se detallan las metodologías de gestión de proyectos y de desarrollo de software que se seguirán. Finalmente se analizan herramientas similares existentes en el mercado y se justifica la realización del presente proyecto.

En el capítulo 2 se presenta el marco conceptual necesario para entendimiento del proyecto y el estado del arte.

En el capítulo 3 se realiza la definición de la metodología de la solución y se identifican los requerimientos de la solución.

En el capítulo 4 se define el diseño conceptual del Data Mart, el diseño lógico del Data Mart, la arquitectura de la solución, y el diseño de extracción.

En el capítulo 5 se construye la solución de Inteligencia de Negocios y se definen las pruebas.

Finalmente, en el capítulo 6 se presentan las conclusiones del presente proyecto y las recomendaciones para trabajos futuros.

ÍNDICE

1	Generalidades _____	8
1.1	Definición del problema _____	8
1.2	Objetivo general _____	11
1.3	Objetivos específicos _____	11
1.4	Resultados esperados _____	11
1.5	Métodos, metodologías y procedimientos. _____	12
1.5.1	Relacionados a la gestión del proyecto _____	12
1.6	BenchMark sobre Herramientas de Inteligencia de negocio. _____	15
1.6.1	Plataforma Bi de Pentaho _____	17
1.7	Alcance y limitaciones del proyecto _____	19
1.7.1	Alcance _____	19
1.7.2	Limitaciones _____	20
1.8	Viabilidad y justificativa del proyecto _____	20
2	Marco Conceptual y Estado del Arte _____	21
2.1	Marco Conceptual _____	21
2.1.1	Conceptos relacionados al problema _____	21
2.1.2	Conceptos relacionados a la propuesta de solución _____	32
2.2	Estado de Arte _____	41
2.2.1	Problemas relacionados _____	41
2.2.2	Conclusiones sobre el estado del arte _____	47
3	Análisis _____	48
3.1	Definición de la metodología de la solución _____	48
3.2	Identificación de requerimientos _____	50
3.2.1	Requerimientos funcionales _____	50
3.2.2	Requerimientos no funcionales _____	53
4	Diseño _____	54
4.1	Diseño conceptual del Data Mart _____	54
4.1.1	Esquema de producción _____	55
4.1.2	Esquema de calidad _____	56
4.2	Diseño Lógico del Data Mart _____	56
4.2.1	Esquema de producción _____	57
4.2.2	Esquema de calidad _____	58
4.3	Arquitectura de la solución _____	58
4.3.1	Fuentes de información _____	60
4.3.2	ETL _____	60
4.3.3	Data Mart _____	60
4.4	Diseño de extracción _____	62
4.4.1	Transformación de la dimensión DIM_LOCAL _____	62

5	Construcción	65
5.1	Construcción	65
5.1.1	Instalación de las herramientas	65
5.1.2	Proceso ETL	66
5.1.3	Configuración del Schema Workbench, base de datos del Data Mart y los usuarios	69
5.1.4	Elaboración del cubo	69
5.1.5	Análisis de los datos	71
5.2	Pruebas	78
5.2.1	Planificación del plan de pruebas.	78
5.2.2	Diseño del plan de pruebas.	79
5.2.3	Determinación de los casos de prueba	79
5.2.4	Ejecución del plan de pruebas	80
5.2.5	Análisis y evaluación del plan de pruebas ejecutado.	80
5.2.6	Caso de Prueba para el ETL_FACT_1 de fact_pro_cantidad	80
6	Observaciones, conclusiones y trabajos Futuros	85
6.1	Observaciones	85
6.2	Conclusiones	86
6.3	Recomendaciones y Trabajos Futuros	87
6.3.1	Implementar herramientas de Inteligencia de Negocios	87
6.3.2	Ampliación a áreas y departamentos	87
7	Referencias	88
	PROJECT MANAGEMENT INSTITUTE	90

Índice de Figuras

Figura 1. Esquema de la plataforma BI [PENTAHO PLATAFORMA DE BI 2013]	17
Figura 2. Esquema de las gerencias de GOR, GRI, GRIAS de RENIEC [RENIEC02 2012]	21
Figura 3. Diagrama de flujo registro de trámite y entrega del DNI [RENIEC07 2012]	26
Figura 4. Esquema General del Proceso en la GRI [RENIEC03 2012]	27
Figura 5. Diagrama de flujo detallado de la línea de procesamiento del DNI[RENIEC05 2012]	28
Figura 6 Diagrama de flujo -línea de procesamiento del DNI [RENIEC06 2012]	29
Figura 7. Procesos de GRIAS [RENIEC04 2012]	30
Figura 8. Procesos de GRIAS [RENIEC04 2012]	31
Figura 9. Entorno básico del Data Warehouse [REEVES 2009]	34
Figura 10. Vista general de la información de un Data Warehouse [PONNIAH 2010]	34
Figura 11. Esquema Estrella [PONNIAH 2010]	36
Figura 12. Esquema Copo de Nieve [PONNIAH 2010]	37
Figura 13. Proceso de Carga (ETL) [MEDINA LA PLATA, 2012]	39
Figura 14. Dashboard [FEW 2006].....	40
Figura 15. Dashboard [FEW 2006].....	40
Figura 16. Página web INEI [INEI01]	46
Figura 17. Página web INEI [INEI01]	46
Figura 18. Fases generales de Desarrollo según Kimball [Elaboración Propia].....	48
Figura 19. Esquema de producción [Elaboración propia]	55
Figura 20. Esquema de calidad [Elaboración propia]	56
Figura 21. Diseño lógico del esquema de producción [Elaboración propia]	57
Figura 22. Diseño lógico del esquema de calidad [Elaboración propia]	58
Figura 23. Arquitectura de la solución [Elaboración propia]	59
Figura 24. Diagrama de componentes de la base de datos del Data Mart [Elaboración propia]	61
Figura 25. Diagrama de componente del servidor de base de datos. [Elaboración propia].....	61
Figura 26. Imagen de carga de una dimensión usando la herramienta Pentaho.	64
Figura 27. Proceso ETL de una dimensión.	68
Figura 28. Proceso ETL de una tabla Hecho	69
Figura 29. Diseño del cubo	70
Figura 30. Diseño del cubo	70
Figura 31. Análisis de la información	72
Figura 32. Análisis de la información	73
Figura 33. Análisis de la información	74
Figura 34. Análisis de la información	75
Figura 35. Análisis de la información	76
Figura 36. Reporte General.....	77
Figura 37. Reporte en Report Designer.	77
Figura 38. Tabla Mov_ficha_registral	80
Figura 39. Tabla Mov_ficha_registral	81
Figura 40. Tabla Lote_impresion.....	81
Figura 41. Tabla Región.....	81
Figura 42. Tabla Despacho_dni	82
Figura 43. Tabla fuente Intermedia 1	82
Figura 44. Tabla fuente Intermedia 2	82
Figura 45. Tabla fuente Intermedia 3	83
Figura 46. ETL Intermedio.....	83
Figura 47. Tabla Final	84

Índice de Tablas

Tabla 1. Comparación de diferentes Herramientas de BI.....	17
Tabla 2. Datos necesarios por gerencia.....	32
Tabla 3. Características del Hardware.	62
Tabla 4. Características del Hardware.	62
Tabla 5. Tabla fuente.	62
Tabla 6. Tabla de estandarización y limpieza de datos.....	63
Tabla 7. Tabla de fuente de datos.....	63
Tabla 8. Tabla de Destino.	63
Tabla 9. Ejemplo de caso de prueba.....	80



Índice de Anexos

Anexo 1: Modelado de los procesos de generación y emisión del DNI. _____	7
Anexo 2: Fuente de datos transaccional _____	12
Anexo 3: Proceso ETL _____	13
Anexo 4: Configuraciones _____	37
Anexo 5: Publicar el archivo _____	43
Anexo 6: Esquemas de análisis y gráficos _____	45
Anexo 7: Gantt Chart del Proyecto de Tesis _____	55
Anexo 8: Documento de Entrevista 001 _____	57
Anexo 9: Documento de Pruebas _____	58
Anexo 10: Constancia de Aceptación _____	67

1 Generalidades

1.1 Definición del problema

En el siglo XXI la información es de suma importancia y es uno de los principales recursos dentro de cualquier empresa. Si se desea maximizar la utilidad que posee la información de un proceso Principal de la empresa, ésta se debe manejar de manera correcta y eficiente. Una forma de hacerlo es obteniendo indicadores que permitan controlarlo y mejorarlo.

Existen diversas empresas que encontraron solución al problema del manejo de la información con el uso de herramientas de inteligencia de negocios, específicamente utilizando la plataforma Pentaho. Como por ejemplo la empresa BANELCO y la empresa BNSF LOGISTIC. La primera tenía un inadecuado control de su rendimiento, debido a las consultas manuales que realizaba y la gran cantidad de información que manejaba. El beneficio logrado fue establecer una ejecución automática de los reportes necesarios para el negocio. La segunda contaba con gran cantidad de información, falta de integridad de la data, problemas para la elaboración y distribución para los reportes dentro de su organización. Los beneficios logrados fueron: facilidad de acceso a los reportes, una rápida y mejor toma de decisiones por parte de los empleados en los procesos de ventas, monitoreo de forma efectiva los procesos del negocio. [PENTAHO CUSTOMERS SUCCESSES 2013].

Situándose específicamente en el caso del RENIEC (Registro Nacional de Identificación y Estado Civil), ésta entidad del estado define al DNI (Documento Nacional de Identidad) como la única cédula de identidad entregada a los peruanos mayores y menores de edad, siendo éste el único título de derecho al sufragio para los mayores de 18 años [RENIEC01]. Debido a la importancia que tiene el DNI para el ciudadano y por ende para el RENIEC, se debe dar la relevancia adecuada a los procesos relacionados al DNI, sobre todo a aquellos que afecten de manera directa a la población.

En el trabajo de campo realizado en el área de informática y a través de entrevistas, se encontró un problema que ocurre de manera frecuente, cada vez que alguna de las Unidades de la Alta Dirección necesita de alguna estadística con respecto al proceso de generación y emisión del DNI, en este caso debe contactarse con el Área de Estadística de la Institución y a la vez ésta debe solicitar la información a la Gerencia de Informática (Base de Datos) , esperar la confirmación, acceder a ésta, y preparar las estadísticas adecuadas. Además, se encontró que el proceso es repetitivo para las diferentes gerencias, desperdiciando horas/hombre en su realización. Es pertinente mencionar que la base de datos de RENIEC no solo trabaja con un sistema transaccional, y en caso sea necesario debe integrar la información antes de presentarla adecuadamente. [RENIEC08 2012]

Una alternativa de solución para cubrir estas necesidades de información, es con el uso de una solución de Inteligencia de Negocios, que cuente con diferentes Data Mart's con información consolidada. El Data Mart es una estructura de datos que está optimizada para la consulta, a diferencia de un sistema normal que puede consumir muchos recursos. Una vez cargado con información, el modelo estaría listo para el análisis, esto se logra mediante la explotación de los datos y mostrando los indicadores adecuados para cada gerencia. Las ventajas de usar esta estructura de datos es que permite acceder más rápido a la información, ya que solo usaría la información relevante a los procesos, integrar la información de diferentes fuentes y reducir el costo de acceder a la información histórica. En cuanto al software a utilizar se eligió la plataforma Pentaho Data Integration, en ésta se diseñan, se ejecutan los procesos de carga de información al Data Mart, también se elaboran diferentes esquemas de análisis de una manera sencilla y directa, y por último permite la elaboración de diferentes tipos de reportes con los cuales se explota la información del Data Mart.

En la presente tesis se han identificado las áreas de producción y rendimiento, para cada una de ellas se ha desarrollado un Data Mart. Con el uso de estos se obtuvo una mejora en el tiempo en la elaboración de reportes, el usuario tiene ahora mayor facilidad para acceder a la información pertinente cuando desee y cuenta con una mayor agilidad en la toma de decisiones.

Los indicadores han sido seleccionados de acuerdo a un análisis de los requerimientos de las Gerencias, que permiten enfocarse en la productividad de los procesos de generación y emisión del DNI. La información estadística junto con la experiencia de cada gerencia, resulta fundamental para la toma de medidas necesarias y de esta manera mejorar dichos procesos. Los indicadores encontrados son:

- Cantidad de DNI's tramitados agrupados por tipo de trámite, local, tipo de DNI, en un periodo determinado.
- Cantidad de DNI's tramitados agrupados por tipo de trámite, UBIGEO de domicilio o de nacimiento, tipo de DNI, en un periodo determinado.
- Cantidad de DNI's emitidos agrupados por tipo de trámite, local, tipo de DNI, en un periodo determinado.
- Cantidad de trámites en un rango de duración predeterminado desde la fecha de trámite hasta la fecha de ingreso a impresiones.
- Cantidad de trámites en un rango de duración predeterminado desde la fecha de ingreso a impresiones hasta las fecha de salida.
- Cantidad de trámites en un rango de duración predeterminado desde la fecha de salida de impresiones hasta la fecha de entrega al local.
- Ranking con los motivos de los trámites observados agrupados por locales, tipo de DNI, en un periodo determinado.

Para la elaboración del presente proyecto se tendrá como base la metodología de Kimball, la cual guía en el proceso de la elaboración de los Data Mart's, los procesos mediante los cuales se extrae la información de las fuentes transaccionales, se transforma la información y se cargan a los Data Mart's.

Los usuarios finales para la herramienta serán: Analistas de información (gerencias GOR, GRI, GRIAS), Power Users (Gerencia de Informática), Consumidores de información (Consultan información gerencias GOR, GRI, GRIAS), Tomadores de

Decisiones (Gerentes, Sub-gerentes), Departamento de TI (Sub-gerencia de gestión de base de datos). [RENIEC08 2012]

1.2 Objetivo general

Analizar, diseñar e implementar una solución de Inteligencia de Negocios orientada a controlar los procesos de generación y emisión del DNI (Documento Nacional de Identidad) en el RENIEC (Registro Nacional de Identificación y Estado Civil).

1.3 Objetivos específicos

- Realizar el análisis y el modelado de los procesos de generación y emisión del DNI.
- Diseñar la arquitectura lógica de la solución y la configuración de los datos a usar.
- Realizar la construcción de los procesos ETL para la extracción, transformación y carga de datos desde un sistema transacción a un sistema de inteligencia de negocios.
- Desarrollar los diferentes reportes para la explotación de la información del Data Mart (esquemas de análisis, gráfico sectorial, gráfico de barras separadas).
- Realizar la prueba de la herramienta configurada con la data de RENIEC.

1.4 Resultados esperados

Resultado 1 para el objetivo 1:

Documento con el modelo de los procesos de generación y emisión del DNI.

Resultado 2 para el objetivo 2:

Documento con el diseño de la arquitectura lógica de la solución y la configuración de los datos a usar.

Resultado 3 para el objetivo 3:

Documento con los procesos ETL para la extracción, transformación y carga de datos desde el sistema transaccional al Data Mart.

Resultado 4 para el objetivo 4:

Prototipos de los diferentes reportes para la explotación de la información del Data Mart (esquemas de análisis, gráfico sectorial, gráfico de barras separadas).

Resultado 5 para el objetivo 5:

Documento con las pruebas de la herramienta configurada.

1.5 Métodos, metodologías y procedimientos.

1.5.1 Relacionados a la gestión del proyecto

1.5.1.1 PMBOK para gestión de proyectos.

Para la gestión del proyecto se empleará una metodología basada en la guía del PMBOK (de las siglas en inglés Project Management Body of Knowledge).

El PMBOK propone las mejores prácticas para la gestión de un proyecto, en este caso un proyecto informático. En el PMBOK se menciona que todo proyecto debe tener en cuenta nueve áreas de conocimiento para una buena gestión, las cuales son: Integración, Alcance, Tiempo, Costes, Calidad, Recursos Humanos, Comunicaciones, Riesgos y Adquisiciones. Estas áreas están a su vez agrupadas en áreas centrales y facilitadoras. Las áreas centrales corresponden a las 4 primeras áreas mencionadas excepto por la de Integración, el resto corresponden a las áreas facilitadoras. [PROJECT MANAGEMENT INSTITUTE 2008]

Las áreas que se tomarán en cuenta son las siguientes:

Gestión de la Integración del Proyecto: sirve para identificar, combinar, unificar y coordinar los diversos procesos y actividades a realizar. Incluye características de unificación, consolidación, articulación, así como las acciones integradoras que son cruciales para la terminación del proyecto, la satisfacción de los interesados y el cumplimiento de los requisitos. Por ello, se elaboró el plan del proyecto para la gestión de la ejecución del proyecto y el control del mismo mediante reuniones con el asesor.

Gestión del Alcance del Proyecto: sirve para asegurar que se desarrolle el trabajo que se requiere, para completar el proyecto de manera exitosa. Por ello, se realiza la definición del alcance y periódicamente se verifica que se esté cumpliendo mediante los entregables. Durante esta etapa se toman los requerimientos en el RENIEC, y se define hasta donde va a abarcar el presente proyecto de tesis.

Gestión del Tiempo del Proyecto: se refiere a los procesos requeridos para asegurar que el proyecto termine en el tiempo planeado. Por ello, se definirán las actividades involucradas en el proyecto y se estimará su duración. A partir de esto, se desarrollará el cronograma del proyecto que será controlado a lo largo del proyecto. En el Anexo 7 se encuentra el Gantt Chart del proyecto de Tesis.

Gestión de Riesgos del Proyecto: se refiere a los procesos involucrados con la identificación, análisis y respuesta a los riesgos del proyecto. Por ello, se desarrollará un plan de riesgos, dentro de este plan se encuentran las acciones que mitiguen el riesgo y se dará seguimiento durante todo el proyecto en caso se identifiquen nuevos riesgos.

Gestión de las Comunicaciones del Proyecto: se refiere a los procesos para garantizar que la generación, recopilación, almacenamiento, distribución y recuperación de la información del proyecto sean oportunos. Por ello, se definirán los interesados y se gestionará la comunicación con el asesor del presente proyecto de fin de carrera. Ello se evidenciará con la planificación de reuniones y se especificarán los medios para distribuir la información.

En cuanto a las áreas de conocimiento que no se van a emplear son:

Gestión de los Costos del Proyecto: no se utilizó, ya que en el proyecto no se presenta un presupuesto asignado para el mismo.

Gestión de Calidad del Proyecto: no se utilizó, debido a que no es un proyecto de gran envergadura.

Gestión de Recursos Humanos: no se utilizó, ya que el equipo que desarrolló el proyecto es de una sola persona.

Gestión de las Adquisiciones del Proyecto: no se utilizó, ya que las herramientas a usar representarán una responsabilidad propia del ejecutor del proyecto adquirirlas y no son muchas como para realizar una gestión como tal.

En el PMBOK se muestran actividades a realizar en cada grupo de procesos, por lo que es una razón más para emplear esta metodología como guía durante el proyecto. Así por todo lo expuesto anteriormente, se confirma el uso de esta guía para la ejecución del proyecto, ya que es la que mejor se adapta para el mismo y porque brinda las mejores prácticas para la gestión del mismo.

1.5.1.2 Relacionados a los resultados esperados

1.5.1.2.1 Metodología BPMN

Resultado Esperado logrado con la metodología

Resultado1: Documento con el modelo de los procesos de generación y emisión del DNI.

Business Process Management (BPM) es un conjunto de métodos, herramientas y tecnologías utilizados para diseñar, representar, analizar y controlar procesos de negocio operacionales. BPM es un enfoque centrado en los procesos para mejorar el rendimiento que combina las tecnologías de la información con metodologías de proceso y gobierno. BPM es una colaboración entre personas de negocio y tecnólogos para fomentar procesos de negocio efectivos, ágiles y transparentes. [BPM, GERENCIA DE PROCESOS DE NEGOCIO 2012]

BPM combina métodos ya probados y establecidos de gestión de procesos con una nueva clase de herramientas de software empresarial. Ha posibilitado adelantos muy importantes en cuanto a la velocidad y agilidad con que las organizaciones mejoran el rendimiento de negocio.

BPMN (Business Process Modeling Notation)

La notación para gestión de procesos del negocio (BPMN) es el estándar más reciente para el modelado de procesos del negocio. BPMN permite hacer un mejor uso de la gestión de procesos del negocio (BPM), ya que normaliza el método de notación que sirve como ayuda en la automatización de procesos. Se encuentra mantenido por el grupo OMG de forma gratuita. [SILVER 2011].

Para el proyecto se usará toda esta metodología debido a que es aceptada como estándar actualmente y se tiene mayor facilidad de acceso a referencias. Además se usará toda la simbología y el principio de que todos los diagramas tienen una sola interpretación. No se llegará a utilizar el segundo principio de salida de archivos ya que el alcance del proyecto solo cubre la representación gráfica.

1.5.1.2.2 Metodología Ralph Kimball

Resultados Esperados logrados con la metodología

Resultado 2: Documento con el diseño de la arquitectura lógica de la solución y la configuración de los datos a usar.

Resultado 3: Documento con los procesos ETL para la extracción, transformación y carga de datos desde el sistema transaccional al Data Mart.

Resultado 4: Prototipos de los diferentes reportes para la explotación de la información del Data Mart (esquemas de análisis, gráfico sectorial, gráfico de barras separadas).

La metodología de Ralph Kimball propone como arquitectura de una solución de soporte de decisiones, armar primero Data Mart's para satisfacer las necesidades específicas de un departamento o área dentro de la empresa, permitiendo así un mejor control de la información que se está abarcando.

La presente metodología tiene como base el ciclo de vida de los Data Warehouses definido por Ralph Kimball. El esquema presentado por Ralph Kimball con el nombre de BDL (Business Dimensional Lifecycle, Ciclo de vida dimensional del negocio) grafica las diferentes etapas por las que debe pasar todo proceso de Data Warehousing. [KIMBALL 2002].

1.6 BenchMark sobre Herramientas de Inteligencia de negocio.

A continuación definiremos algunas herramientas de Inteligencia de negocio que se encuentran en el mercado.

Oracle Enterprise BI server: Es una plataforma de inteligencia de negocio que maneja una gran capacidad de análisis. Diseñada para ser escalable, confiable y para que su rendimiento sea el adecuado. La plataforma está basada en una moderna arquitectura de servicio web. [ORACLE 2013].

Business Objects Enterprise XI: Esta plataforma fue diseñada para integrarse perfectamente con las webs existentes y beneficiar a las aplicaciones del cliente, sin imponer un nuevo conjunto de normas y procesos. Esta plataforma puede abarcar a toda la organización. Permite explorar, extraer, transformar y distribuir datos para que

los usuarios puedan disponer de información precisa, puntual y fiable. [BUSINESS OBJECTS 2013].

Microsoft BI Tools: Esta plataforma satisface las necesidades de inteligencia de negocios de ejecutivos y empleados, minimizando los problemas de gestión de TI y de costos con Microsoft Intelligence Solution. Se encuentra disponible como parte de la familia de Microsoft office y productos de Microsoft SharePoint Server, y construido sobre la plataforma del poderoso Microsoft SQL Server Plataform. [MICROSOFT 2013]

Pentaho- Kettle: Programa ETL que maneja un conjunto de herramientas para realizar la extracción, transformación y carga. Uno de los principales objetivos es que el proyecto de ETL sea fácil de generar, mantener y desplegar. Es una herramienta Open Source.[PENTAHO 2013]

Algunas ventajas y beneficios del uso de software Open Source son:

- Al construir, integrar y desarrollar componentes Open Source dentro de una sola plataforma el coste de implementar soluciones Business Intelligence se reduce.
- El sistema es fácil y óptimamente escalable, pudiendo extender la misma entre diferentes servidores sin limitaciones del número de usuarios concurrentes, es decir, no existe un número limitado de licencias para su uso.

En la tabla 1 se puede apreciar una tabla con las características relevantes de cada herramienta.

Características de la herramienta	Oracle Enterprise BI server	Business Objects Enterprise XI	Microsoft BI Tools	Pentaho-Kettle
¿Forma parte de una plataforma integrada a inteligencia de negocios?	Si	Si	Si	Si
¿La herramienta de extracción posee una interface gráfica de uso?	Si	Si	Si	Si
¿Soporta diversos tipos de base de datos?	Si	No	Si	Si
¿Requiere una fácil instalación de la herramienta?	No	Si	Si	No

¿La plataforma posee una herramienta de explotación, herramientas de reporting, herramientas de consultas y análisis?	Si	Si	Si	Si
---	----	----	----	----

Tabla 1. Comparación de diferentes Herramientas de BI

Luego de un análisis y comparación entre las diversas herramientas de extracción, se concluye que al cumplir Pentaho con la mayoría de las características, y ser una herramienta de Open Source es la herramienta más apropiada para el desarrollo de éste proyecto.

1.6.1 Plataforma Bi de Pentaho

Compuesta por componentes OpenSource provee la arquitectura y la Infraestructura a la vez. En la figura 1 se puede apreciar que forma un proceso centralizado bajo un marco de trabajo orientado a la solución de problemas de soporte de decisión empleando componentes de BI(Business Intelligence) y permitiendo desarrollos completos para atender a soluciones de Inteligencia de Negocios.[PENTAHO PLATAFORMA DE BI 2013]

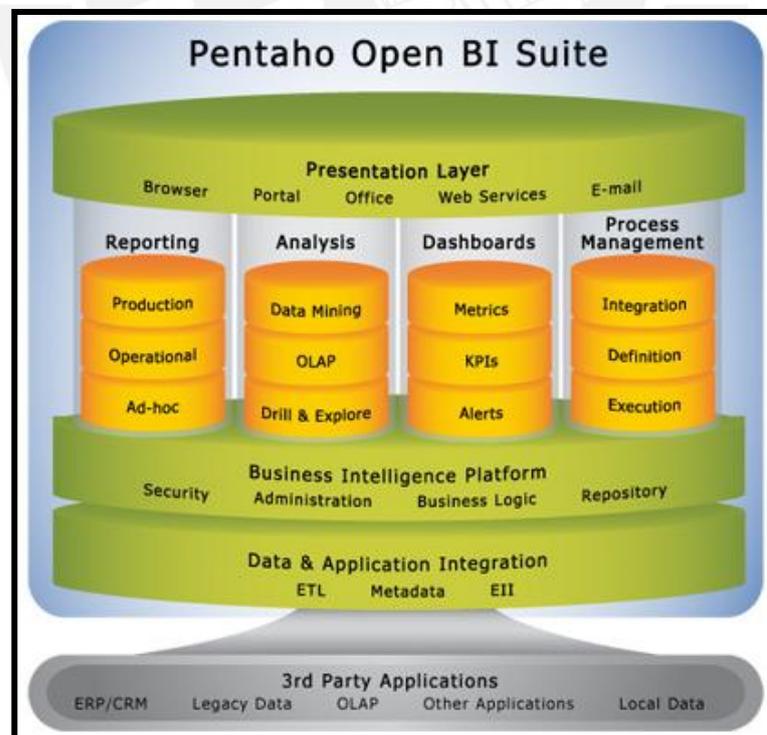


Figura 1. Esquema de la plataforma BI [PENTAHO PLATAFORMA DE BI 2013]

En resumen la plataforma BI integra componentes OpenSource mostrando una combinación de flujos de trabajo y administración de procesos.

Pentaho Data Integration (Kettle)

Es una poderosa herramienta de ETL (Extracción, transformación y carga del Data Warehouse), llamado por sus siglas en ingles “K Extraction Transformation Transportation Load E”. Utiliza un entorno gráfico de desarrollo, y es multiplataforma. Principalmente se basa en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y trabajos o “Jobs” (colección de transformaciones). [PENTAHO DATA INTEGRATION 2013]

Incluye 4 herramientas:

- SPOON: Se utiliza para diseñar transformaciones empleando el entorno gráfico.
- PAN: Se utiliza para ejecutar transformaciones empleando el entorno gráfico.
- CHEF: Se utiliza para crear trabajos
- KITCHEN: Se utiliza para ejecutar “Jobs” planificados en modo “Batch” (serie, lote) para correr automáticamente en intervalos de tiempo definido.

Pentaho Analysis services

Esta herramienta forma parte del motor OLAP integrado en el suit BI de Pentaho. Las principales características de esta herramienta es que permite un modelado multidimensional de datos, para esto utiliza MONDRIAN como servidor OLAP del tipo ROLAP, el modelo de cubos se guarda en archivos XML y el lenguaje de consulta es MDX.[PENTAHO ANALYSIS SERVICES 2013]

En el caso de la navegación y la exploración de datos utiliza JPivot, el cual permite el análisis ad-hoc de los cubos de MONDRIAN, este componente como otros puede ser reemplazado por otras soluciones que integran Pentaho. [PENTAHO ANALYSIS SERVICES 2013]

Pentaho Reporting

Consiste en un motor de presentación, con la capacidad de generar reportes de manera flexible, esta herramienta permite elaborar desde sencillos reportes iniciales

hasta complejos reportes ajustados a la necesidad del negocio. [PENTAHO REPORTING 2013]

Pentaho Cube Designer

Con esta aplicación se llegan a crear los cubos, cube designer es el fichero de esquema del cubo MONDRIAN, en él se definen dimensiones, niveles de jerarquía de las dimensiones, hechos, y conexión a la base de datos transaccional, los cuales sirven de dato para diseñar el cubo OLAP. La aplicación Cube Designer es un buen apoyo para la creación de cubos. Se recomienda usar esta aplicación para modelos de complejidad media.

1.7 Alcance y limitaciones del proyecto

1.7.1 Alcance

1.7.1.1 Alcance del proyecto:

Con respecto al alcance del proyecto solo se llegó a realizar el análisis, diseño e implementación. No se llegó a realizar la implantación de la solución de Inteligencia de Negocios. Se verificó el funcionamiento adecuado de la solución, utilizando data histórica.

1.7.1.2 Alcance del producto:

En el presente proyecto de tesis se cubrieron principalmente 3 gerencias, la Gerencia de Operaciones Registrales (GOR), Gerencia de Registro e Identificación (GRI), Gerencia de Restitución de la Identidad y Apoyo Social (GRIAS).

Se han elegido estas Gerencias debido a la importancia que tienen dentro de la institución, y también debido a la facilidad de acceso a su información y a sus requerimientos.

El presente proyecto de tesis se centra en los procesos de generación y emisión del DNI. En el área de Producción se tomó en cuenta cada etapa por la cual pasa un trámite, desde el registro en el sistema del trámite, luego el ingreso a impresiones, después el despacho desde el local de impresiones al local de entrega, y finalmente la recepción de los trámites en el local. En el área de Calidad se tomó en cuenta los trámites que fueron observados

El presente proyecto de tesis se centra en 3 gerencias, el modelo ER (Entidad-Relación) que se utilizó es limitado solo a la información necesaria por cada Gerencia, no se tomó todo el modelo ER de RENIEC. Debido a esto se mantuvo un contacto frecuente con el RENIEC para que el modelo utilizado vaya de acuerdo al que utilizan en RENIEC.

1.7.2 Limitaciones

En el presente proyecto de tesis, sólo se tomó en cuenta el DNI convencional, no abarcó el DNI electrónico. A pesar de que los procesos básicamente son los mismos desde la captura hasta la aprobación y solo cambia la impresión en la parte operativa, no se tomó en cuenta debido a que la información oficial de los procesos de RENIEC es sobre el DNI convencional, y no se encuentran formalizados los procesos del DNI Electrónico porque todavía se encuentra en proceso de implementación.

1.8 Viabilidad y justificativa del proyecto

Con el uso de la Solución de Inteligencia de Negocios se podrá brindar indicadores de gestión necesarios para las gerencias. Estos indicadores se muestran en esquemas de análisis elaborados en el Software Pentaho, lo que permite elaborar dinámicamente diferentes consultas para lograr un oportuno control de los procesos, lo que conlleva al mejoramiento del servicio que brinda al ciudadano. Usando las herramientas de Report Designer se elaboran reportes fácilmente lo que permite ahorrar recursos humanos que pueden estar involucrados en la elaboración de reportes y gráficas para la gerencia. Debido a que el Data Mart posee una estructura especial para el procesamiento masivo de datos, permite ahorrar tiempo en el proceso de la información. Finalmente la solución de Inteligencia de Negocios será de fácil uso para los empleados del RENIEC.

La entidad del estado RENIEC, tiene en sus planes realizar un Data Warehouse de todas las dependencias orgánicas, vinculando también las áreas financieras. Es un proyecto de gran magnitud, ya que piensa mostrar los indicadores de cada Gerencia, e integrarlos para obtener información esencial. Esta iniciativa muestra la importancia que tiene la inteligencia de negocios en las empresas y cómo su uso es de mucha ayuda para el mejoramiento en la atención al público. Al finalizar el proyecto de tesis ésta puede servir de mucha ayuda para el futuro proyecto de un Data Warehouse para RENIEC.

2 Marco Conceptual y Estado del Arte

2.1 Marco Conceptual

2.1.1 *Conceptos relacionados al problema*

Los procesos relacionados con el DNI son del interés de 3 gerencias: la Gerencia de Operaciones Registrales (GOR), Gerencia de Registro e Identificación(GRI), Gerencia de Restitución de la Identidad y Apoyo Social(GRIAS). Esta organización se puede apreciar en la figura 2.

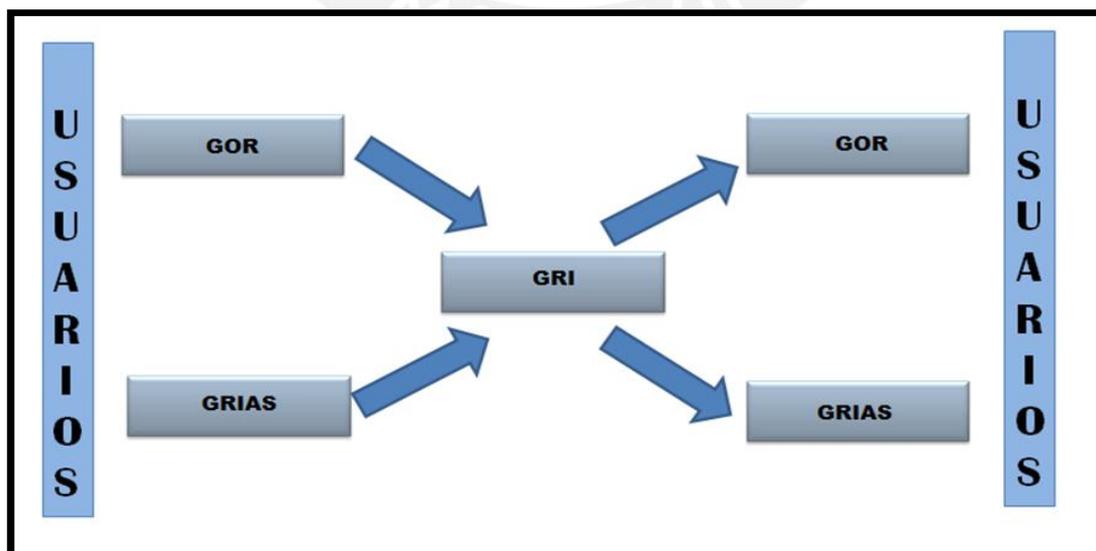


Figura 2. Esquema de las gerencias de GOR, GRI, GRIAS de RENIEC [RENIEC02 2012]

2.1.1.1 Gerencia de Operaciones Registrales (GOR)

Es la encargada de planear, organizar, ejecutar, supervisar y controlar las acciones relacionadas con el proceso de identificación de las personas, así como las inscripciones de los hechos vitales y modificatorios del estado civil. Asimismo, coordina la actualización del Padrón Electoral y verificación de firmas de adherentes y supervisa el desempeño de las Jefaturas Regionales.

A continuación se describirá el proceso de identificación, el cual está relacionado con el Registro de Trámite y entrega del DNI según el documento RENIEC en su “Guía de Procedimientos. Registro del trámite y entrega del documento nacional de identidad”.

El flujo completo del proceso de identificación se puede apreciar en la figura 3 y en el Anexo 1 se puede apreciar en detalle el proceso de identificación.

2.1.1.1.1 Proceso de identificación

El ciudadano solicita el trámite del DNI, el registrador Captura los datos del ciudadano y procede a verificar los documentos según el tipo de trámite, señalados por el TUPA (Texto Único de Procedimientos Administrativos), si no es conforme devuelve los documentos al ciudadano y orienta debidamente, si se encuentra conforme selecciona el tipo de trámite a realizar y registra el tipo de trámite. En caso sea una inscripción se le pregunta al ciudadano si la partida de nacimiento es incorporada, si su respuesta es afirmativa se rectifica en el sistema, se registran nuevos datos rectificados y se actualiza la foto, o en caso contrario se registran los datos según partida de nacimiento y se registra la información de los documentos de sustento (Recibos, Actas). En caso no sea una inscripción se registra el número de DNI en el sistema, si es duplicado se registra la información de los documentos de sustento (recibo, actas), en caso contrario se registran nuevos datos rectificados y/o se actualiza la foto.

El registrador verifica los datos en pantalla con la aprobación del ciudadano en señal de conformidad, se imprime la ficha registral, se procede a registrar los datos en la ficha registral. Primero se estampan las huellas del ciudadano en el formulario, luego se solicita la firma del documento en señal de conformidad, después se verifican las impresiones dactilares y firma en la ficha registral del documento; si es una inscripción o renovación del DNI se pega la foto en la ficha registral y se consigna el nombre, DNI y ficha del registrador, en caso contrario solo se consigna el nombre, DNI y ficha del registrador. Finalmente se desglosa el ticket, se le entrega al ciudadano y se plasman

las impresiones dactilares en el desglosable.

Se entregan las fichas al supervisor, quien se encarga de realizar el control de calidad de los trámites, si no es conforme se procede a observar el trámite, en caso contrario se genera el envío de sobres. El envío de sobres consiste en pistolear la ficha registral según el tipo de trámite, generar cargo de lotización, ensobrar las fichas registrales conforme el reporte de cargo de lotización generado, adosar la cara externa sobre una copia de cargo de lotización, agrupar sobres y generar conforme la hoja de ruta. Y finalmente se depositan en una valija para el recojo del Courier.

En el área de trámite documentario (GOR) se recogen las valijas para ser transportadas al área de procesamiento. Una vez en la gerencia de Registro de Identificación (GRI) se reciben envíos de fichas registrales en trámites documentarios, se reciben sobres y fichas registrales, y se evalúan si se trata de trámites manuales o semiautomáticos. Si es semiautomático se reciben los sobres, en caso contrario se ingresan los datos del trámite al sistema y se procede a lotizar. Si se actualiza la imagen se digitaliza el formulario para la (INDRA), en caso contrario se asignan sobres y fichas registrales a evaluadores. Si las evaluaciones de fichas registrales están conformes se aprueba el trámite, en caso contrario se desaprueba el trámite y se envía la ficha a archivo registral.

En la gerencia de informática se verifica la huella digital (AFIS) si es conforme se actualiza el RUIPN y se generan los lotes de impresión, en caso contrario se evalúa con el perito dactiloscópico, si es conforme se actualiza el RUIPN y se generan los lotes de impresión, en caso contrario se verifica si es una suplantación; si es una suplantación se habilita, se cancela y se le notifica al ciudadano, en caso no sea una suplantación se desaprueba el trámite y se ficha el archivo registral.

Una vez que se generaron los lotes de impresión, se imprimen los lotes de DNI (ENOTRIA) en la Gerencia de Registro de Identificación, y se envían al despacho de DNI. En el Área de Despacho se reciben los lotes, se ordenan y se distribuyen los lotes por DNI, se generan los envíos de DNI. En el área de Trámite documentario se recogen las valijas de trámites de DNI, se transportan a áreas de supervisión. En el área de supervisión se recibe la valija del Courier y se extrae el paquete del DNI. Se procede a validar el despacho y los lotes del DNI, primero se verifica la cantidad de lotes de DNI con la registrada en la guía de despacho, si no coincide se comunica con la oficina de despacho de DNI para regularizar la omisión y se sella la guía de

despacho en señal de conformidad, en caso coincida solo se sella la guía de despacho en señal de conformidad. Y finalmente se entrega el DNI al encargado del área de entregas.

En el área de Entregas se pistolea el DNI y se verifica con relación al lote de impresión, si no coincide se registra en cargo de entrega los DNI faltantes, se comunica a la oficina de despacho DNI para regularizar la omisión, se entrega el cargo al supervisor. En caso coincida se recibe el DNI en aplicativo de entrega, se entrega cargo al supervisor. Finalmente se intercala el DNI por género y por apellido.

El ciudadano solicita la entrega del DNI, si no tiene ticket firma la declaración jurada de pérdida de ticket y se la entrega al registrador, en caso tenga ticket se le pregunta si cuenta con el DNI anterior, si su respuesta es afirmativa se le entrega al registrador, en caso contrario firma una declaración jurada de pérdida del DNI anterior. Se procede a buscar el DNI en agencia, si no se encuentra se coloca como trámite observado, en caso se encuentre se ubica el DNI, se ingresa al aplicativo de entrega, si es un DNI tramitado por la web se valida la huella del titular y se imprime el formato web, en caso contrario se valida el recibo y se genera la entrega del DNI al ciudadano.

2.1.1.1.2 Estado de los trámites

Trámites Observados

En caso existan trámites observados, se identifica y evalúa el motivo de la observación y se determina si existe responsabilidad del registrador, y se registra en el aplicativo o control manual.

En caso que el trámite se considere que no amerita ninguna observación, se efectúa la devolución inmediata al área correspondiente. En caso el ciudadano haya declarado su número telefónico o correo se le contacta, y se le comunica la observación de su trámite para su subsanación. Se procede a clasificar y archivar las fichas registrales por orden alfabético, según apellido del titular.

Si el ciudadano al momento de acercarse a subsanar el trámite, no cuenta con los documentos requeridos, se notifica el motivo de la observación en el formato de comunicación de observación de trámite.

Trámites Recuperados

Si el ciudadano se apersona a la Oficina registral, agencia o punto de atención y presenta la documentación requerida, se procede a recuperar su trámite. Se recupera el trámite en una nueva ficha registral, en los siguientes casos:

- Cuando se modifique los campos y/o requisitos.
- Cuando se requiera nueva captura de imágenes (huella y firma).
- Cuando la ficha registral se encuentre deteriorada o manchada.
- Otros que se consideren necesarios.

De recuperar el trámite en una ficha registral manual, se coloca en el campo de observaciones el número de ficha registral y fecha del trámite anterior.



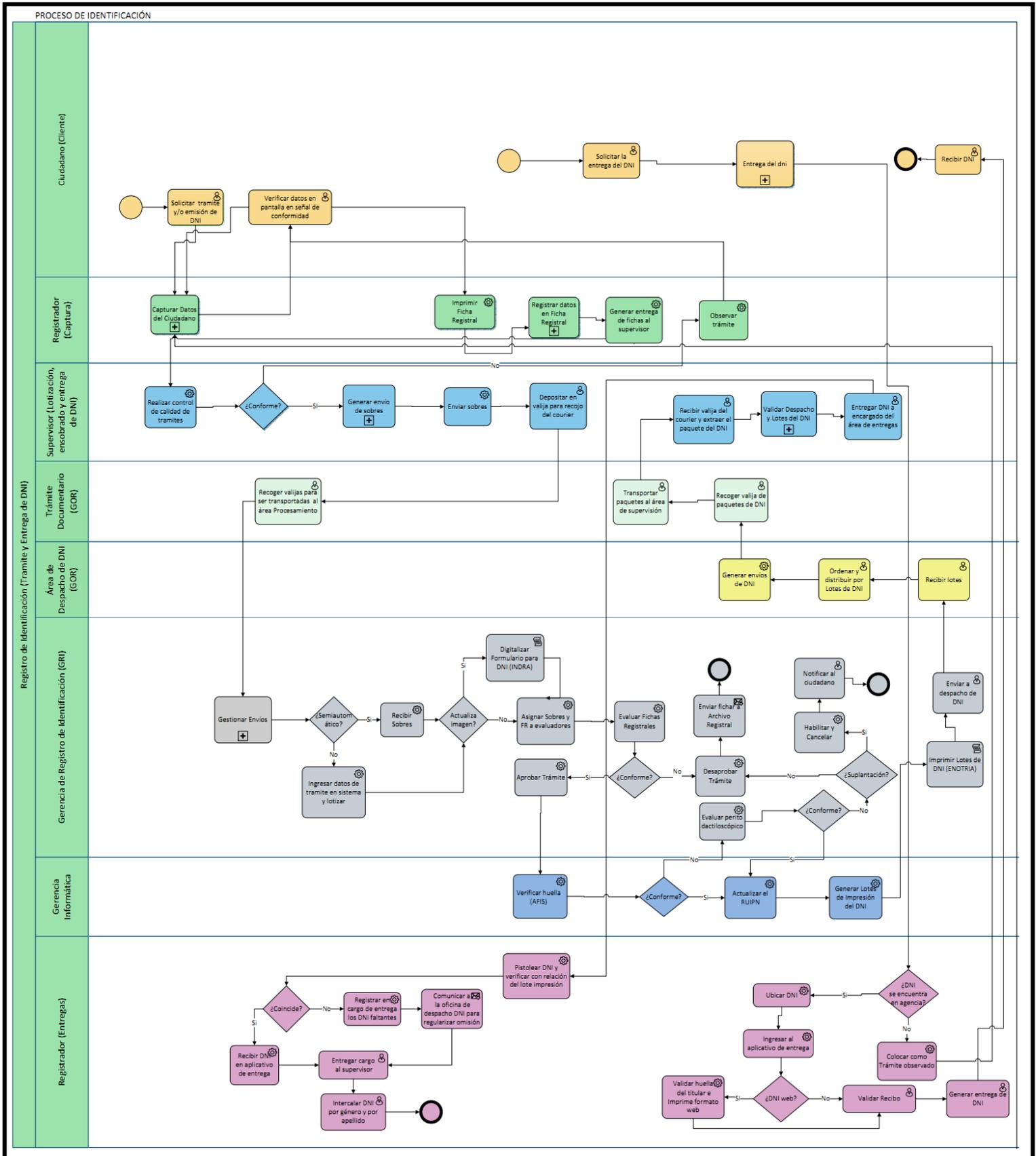


Figura 3. Diagrama de flujo registro de trámite y entrega del DNI [RENIEC07 2012]

2.1.1.2 Gerencia de Registro e Identificación (GRI)

La Gerencia de Registros de Identificación es el órgano de la línea encargado de conducir los procesos, modificación del estado civil, evaluación, depuración y actualización del Registro Único de Identificación de las Personas Naturales, así como el mantenimiento del archivo registral de la institución referente al Registro Único de Identificación de las Personas Naturales y de los Registros Civiles.

El esquema general se puede apreciar en la figura 4, así como el diagrama de flujo detallado en la figura 5 y el diagrama de flujo de la línea de procesamiento del DNI.

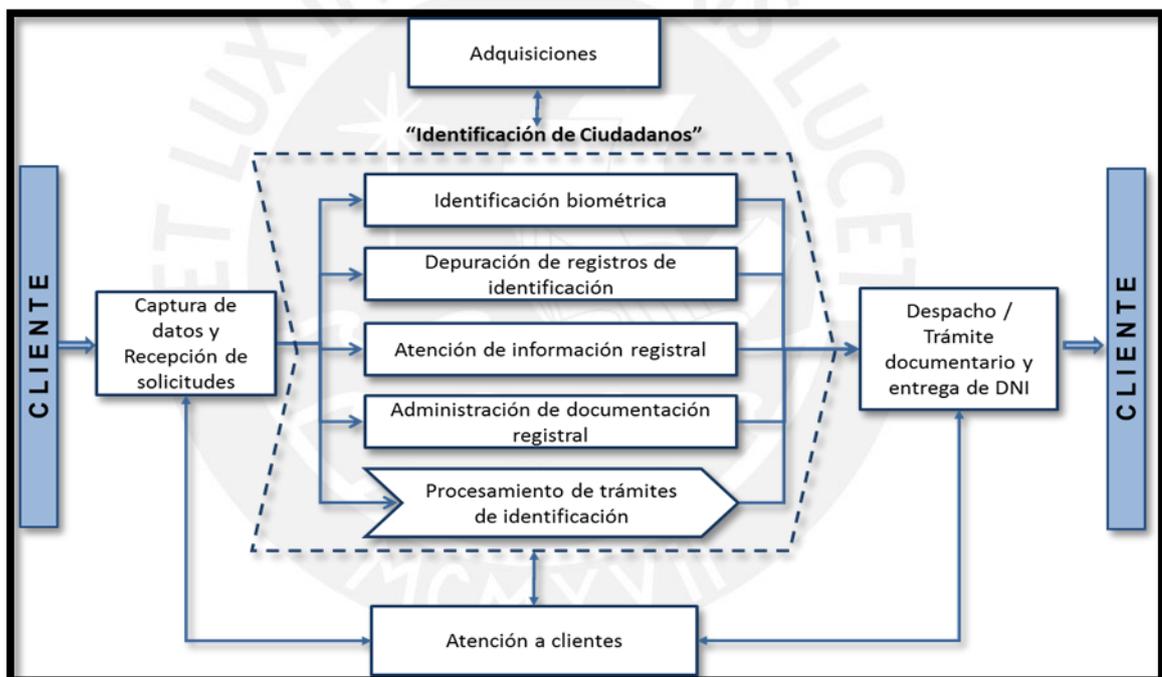


Figura 4. Esquema General del Proceso en la GRI [RENEC03 2012]

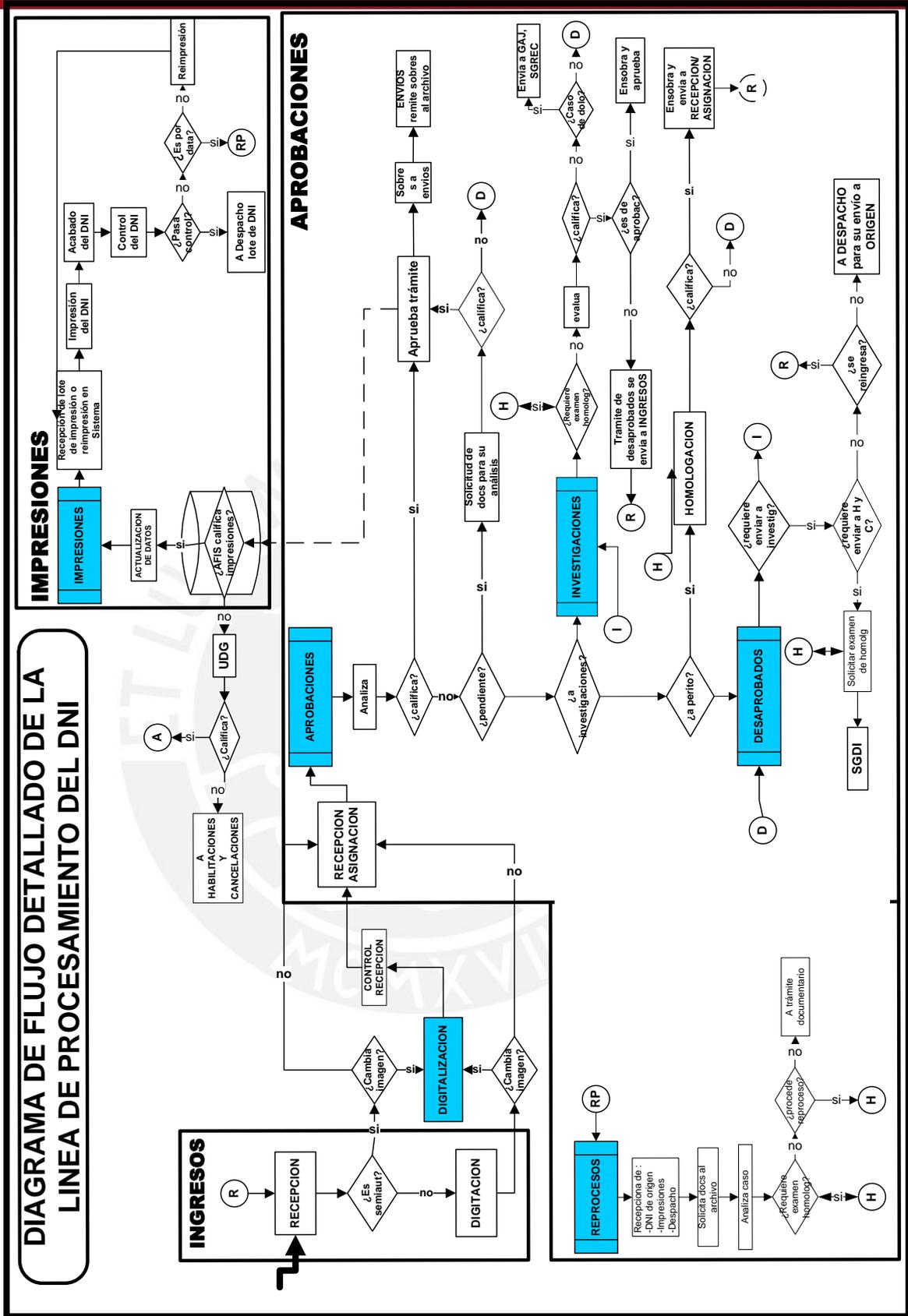


Figura 5. Diagrama de flujo detallado de la línea de procesamiento del DNI [RENIEC05 2012]

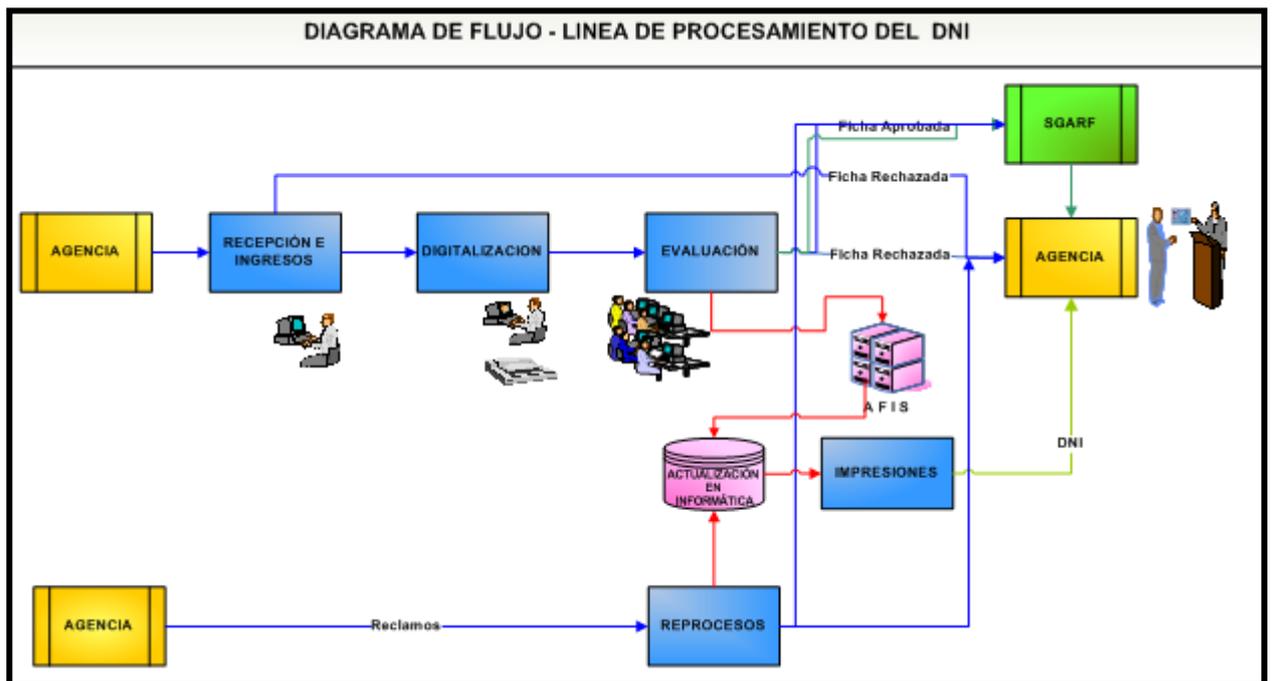


Figura 6. Diagrama de flujo -línea de procesamiento del DNI [RENIEC06 2012]

2.1.1.3 Gerencia de Restitución de la Identidad y Apoyo Social (GRIAS)

Encargada de proponer los lineamientos y políticas institucionales en materia de apoyo social a los grupos poblacionales menos favorecidos, considerando los enfoques de derecho, género e interculturalidad; con especial énfasis en la población indocumentada. Tiene a su cargo la formulación y ejecución del Plan Nacional de Restitución de la Identidad y Apoyo Social. Se puede apreciar el proceso de GRIAS en la figura 7 y 8.

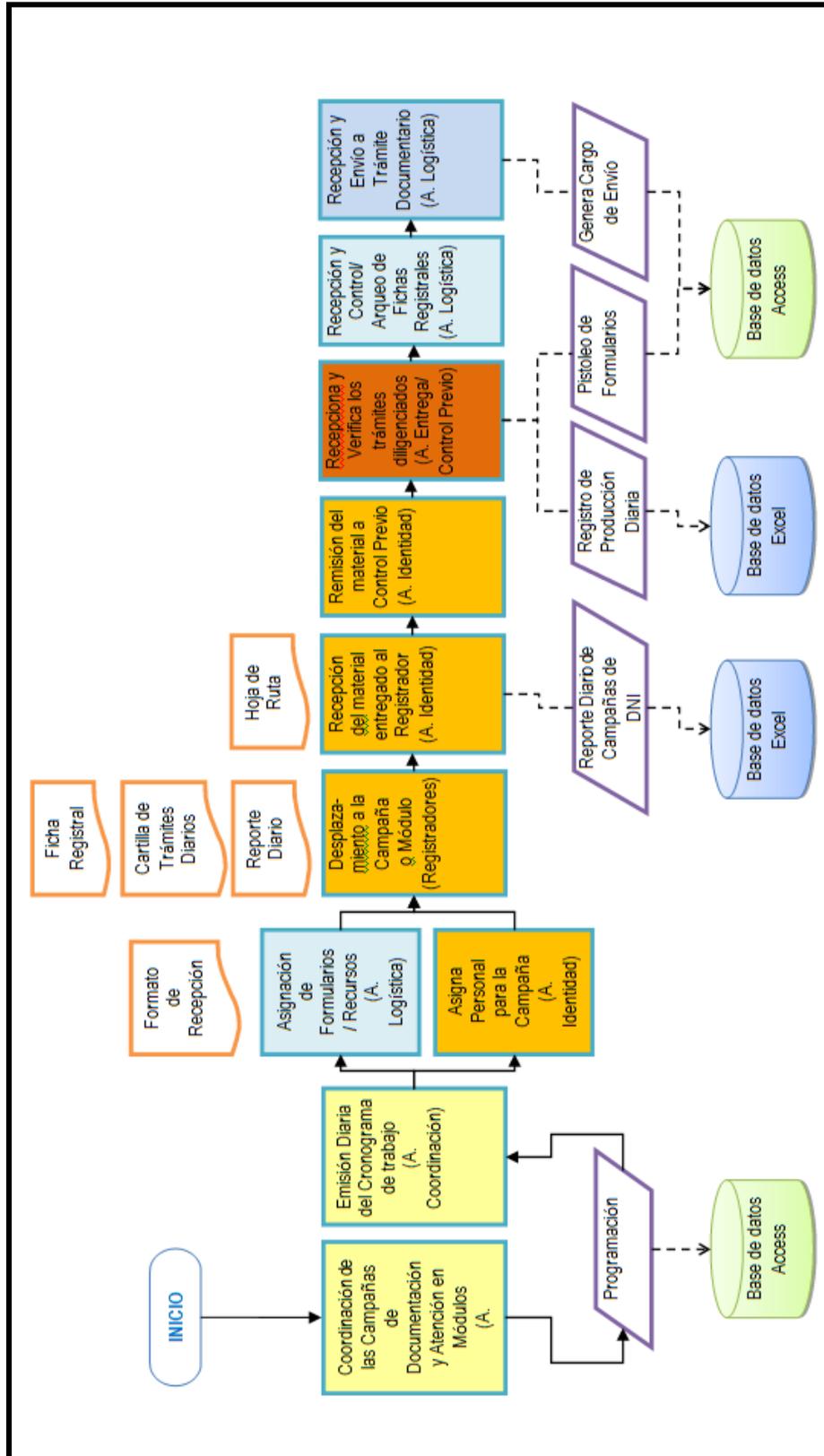


Figura 7. Procesos de GRIAS [RENIEC04 2012]

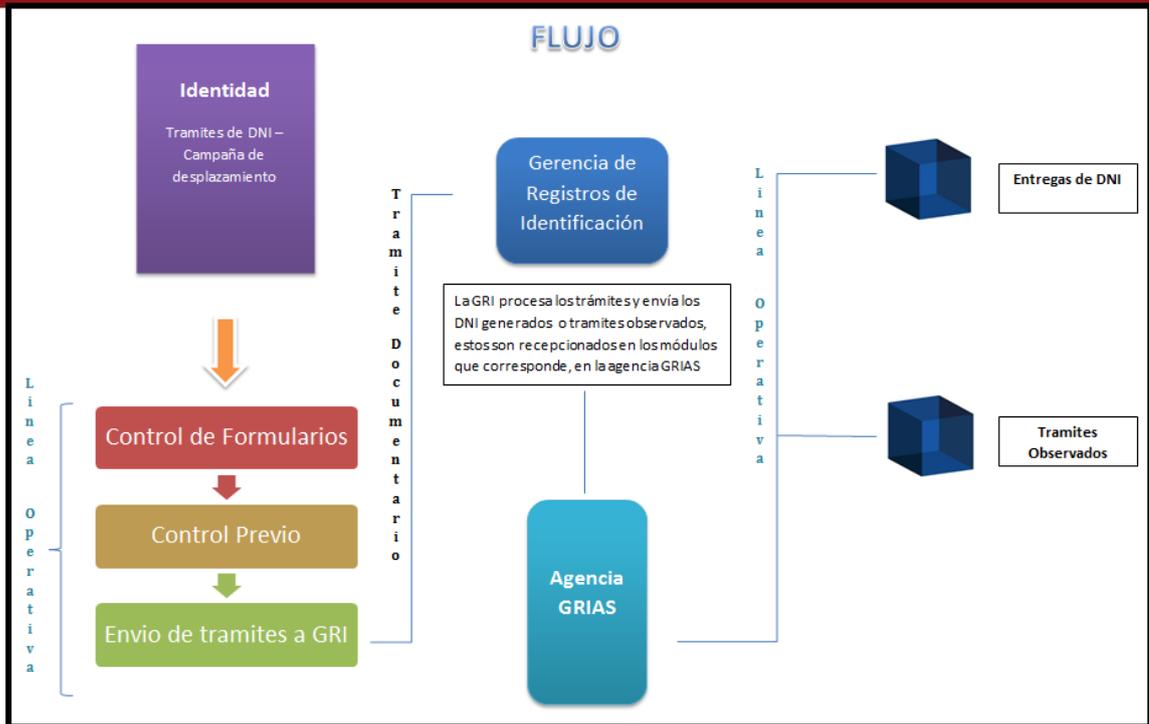


Figura 8. Procesos de GRIAS [RENIEC04 2012]

2.1.1.4 Datos necesarios por gerencia

Los datos que más necesitan las gerencias según los documentos de requerimientos de las gerencias de Operaciones Registrales, de Registro e Identificación, de Restitución de la Identidad y Apoyo Social se pueden apreciar en la tabla 2. En esta tabla también se incluyen los requerimientos tomados en una entrevista con personal de RENIEC, el detalle de esta entrevista se encuentra en el Anexo 8.

GOR (Gerencia de Operaciones Registrales)	
Trámites observados	
Eficacia del tiempo de recepción de DNI en agencia	
Trámites de DNI	
Trámites de DNI observados en agencia (mejorar el servicio de trámite realizando un registro más efectivo)	
GRI (Gerencia de Registro e Identificación)	
Inscripciones ordinarias	
Rectificaciones Administrativas de estado civil	
Apelaciones y Reconsideraciones en el SIHABCAN	
GRIAS (Gerencia de Restitución de la Identidad y Apoyo Social)	
Con los datos que pide esta gerencia, se calcula el tiempo en que se demora en entregar un DNI, la cantidad de reimpresiones, entre otras cosas.	UBIGEO RENIEC
Fecha del trámite	UBIGEO RENIEC residencia
Fecha de ingreso al Sistema	Código de local
Fecha de ingreso a GRI	Centro poblado
	Trámite observado
	Fecha de nacimiento
	Jefatura regionales

Fecha de evaluación del trámite Fecha de impresión del trámite Fecha de emisión (entrega a despacho) Fecha de recepción en agencia Numero de Ficha Tipo de captura Cantidad de Recibos	Jefatura regionales responsable Población objetivo Tipo de captura
--	--

Tabla 2. Datos necesarios por gerencia.

Ésta información es la principal fuente que se usaría en el Data Mart, y con las cuales se prepararía información crítica para la gerencia.

2.1.2 Conceptos relacionados a la propuesta de solución

2.1.2.1 Teoría de Inteligencia de Negocios

La Inteligencia de Negocios es un enfoque y un conjunto de herramientas para la gestión empresarial que permiten a una organización determinar qué información es relevante y útil para la toma de decisiones corporativas. La Inteligencia empresarial se encarga de realizar la gestión de los datos, consultar estos y generar informes/reportes, realizar todo el procesamiento para convertir la lógica del negocio a la lógica de sistemas empresariales. [KIMBALL 2002, 393]

Los beneficios de un sistema de Inteligencia de Negocios abarcan la reducción de costos, aumento de beneficios, mejora en la satisfacción del cliente y mejoras en la Comunicación empresarial.

2.1.2.2 Herramientas de Inteligencia de Negocios

Uno de los principales mecanismos de descubrimiento del conocimiento es el análisis eficiente de los datos usando sistemas modernos de análisis. [RUSSEL 2010]

Con herramientas de Inteligencia de Negocios, las compañías pueden darles a sus Usuarios autorizados acceso a un único punto de información que acceda a todos los sistemas de la empresa: intranet, extranet, o E-Business, Data Warehouse, sistemas ERP's, Data Mart o aplicaciones a la medida, de una manera fácil sin necesidad de conocimiento técnico y lo más gráfico de entender posible. [KIMBALL 2002]

2.1.2.2.1 Data Warehouse

Es un conjunto finito de componentes y procesos que forman parte de una arquitectura diseñada para guardar la información y cuyo fin es satisfacer las necesidades de análisis y consultas que realizan los actores. [KIMBALL 2002, 397]

El Data Warehouse es un conjunto de datos que tienen relación y están orientados sobre una materia en especial, tienen cambios durante el paso del tiempo y son fundamentales en el proceso de toma de decisiones para la parte administrativa. Se puede apreciar el entorno básico en la figura 9 y una visión general del Data Warehouse en la figura 10.

El propósito que tiene es darle las herramientas adecuadas a los responsables ejecutivos de la empresa para la comprensión de las realidades del pasado y adoptar elementos para la planificación y toma de decisiones de futuro a corto, medio y largo plazo. [HARJINDER & PRAKASH 1996]

El Data Warehouse ofrece algunas de las siguientes funcionalidades:

- Integración de bases de datos del mismo tipo de orígenes diferentes.
- Ejecución de consultas complejas, con diferentes niveles de agrupamiento y ofreciendo la visualización de la información de la manera más adecuada.
- Agrupación y análisis de los datos de forma interactiva.
- Control de calidad de los datos para asegurar consistencia, homogeneidad y relevancia.

Una de las principales ventajas para quien lo aplica es que proporciona una herramienta para la toma de decisiones en cualquier área funcional, teniendo como base información integrada y global del negocio. De esta manera los procesos empresariales pueden llegar a ser optimizados al poder analizar cuáles son sus fallas puntuales.

Presenta algunas desventajas por ejemplo cuando el Data Mart crece en tamaño se vuelve más complejo, lo que ocasiona un bajo rendimiento de las consultas y el modelo se hace cada vez menos ineficiente.

Muchas veces al tener una administración central puede ocasionar que los requerimientos de información de unidades se vean afectados entre sí.

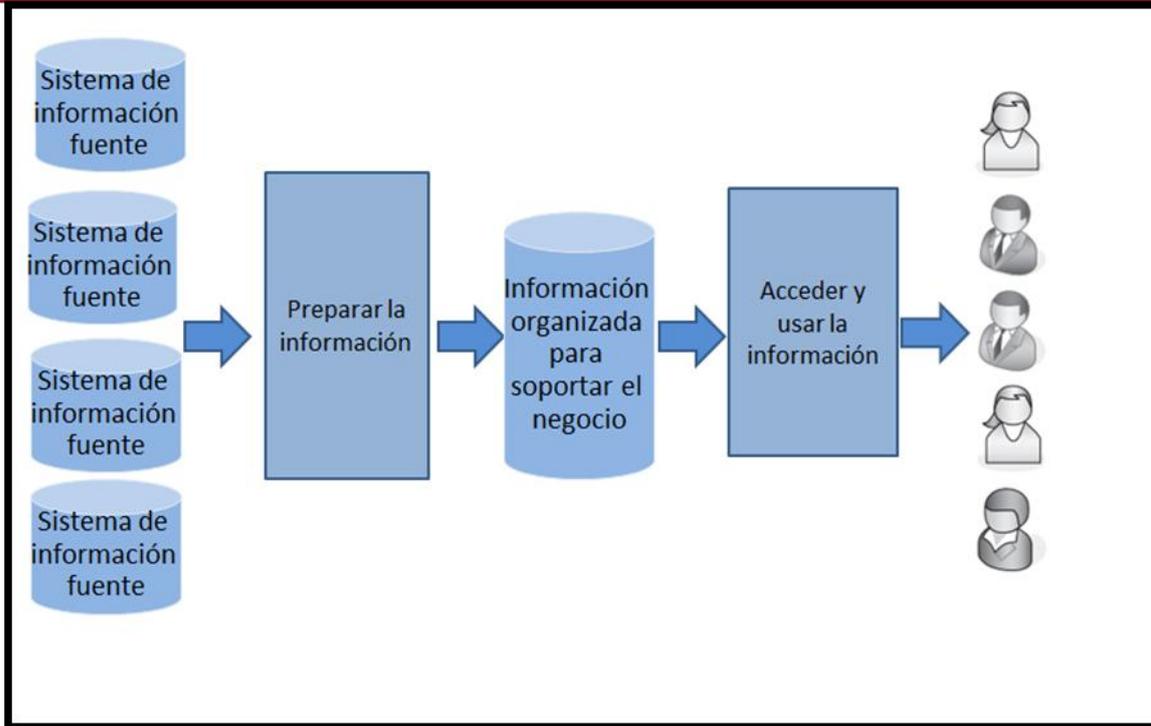


Figura 9. Entorno básico del Data Warehouse [REEVES 2009]

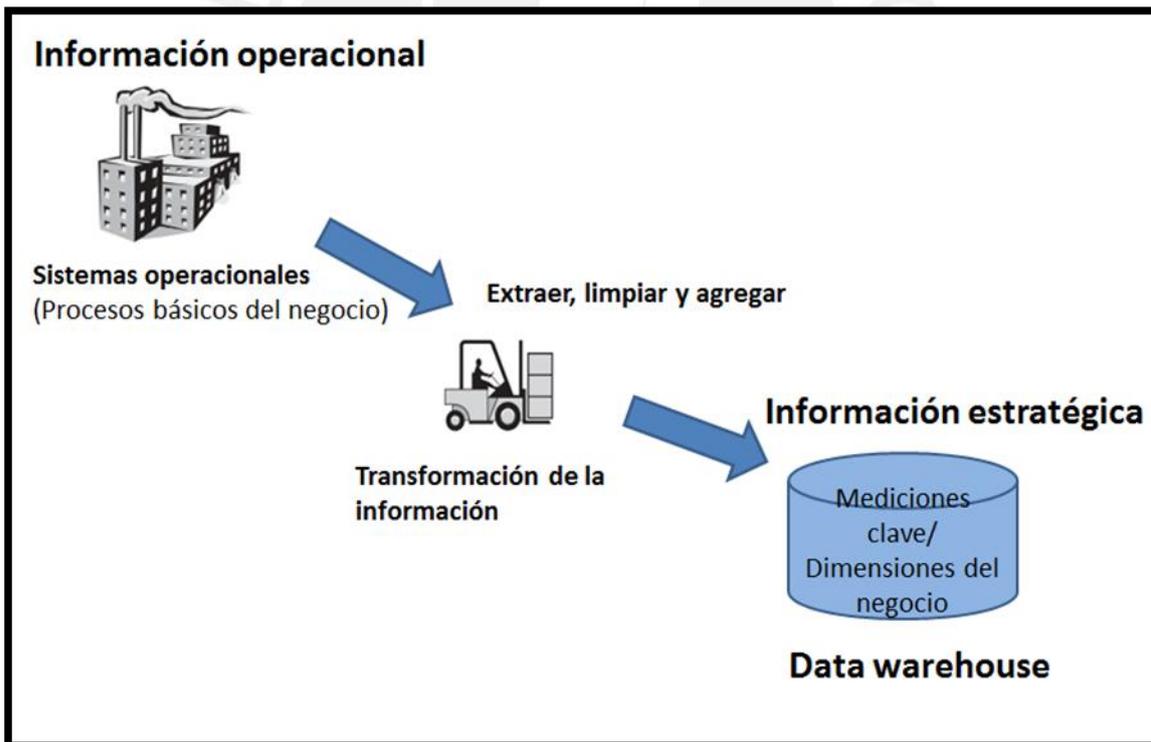


Figura 10. Vista general de la información de un Data Warehouse [PONNIAH 2010]

2.1.2.2.2 Data Mart

Es un subconjunto lógico y físico del área de presentación de un Data Warehouse. Originalmente, los Data Mart's eran definidos como un subconjunto altamente organizado de información, usualmente escogido para responder cuestiones específicas en el negocio. Esta definición no era exacta debido a que afirmaba que los Data Mart's eran inflexibles y no podían combinarse unos a otros. La primera definición ha sido remplazada y ahora el Data Mart es definido como un juego flexible de datos, basado en la forma de dato más atómica posible para que se extraiga de una fuente operacional, y se presenta en un modelo simétrico que es más resistente cuando se tienen consultas inesperadas de los usuarios. Decimos que los Data Mart's están conectados al bus del Data Warehouse. Y en su definición más simple un Data Mart representa información de un solo proceso del negocio, en cambio el Data Warehouse se encarga de tener como alcance a toda la organización. [KIMBALL 2002, 396] [INEIN]

Modelamiento dimensional.

Después de hablar de BI y algunas de sus estructuras es pertinente mencionar el Modelamiento dimensional.

Es una técnica de diseño lógico utilizada para soluciones de Inteligencia de Negocios, que tienen como propósito presentar los datos en una arquitectura estándar y de esta manera los usuarios finales puedan tener un alto rendimiento al momento de acceder a la información. El modelamiento se basa en esquemas estrella, conformados por tablas hechos y tablas dimensionales. [KIMBALL 2002, 399]

Esquema Estrella

Se caracteriza por tener como único elemento central a una tabla llamada Tabla de hechos (Fact Table), que está conectada radialmente a varias tablas dimensiones. Este modelo es fácil de entender y responde a las consultas de los usuarios optimizando y reduciendo el número físico de uniones que se necesitan entre la tabla hechos y la tabla dimensiones. [KIMBALL 2002, 13]. Se puede apreciar un ejemplo del modelo estrella en la figura 11.

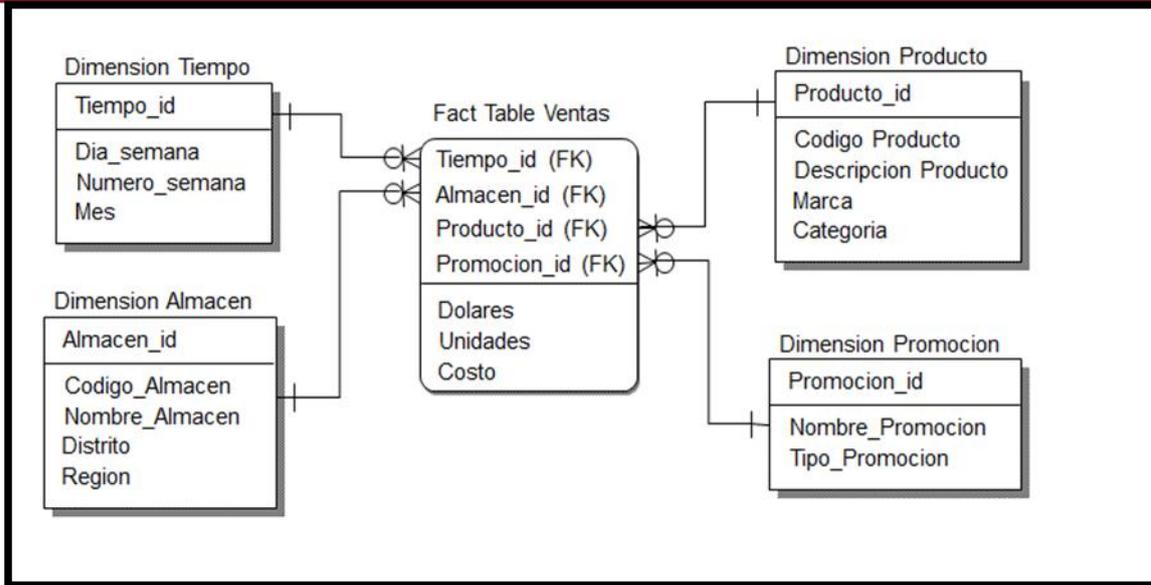


Figura 11. Esquema Estrella [PONNIAH 2010]

Fact Table (Tablas Hecho)

Las Tablas de Hechos forman parte del Data Mart y son las tablas principales dentro del esquema en estrella que representa datos numéricos en el contexto de las entidades del negocio. Dicha tabla se encuentra constituida por medidas o métricas, que son datos cuantitativos acerca de los hechos de un tema del negocio y por Foreign Keys. Las tablas de hechos están relacionadas con cada una de sus tablas dimensiones. [KIMBALL 2002, 16-18] [MOSS, ATRE 2003]

Tablas dimensiones

Son aquellas tablas que se relacionan con las tablas de hechos en un modelo dimensional a través de su llave primaria. Las dimensiones son objetos del negocio con los cuales se puede analizar la tendencia y su comportamiento. Las definiciones de las dimensiones se basan en políticas propias de la compañía e indican la manera en que la organización interpreta su información para segmentar el análisis facilitando la observación de los datos. [KIMBALL 2002, 19-21] [BIR]

Esquema copo de nieve

Es una extensión del esquema estrella donde cada punta de la estrella se llega a expandir en más puntas y su nombre se debe a que el diagrama se asemeja a un copo de nieve. Este tipo de esquema normaliza dimensiones para eliminar la redundancia y

los datos de las tablas dimensiones se agrupan en varias tablas en lugar de una tabla grande. [KIMBALL 2002, 55]. Se puede apreciar un ejemplo del modelo copo de nieve en la figura 12.

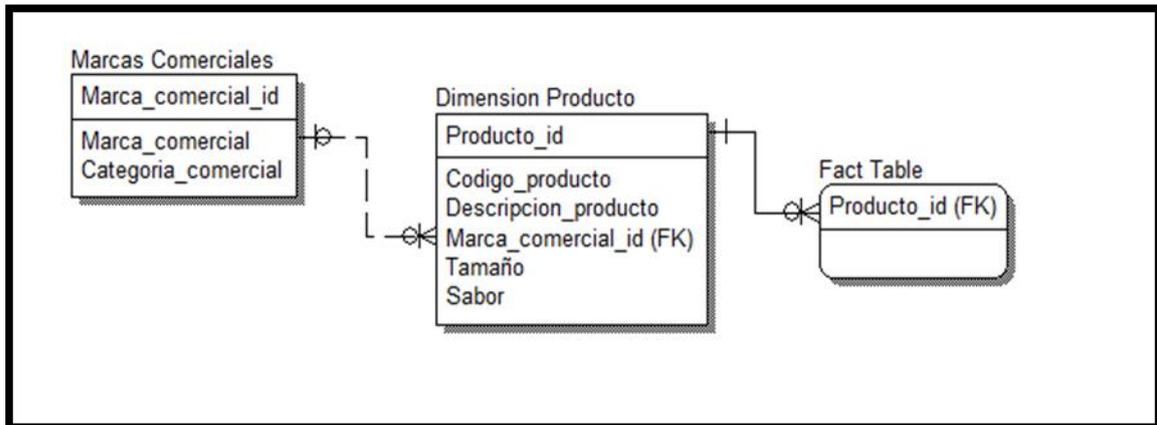


Figura 12. Esquema Copo de Nieve [PONNIAH 2010]

Ahora vamos a mencionar algunos aspectos de las técnicas de diseño y las estructuras con las que trabaja el Data Mart.

Sistemas fuentes

Una parte importante del Data Mart son los sistemas fuentes, estos son sistemas transaccionales que han sido diseñados para el soporte de las operaciones del negocio. La base de datos de estos sistemas se llama base de datos operacionales (On Line Transactional Process: OLTP) y de ésta se extraerán los datos que permitirán generar información para la toma de decisiones a nivel operacional. Estas bases de datos persiguen fundamentalmente el registro de transacciones y la consistencia de los datos. [MOSS, ATRE 2003]

Tecnología OLAP (On Line Analytical Process)

Esta tecnología permite aprovechar la forma en cómo está estructurada la información de un Data Mart o un Data Warehouse. Fundamentalmente permitirá analizar información dinámicamente a los niveles táctico y estratégico basados en Cubos que contienen las medidas y las Dimensiones. Algunas de las técnicas más conocidas son: MOLAP, ROLAP, HOLAP. [MOSS, ATRE 2003]

La técnica MOLAP es el camino más tradicional del análisis OLAP. En MOLAP, la información no es guardada en una base de datos relacional, en cambio se guarda en un cubo multidimensional.

La técnica ROLAP trabaja directamente con la base de datos relacional, ésta es guardada como tablas relacionales y nuevas tablas son creadas para sostener la información agregada.

La técnica HOLAP intenta combinar las ventajas de MOLAP y ROLAP. En este caso la base de datos se dividirá en relacional para sostener las grandes cantidades de data detallada y un almacenamiento especial para menores cantidades de data menos detallada.

Herramientas de extracción, transformación y carga (Procesos ETL)

Una vez que se tienen las estructuras bien definidas se usan los procesos ETL (Extracción, Transformación y Carga), estos procesos son los encargados de alimentar el Data Warehouse o Data Mart de datos. Estos tres pasos se realizan con el objetivo de mover datos desde diversas fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, Data Mart o Data Warehouse para el análisis que le corresponde. [ETLBIAD]

El primer paso es la extracción y tiene como meta obtener los datos desde las fuentes de información. Para eso se leen los datos, se entienden los datos fuentes y se copian los necesarios para el Data Warehouse en el Staging área (área de ETL) para su manipulación posterior.

El segundo paso es la transformación, una vez que la data es extraída del Staging Area, se realizan una serie de reglas del negocio o funciones sobre los datos extraídos para convertirlos en datos que posteriormente serán cargados.

Finalmente después de haber depurado y transformado los datos, estos van a ser cargados al área de presentación del Data Warehouse. Se puede apreciar un esquema del proceso de carga en la figura 13.

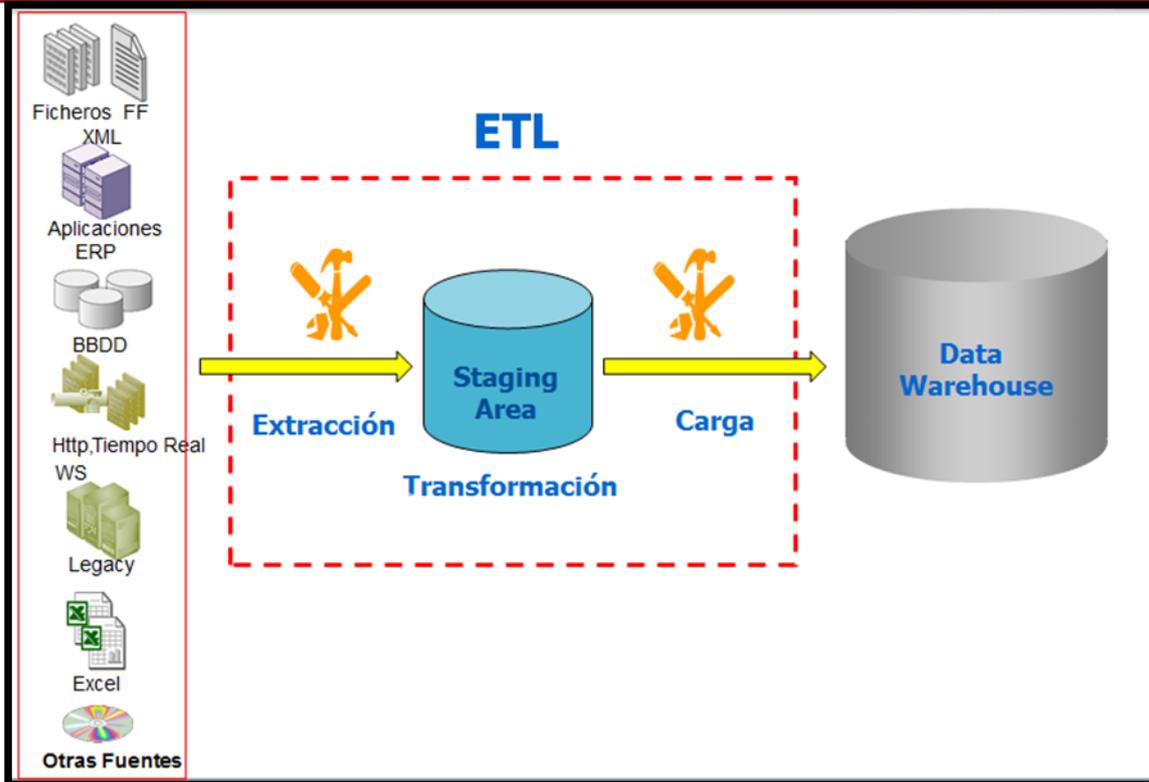


Figura 13. Proceso de Carga (ETL) [MEDINA LA PLATA, 2012]

Herramientas de Explotación

Una solución de inteligencia de negocios se completa mostrando de manera adecuada al usuario la información que necesite en el momento oportuno. Se realizan tareas como construcción de interfaces, desarrollo de consultas, instalación de herramientas, capacitación, soporte. [KIMBALL 2002, 13-15]

Un Dashboard es una interfaz de explotación de información que de forma resumida muestra en una sola vista la información crucial que describe la forma en que se van alcanzando los objetivos. En esta interfaz es muy importante centrarse en el diseño y debe estar personalizado de acuerdo al usuario. Se debe tener un buen conocimiento del negocio, conocer al tipo de cliente, identificar información clave para el usuario, definir bosquejos del dashboard, validar los dashboards. Se puede apreciar ejemplos de dashboards en las figuras 14 y 15.

También se puede explotar la información a través de reportes, son básicamente consultas analíticas, mediante las cuales se busca tener una visión general del negocio entre los diferentes sistemas y departamentos de la organización. Se busca la visión integral del negocio

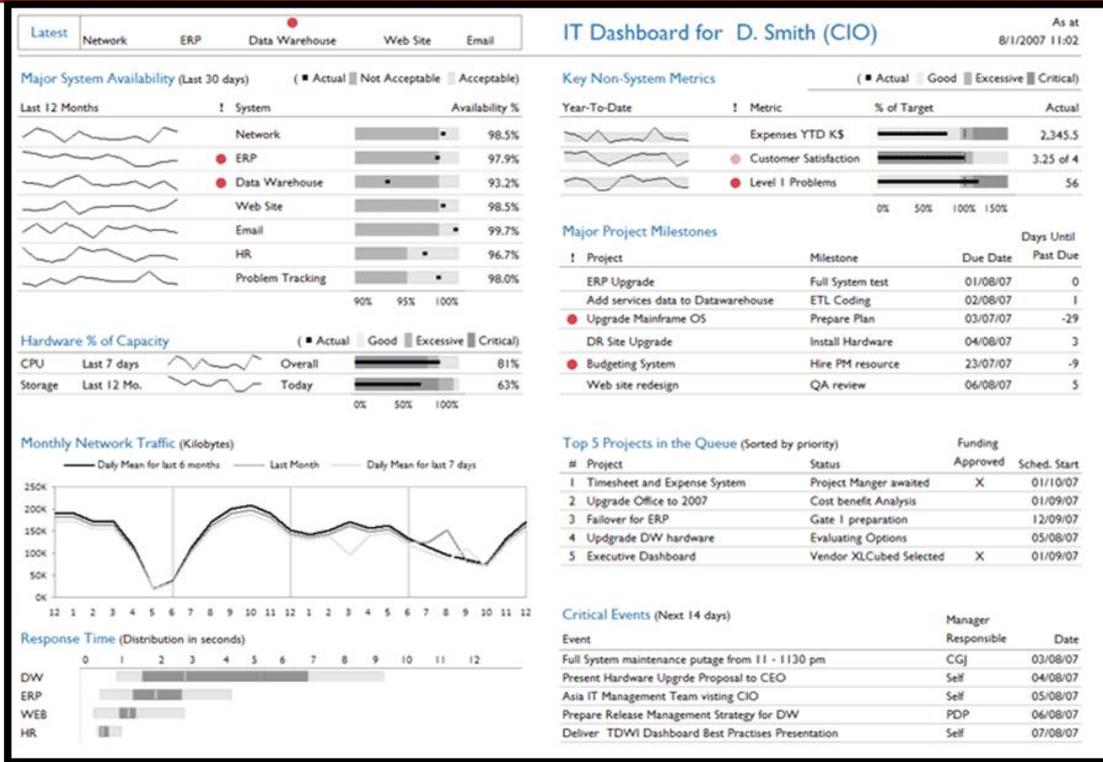


Figura 14. Dashboard [FEW 2006]

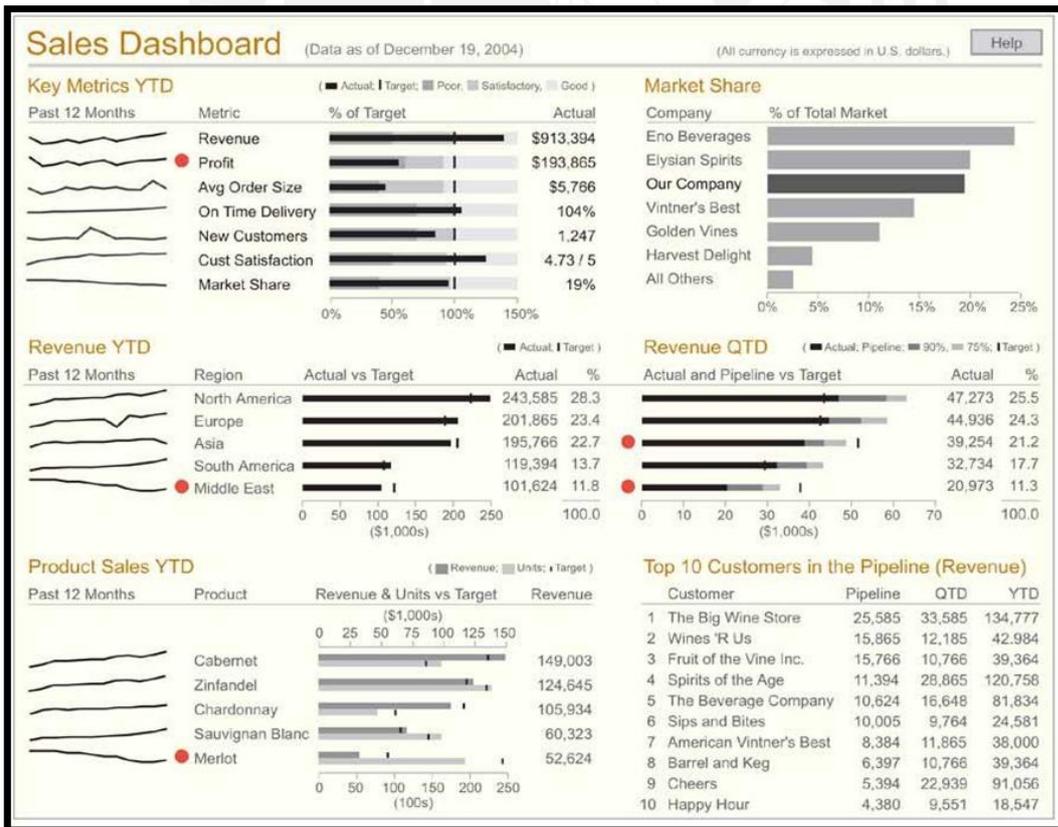


Figura 15. Dashboard [FEW 2006]

2.2 Estado de Arte

2.2.1 Problemas relacionados

En el estado del Arte se buscó diferentes situaciones en donde la problemática haya sido resuelta con una herramienta de inteligencia de negocios, específicamente con un Data Mart. De esta manera se puede ver la semejanza en cómo se afrontan estos problemas:

- Falta de calidad de la información.
- Masa crítica de datos.
- Es necesario automatizar los procesos de extracción y distribución de la información.

La mayoría de los casos encontrados son tesis, en las referencias se encuentran debidamente identificadas.

2.2.1.1 Caso solución de inteligencia de negocios para el área de finanzas de la Municipalidad Metropolitana de Lima

Título: Análisis, diseño e implementación de una solución de inteligencia de negocios para el área de finanzas de la Municipalidad Metropolitana de Lima [NÚÑEZ 2010]

El proyecto surge por la necesidad que tienen los usuarios del Área de Finanzas de la Municipalidad Metropolitana de Lima de acceder de manera rápida a información confiable relacionada con su trabajo. Al tener esta información disponible los usuarios se pueden centrar en realizar el análisis desde diversos puntos de vista y determinar las acciones que crean conveniente para mejorar su gestión.

El procedimiento a realizar por los usuarios para acceder a esta información, era solicitar al Área de sistemas la exportación de los datos de la Base de Datos, vaciar estos datos en un archivo Excel, depurar, ordenar, organizar y clasificar dichos datos para obtener la información que requieren para realizar sus labores. Este procedimiento manual genera diversos problemas tales como: dependencia del Área de Sistemas, información generada con una alta posibilidad de error (proceso manual), inversión de tiempo en procedimientos mecánicos e información dispersa en archivos de Excel, dificultando la consulta de datos históricos.

El proyecto tiene como objetivo el análisis, diseño e implementación de una Solución de Inteligencia de Negocios para el Área de Finanzas de la Municipalidad

Metropolitana de Lima, relacionados a la ejecución de los ingresos, la ejecución de los gastos y de los indicadores que permitan llevar un adecuado control de la gestión municipal. De modo que esta herramienta automatice el procedimiento que ha sido explicado en el párrafo anterior, y, que en un entorno amigable, permita a los usuarios acceder a información de mejor calidad, más confiable, en menor tiempo y en un repositorio que permita acceder a información histórica.

2.2.1.2 Caso Solución Integral para explotar eficientemente la Información de los contactos con los clientes utilizando Data Mart en Telefónica del Perú

Título: Solución Integral para explotar eficientemente la Información de los contactos con los clientes utilizando Data Mart en Telefónica del Perú. [GAMARRA 2010]

La carencia de integración de la información en Telefónica representa un problema para los jefes, analistas, supervisores, operadores e incluso para los propios clientes. No existe un ente integrador que reúna los contactos con el cliente y los publique en un único lugar al cual puedan acceder los usuarios internos y utilizar esta información para análisis, seguimientos, búsquedas, cruce de información, etc.

El problema se da cuando un usuario interno tiene que realizar algún cruce de información con estas fuentes, ya que debe buscar a las personas responsables de dicha información, solicitarla, esperar la confirmación, acceder a ésta y exportarla a algún motor de base de datos, analizarla, entenderla, depurarla, para finalmente poder realizar el cruce que se necesitaba inicialmente. En caso se tenga que realizar este proceso con 2 o más fuentes, significa realizar el mismo procedimiento multiplicado por el número de fuentes que se necesite.

Un simple cruce de información con fuentes de múltiples contactos con el cliente toma para un usuario promedio alrededor de 1 semana, sólo para conseguir integrar esta información. En caso esta simple tarea se deba realizar periódicamente, se convertiría prácticamente en una función diaria; función que abarcaría todo el tiempo del analista en su día de trabajo.

La implementación de un Data Mart de Contactos, resolverá la necesidad de los usuarios de Telefónica de contar con la información de los contactos con los clientes

completa, validada y disponible cuando ellos lo requieran, asimismo, los análisis que se realicen de ésta incrementará el conocimiento de las necesidades del cliente.

2.2.1.3 Caso Implementación de una herramienta de inteligencia de negocios para la administración de justicia sobre una metodología ad-hoc

Título: Implementación de una herramienta de inteligencia de negocios para la administración de justicia sobre una metodología ad-hoc. [EGUILA PARCO 2007]

En este proyecto de tesis se desarrolla el Análisis y Diseño para el desarrollo e implementación futura de una herramienta de Inteligencia de Negocios para la toma de decisiones en el Área de Defensoría de Oficio del Ministerio de Justicia, el propósito de la implementación de dicha herramienta es tener un mejor control y gestión de la información del sistema de defensores de oficio de forma que, ayude a mejorar la calidad del servicio que presta dicha entidad, por la toma de decisiones eficientes conseguidas a partir de la información que este sistema proporcione a los directivos de dicha entidad. Este trabajo muestra la metodología utilizada para el desarrollo del Data Mart, el cual planteara cómo desarrollarlo y se muestra mediante prototipos su funcionamiento.

La información consolidada permitirá realizar un mejor control con la aplicación de tecnologías de información, en este caso, el uso de Data Mart, esto contribuirá en beneficio de todos los peruanos con lo cual permitirá disponer de herramientas de gestión y a su vez de control para el Sector de justicia.

2.2.1.4 Caso Data Mart del área de Riesgos de la Banca Personal de una institución Financiera.

Título: Análisis, diseño e implementación de un Data Mart del área de Riesgos de la Banca Personal de una institución Financiera. [CHIPANA 2004]

Las Entidades Financieras tiene como principal objetivo el otorgar créditos para luego cobrarlos. La complejidad de esta actividad se centra principalmente en hacer una buena elección del cliente, con el propósito de garantizar que cumpla debidamente con

el pago del crédito que se le otorgue, y de esta manera obtener las ganancias esperadas en el tiempo provisto.

La mayoría de las entidades financieras cuentan con información obtenidas de bases de datos transaccionales, de suma importancia para registrar las operaciones que se realizan. Estos datos son de suma importancia pero no se puede acceder a ellos con la facilidad debida cuando se necesita contar con información estratégica.

El desarrollo de un Data Mart para el área de Riesgos de la Banca Personal permite obtener la información estratégica requerida, como podría ser el acceso directo a los indicadores del negocio. Señalar cuales son los factores que realmente inciden en el buen o mal funcionamiento, detectar situaciones fuera de lo normal, encontrar los factores que maximicen el beneficio y predecir el funcionamiento futuro con un alto porcentaje de certeza

2.2.1.5 Caso Data Mart de gestión de negocios aplicado a una institución financiera.

Título: análisis de la implementación de un Data Mart de gestión de negocios aplicado a una institución financiera. [SALAZAR 2003]

Esta tesis analiza la propuesta que está dirigida a implementar un sistema informático eficiente y confiable para la evaluación de la gestión financiera diaria, y para la supervisión del nivel de cumplimiento de las metas asignadas a cada ítem de rendimiento. Esto, con el propósito de buscar que se alcancen los objetivos estratégicos fijados.

La implementación de esta solución involucra la elaboración de un modelo de datos que soporte las consultas requeridas por los usuarios para un mejor análisis del desenvolvimiento de la entidad financiera.

El producto final será un sistema de análisis de gestión y otro de seguimiento de objetivos, los cuales deben ser lo suficientemente flexibles para permitir implementar todo tipo de vistas requeridas.

2.2.1.6 Caso Data Warehouse INEI (Instituto nacional de estadística e informática)

Los Censos Nacionales de Población y Vivienda, constituyen la operación estadística más grande que desarrollan las oficinas de estadística y son a la vez, la fuente más eficaz para obtener información sobre la magnitud, distribución espacial y composición de la población, así como de las características de las viviendas y de los hogares.

En base a esta operación estadística es posible desagregar la información a nivel nacional, departamental, provincial, distrital y de centros poblados urbanos y rurales del país y obtener una respuesta inmediata. En este sistema, en su presentación inicial cuenta con la información del Censo Nacional 2007: XI de Población y VI de Vivienda (CPV 2007) la cual está organizada en tres conjuntos de temas de consulta:

- Cuadros Estadísticos Predefinidos: muestra los cuadros estadísticos más consultados al INEI, son cuadros que el usuario puede modificar su presentación y niveles de desagregación geográfica.
- Consultas Personalizadas: permite a los usuarios elaborar cuadros a medida de sus necesidades.
- Consulta de Indicadores: muestra los principales indicadores demográficos, sociales, económicos y de vivienda y hogar, construidos con la información censal.
- Mapa Temático de Indicadores: permite realizar el análisis espacial de los indicadores seleccionados

Se puede apreciar consultas del Warehouse de INEI en las figuras 16 y 17.

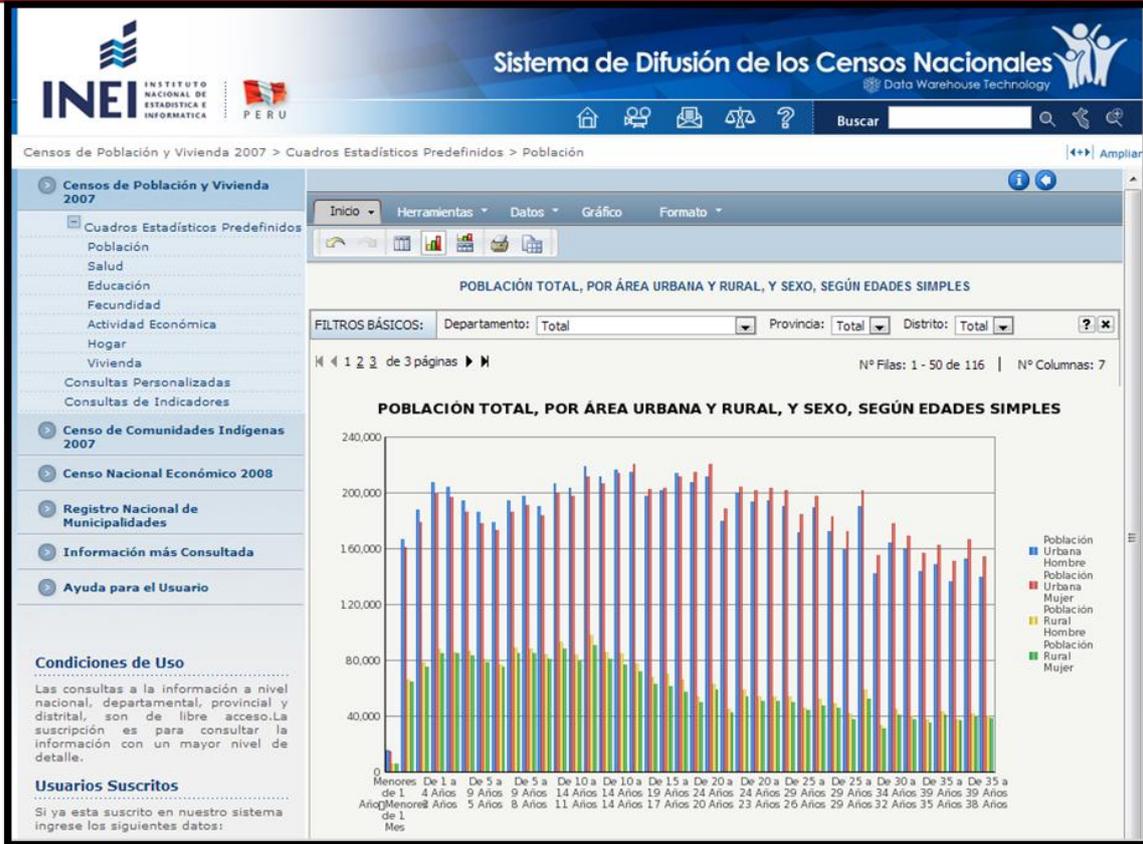


Figura 16. Página web INEI [INEI01]

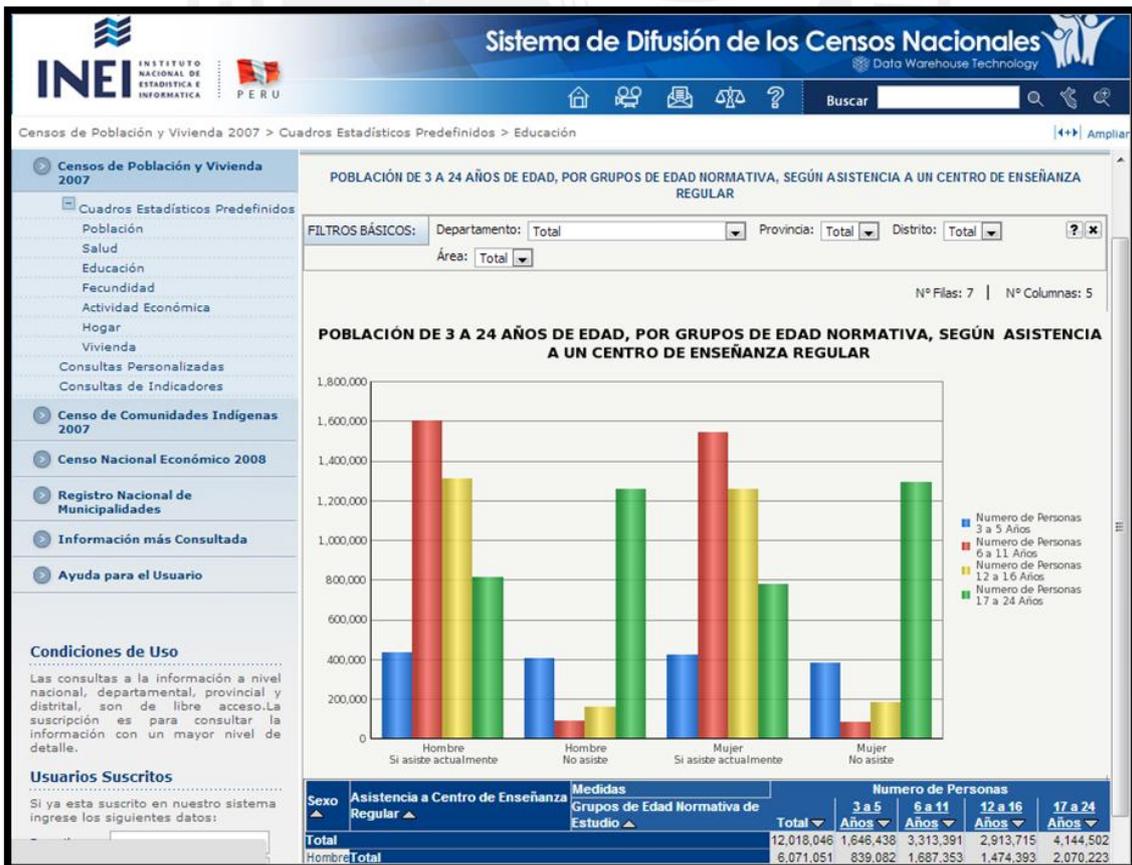


Figura 17. Página web INEI [INEI01]

2.2.2 Conclusiones sobre el estado del arte

Luego de analizar las tesis pasadas se llega a la conclusión de que los Data Mart's, han sido fundamentales para mostrar los indicadores de los procesos clave del negocio, esto ha permitido que la alta dirección tome las medidas correctivas necesarias y mejore sobremanera los procesos. Con el debido análisis y organización de la estructura del Data Mart se puede facilitar mucho el trabajo y brindar información necesaria a la alta dirección.

El RENIEC dado que es una entidad del estado, que no tiene fines de lucro, y se debe solo a los peruanos, su único fin es mejorar la calidad de sus servicios. En esta institución los procesos clave son el registro del trámite y entrega del DNI. El desarrollo de una solución de Inteligencia de Negocios que permita explotar la información estratégica requerida por la alta dirección mediante el uso de Data Mart's, facilitaría tener un acceso directo a los indicadores de los procesos, esto permitiría señalar cuales son los factores que realmente inciden en el buen o mal funcionamiento, detectar situaciones anormales, llegar a ubicar los factores que ayuden al proceso y predecir el funcionamiento futuro con un alto porcentaje de certeza.

3 Análisis

3.1 Definición de la metodología de la solución

Se eligió la metodología que tiene como base el ciclo de vida de los Data Warehouses definido por Ralph Kimball. El esquema usado por Ralph Kimball con el nombre de BDL (Business Dimensional Lifecycle) se encarga de graficar las distintas etapas por las que debe pasar todo proceso de Data Warehousing.

La metodología de Kimball divide el desarrollo del almacén de datos en 5 fases generales: Marco conceptual, análisis, diseño, construcción, implantación. Como se puede apreciar en la figura 18.

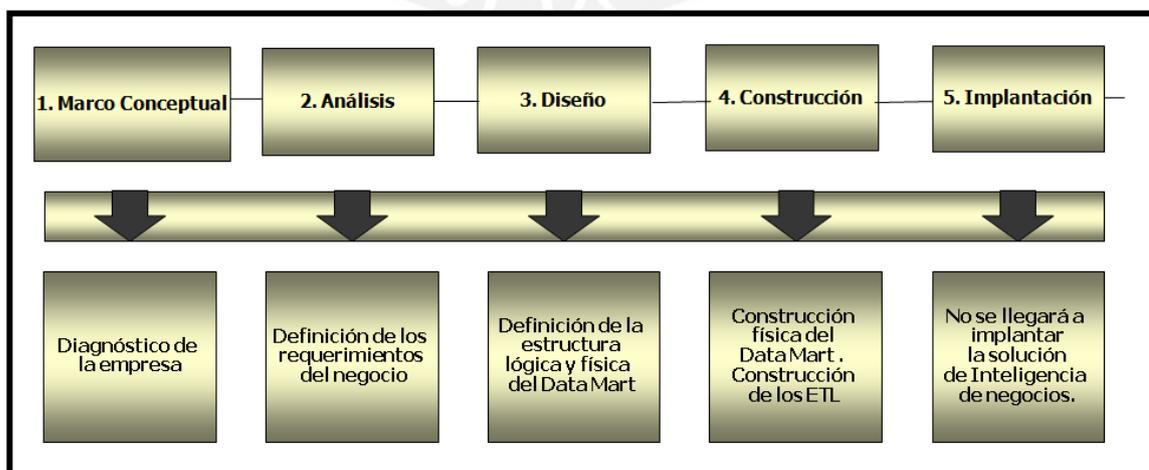


Figura 18. Fases generales de Desarrollo según Kimball [Elaboración Propia]

Marco Conceptual: La primera etapa de todo proyecto de inteligencia de negocios es el diagnóstico de la empresa, en la cual se realiza el levantamiento de información de la empresa, se identifica sus necesidades, se diagnostica el problema y se plantea una solución adecuada. El desarrollo de todas las etapas del proyecto se logra a partir de las diversas entrevistas y reuniones con los diferentes encargos de RENIEC. En esta etapa también se definen las actividades, alcance y cronograma del proyecto, incluyendo tiempo y recursos.

Análisis: Se realiza la definición de los requerimientos del negocio, este es el punto de inicio del proyecto y a la vez el cimiento para las siguientes etapas del proyecto, las cuales se enfocan en la tecnología, los datos y las aplicaciones. En esta etapa se realiza la entrevista a los usuarios finales lo que permite entender los procesos del negocio, sus necesidades y requerimientos. En la presente tesis se entrevistó al personal de RENIEC, en el Anexo 8 se encuentra el detalle de la entrevista. Durante esta etapa también se identifican los sistemas transaccionales fuente, se conoce la arquitectura de los sistemas transaccionales, se realiza un inventario y se analiza la calidad de los datos. Para el caso de los sistemas transaccionales fuente, se contó con una comunicación fluida con el Área de Base de datos, de esta manera se utilizó data real de RENIEC, y se utilizó un modelo similar a la base de datos Transaccional de RENIEC.

Diseño: Se ha usado como referente en esta etapa a Ralph Kimball. En el libro “The Data Warehouse Toolkit” (“El kit de herramientas Data Warehouse”), Kimball dedica tres capítulos a la etapa del modelado dimensional, donde describe las diferentes arquitecturas de construcción, esquemas de modelos y de tablas de hechos, definiendo el concepto de Data Mart.

Se define la estructura del Data Mart tanto a lógico como físico. Además, se realiza un Diagrama de Mapeo de Datos, el cual muestra la relación de cómo los datos fuentes están relacionados con los datos del modelo multidimensional. En esta etapa se tuvo un feedback con el personal de Base de datos y el área de TI de RENIEC, para elaborar modelos adecuados de Data Mart's.

Construcción: Durante esta etapa se construye la estructura física del Data Mart. También, se desarrollan los procesos de extracción, transformación y carga al Data Mart.

Durante esta etapa se configura e instala la base de datos a utilizar, así como también la plataforma de Inteligencia de Negocios de software libre: Pentaho. La etapa de construcción de ETL consta de tres sub etapas principales: extracción, transformación y carga de datos (procesos ETL). Luego de la construcción del ETL y los reportes se pone en práctica el plan de pruebas que permite identificar los errores originados en las etapas anteriores. Las pruebas son realizadas en el presente proyecto utilizando diferentes parámetros de entrada.

Implantación: Durante esta etapa se implanta la solución de Inteligencia de Negocios en la Empresa. En este caso con data de prueba real suministrada por la RENIEC se comprobó el correcto funcionamiento de la solución de Inteligencia de Negocios. No se llegó a implantar la solución de Inteligencia de Negocios en la empresa.

3.2 Identificación de requerimientos

La solución planteada tiene como objetivo principal ofrecer una herramienta que mediante su uso adecuado mejore la toma de decisiones en las gerencias GOR, GRI, GRIAS en el RENIEC. Es decir el requerimiento principal es que la solución de Inteligencia de Negocios propuesto brinde un conjunto de facilidades que permitan utilizar la información disponible en la RENIEC para hacer un mejor análisis a nivel de producción, y de emisión del DNI. De esta premisa general se puede identificar los requerimientos funcionales y no funcionales que debe cumplir la solución.

3.2.1 *Requerimientos funcionales*

Los requerimientos funcionales en el desarrollo del Data Mart constituyen las necesidades de información de los usuarios finales que en este caso es el personal de la gerencia de RENIEC. Específicamente los usuarios finales para la herramienta son: Analistas de información (gerencias GOR, GRI, GRIAS), Power Users (Gerencia de Informática), Consumidores de información (Consultan información gerencias GOR, GRI, GRIAS), Tomadores de Decisiones (Gerentes, Sub-gerentes), Departamento de TI (Sub-gerencia de gestión de base de datos). [RENIEC08 2012]

El desarrollo del presente proyecto se basó en las necesidades de información de la gerencia de informática del RENIEC. Se estima que anualmente maneja más de 8 millones de registros

Luego de algunas reuniones con el personal de la gerencia de informática, se obtuvieron las necesidades generales de la entidad.

-Contar con una herramienta para brinde información necesaria para facilitar el control de los procesos: Es la necesidad principal dentro de toda el área y a la vez el objetivo principal de la solución planteada. Actualmente el área de estadísticas utiliza Microsoft Excel para la elaboración de reportes. Sin embargo al utilizar la solución de Inteligencia de negocios dentro de las diferentes aéreas, la persona encargada de tomar las decisiones adecuadas, podrá acceder a la información requerida lo más rápido posible sin necesidad de solicitarla al área de estadística, evitando así que un funcionario del área de estadística pida la información al área de base de datos y elabore un reporte manualmente.

-Tener mejor organizada la información del área: El personal de gerencia de informática maneja una gran cantidad de información en el día a día debido a la generación y emisión del DNI. Por ello, requieren contar con la información actualizada y organizada ya sea para tomar decisiones u otras actividades. La RENIEC no solo maneja un sistema transaccional, es por eso que se tiene que integrar la información para luego poder analizar la información sumariada o agrupada es mediante los reportes. La solución de Inteligencia de Negocios permitirá contar con la información organizada de manera intuitiva de tal forma que el personal pueda acceder a ella.

Los requerimientos funcionales que se presentan a continuación han sido agrupados por módulos logísticos para un mejor entendimiento en: Producción, Calidad.

3.2.1.1 Producción

Para las gerencias en el RENIEC es de mucha utilidad saber la cantidad de trámites que se han tramitado (Gerencia de Operaciones Registrales) y la cantidad de trámites que se han emitido (Gerencia de Registro e Identificación). Con esta información se puede analizar la información de los trámites y se podrá responder preguntas como ¿Cuántos DNI's son tramitados, por tipo de trámite, por local, por tipo de DNI, en una periodo determinado? , ¿Cuántos DNI's son emitidos, por tipo de trámite, por local, por

tipo de DNI, en un periodo determinado? De esta manera se podrán tomar las acciones correctivas necesarias. [RENIEC08 2012]

Los esquemas de análisis requeridos en este módulo son:

- Cantidad de DNI's tramitados agrupados por tipo de trámite, local, tipo de DNI, en un periodo determinado.
- Cantidad de DNI's tramitados agrupados por tipo de trámite, UBIGEO de domicilio o de nacimiento, tipo de DNI, en un periodo determinado.
- Cantidad de DNI's emitidos agrupados por tipo de trámite, local, tipo de DNI, en un periodo determinado.

También se tomó en cuenta el tiempo de duración del trámite en sus diferentes fases. Con esta información se responderá preguntas como: ¿Cuántos trámites se demoraron un rango de tiempo predeterminado, en cada una de sus fases, agrupados por fecha, por tipo de trámite, por local, por tipo de DNI?

Los esquemas de análisis requeridos en este módulo son:

- Cantidad de trámites acorde a un rango de duración tiempo predeterminado, abarca desde la fecha de trámite hasta la fecha de ingreso a impresiones.
- Cantidad de trámites acorde a un rango de duración tiempo predeterminado, abarca desde la fecha de ingreso a impresiones hasta la fecha de salida de impresiones.
- Cantidad de trámites acorde a un rango de duración tiempo predeterminado, abarca desde la salida de impresiones hasta la fecha de entrega al local.

3.2.1.2 Calidad

En esta parte se tomó en cuenta los trámites que han sido observados. Con esta información se responderán preguntas como: ¿Cuál es el ranking de los trámites rechazados?

Los esquemas de análisis requeridos en este módulo son:

- Ranking con los motivos de los trámites observados agrupados por locales, tipo de DNI, en un periodo determinado.

Los requerimientos presentados podrán satisfacer las necesidades de información de las diferentes áreas. De esta manera el personal involucrado podrá contar con la información adecuada, organizada en el tiempo adecuado de tal forma que le ayude en la toma de decisiones. [RENIEC08 2012]

3.2.2 Requerimientos no funcionales

Los requerimientos no funcionales no pueden asociarse a un Caso de Uso en particular, pero son de mucha importancia en el desarrollo de una solución de Inteligencia de Negocios. A continuación se presenta los requerimientos no funcionales identificados:

- El modelo multidimensional de la solución debe tener escalabilidad.
- La herramienta de explotación seleccionada debe permitir la realizar las técnicas de consulta multidimensional como Slicing, Dicing, Drilling, etc.
- La solución debe contener esquemas de análisis y diferentes gráficos elaborados con la herramienta de explotación seleccionada. Además de la posibilidad de exportarlos a archivos de formatos estándar como archivo de Excel, Pdf, CSV.
- La herramienta de explotación debe permitir mostrar la información de los Data Mart's en base a los 3 siguientes tipos: Reportes tradicionales (basados en columnas), tablas pivote (matriz de doble entrada) y Gráficos.

- Poder acceder a la información deseada de manera rápidamente: El tener acceso a la información requerida es una necesidad de suma importancia para la gerencia. Los sistemas actuales están diseñados para el almacenamiento de la información de manera rápida, pero no para la consulta de ella. Los reportes muchas veces no satisfacen las necesidades de información de la gerencia y consumen trabajo y tiempo en la elaboración de reportes que permiten contar con la información rápidamente. La solución de Inteligencia de Negocios permitirá acceder a la información de manera rápida y además la elaboración de reportes, ahorrando tiempo y carga a las aéreas de informática, estadística y base de datos.

- Contar con una herramienta fácil de usar: La solución planteada no sirve si es que los usuarios no la utilizan. El personal está acostumbrado a emplear el Microsoft Excel como herramienta de análisis de datos. La nueva herramienta debe ser fácil de usar de tal forma que todo el personal pueda emplearlo.

4 Diseño

En el presente capítulo se describirá el diseño de los Data Mart's, que van acorde a los requerimientos expuestos en la sección anterior.

En este caso para el diseño de la base de datos se realizó el diseño lógico y físico. El primero muestra las tablas y la relación entre ellas. El segundo, diseño físico, muestra las tablas con sus campos y sus respectivos tipos de datos

A continuación se detallaran los aspectos que se tomaron en cuenta en el diseño del modelo de datos del Data Mart.

4.1 Diseño conceptual del Data Mart

Para satisfacer las necesidades de información de los usuarios se definió 2 esquemas de modelos estrella: Producción y calidad. Cada esquema abarca un tema de análisis de información en particular organizando las variables y dimensiones necesarias para satisfacer la demanda de información. Es pertinente explicar algunos conceptos relacionados al diseño lógico

-Nivel de granularidad: Indica hasta qué nivel de detalle puede abarcar el análisis del modelo estrella. Está relacionado directamente con la tabla de hechos, pues un registro de esta tabla indica el nivel de granularidad del modelo estrella.

-Nivel de agregación: Indica cómo está agrupada o sumariada la información del esquema. También está relacionado a la tabla de hechos pues contiene cantidades o montos sumariados.

-Jerarquía de las dimensiones: Algunas dimensiones tienen campos ordenados de una manera jerárquica que permite un análisis de datos más generales a específicos y viceversa. Las jerarquías se encuentran en las tablas dimensiones y cada una de ellas puede tener distintos niveles de jerarquía.

4.1.1 Esquema de producción

Este esquema tiene como objetivo permitir el análisis de los trámites que se han tramitado y emitido en el RENIEC a un nivel detallado. Satisface los requerimientos del análisis de los trámites. Las dimensiones son: Locales, Ubigeos, Fechas, DNI's Gratuitos, Tipos de DNI's, Tipos de trámites, Rango de tiempo, Región. Se puede apreciar el esquema de producción en la figura 19.

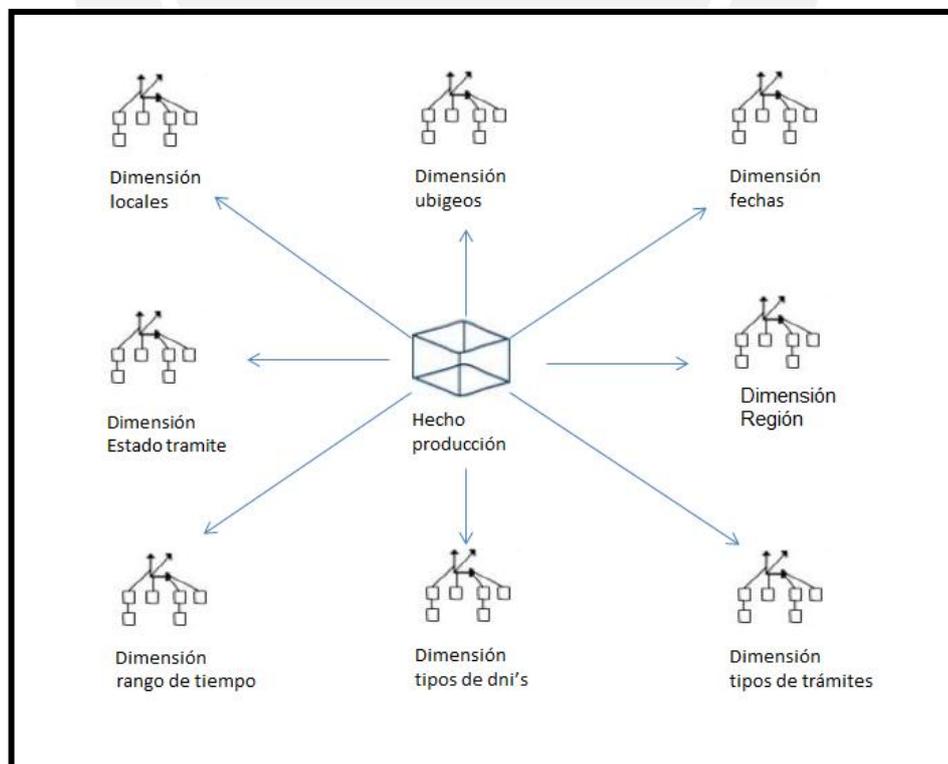


Figura 19. Esquema de producción [Elaboración propia]

4.1.2 Esquema de calidad

Este esquema tiene como objetivo permitir el análisis de los trámites que han sido observados, a un nivel detallado. Las dimensiones son Locales, Ubigeos, Fechas, Tipos de DNI's, Tipos de trámites, Estado de trámite, Región. Se puede apreciar el esquema de calidad en la figura 20.

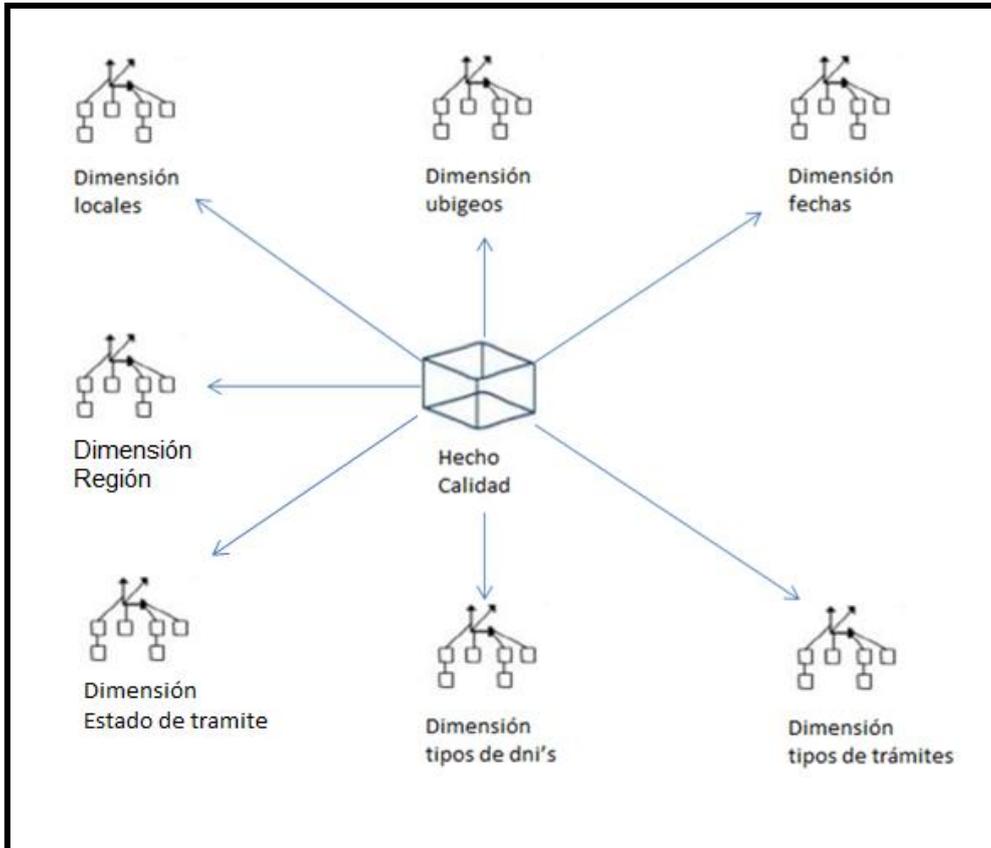


Figura 20. Esquema de calidad [Elaboración propia]

4.2 Diseño Lógico del Data Mart

En este caso para diseñar el modelo lógico primero definimos los campos de las tablas con sus tipos de datos más adecuados. Todo campo a emplear en el modelo de Data Mart está entre alguno de los siguientes grupos de tipos de datos: Numéricos exactos, numéricos aproximados, fecha y cadena de caracteres.

-Numérico exacto: Se empleó el tipo de dato INT para los identificadores de todas las tablas tanto de dimensiones como de hechos. Esto debido a que las búsquedas y consultas son más rápidas cuando la llave primaria es un número entero.

-Cadena de caracteres: Se empleó el tipo de dato VARCHAR para los valores de texto debido a que estos pueden de ser longitud muy variable

A continuación el diseño lógico de los esquemas estrellas.

4.2.1 Esquema de producción

En la figura 21 se puede apreciar el esquema de producción, éste esquema brinda información detallada de los trámites realizados en RENIEC.

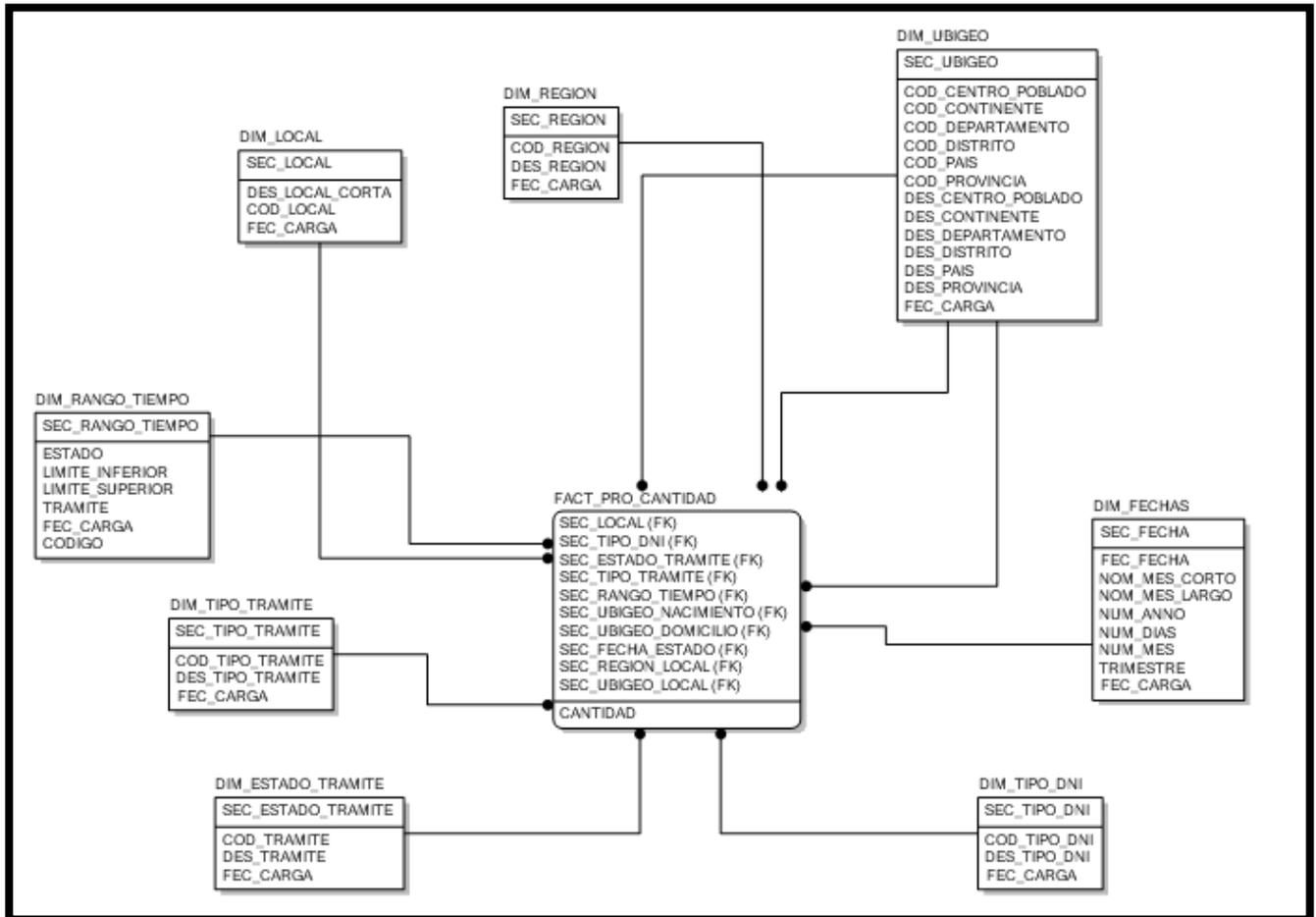


Figura 21. Diseño lógico del esquema de producción [Elaboración propia]

A manera de explicación, la dimensión Rango Tiempo tiene un estado que puede ser bajo, medio o alto, que corresponde a la cantidad de días que se ha demorado un trámite, el límite inferior y límite superior esta dado para cada uno de los estados, y estos criterios se repiten para cada uno de los estados por los que pasa el trámite, que puede ser: ingreso al sistema, ingreso a impresiones, envío desde el local de impresiones al local de recojo del DNI, y recepción de los trámites en el local del recojo del DNI.

4.2.2 Esquema de calidad

En la figura 22 se puede apreciar el esquema de calidad, éste esquema brinda información detallada de los trámites que han sido observados.

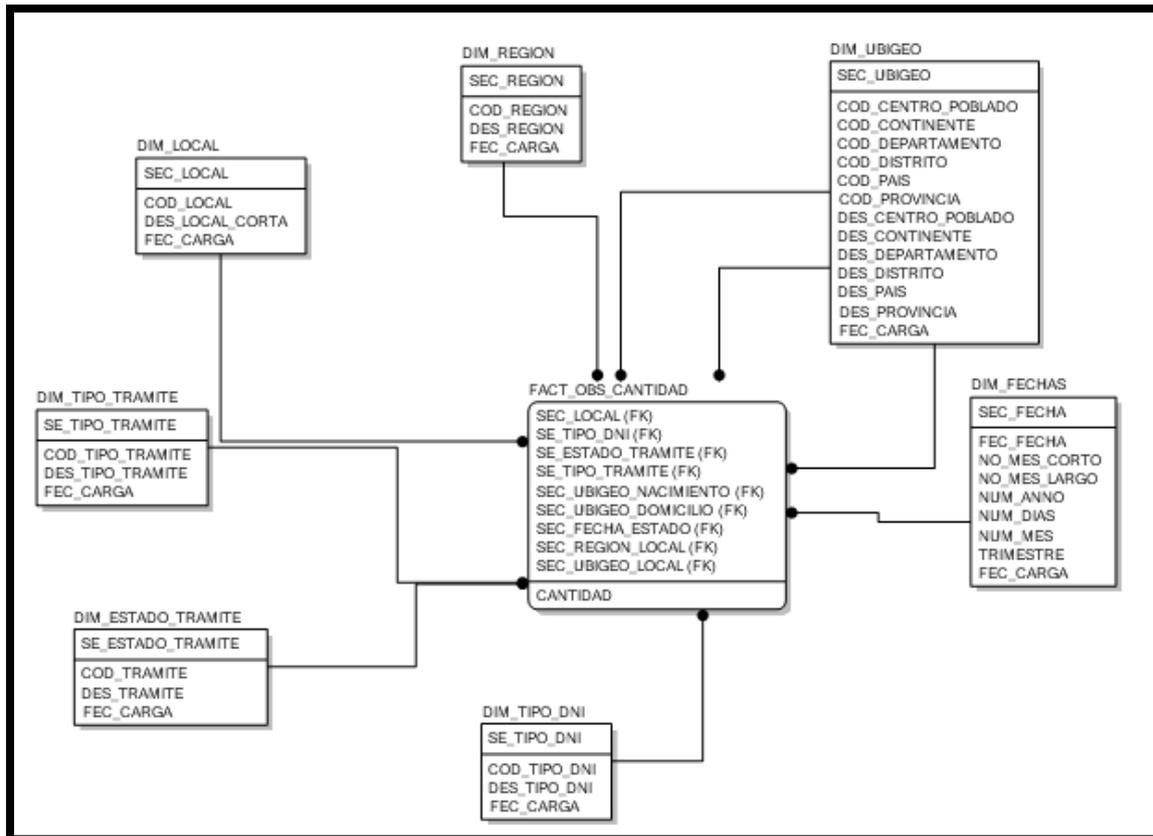


Figura 22. Diseño lógico del esquema de calidad [Elaboración propia]

4.3 Arquitectura de la solución

La arquitectura que permite la implementación de la Solución de Inteligencia de Negocios estuvo formada por diversos componentes.

- **Componente Fuentes de información:** Está formada por las fuentes de información del presente proyecto y por ello será la fuente de datos para el Data Mart. Estuvo formado por la base de datos que contiene la información transaccional de los trámites del RENIEC.
- **Componente ETL:** Implementa los procesos necesarios para extraer la información de los datos fuentes, transformarlos y cargarlos en la base de datos del Data Mart. Se utilizó la herramienta Pentaho Integration para la elaboración de los procesos ETL.

- **Componente Data Mart:** Representa a la base de datos multidimensional del Data Mart. Tiene una estructura especialmente diseñada para cargarse en una herramienta de explotación y permitir el análisis de la información. Este componente así como el de las Fuentes de información estuvo en la base de datos PostgreSQL.
- **Componente OLAP:** Contiene los servicios OLAP que sirviéndose de la base de datos del Data Mart permitió mostrar los datos en Cubos multidimensionales para su mejor análisis. Se utilizó Pentaho Schema Workbench para crear los cubos.
- **Componente Interfaz de Usuario:** Representa la interfaz con la que interactúa el usuario. A través de ella es posible tomar las decisiones correspondientes en base a la información que se les muestra. Se utilizó la consola de usuario de Pentaho como herramienta de explotación e interfaz de usuario. En esta interfaz se pueden elaborar las tablas de doble entrada, esquemas de análisis, entre otros gráficos.

En la Figura 23 se muestra la interacción de los componentes que conforman la arquitectura de la solución.

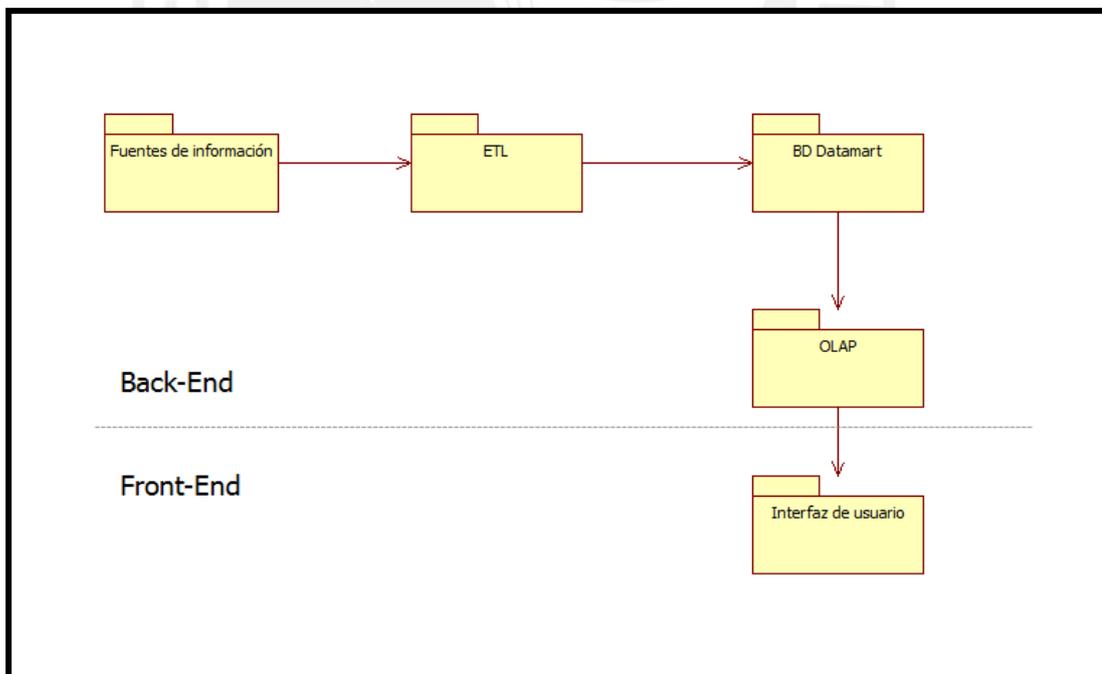


Figura 23. Arquitectura de la solución [Elaboración propia]

A continuación se desarrollará a mayor detalle los principales componentes que forman parte del Back-End.

4.3.1 Fuentes de información

Se tomó en cuenta a la base de datos de RENIEC, que se encuentra con los filtros pertinentes, para que no se tomen en cuenta las tablas que son innecesarias para la elaboración del Data Mart.

Esta base de datos es la que proporcionó de información al Data Mart. Se utilizó información de prueba para probar la funcionalidad del Data Mart. La estructura de la base de datos se encuentra indicada en el Anexo 2.

4.3.2 ETL

El proceso ETL se encarga principalmente del paso de la información de un modelo de base de datos relacional a un modelo multidimensional. El proceso tiene como tarea extraer la información que se encuentra almacenada en el modelo relacional, realiza la transformación de los tipos de datos de los campos correspondientes y carga los datos a las tablas de dimensiones y hechos de la base de datos multidimensional.

Para empezar el diseño del proceso ETL se realizó un mapeo de los datos de la base de datos fuente con los de la base de datos del Data Mart. Es así como se puede establecer la relación de los campos en ambas bases de datos.

Una vez realizado el mapeo de datos se diseñó el flujo del proceso de ETL para cargar cada dimensión o hecho para tener una idea clara sobre las transformaciones que sufrirán los valores al pasar de una base de datos a otra. Esto se debió realizar para facilitar la implementación de los procesos ETL en la fase de construcción.

Los diagramas de los procesos ETL se encuentran detallados en el Anexo 3.

4.3.3 Data Mart

El Data Mart estuvo formado por la entidad de base de datos. A nivel físico, las tablas están almacenadas en diversos archivos planos, pero podemos dividirlos en 2 grupos. Los dos grupos son: Dimensiones y Hechos. Esto permitió una mejor organización, además de permitir agregar de manera más simple futuras tablas a medida que crezca la base de datos. En la figura 24 se observa el diagrama de componentes que conforma la base de datos del Data Mart.

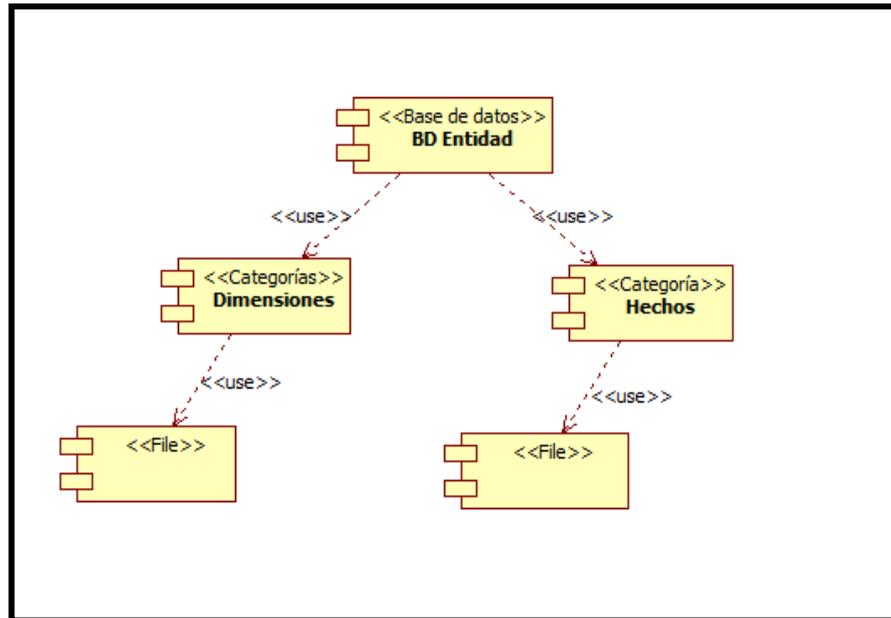


Figura 24. Diagrama de componentes de la base de datos del Data Mart [Elaboración propia]

Es importante mencionar que los componentes del Back-End se encontrarían en el servidor de la RENIEC y sólo la interfaz del usuario estaría en cada computadora cliente de los usuarios finales cuando se llegue a implementar. En la figura 25 se muestra el diagrama de componente de la computadora que sería el servidor que contenga la base de datos del Data Mart.

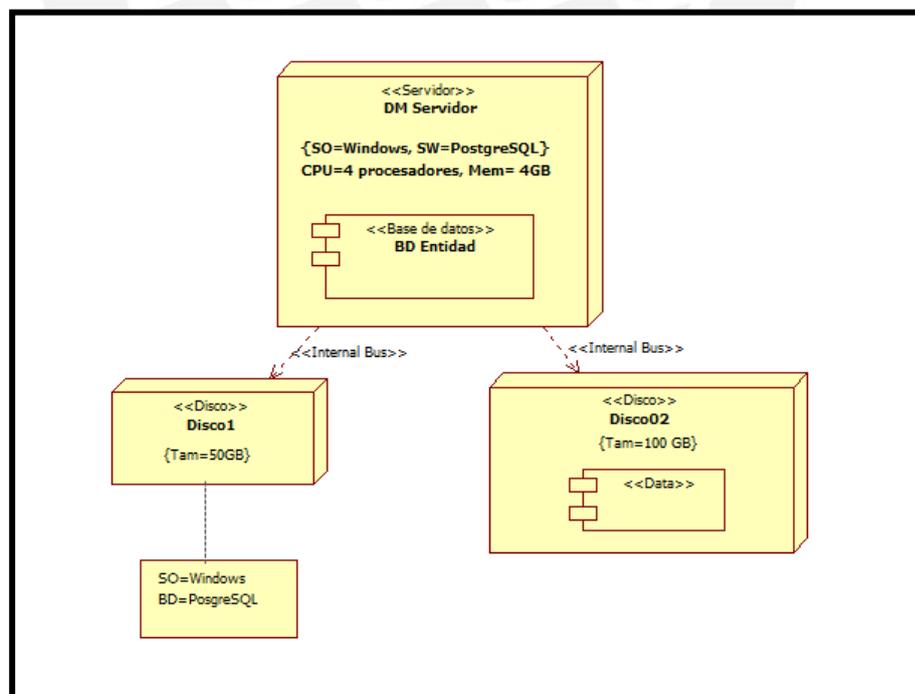


Figura 25. Diagrama de componente del servidor de base de datos. [Elaboración propia]

La computadora que almacenará el Data Mart debe tener como mínimo las siguientes características de hardware:

Procesador	Procesador de 4 núcleos
Memoria RAM	4 GB
Disco Duro	150 GB

Tabla 3. Características del Hardware.

Además, debe tener el sistema operativo Windows Vista y tener instalado PostgreSQL y las herramientas de Pentaho: Servidor de BI, Data Integration, Schema Workbench, Report Designer. Por otro lado, las computadoras con las que interactuara el usuario deben tener como mínimo las siguientes características de hardware:

Procesador	Procesador de 2 núcleos
Memoria RAM	2 GB
Disco Duro	50 GB

Tabla 4. Características del Hardware.

Además, debe contar con Windows Vista y la herramienta para la visualización de datos llamada Pentaho.

4.4 Diseño de extracción

A continuación se presentará el mapeo de cómo se cargaron las tablas destino y sus tipos de datos, indicando las diversas fuentes, y finalmente se muestra en la figura 26, el diagrama para cargar una dimensión en el Pentaho.

4.4.1 Transformación de la dimensión DIM_LOCAL

Descripción de la tabla fuente

Tipo de fuente	Nombre de tabla	Descripción
BD	LOCAL	Registro de locales
BD	UBIGEO	Registro de ubigeos

Tabla 5. Tabla fuente.

Estandarización y limpieza de datos

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
SEC_LOCAL	PK	Integer	Numérico	No debe ser Nulo	No tiene
DES_LOCAL_CORTA		Varchar(60)	Texto	No debe	No tiene

				ser Nulo	
COD_LOCAL		Varchar(6)	Texto	No debe ser Nulo	No tiene
FEC_CARGA		Date	Date	No debe ser Nulo	No tiene

Tabla 6. Tabla de estandarización y limpieza de datos.

Fuente de datos

Tabla LOCAL				
Nombre	Llave	Tipo	Formato	Consideración Importante
COD_LOCAL	PK	Varchar(6)	Texto	No tiene espacios en blanco.
COD_CENTRO_POBLADO	FK	Char(3)	Texto	No tiene espacios en blanco.
COD_CONTINENTE	FK	Char(2)	Texto	No tiene espacios en blanco.
COD_DEPARTAMENTO	FK	Char(2)	Texto	No tiene espacios en blanco.
COD_DISTRITO	FK	Char(2)	Texto	No tiene espacios en blanco.
COD_PAIS	FK	Char(2)	Texto	No tiene espacios en blanco.
COD_PROVINCIA	FK	Char(2)	Texto	No tiene espacios en blanco.
COD_REGION	FK	Varchar(2)	Texto	No tiene espacios en blanco.
COD_TIPO_LOCAL		Char(1)	Texto	No tiene espacios en blanco.
DES_LOCAL_CARGA		Varchar(100)	Texto	
DES_LOCAL_CORTA		Varchar(60)	Texto	

Tabla 7. Tabla de fuente de datos.

Tabla Destino

Tabla:	DIM_LOCAL	
Campo	Tipo	Mapeo
SEC_LOCAL	Integer	Autogenerado
DES_LOCAL_CORTA	Varchar(60)	LOCAL.DES_LOCAL_CORTA
COD_LOCAL	Varchar(6)	LOCAL.COD_LOCAL
FEC_CARGA	Date	Fecha de carga de datos

Tabla 8. Tabla de Destino.

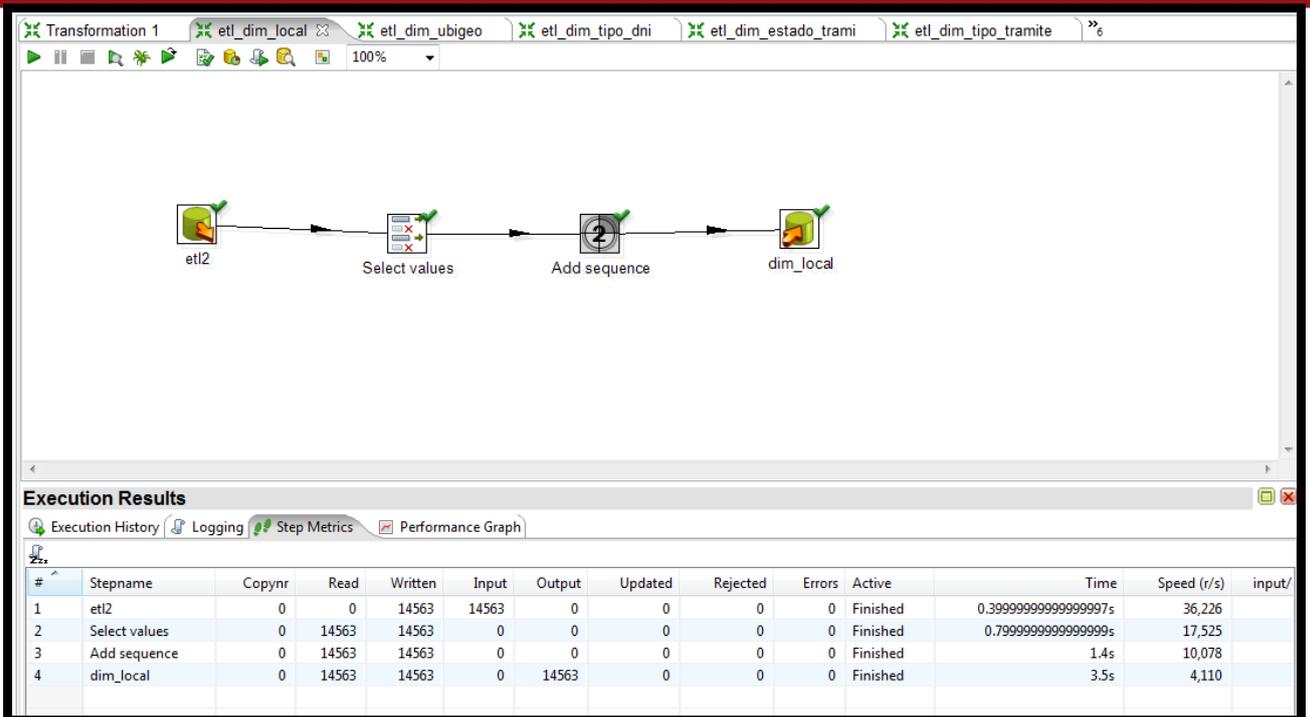


Figura 26. Imagen de carga de una dimensión usando la herramienta Pentaho.



5 Construcción

En el presente capítulo se mostrará los aspectos y criterios tomados en cuenta en la etapa de construcción

5.1 Construcción

La fase de Construcción del proyecto está definida fundamentalmente por la implementación de los procesos ETL debido a que los modelos de base de datos fuentes y del Data Mart ya están definidos y explicados en los capítulos anteriores. En el caso de la interfaz gráfica de usuario se utilizó como herramienta de explotación a Pentaho.

5.1.1 *Instalación de las herramientas*

Es necesario instalar y configurar adecuadamente las herramientas a utilizar para la puesta en producción de la solución. A continuación los programas que debieron ser instalados en el servidor.

- Motor de base de datos PostgreSQL (para el Data Mart)
- Pentaho Data Integration (para el proceso ETL)
- Pentaho Schema Workbench (para la elaboración de los cubos)
- Pentaho Business Intelligence Server (para la explotación de la información)
- Pentaho Report Designer (para la elaboración de los reportes)

5.1.2 Proceso ETL

Se implementó los procesos ETL a través de transformaciones y trabajos del Pentaho Data Integration.

En términos generales deben implementarse dos tipos de proceso ETL. Un proceso que permite cargar toda la información histórica de la base de datos fuente a la base de datos del Data Mart, el cual deberá ser ejecutado primero y por única vez. Luego una vez que se tiene el Data Mart con información histórica ésta debe ser actualizada. Este proceso se conoce como carga periódica y debe ser un proceso con una frecuencia que debe definir el área de informática.

5.1.2.1 Proceso ETL para carga histórica

Este proceso debe cargar todas las dimensiones y hechos de la base de datos del Data Mart desde una fecha razonable en el RENIEC, para que se pueda apreciar la información deseada.

Aspectos a tomar en cuenta:

- El proceso será implementado usando programación descendente, es decir, habrá un trabajo (Job en inglés) principal el cual deberá ser invocado para realizar la carga. El trabajo mencionado llama a dos trabajos, uno dedicado exclusivamente a cargar las tablas dimensiones y el otro la de hechos. Cada uno llamará a su vez a otros y así sucesivamente se implementará las transformaciones de las tablas en particular.
- El proceso cuenta con una base de datos intermedia que tiene una estructura similar a la base de datos del Data Mart. De esta manera no se llega a saturar a la base de datos transaccional durante la ejecución del proceso ETL.

- Dado que se va a cargar la información histórica, el proceso debe ser rápido, pero sobre todo debe permitir monitorear la carga de información. En caso se presente alguna eventualidad se debe poder identificar hasta qué parte de la información pudo ser cargada a la base de datos del Data Mart. La herramienta Pentaho Data Integration ya cuenta con herramientas que permiten hacer el seguimiento de sus procesos.

5.1.2.2 Proceso ETL de carga periódica

Este proceso debe cargar la nueva información de los hechos y en caso se haya agregado información nueva actualizar las dimensiones. Este proceso debe ser repetitivo teniendo en cuenta una frecuencia variable.

Aspectos a tomar en cuenta:

- Su implementación es muy parecida al proceso de carga histórica debido a que se deben cargar dimensiones y hechos. No obstante, este proceso tiene parámetros de entrada que permitan cargar solo la información del día. Estos parámetros están conformados por la fecha inicial (año, mes y día) y fecha final, es decir, el rango de fechas del cual se desea cargar la información, al ejecutarse el proceso se debe indicar estos valores.
- El proceso tiene una base de datos intermedia que tiene una estructura similar a la base de datos del Data Mart. Esto permite realizar un comparativo con la información del Data Mart e insertar o actualizar según sea conveniente.
- En el caso de las dimensiones, el proceso deberá validar si ya existe el registro y sólo agregar los datos en caso no existan. Si bien los datos de las dimensiones podrían borrarse y volverse a cargar en su totalidad. Esto genera un error de llave foránea con los datos ya registrados en los hechos.
- En el caso de los hechos, el proceso deberá borrar los registros del rango de fechas ingresado como parámetro y volverlos a cargar. Esto se realiza para evitar registros duplicados y en caso sea necesario un reproceso.

A continuación se presentan algunos ETL para la carga periódica.

-El proceso de la figura 27 corresponde al proceso ETL que realizó la extracción y carga de la dimensión local del Data Mart. En este caso la Tabla Input utilizó información de la base de datos transaccional, después se seleccionaron los valores a utilizar y en caso se necesitó de alguna transformación de dato se hizo en esta etapa, luego se le asignó una secuencia es decir una Primary Key a la dimensión y finalmente se cargó en la tabla temporal de dimensión local.

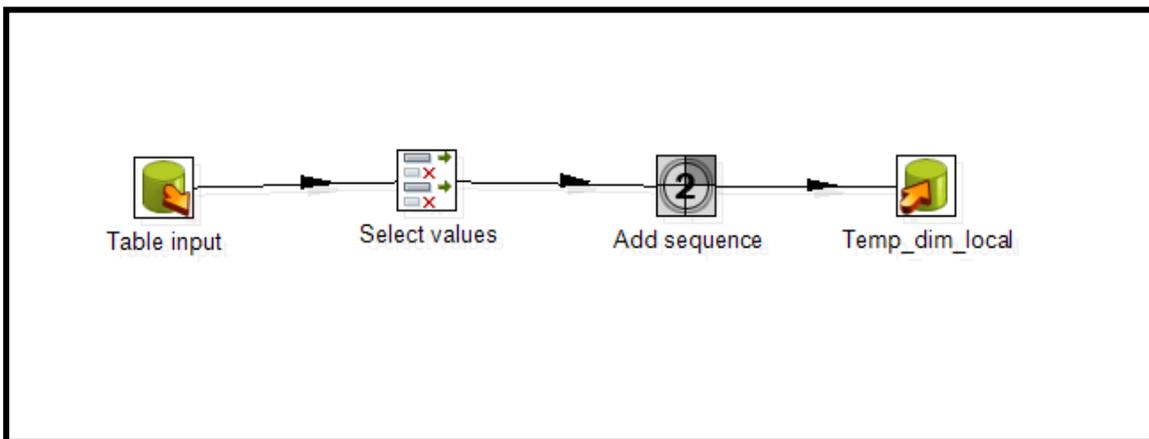


Figura 27. Proceso ETL de una dimensión.

El proceso de la figura 28 corresponde al ETL que realizó la extracción, transformación y carga de la información de los trámites. Se realizó la extracción de la información de diferentes tablas mediante diferentes consultas a la base de datos y transformaciones necesarias para cargarla a la tabla hecho de producción (FACT_PRO_CANT). Primero se utilizó una tabla temporal con todos los valores de los trámites, después con el símbolo de Stream Lookup se comparó cada una de las filas de la tabla temporal con una tabla dimensión, para de esta manera sacar la secuencia de la dimensión y almacenarla temporalmente para cada una de las filas de la tabla temporal, este proceso se realizó con todas las tablas dimensión del Data Mart. Al final del flujo se obtuvieron los valores de los trámites con sus respectivas secuencias, después se seleccionaron los valores necesarios para la tabla Hecho y se insertaron al Data Mart.

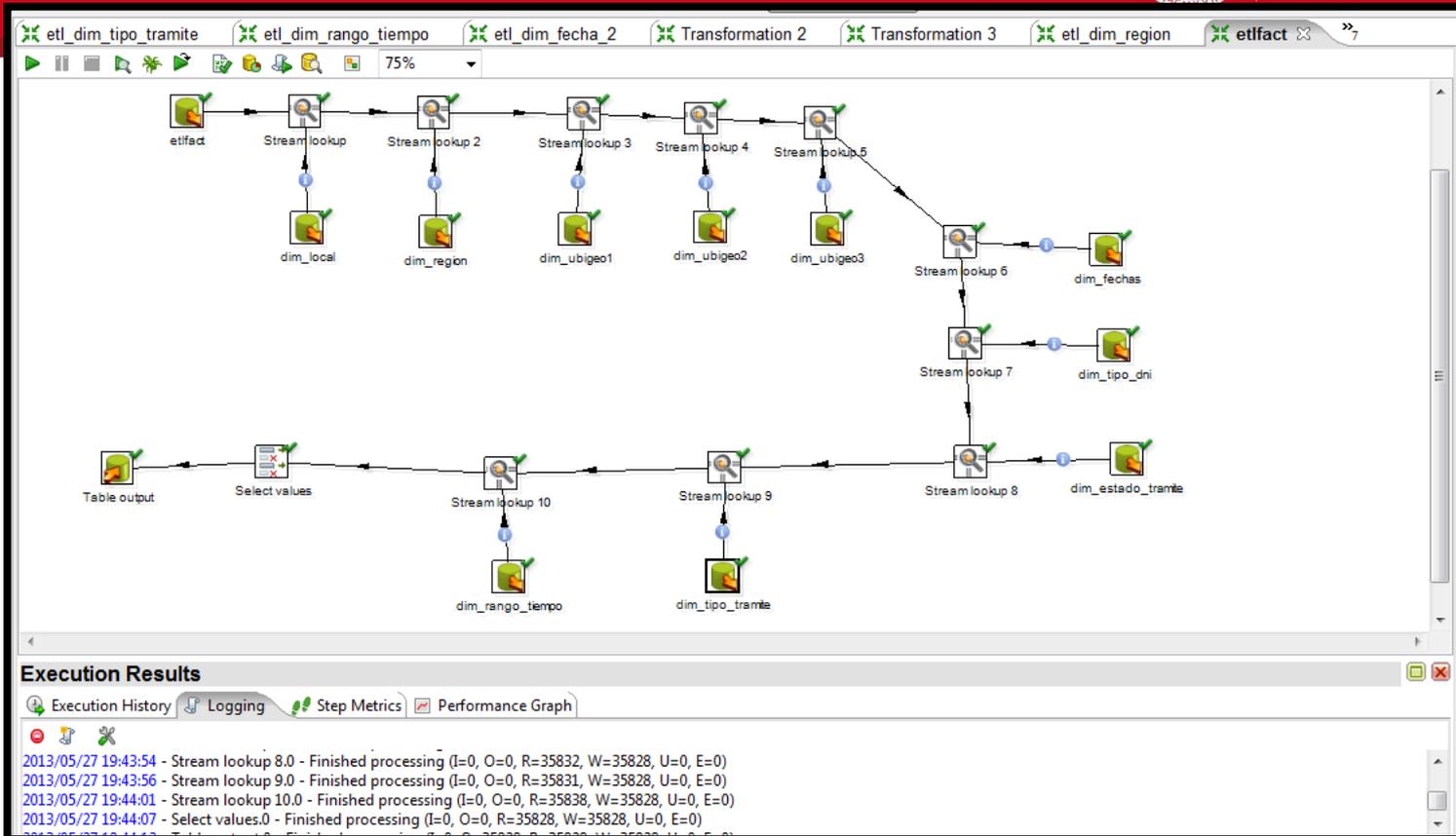


Figura 28. Proceso ETL de una tabla Hecho

El detalle y especificación de los procesos ETL se encuentran en el Anexo 3.

5.1.3 Configuración del Schema Workbench, base de datos del Data Mart y los usuarios

En esta fase se configuraron los usuarios y el administrador, después se configuró la base de datos del Data Mart, y finalmente se relacionó la base de datos del Data Mart con la herramienta Schema Workbench. Todos estos pasos están detallados en el Anexo 4.

5.1.4 Elaboración del cubo

La fase de elaboración del Cubo consiste principalmente en la construcción del cubo con las tablas dimensiones y la tabla hecho. Se utilizó la herramienta Schema Workbench. Se puede apreciar el diseño de los cubos en las figuras 29 y 30.

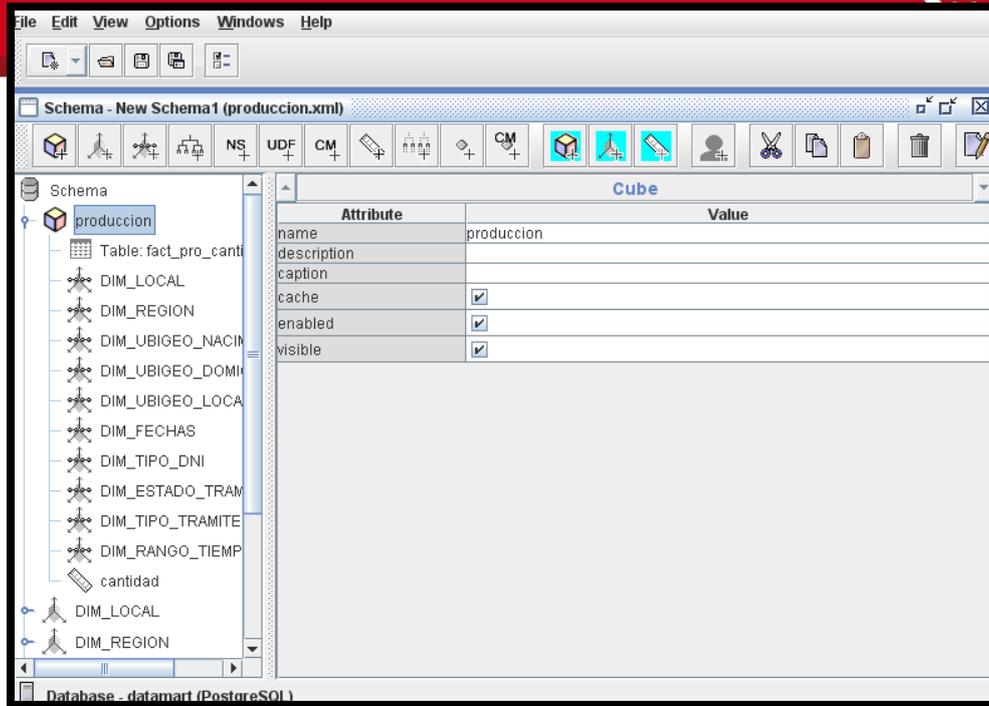


Figura 29. Diseño del cubo

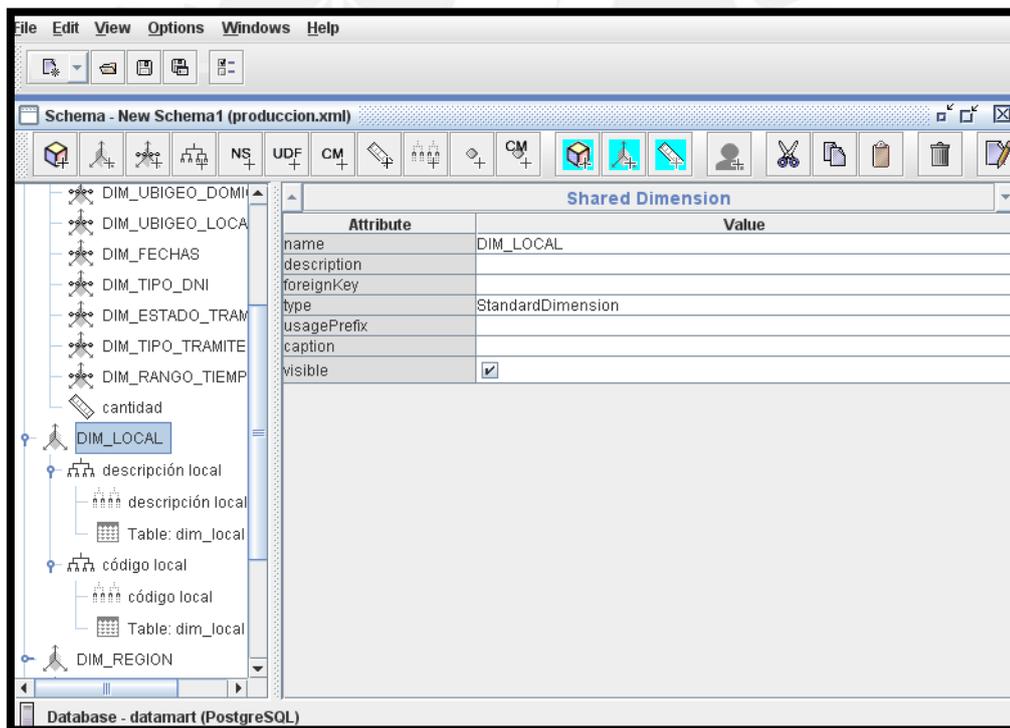


Figura 30. Diseño del cubo

Una vez que se diseñó el cubo, se publicó el archivo para que se pueda apreciar el análisis de los datos. Los pasos para publicar el archivo están detallados en el Anexo 5.

5.1.5 Análisis de los datos

En esta fase fue necesario que el usuario ingrese a la opción de User Console Login en Pentaho. Para elaborar un nuevo análisis de datos el usuario seleccionó la opción de New Analysis.

En la ventana aparecen todos los atributos de las dimensiones y solo arrastrando los atributos a la sección de Layout, se pueden colocar tanto en una columna o una fila. En la sección de Measures se selecciona el atributo principal de la tabla Hecho que en este caso concreto del Data Mart de producción sería Cantidad.

Realizando un click en la fecha derecha de cada campo se puede filtrar las filas, de esta manera se pueden restringir campos que no se deseen en la consulta. De esta manera se pueden realizar una gran cantidad de consultas dinámicamente.

En la figura 31 se puede observar el esquema de análisis, que se obtuvo al explotar la información del Data Mart de Producción. En éste se aprecia la descripción del tipo de DNI, la región de inscripción del trámite, la descripción de los locales involucrados, el estado del tiempo de duración en que se encuentran esos trámites y cuantos trámites se han realizado en cada una de las etapas de la generación y emisión del DNI. En el Anexo 6 se puede observar todos los esquemas de análisis, y los gráficos para los Data Mart's de Producción y Calidad.

The screenshot displays the Pentaho Business Intelligence (BI) interface. At the top, the Pentaho logo is visible. The main area shows a data table with the following columns: estado, descripción tipo de DNI, descripción región, descripción local, trámite, envío cantidad, impresión cantidad, ingreso cantidad, and recepción cantidad. The table contains 24 rows of data, with the first row showing 'CASOS SOCIALES' and a 'recepción cantidad' of 91. The interface also includes a sidebar with various tools and a top bar with the text 'Rows: 247 out of 370 Cols: 4'.

estado	descripción tipo de DNI	descripción región	descripción local	trámite	envío cantidad	impresión cantidad	ingreso cantidad	recepción cantidad
			CASOS SOCIALES		-	-	42	-
			CIENEGUILLA		-	-	105	91
			COMAS		-	-	756	703
			COMAS III		-	-	319	305
			Comas II		-	-	194	162
			GRIAS-LIMA		-	-	457	408
			HUACHO		-	-	492	138
			HUARAL		-	-	344	143
			INDEPENDENCIA		-	-	1.423	1.268
			JESUS MARIA		-	-	1.846	1.720
			LIMA		22.889	39.969	23.321	3.268
			LURIGANCHO(CHOSICA)		-	-	327	287
			MAC CALLAO		-	-	303	294
			MAYOR		-	-	950	241

Figura 31. Análisis de la información

La información también se puede explotar en forma de diferentes gráficos como se puede apreciar en las figuras 33, 34 y 35.

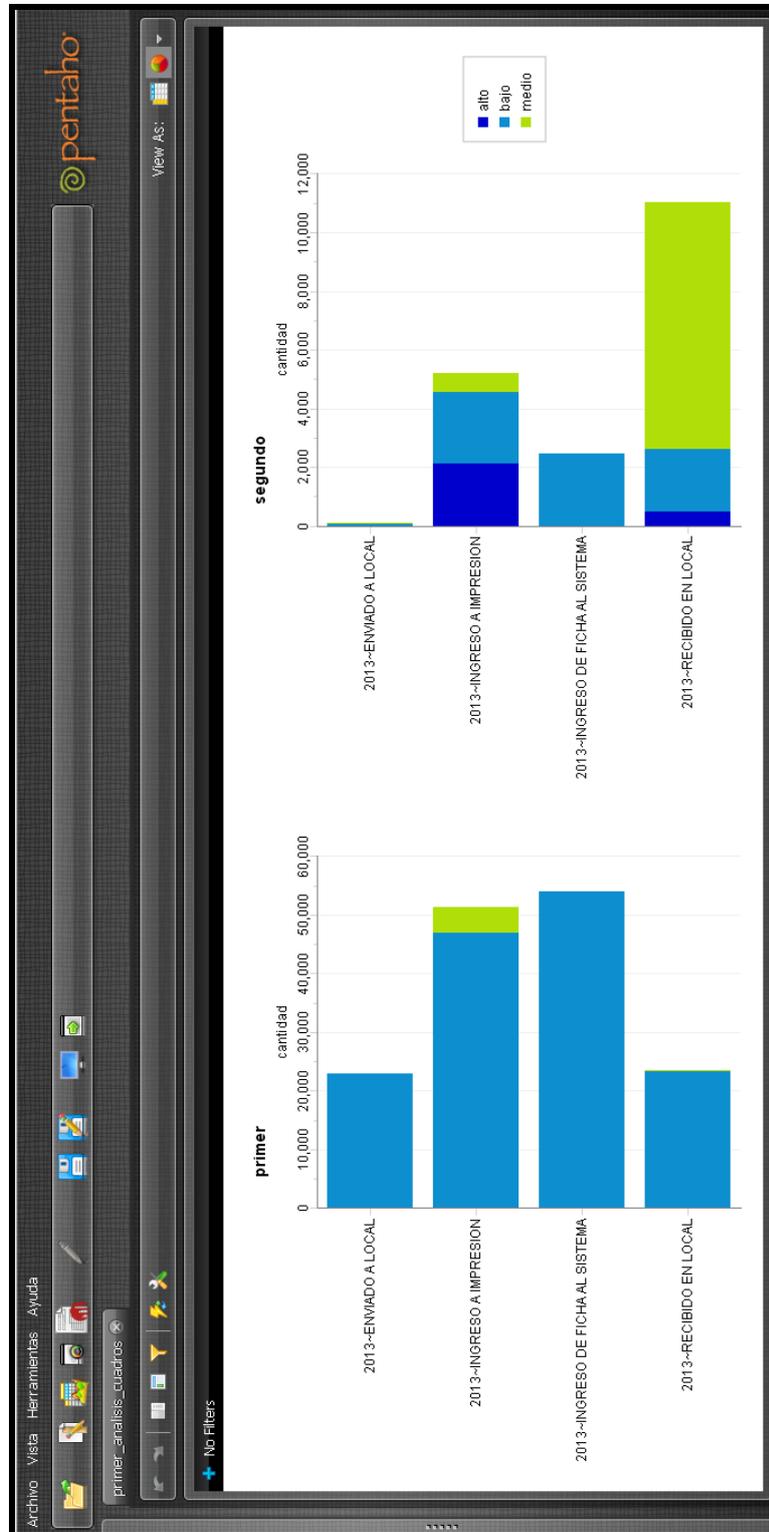


Figura 33. Análisis de la información

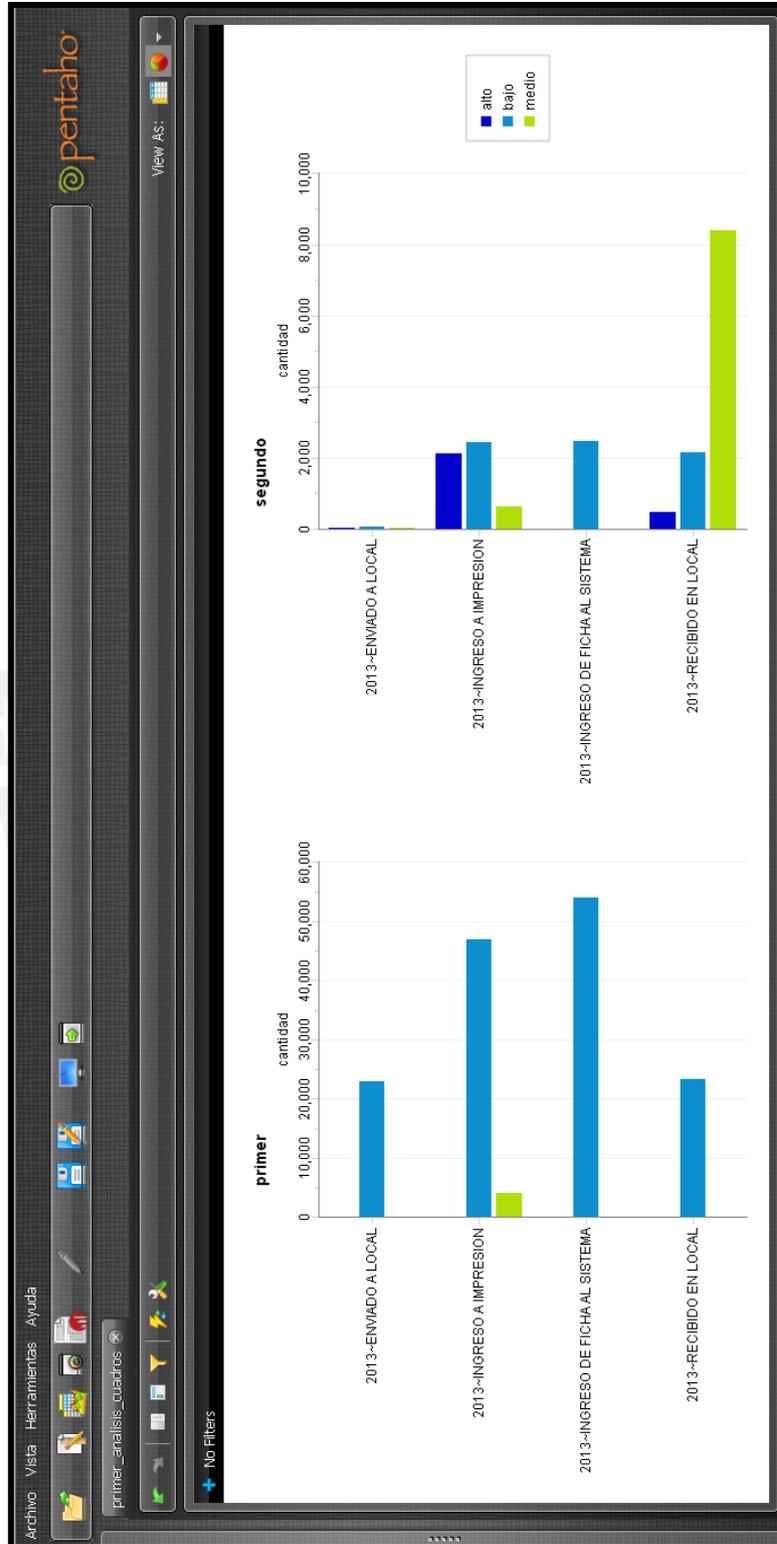


Figura 34. Análisis de la información

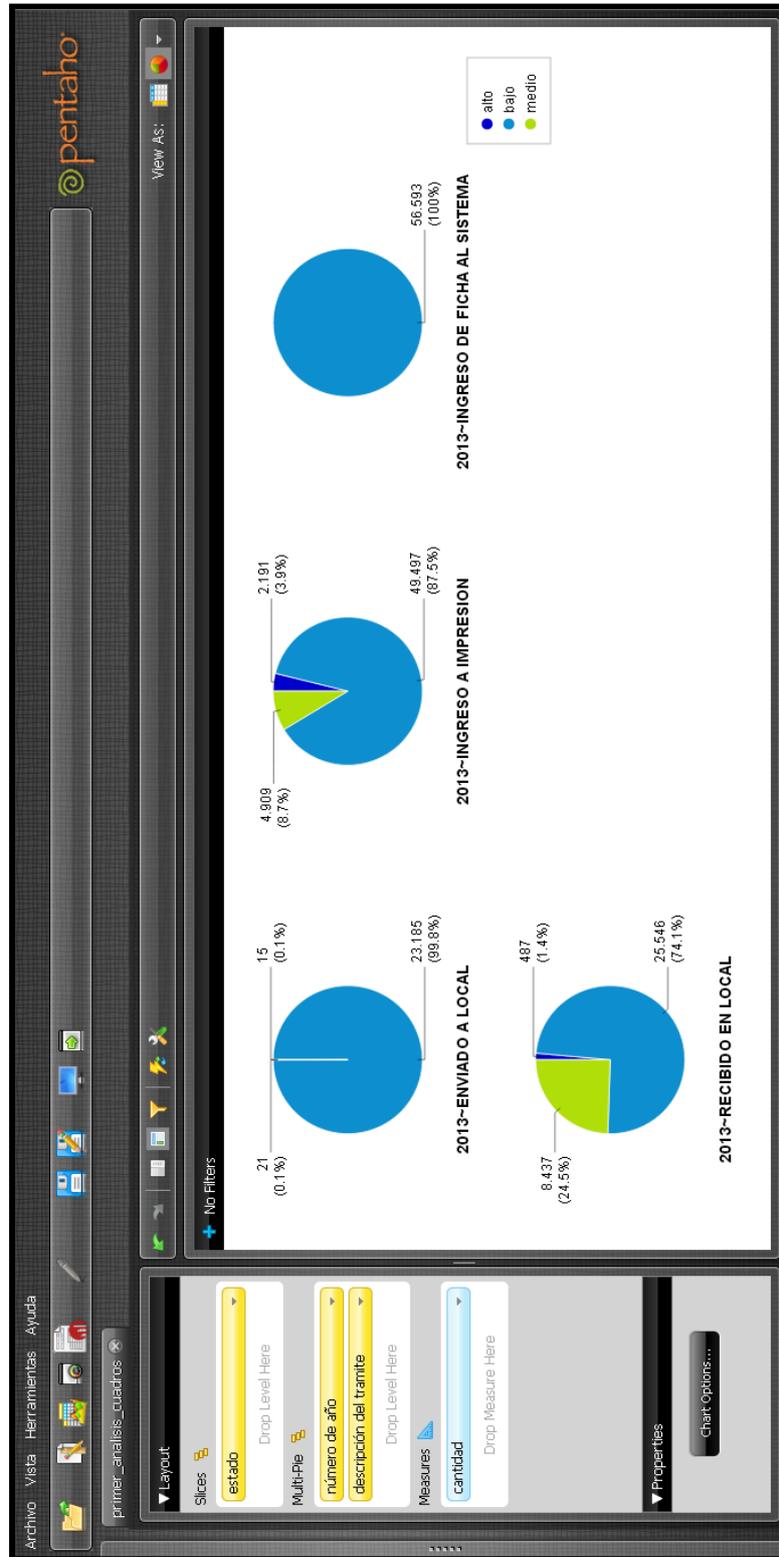


Figura 35. Análisis de la información

Adicionalmente se explotó la data del Data Mart de Producción mediante los clásicos Reportes. Para elaborar los reportes se seleccionó el Report Designer de Pentaho. En este programa se configuró la base de datos a utilizar, y se elaboraron los las consultas a la base de datos para la información requerida.

En la figura 36 se puede observar un reporte general elaborado por Report Designer.

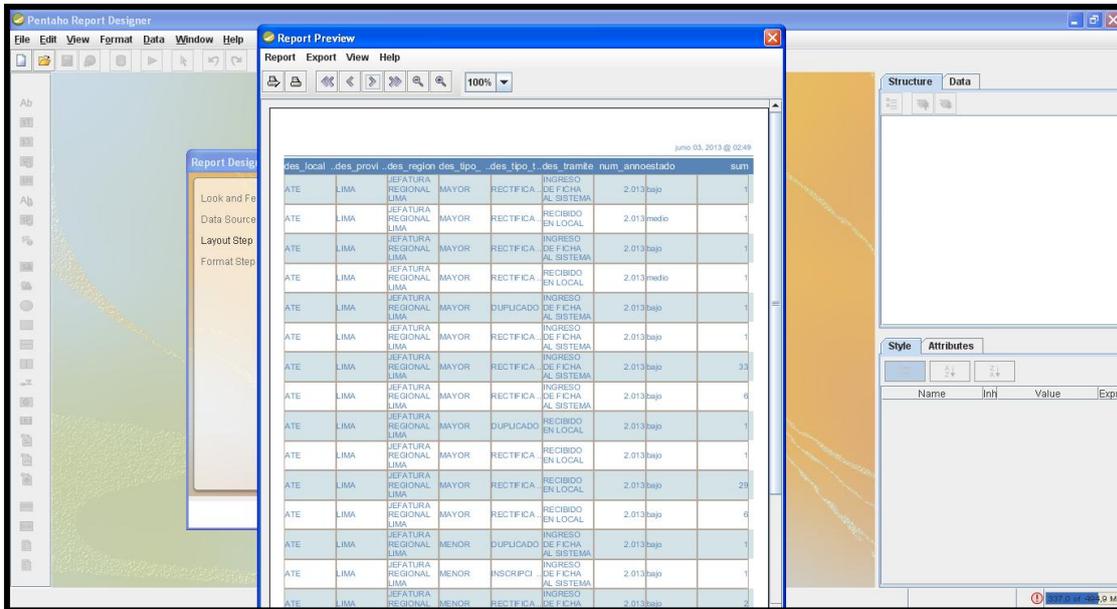


Figura 36. Reporte General

En la figura 37 se puede apreciar que también es posible darle formato al reporte con la herramienta de Pentaho Report Designer

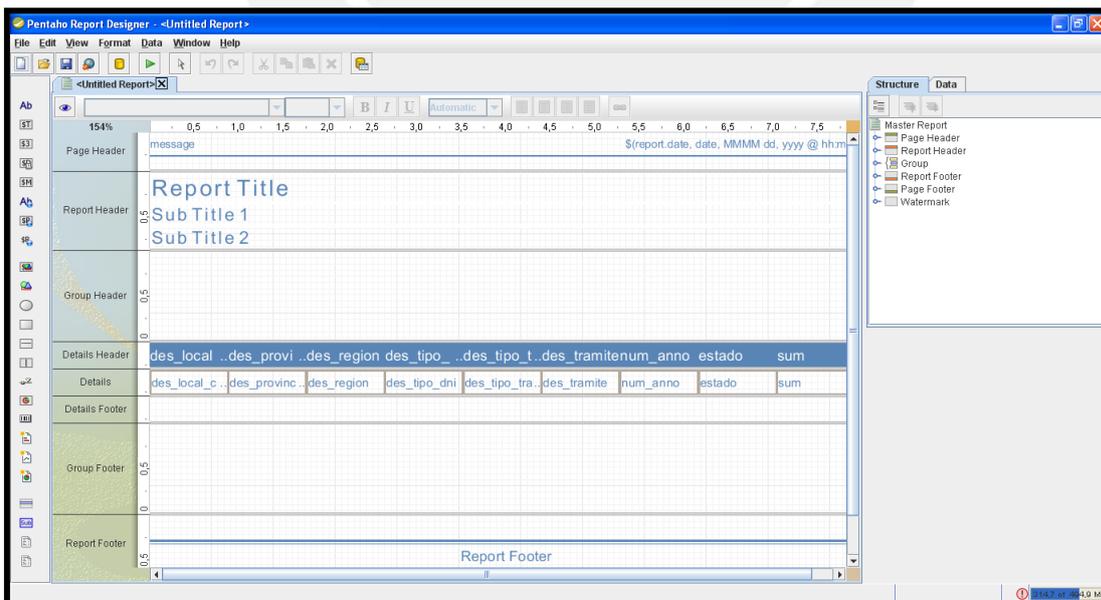


Figura 37. Reporte en Report Designer.

5.2 Pruebas

Dentro del proceso de la construcción de la solución de Inteligencia de Negocios, las pruebas constituyen una de las actividades indispensables debido a que a través de ellas se asegura la calidad del producto. Por ello es necesario definir adecuadamente una metodología de plan de pruebas que permita verificar los componentes de la solución y así asegurar que se ha construido un producto que cumpla con los requerimientos planteados.

La metodología de plan de pruebas está compuesta por:

- Planificación de las pruebas.
- Diseño del plan de pruebas.
- Determinación de los casos de prueba.
- Ejecución del plan de pruebas.
- Análisis y evaluación del plan de pruebas ejecutado.

A continuación se detallara cada paso de la metodología definida.

5.2.1 *Planificación del plan de pruebas.*

El primer paso es definir el objetivo del plan de pruebas. Este objetivo es diferente de los planes de prueba tradicionales elaborados para los proyectos de desarrollo de software debido a que el objetivo del presente plan está relacionado con la calidad la información que se procesa y entrega al usuario. Por ello, el objetivo está dado por asegurar la calidad de los datos del Data Mart cubriendo específicamente los procesos ETL. Significa poder verificar que la información que se encuentra en los sistemas fuentes es igual a la información mostrada a través de la explotación.

Para llevar a cabo el objetivo del plan se deben identificar los objetos sobre los que se ejecutaran el plan de pruebas. Los objetos identificados son:

- Los procesos ETL, desde la base de datos fuente hasta la base de datos del Data Mart.
- La base de datos del Data Mart.

Sobre estos objetos se realizará la siguiente prueba empleando el método de caja negra.

-Prueba de Extracción, Transformación y Carga desde la base de datos fuente a la base de datos del Data Mart.

Una vez ejecutado el plan de pruebas se registrará los resultados en un informe el cual es definido en el Diseño del plan de pruebas.

5.2.2 Diseño del plan de pruebas.

El plan tendrá la siguiente estructura: se definirán los casos de prueba los cuales se ejecutarán sobre los objetos de prueba. Se registrará la salida obtenida y se comparará dicho resultado con la salida esperada para dicha entrada. De esta manera se determinará si cada objeto cumple con los requisitos de calidad.

5.2.3 Determinación de los casos de prueba

Se ha elaborado una plantilla que sirve para especificar los casos de prueba. Esta plantilla tendrá a grandes rasgos, la siguiente información.

- Nombre del caso de prueba
- Objeto a analizar en el caso de prueba.
- Objeto del caso de prueba
- Datos de entrada del caso de prueba.
- Resultado y salida del caso de prueba.

En la tabla 9 muestra un ejemplo de caso de prueba, en el Anexo 9 se encuentran detallados los casos de prueba.

Caso de prueba Nro. 1
Objeto a analizar del caso de prueba:
Proceso ETL y Base de datos del Data Mart.
Objetivo del caso de prueba:
Verificar que la tabla FACT_PRO_CANTIDAD sea correctamente cargada por medio del ETL teniendo como fuente la tabla auxiliar.
Pre-requisitos del caso de prueba:
-Disponibilidad de la BD Fuente -BD intermedia creada -BD Data Mart creada -El Job ELT_FACT_1 debe estar creado y listo para usarse
Parámetros requeridos:
Datos de entrada:
Tabla Mov_ficha_registral Tabla Lote_impresion Tabla Region Tabla Despacho_dni Tabla fuente Intermedia

Nro.	Función a probar	Acción	Resultado Esperado	Resultado Obtenido
1	Cantidad de registros por la tabla en la BD Fuente frente a la cantidad de registros por tabla en BD Data Mart	Ejecutar el Job ETL_FACT_1	La tabla FACT_PRO_CANTIDAD muestre valores consistentes y adecuados.	Se cargó los datos correctamente.

Tabla 9. Ejemplo de caso de prueba.

5.2.4 Ejecución del plan de pruebas

Una vez definidos todos los casos de prueba, se debe empezar con la ejecución de las mismas. El probador analiza el caso de pruebas siguiendo las instrucciones paso a paso indicadas en la especificación del caso de prueba. En caso se produzca un resultado incorrecto o inesperado se deben registrar en la parte de Resultad Obtenido.

5.2.5 Análisis y evaluación del plan de pruebas ejecutado.

Al culminar la ejecución de todos los casos de prueba es posible determinar si los objetos de prueba cumplen con los requerimientos de aceptación y por ende si el producto final cumple con los requerimientos. En caso hayan surgido casos de prueba con resultados incorrectos estos deben ser corregidos y atendidos hasta obtener los resultados esperados en todos los casos.

5.2.6 Caso de Prueba para el ETL_FACT_1 de FACT_PRO_CANTIDAD

Datos de entrada:

Tabla Mov_ficha_registral: Se listan las columnas significativas para el proceso y una porción del número de registros.

	tip_doc_reg_ficha character(1)	cod_continente_naci character varying(2)	cod_continente_domicilio character varying(2)	cod_pais_naci character varying(2)	cod_pais_domicilio character varying(2)	cod_departamento_naci character varying(2)	cod_departamento_domicilio character varying(2)	cod_provincia_naci character varying(2)	cod_provincia_domicilio character varying(2)	cod_distrito_ character va
1	E	92	92	02	02	02	02	00	00	00
2	E	92	92	02	33	02	14	00	01	00
3	E	92	92	02	33	02	14	00	01	00
4	E	92	92	02	33	02	14	00	01	00
5	E	92	92	02	33	02	14	00	01	00
6	E	92	92	02	33	02	14	00	01	00
7	E	92	92	02	33	02	14	00	01	00
8	E	92	92	02	33	02	14	00	01	00
9	E	92	92	02	33	02	14	00	01	00
10	E	92	92	02	33	02	14	00	01	00
11	E	92	92	02	33	02	14	00	01	00
12	E	92	92	02	33	02	14	00	04	00
13	E	92	92	02	33	02	14	00	05	00
14	E	92	92	02	33	02	14	00	09	00

Figura 38. Tabla Mov_ficha_registral

cod_distrito_naci character varying(2)	cod_distrito_domicilio character varying(2)	cod_centro_poblado_naci character varying(3)	cod_centro_poblado_domicilio character varying(3)	est_ficha_reg character(1)	cod_tipo_tramite character(1)	fec_crea_audi timestamp without time zone
00	00	000	000	9	I	2013-04-02 00:00:00
00	05	000	000	9	R	2013-03-19 00:00:00
00	06	000	000	9	I	2013-03-21 00:00:00
00	12	000	000	9	I	2013-03-21 00:00:00
00	19	000	000	9	I	2013-03-21 00:00:00
00	19	000	000	9	R	2013-03-21 00:00:00
00	27	000	000	9	I	2013-03-21 00:00:00
00	35	000	000	9	R	2013-03-19 00:00:00
00	37	000	000	9	I	2013-03-19 00:00:00
00	37	000	000	9	I	2013-03-21 00:00:00
00	43	000	000	9	I	2013-03-20 00:00:00
00	01	000	000	9	I	2013-03-18 00:00:00
00	06	000	000	9	I	2013-03-19 00:00:00
00	01	000	000	9	I	2013-03-21 00:00:00

Figura 39. Tabla Mov_ficha_registral

Tabla Lote_impresion: Se listan las columnas significativas para el proceso y una porción del número de registros.

	num_lote numeric	fec_ini_impr timestamp without time zone	fec_fin_impr timestamp without time zone
1	152525	2013-03-26 00:00:00	2013-03-26 00:00:00
2	153103	2013-04-04 00:00:00	2013-04-04 00:00:00
3	152400	2013-03-25 00:00:00	2013-03-25 00:00:00
4	152241	2013-03-22 00:00:00	2013-03-22 00:00:00
5	153442	2013-04-08 00:00:00	2013-04-09 00:00:00
6	153200	2013-04-05 00:00:00	2013-04-05 00:00:00
7	152097	2013-03-21 00:00:00	2013-03-21 00:00:00
8	152211	2013-03-22 00:00:00	2013-03-22 00:00:00
9	152328	2013-03-23 00:00:00	2013-03-23 00:00:00
10	153368	2013-04-08 00:00:00	2013-04-08 00:00:00
11	153829	2013-04-12 00:00:00	2013-04-12 00:00:00
12	152050	2013-03-20 00:00:00	2013-03-20 00:00:00
13	153536	2013-04-09 00:00:00	2013-04-10 00:00:00
14	152645	2013-03-28 00:00:00	2013-03-28 00:00:00
15	152452	2013-03-26 00:00:00	2013-03-26 00:00:00

Figura 40. Tabla Lote_impresion

Tabla Región: Se listan las columnas significativas para el proceso y una porción del número de registros.

	cod_region character varying(2)	des_region character varying(100)
5	16	PAR
6	13	JEFATURA REGIONAL PUCALLPA
7	12	JEFATURA REGIONAL ICA
8	0	GENERICO
9	1	JEFATURA REGIONAL PIURA
10	2	JEFATURA REGIONAL TRUJILLO
11	3	JEFATURA REGIONAL TARAPOTO
12	4	JEFATURA REGIONAL IQUITOS
13	5	JEFATURA REGIONAL CHIMBOTE
14	6	JEFATURA REGIONAL HUANCAYO
15	7	JEFATURA REGIONAL AYACUCHO
16	8	JEFATURA REGIONAL AREQUIPA
17	9	JEFATURA REGIONAL CUSCO
18	10	JEFATURA REGIONAL LIMA
19	11	JEFATURA REGIONAL PUNO
20	14	JEFATURA REGIONAL HUANCAVELICA

Figura 41. Tabla Región

Tabla Despacho_dni: Se listan las columnas significativas para el proceso y una porción del número de registros.

	cod_proceso character(1)	num_anio character varying(4)	num_envi character varying(10)	cod_area_ori character varying(5)	num_lote numeric	num_dni character varying(8)	fec_rece timestamp without time zone	est_dni_desp character(1)	cod_local_ori character varying(6)	num_ficha_reg character varying(8)
1	D	2013	16432	64019	151853	10077768	2013-03-19 00:00:00	4	000371	96201511
2	D	2013	16432	64019	151853	43629877	2013-03-19 00:00:00	4	000371	96201541
3	D	2013	16432	64019	151853	44599277	2013-03-19 00:00:00	4	000371	96201549
4	D	2013	16432	64019	151853	44782335	2013-03-19 00:00:00	4	000371	96201391
5	D	2013	16432	64019	151853	45261209	2013-03-19 00:00:00	4	000371	96201501
6	D	2013	16432	64019	151853	46421820	2013-03-19 00:00:00	4	000371	96201534
7	D	2013	16432	64019	151853	46728792	2013-03-19 00:00:00	4	000371	96201522
8	D	2013	16432	64019	151853	70685746	2013-03-19 00:00:00	4	000371	96201413
9	D	2013	16432	64019	151853	73311838	2013-03-19 00:00:00	4	000371	96201397
10	D	2013	16432	64019	151853	9001403	2013-03-19 00:00:00	4	000371	96201489
11	D	2013	16437	64019	151853	40687164	2013-03-19 00:00:00	4	000371	96201550
12	D	2013	16437	64019	151853	43490040	2013-03-19 00:00:00	4	000371	96201427
13	D	2013	16437	64019	151853	45304377	2013-03-19 00:00:00	2	000371	96201503
14	D	2013	16439	64019	151853	25554899	2013-03-19 00:00:00	2	000371	96201490
15	D	2013	16439	64019	151853	25557262	2013-03-19 00:00:00	4	000371	96201548

Figura 42. Tabla Despacho_dni

Resultado Intermedio:

Tabla fuente Intermedia: Se listan todas las columnas significativas para el proceso de ETL y una porción del número de registros.

	tip_doc_reg_ficha character(1)	cod_continente_naci character varying(2)	cod_continente_domicilio character varying(2)	cod_pais_naci character varying(2)	cod_pais_domicilio character varying(2)	cod_departamento_naci character varying(2)	cod_departamento_domicilio character varying(2)	cod_provincia_naci character varying(2)	cod_provincia_domicilio character varying(2)
1	E	92	92	33	33	14	14	01	01
2	F	92	92	33	33	14	14	01	01
3	F	92	92	33	10	14	06	05	00
4	F	92	92	33	33	14	14	01	01
5	F	92	92	33	33	14	14	01	08
6	E	92	92	33	33	14	14	04	04
7	F	92	92	33	33	13	13	01	01
8	F	92	92	33	33	09	14	07	01
9	F	92	92	33	33	07	04	07	01
10	F	92	92	33	33	14	14	01	01
11	F	92	92	33	33	20	17	01	01
12	F	92	92	33	33	14	14	09	01
13	F	92	92	33	33	02	14	07	01
14	F	92	92	33	33	14	14	01	01

Figura 43. Tabla fuente Intermedia 1

	cod_distrito_naci character varying(2)	cod_distrito_domicilio character varying(2)	cod_centro_poblado_naci character varying(3)	cod_centro_poblado_domicilio character varying(3)	cod_region character varying(2)	cod_local_ori character varying(6)	cod_local_continente character varying(2)	cod_local_pais character varying(2)	cod_local_departamento character varying(2)
1	22	26	000	000	10	000371	92	33	14
2	09	35	000	000	10	000371	92	33	14
3	13	00	000	000	10	000371	92	33	14
4	13	30	000	000	10	000371	92	33	14
5	22	01	000	000	10	000371	92	33	14
6	01	05	000	000	10	000371	92	33	14
7	16	01	000	000	10	000371	92	33	14
8	02	37	000	000	10	000371	92	33	14
9	06	28	000	000	10	000371	92	33	14
10	01	04	000	000	10	000371	92	33	14
11	01	01	000	000	10	000371	92	33	14
12	01	04	000	000	10	000371	92	33	14
13	01	06	000	000	10	000371	92	33	14
14	33	30	000	000	10	000371	92	33	14

Figura 44. Tabla fuente Intermedia 2

cod_local_provincia character varying(2)	cod_local_distrito character varying(2)	cod_local_centro_poblado character varying(3)	est_ficha_reg character(1)	cod_tipo_tramite character(1)	fec_crea_audi timestamp without time zone	fec_ini_impr timestamp without time zone	dias interval	cantidad integer	rango integer	cantidad integer
01	01	000	9	R	2013-03-20 00:00:00	2013-03-27 00:00:00	7 days	1	12	1
01	01	000	9	R	2013-03-18 00:00:00	2013-03-21 00:00:00	3 days	2	11	1
01	01	000	9	D	2013-03-18 00:00:00	2013-03-18 00:00:00	00:00:00	1	11	1
01	01	000	9	R	2013-03-18 00:00:00	2013-03-20 00:00:00	2 days	1	11	1
01	01	000	9	R	2013-03-18 00:00:00	2013-03-25 00:00:00	7 days	1	12	1
01	01	000	9	R	2013-03-18 00:00:00	2013-03-20 00:00:00	2 days	1	11	1
01	01	000	9	D	2013-03-21 00:00:00	2013-03-20 00:00:00	00:00:00	1	11	1
01	01	000	9	D	2013-03-22 00:00:00	2013-03-22 00:00:00	00:00:00	1	11	1
01	01	000	9	D	2013-03-18 00:00:00	2013-03-18 00:00:00	00:00:00	1	11	1
01	01	000	9	D	2013-03-21 00:00:00	2013-03-21 00:00:00	1 day	2	11	1
01	01	000	9	D	2013-03-20 00:00:00	2013-03-21 00:00:00	1 day	1	11	1
01	01	000	9	R	2013-03-18 00:00:00	2013-03-20 00:00:00	2 days	1	11	1
01	01	000	9	R	2013-03-18 00:00:00	2013-03-21 00:00:00	3 days	1	11	1
01	01	000	9	D	2013-03-19 00:00:00	2013-03-20 00:00:00	1 day	5	11	1

Figura 45. Tabla fuente Intermedia 3

ETL Intermedio:

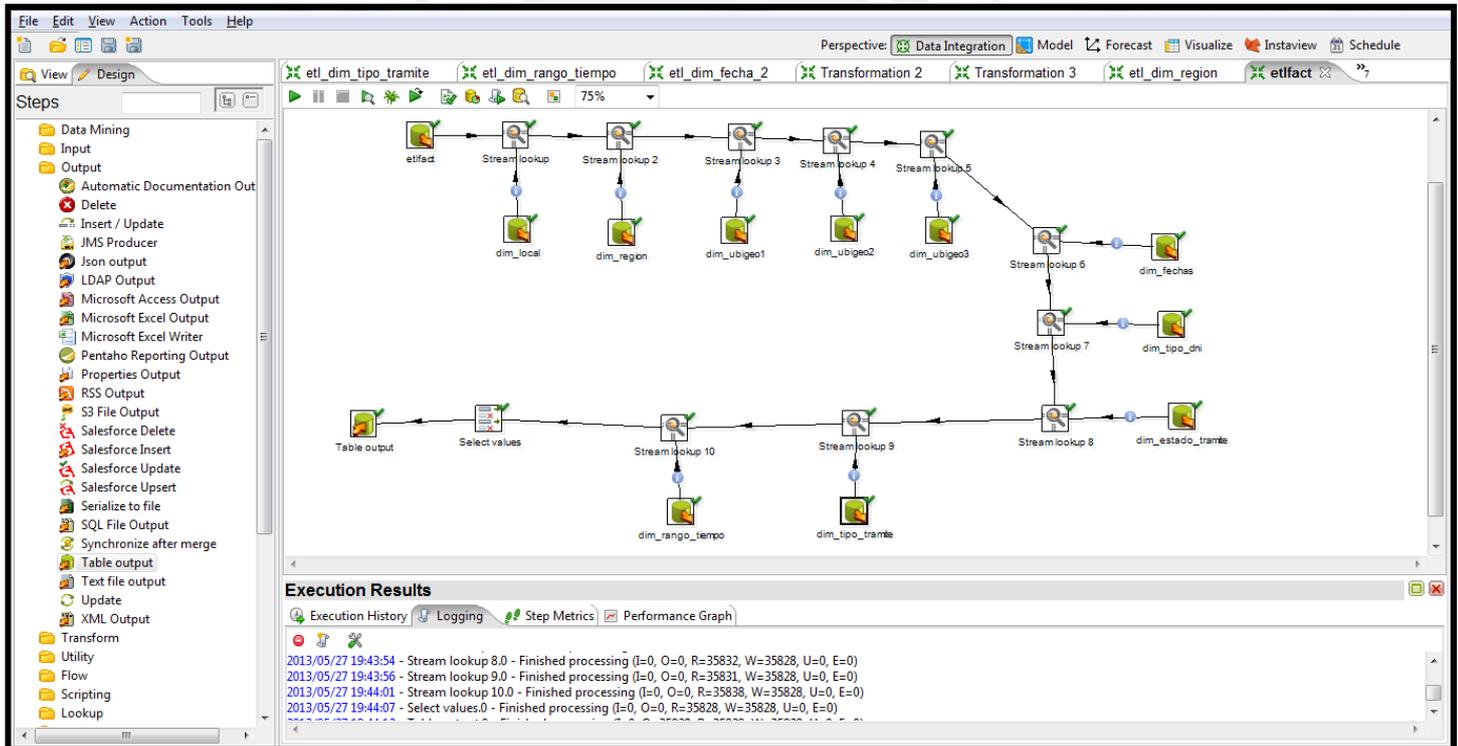


Figura 46. ETL Intermedio

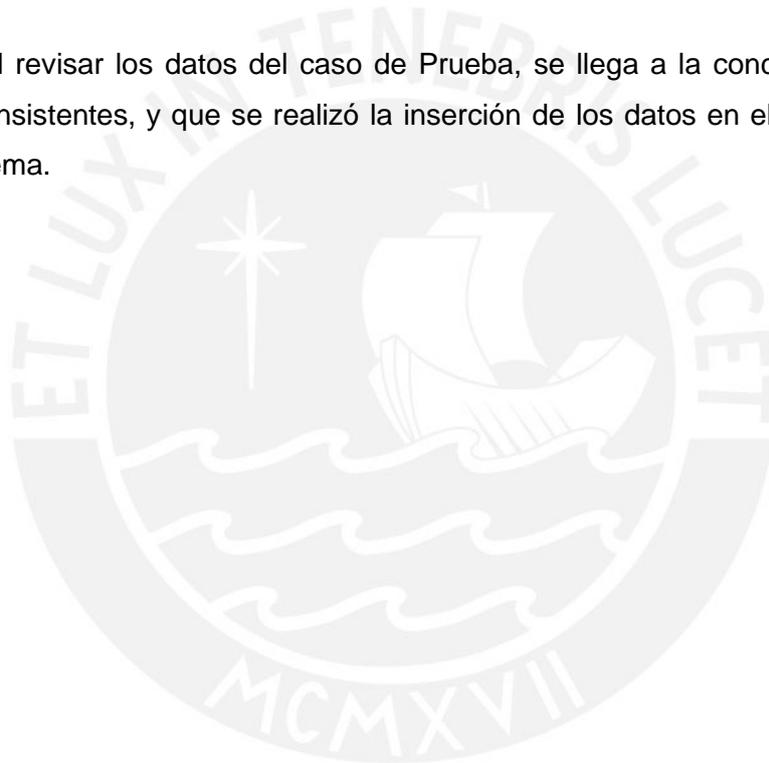
Resultado Final

Tabla Final de FACT_PRO_CANTIDAD

	cantidad integer	sec_local integer	sec_tipo_dni integer	sec_estado_tramite integer	sec_tipo_tramite integer	sec_rango_tiempo integer	sec_ubigeo_nacimiento integer	sec_ubigeo_domicilio integer	sec_fecha_estado integer	sec_region_local integer	sec_ubigeo_local integer
1	2	9311	2	1	3	1	2442	2442	16	22	2442
2	2	10457	2	1	2	1	2442	2264	2	18	2265
3	1	11867	2	1	3	1	2442	2265	4	18	2265
4	1	5298	2	1	3	1	2442	2271	4	18	2271
5	1	4225	2	1	3	1	2442	2277	4	18	2277
6	1	4225	2	1	2	1	2442	2277	4	18	2277
7	1	1326	2	1	3	1	2442	2284	4	18	2389
8	1	6685	2	1	2	1	2442	2391	2	18	589
9	1	6685	2	1	3	1	2442	2393	2	18	589
10	1	3310	2	1	3	1	2442	2393	2	18	2393
11	1	3310	2	1	3	1	2442	2393	4	18	2393
12	1	9508	2	1	3	1	2442	2399	3	18	2399

Figura 47. Tabla Final

Finalmente al revisar los datos del caso de Prueba, se llega a la conclusión que los datos son consistentes, y que se realizó la inserción de los datos en el Data Mart sin ningún problema.





6 Observaciones, conclusiones y trabajos Futuros

6.1 Observaciones

-Los procesos de ETL fueron implementados mediante procesos de Data Integration y para poder ejecutar estos procesos se debe ingresar a la interfaz de ejecución de la herramienta de inteligencia de negocios.

-Las herramientas de Pentaho que se hacen mención en todo el documento corresponden a la versión Pentaho Enterprise Edition, es decir son herramientas elaboradas por el grupo desarrollador de la comunidad Pentaho, esta edición requiere licencia. Por otro lado Pentaho ofrece una versión que no requiere licencia llamada Pentaho Community. Se utilizó la versión Enterprise Edition porque ofrece algunas funcionalidades adicionales.

6.2 Conclusiones

Luego de haber culminado el proyecto se puede llegar a las siguientes conclusiones:

-La metodología BPMN, logró mostrar una visión muy detallada y clara de los procesos de generación y emisión del DNI.

-Las necesidades de información de las gerencias y del área de estadística sobre los procesos de generación y emisión del DNI, fueron identificadas satisfactoriamente debido a que se consideró los indicadores con mayor demanda entre las gerencias. Esto contribuyó a identificar requerimientos claros y precisos que fueron documentados y utilizados para la construcción del modelo multidimensional.

-El modelo multidimensional de la solución de Inteligencia de Negocios logró abarcar las necesidades de información identificadas y fue presentado utilizando diagramas de fácil comprensión que permitieron una correcta validación del mismo.

-Los procesos de extracción, transformación y carga de los datos lograron contar con Data Mart's de datos correctos y coherentes provenientes de la base de datos transaccional.

-Los esquemas de análisis, gráficos de barra, reportes elaborados permitieron mostrar los indicadores que brindan información relevante a los usuarios respectivos. Y la forma dinámica en que se pueden ir modificando ayuda mucho al análisis de los datos. En el Anexo 10 se puede apreciar la Constancia de Aceptación por parte de RENIEC de la información brindada.

-La metodología de Kimball logró exitosamente la construcción de la solución. Debido a que la elaboración de los Data Mart's se realizó para áreas específicas.

-La elección de la herramienta de explotación fue la adecuada debido a que permitió una fácil interacción con los usuarios.

-La selección de la suite Pentaho como plataforma de la solución fue la adecuada porque a través de sus herramientas se pudo abarcar las diferentes etapas de una solución de Inteligencia de Negocios.

6.3 Recomendaciones y Trabajos Futuros

El presente proyecto puede tomarse como base para futuros proyectos o soluciones similares.

6.3.1 *Implementar herramientas de Inteligencia de Negocios*

Se pueden usar diferentes herramientas de inteligencia de negocios que se incorporan al Data Mart elaborado. El objetivo principal de estas herramientas es brindar medios más efectivos y fáciles de utilizar para la generación de información útil de manera rápida, integrada y con la seguridad de contar con información consistente. A continuación se listan algunas herramientas a integrarse con el Data Mart.

-Minería de datos (Data Mining): Herramienta que explora las diferentes bases de datos en busca de tendencias, patrones ocultos y comportamientos encontrando predecibles que ni un experto puede llegar a encontrar fácilmente.

-Mantenimiento de Metadatos: Herramienta que ayudan a mantener la información y documentación acerca de las estructuras de datos de los sistemas transaccionales, procesos ETL y Data Mart's. Permite tener un repositorio de los datos acerca de los propios datos y en caso exista algún cambio en las bases de datos puede mostrar los impactos que implica dicho cambio.

6.3.2 *Ampliación a áreas y departamentos*

Se puede desarrollar Data Mart's para las otras áreas de la empresa e ir construyendo un Data Warehouse Corporativo. De esta manera se pueden abarcar otras gerencias de RENIEC, así como se puede abarcar otros procesos.

En caso se desee incorporar esta ampliación, se debe tener en cuenta la integración de todos los Data Mart's a construir, para no tener Data Mart's independientes y evitar la redundancia de información. Para estos casos se recomienda elaborar un solo diseño integral del Data Warehouse e ir construyendo Data Mart's según las necesidades de los usuarios, o en su defecto buscar integración como criterio principal entre los Data Mart's que se vayan desarrollando.

7 Referencias

- Moss, Larissa Terpeluk ; Atre, Shaku.
2003 "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications"
Addison Wesley, Boston
- Kimball, Ralph.
2002 "The Data Warehouse toolkit: the complete guide to dimensional modeling."
Willey, New York
- Harjinder & Prakash.
1996 "Data Warehousing, la integración de información para la mejor toma de decisiones"
Prentice Hall, México
- Russel Pears, Muhammad Usman
2010 "Integration of Data Mining and Data Warehousing: A Practical Methodology".
School of Computing and Mathematical Sciences, Auckland University of Technology, New Zealand
- Reeves, Laura
2009 "A Manager's Guide to Data warehousing"
Wiley Publishing, Indianapolis, Indiana
- Ponniah, Paulraj
2010 "Data warehousing fundamentals for IT professionals"
John Wiley & Sons, Hoboken, New Jersey
- Medina La Plata, Edison
2012 Curso en la Escuela de Postgrado de la UPC
"Gestión de la Información con Business Intelligence"

Few, Steven

2006

“Information Dashboard Design
The Effective Visual Communication of Data”
O'Reilly Media, Sebastopol, California

✓ Tesis

Núñez Soto, Grace Isabel

2010

Análisis, diseño e implementación de una solución de
inteligencia de negocios para el área de finanzas de la
Municipalidad Metropolitana de Lima
Tesina, Ingeniería Informática, Licenciatura, PUCP

Gamarra Ramirez, Alejandro Javier

2010

Solución Integral para explotar eficientemente la Información de
los contactos con los clientes utilizando Datamart en Telefónica
del Perú
Tesina, Facultad de Ingeniería de Sistemas, Pregrado,
Universidad Nacional Mayor De San Marcos

Eguila Canales, Arturo

Parco Iquiapaza, Alex Alsey

2007

Implementación de una herramienta de inteligencia de negocios
para la administración de justicia sobre una metodología ad-hoc
Tesina, Facultad de Ingeniería de Sistemas, Pregrado,
Universidad Nacional Mayor De San Marcos

Chipana Peceros, Estanislao

2004

Análisis, diseño e implementación de un Data Mart del área de
Riesgos de la Banca Personal de una institución Financiera
Tesina, Ingeniería Informática, Licenciatura, PUCP

Salazar Vázquez, Arturo

2003 Análisis de la implementación de un Data Mart de gestión de negocios aplicado a una institución financiera
Tesina, Ingeniería Informática, Licenciatura, PUCP

PROJECT MANAGEMENT INSTITUTE

2008 A Guide to the Project Management Body of Knowledge (PMBOK® Guide)
Third Edition

Silver, Bruce

2011 BPMN Method & Style. Segunda edición. California: Cody-Cassidy Press.

✓ Referencias web

Empresa Inteligencia de Negocios S.A.

2000 “Inteligencia de Negocios”
Consulta: Viernes 19 de septiembre del 2012.
< <http://www.idensa.com/> >

ETL-Tools.Info

2006 “Business Intelligence – Almacén de Datos - ETL”.
Consulta: Viernes 19 de octubre del 2012.
< <http://etl-tools.info/es/> >

BPM (GERENCIA DE PROCESOS DE NEGOCIO)

2012 Tomado del Libro BPM
Consulta: Viernes 16 de Noviembre del 2012
<http://www.konradlorenz.edu.co/images/publicaciones/suma_digital_sistemas/bpm.pdf/>

Becker, Bob

2012 “Kimball University: Three ETL Compromises to Avoid”
Informationweek software

Consulta: Miércoles 17 de Octubre del 2012

<<http://www.informationweek.com/software/information-management/kimball-university-three-etl-compromises/223101036/>>

Página oficial de INEI. Censo de población y Vivienda 2007.

Consulta: 19 de Setiembre del 2012

< <http://ineidw.inei.gob.pe/ineidw/> >

Página oficial de Pentaho, Customer Successes

Consulta: 28 de Abril del 2013

< <http://www.pentaho.com/customers/success-stories/>>

Página oficial de Pentaho, Open Source

Consulta: 25 de Marzo del 2013

<<http://www.pentaho.com>>

Página de la comunidad Pentaho, Plataforma de Bi

Consulta: 25 de Marzo del 2013

< http://community.pentaho.com/projects/bi_platform/>

Página de la comunidad Pentaho, Pentaho Analysis services

Consulta: 25 de Marzo del 2013

< <http://mondrian.pentaho.com/>>

Página de la comunidad Pentaho, Pentaho Reporting

Consulta: 25 de Marzo del 2013

< <http://reporting.pentaho.com/>>

Página de la comunidad Pentaho, Pentaho Data Integration (Kettle)

Consulta: 25 de Marzo del 2013

< <http://kettle.pentaho.com/>>

Página oficial de Oracle

Consulta: 05 de Abril del 2013

<http://www.oracle.com/technetwork/middleware/bi-enterprise-edition/overview/index.html>

Página oficial de Business Objects

Consulta: 05 de Abril del 2013

http://www.businessobjects.com/pdf/products/platform/enterprise_tech_overview.pdf

Página oficial de Microsoft

Consulta: 05 de Abril del 2013

<http://www.microsoft.com/en-us/bi/Products/bi-solution-builder.aspx>

✓ Documentos brindados por RENIEC

RENIEC01

2012 Guía de Procedimientos. Registro del trámite y entrega del documento nacional de identidad

RENIEC02

2012 Documento de requerimientos de la Gerencia de Operaciones Registrales (GOR), Gerencia de Registro e Identificación (GRI), Gerencia de Restitución de la Identidad y Apoyo Social (GRIAS)

RENIEC03

2012 Requerimientos de Gerencia de Registro e Identificación (GRI)

RENIEC04

2012 Descripción de procesos. Gerencia de Restitución de la Identidad y Apoyo Social (GRIAS)

RENIEC05

2012 Diagrama de flujo detallado de la línea de procesamiento del DNI

RENIEC06

2012 Diagrama de flujo -línea de procesamiento del DNI

RENIEC07

2012 Diagrama de flujo registro de trámite y entrega del DNI

RENIEC08
2012

Documento de entrevista 001

