

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

SISTEMA DE APOYO A LA TOMA DE DECISIONES EN
ANÁLISIS DE MORAS TRIBUTARIAS

TRABAJO DE TITUCIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

RICHARD PATRICIO BURNEO AGUILERA
richard.burneo@epn.edu.ec

Directora: PHD. María Asunción Hallo Carrasco
maria.asuncion@epn.edu.ec

2022

DECLARACIÓN

Yo, Richard Patricio Burneo Aguilera, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Richard Patricio Burneo Aguilera

DECLARACIÓN

Certifico que el presente trabajo fue desarrollado por Richard Patricio Burneo Aguilera, bajo mi supervisión.

Dra. María Hallo.
DIRECTORA

AGRADECIMIENTO

Agradezco a Dios por estar conmigo en lo largo de esta etapa, cuidando de mi vida y mi voluntad.

A mis padres, Richar y Rosalba, quienes siempre me han apoyado de manera incondicional.

A mis abuelitos mi papi Rommel y Chelita, que me han cuidado y me han dado muchos consejos y enseñanzas.

A mis hermanas, Andrea y Gaby por unirnos para superar las batallas que la vida nos ha dado.

A mi abuelita Dori, por sus mensajes de apoyo.

A la Dra. María Hallo por confiar en mí y ser apoyo en la realización de este proyecto.

A mis amigos, Alex, Jorge, Javi, David, Andrés, Jaime Daniel, Nicolas, Josué que siempre estuvieron presentes para apoyarme y escucharme.

A mi padrino Jairo, Andrés y amigos del grupo que me han ayudado a encontrar el camino para salir adelante.

A mis primos que siempre estuvieron para mí, apoyando y compartiendo buenos momentos.

DEDICATORIA

Dedicado a mi papi Rommel, quien hoy es mi orgullo, mi fuerza y motivación. Avanzamos papi Rommel.

Hay leyes de la naturaleza que son imposibles de superar, mientras tanto construimos nuestra propia suerte, sonriéndole a la vida.

ÍNDICE DE CONTENIDO

ÍNDICE DE CONTENIDO	5
LISTA DE TABLAS	7
LISTA DE FIGURAS	8
RESUMEN.....	9
ABSTRACT	10
1 INTRODUCCIÓN.....	11
1.1 Formulación del problema.	12
1.2 Justificación.....	13
1.3 Objetivos	14
1.4 Alcance.....	15
1.5 Marco teórico.....	15
1.6 Sistema de soporte de decisiones (DSS).....	22
2 METODOLOGÍA.....	26
2.1 Planificación del proyecto.	27
2.2 Definición de requerimientos del negocio	27
2.3 Modelado dimensional.....	28
2.4 Diseño físico.....	29
2.5 Diseño e implementación del subsistema de extracción, transformación y..... carga (ETL).....	29
2.6 Implementación	30
2.7 Mantenimiento y crecimiento del <i>data warehouse</i>	30
2.8 Especificación de aplicaciones de BI	30
2.9 Diseño de la arquitectura técnica.....	31
3 IMPLEMENTACIÓN	32
3.1 Planificación	32
3.2 Definición de requerimientos del negocio	32
3.3 Dirección de datos preprocesamiento.....	35
3.4 Modelo dimensional.....	37
3.5 Descripción de dimensiones.....	39
3.6 Arquitectura	41
3.7 Diseño e implementación del subsistema del ETL.....	48
4 RESULTADO OBTENIDOS Y DISCUSIÓN.....	55
5 CONCLUSIONES Y RECOMENDACIONES.....	57
6 BIBLIOGRAFÍA	58
7 ANEXOS	60

Anexo 1	60
Anexo 2	61

LISTA DE TABLAS

Tabla 1 Requerimientos.....	32
Tabla 2 Atributos de las fuentes de datos	34
Tabla 3 Base de datos Cobranza.....	35
Tabla 4 Base de datos ingresos.....	36
Tabla 5 Base de datos remisión.....	36
Tabla 6 Base de datos difícil recaudo	36
Tabla 7 Dimensiones	38
Tabla 8 Requerimiento N°1.....	61
Tabla 9 Requerimiento N°2.....	61
Tabla 10 Requerimiento N°3.....	61
Tabla 11 Requerimiento N°4.....	61
Tabla 12 Requerimiento 5.....	62
Tabla 13 Requerimiento N°6.....	62
Tabla 14 Requerimiento N°7.....	62
Tabla 15 Requerimiento N°8.....	62
Tabla 16 Requerimiento N°9.....	62
Tabla 17 Requerimiento N°10.....	64
Tabla 18 Requerimiento N°11.....	64
Tabla 19 Requerimiento N°12.....	64
Tabla 20 Requerimiento N°13.....	64
Tabla 21 Requerimiento N°14.....	64
Tabla 22 Requerimiento N°15.....	66
Tabla 23 Requerimiento N°16.....	66
Tabla 24 Requerimiento N°17.....	66
Tabla 25 Requerimiento N°18.....	66
Tabla 26 Requerimiento N°19.....	66
Tabla 27 Requerimiento 20.....	66
Tabla 28 Requerimiento N°21.....	67
Tabla 29 Requerimiento N°22.....	67
Tabla 30 Requerimiento N°23.....	67
Tabla 31 Requerimiento N°24.....	67

LISTA DE FIGURAS

Figura 1 Componentes de la inteligencia de negocios	20
Figura 2 Tareas de la metodología de Kimball, denominada Business Dimensional Lifecycle.....	26
Figura 3 Fuente de datos	33
Figura 4 Gráfico dimensional de alto nivel	41
Figura 5 Arquitectura técnica	42
Figura 6 Modelo lógico.....	43
Figura 7 Modelo Físico	44
Figura 8 Dimensión contribuyente.....	44
Figura 9 Dimensión actividad.....	45
Figura 10 Dimensión fecha	45
Figura 11 Dimensión impuesto	45
Figura 12 Dimensión proceso revisión	46
Figura 13 Dimensión dificultad recaudación.....	46
Figura 14 Dimensión ubicación.....	46
Figura 15 Dimensión Obligación	47
Figura 16 Dimensión pago en remisión.....	47
Figura 17 Tabla de hechos	48
Figura 18 ETL DSA.....	48
Figura 19 ETL DSA Cobranza	49
Figura 20 ELT DSA Ingreso	49
Figura 21 ETL DSA Remisión	49
Figura 22 ETL DSA difícil recaudo	49
Figura 23 ETL DWH.....	50
Figura 24 ETL DWH Dimensión contribuyente.....	50
Figura 25 ETL DWH Dimensión actividad	50
Figura 26 ETL DWH difícil recaudo.....	51
Figura 27 ETL DWH Dimensión impuesto.....	51
Figura 28 ETL DWH Dimensión revisión de proceso	51
Figura 29 ETL DWH Dimensión fecha	51
Figura 30 ETL DWH Dimensión ubicación.	52
Figura 31 ETL DWH Dimensión obligación.	52
Figura 32 ETL DWH Dimensión Pago remisión.	52
Figura 33 ETL DWH Pre-tabla hechos	52
Figura 34 ETL DWH Tabla de hechos	53
Figura 35 ETL DWH.....	54
Figura 36 Dashboard descriptivo	55
Figura 37 Dashboard predictivo	56

RESUMEN

En la presente tesis se realiza el diseño e implementación de sistema de apoyo a la toma de decisiones en análisis de moras tributarias. Se utiliza la metodología de Kimball [1] para la construcción del *data mart* (DM) tomando en cuenta sus cuatro aspectos básicos: centrarse en el negocio, construir una infraestructura de información adecuada, realizar entregas en incrementos significativos, ofrecer la solución completa. Las perspectivas e indicadores utilizados para el análisis de los datos son generados a partir de la información recolectada durante el proceso de análisis de requerimientos de tal forma que respondan las preguntas de negocio y que satisfagan las necesidades específicas de la institución. Dentro del aspecto de integración de datos, se realiza un proceso de limpieza y calidad de los datos obtenidos de las diferentes fuentes de información que posteriormente son analizados para la presentación de los resultados finales. El resultado de este trabajo permite generar datos estadísticos y cuantitativos del proceso de cobros tributarios que se podrán visualizar. La parte final de este trabajo presenta la interfaz final que es utilizada para guiar, facilitar y apoyar el proceso de toma de decisiones en el análisis de moras tributarias.

Palabras clave: *Data mart*, toma de decisiones, moras tributarias.

ABSTRACT

In this thesis, the design and implementation of a decision support system for the analysis of tax default are carried out. The Kimball methodology [1] is used for the construction of the data mart (DM) considering its four basic aspects: focus on the business, build an adequate information infrastructure, deliver insignificant increments, and offer the complete solution. The perspectives and indicators used for the analysis of the data are generated from the information collected during the requirements analysis process in such a way that they answer the business questions and meet the specific needs of the institution. Within the data integration aspect, a process of cleaning and quality of the data obtained from the different sources of information is carried out and which are subsequently analyzed for the presentation of the results. The result of this work allows for generation of rate statistical and quantitative data of the tax collection process that can be visualized. The final part of this paper presents the final interface that is used to guide, facilitate, and support the decision-making process in the analysis of tax default.

Keywords: *Datamart, decision making, tax default.*

1 INTRODUCCIÓN.

La tributación en América Latina toma relevancia en la década de los noventa y ha evolucionado en los diez últimos años, los recursos recaudados no han sido suficientes para solventar los gastos públicos y lograr la sostenibilidad fiscal. [2]

Al comparar la carga fiscal de países de primer mundo con países de Latinoamérica se evidencia que los ingresos tributarios en los países de América Latina corresponde al 18,4% en el PIB en el año 2015 equivalentes alrededor de la mitad del porcentaje que representa en los países desarrollados, en los países miembros de la Organización para la Cooperación y el Desarrollo Económico (OCDE) su equivalente fue de 34,8% y en los países de la unión europea de 39,2 puntos porcentuales.[3]

La Comisión Económica para América Latina y el Caribe (CEPAL) es responsable de promover el desarrollo económico y social, por lo que para el 2019 tiene el reto de aumentar la recaudación de los impuestos más progresivos. De esta manera, para aumentar los ingresos al corto plazo, se debe aplicar cambios fiscales con fines recaudatorios. Sin embargo, se debe considerar que este escenario a mediano y largo plazo afectará a los sectores económicos, lo que conlleva a la evasión y la elusión fiscal, que es otro reto que enfrentan los países de Latinoamérica.[4]

América Latina en la última década ha pasado un proceso de transformación respecto a su situación tributaria con una infinidad de cambios, los mismos que han sido frecuentes, lo cual ha provocado un aumento en la presión fiscal y la necesidad de mejorar la eficiencia recaudadora de los gobiernos. Gómez y Morán mencionan que “entre los años 1990 y 2010 la carga tributaria promedio de la región aumentó más del 42%, pasando de 13,4% al 19,1% del PIB” [2]. Según aumenta la presión fiscal, los contribuyentes se ven obligados en buscar alternativas que permitan optimizar su carga tributaria. La planificación tributaria es percibida como una herramienta que puede ahorrar recursos, además, permite analizar el impacto fiscal en la toma de decisiones al momento de una negociación.

En este contexto, la finalidad del presente trabajo es usar sistemas de apoyo en la toma de decisiones como solución a las moras tributarias como una herramienta útil para la toma de decisiones en una Institución tributaria, en donde busca la optimización de recursos a través de una correcta carga fiscal.

En la actualidad, normalmente los países poseen una institución que se encarga de la recaudación de los impuestos a sus ciudadanos. Los impuestos son “tributos” que se

exige a los habitantes según su capacidad económica. En las instituciones tributarias existen contribuyentes que no realizan los pagos en el tiempo reglamentario e inclusive hay quienes evaden las responsabilidades tributarias por diferentes causas, provocando pérdidas monetarias a la institución como al Estado al que pertenecen.[5]

1.1 Formulación del problema.

Las moras tributarias desempeñan un papel importante en el proceso de recaudación de impuestos. Por ejemplo, en los Estados Unidos, donde se cree que las herramientas de cobro de deudas son efectivas, los impuestos morosos aún representan más del 25 % de la brecha fiscal bruta total en el 2016. Además, las moras fiscales son los ingresos potenciales que están fácilmente disponibles para instituciones tributarias, donde las agencias tributarias invierten recursos sustanciales en políticas dirigidas a reducir la morosidad fiscal. [5]

En la actualidad, existen una gran cantidad de datos históricos acumulados, y el desafío hoy en día para las agencias tributarias, es mejorar el proceso de recaudación de impuestos y formular optimas políticas legislativas para brindar mejores servicios al contribuyente. [6]

Históricamente, el 14 por ciento de los impuestos adeudados no se recaudan, sin embargo, no es factible física ni políticamente aumentar el cumplimiento a través de más auditorías. El enfoque emergente de un Servicio de Rentas Internas es esforzarse en aumentar el cumplimiento voluntario al enfatizar el análisis de segmentación del mercado (identificando grupos enteros de contribuyentes que no cumplen) y tomar medidas que afectan a todo el grupo, en lugar de su enfoque histórico de auditar a los contribuyentes de manera individual. La implementación de este enfoque depende de un data warehouse (DW), pero también requiere un cambio en la cultura del servicio de rentas internas. [7]

El principal problema con varias agencias tributarias es su dificultad para acceder fácilmente y analizar el conjunto completo de información de recaudación de impuestos. Además, la falta de un sistema de apoyo de decisiones impide al responsable de las decisiones resolver de manera eficaz los problemas relacionados a procesos de cobros tributarios.

1.2 Justificación

Justificación teórica

En la implementación del sistema de apoyo a la toma de decisiones se desarrollará un data mart que posean datos que permitan crear dashboards en una herramienta de visualización de datos que servirá a los usuarios con información necesaria y óptima para tomar decisiones en diferentes situaciones o escenarios que enfrenten. [6]

En el artículo "*Decision making and reduction techniques for tax revenue using data warehouse*", los autores recomiendan utilizar un sistema de apoyo a la toma de decisiones fiscales y data warehouse (DW) para decisiones estratégicas con el fin de lograr un mejor modelo optimizado en la institución tributaria. [7]

En el artículo "*The future is bright*" escrito por Ralph Kimball se menciona que:

"Un almacén de datos (DW) nunca fue más valioso e interesante de lo que es ahora. Tomar decisiones basadas en datos es fundamental y es obvio, que la actual generación de usuarios de negocios y diseñadores de almacenamiento, no pueden imaginar un mundo sin acceso a los datos". [8]

Los DW por la manera en que organizan los datos y elementos de uso, están destinados a proporcionar herramientas útiles para ayudar a los gerentes en el proceso de toma de decisiones. Adicionalmente permiten realizar reingeniería dentro de la organización, consolidar la información, medir los resultados del negocio de una manera más efectiva y dar un mejor servicio. [9]

El objetivo principal que se desea lograr cuando se desarrolla un data warehouse en inteligencia de negocios (BI), es la aspiración de optimizar el proceso de toma de decisiones para lograr mejores resultados organizacionales, ya que estas permiten realizar análisis más detallados para descubrir los problemas y oportunidades a tiempo.

En cuanto a los data marts se los define como un subconjunto de un data warehouse o como un data warehouse pequeño para organizar la información que se encuentra dispersa facilitando la consulta por diferentes criterios y para facilitar el proceso de toma de decisiones es necesario el desarrollo de un componente de visualización [10], por lo que se pondrá énfasis en este componente.

El desarrollo de un data mart se lo realizará orientado a procesos, es decir se buscará la integración de BI con los sistemas que soportan los diferentes procesos para evitar la dispersión y redundancia de datos. [11]

Justificación Metodológica

El enfoque escogido para el desarrollo del sistema de apoyo de toma de decisiones usando data mart es el propuesto por Kimball con énfasis en análisis y diseño de requerimiento en la toma de decisiones. El desarrollo se realizará con un diseño *top down* (de arriba hacia abajo) y la implementación *bottom up* (de abajo hacia arriba). Al aplicar esta orientación al proyecto propuesto funciona bien porque se puede implementar data marts en áreas específicas para integrarlos en un data warehouse. [8]

En la etapa de definición de requerimientos del negocio se realizará el análisis de apoyo a la toma de decisiones. El análisis de apoyo a la toma de decisiones proporciona la dirección inicial de todo el proceso de desarrollo del sistema de apoyo a la toma de decisiones en procesos de cobros tributarios.

La metodología de Kimball permitirá construir una arquitectura que se adapte fácilmente al cambio ya que a medida que avanza el tiempo, se presentan nuevos requerimientos, por lo tanto, es necesario dejar abierta la posibilidad de que en el futuro se desarrollen nuevos data marts de acuerdo con las necesidades de la organización. [7]

Justificación Práctica

El sistema de apoyo a la toma de decisiones en análisis de moras tributarias usando un data mart, facilitará a una institución tributaria a acceder y analizar un conjunto completo de información histórica, ofreciendo al responsable de la toma de decisiones alternativas para la solución eficaz de problemas y la disminución de moras que posean los contribuyentes con dicha institución tributaria. Además, el sistema servirá para mejorar la recaudación de impuestos.

1.3 Objetivos

Objetivo General

Implementación de un data mart como solución de inteligencia de negocios, bajo la metodología de Ralph Kimball para optimizar la toma de decisiones respecto al manejo de indicadores de contribuyentes en mora de una institución tributaria.

Objetivos Específicos

- Identificar los procesos de toma de decisiones y analizar los requerimientos de información de acuerdo con las perspectivas y necesidades de la institución financiera.
- Analizar la información de la base de datos transaccional e identificar las fuentes de datos requeridas.

- Construir el data mart en base a la metodología Ralph Kimball que cumpla con los requerimientos necesarios del departamento tributario.

1.4 Alcance

La investigación se delimita a las deudas tributarias de los contribuyentes. En este trabajo, se propone un sistema de apoyo a la toma de decisiones usando un data mart con información relevante de contribuyentes en procesos de cobros tributarios que permita a futuro la búsqueda de patrones de deudores de alto nivel de riesgo de recaudación.

1.5 Marco Terrico

La tecnología de información (TI).

Según lo definido por la asociación de la tecnología de información de América (ITAA) es “el estudio, diseño, desarrollo, implementación, soporte o dirección de los sistemas de información computarizados, en particular de software de aplicación y hardware de computadoras.”[12] Se ocupa del uso de las computadoras y su software para convertir, almacenar, proteger, procesar, transmitir y recuperar la información. En el artículo “El impacto de las herramientas de inteligencia de negocios en la toma de decisiones de los ejecutivos” [13], los autores describen que: “el comienzo de una nueva economía basada en la información se inició en la década de los setenta en los EUA. En estos años la aplicación de la informática al mundo de los negocios se consideraba una cuestión meramente técnica y orientada al tratamiento automático de los datos”.

Es a partir de los setenta cuando los investigadores centran su atención en la gestión de la información que puede extraerse de los datos para utilizarla en actividades de planificación y control y como soporte a la toma de decisiones empresariales. Los científicos sociales, técnicos y políticos han debatido ampliamente y desde distintos puntos de vista los impactos que en la sociedad está produciendo el desarrollo de las tecnologías de la información y aquellos que se van a producir como consecuencia de su amplia difusión en la sociedad.

Los sistemas de información empresariales.

En el libro “Sistemas de información para la toma de decisiones” el autor Daniel Cohen describe a un sistema de información como conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio, el cual realiza

cuatro actividades básicas: entrada, almacenamiento, procesamiento y salida de información. [14]

- **Entrada de Información:** Es el proceso mediante el cual el sistema de información toma los datos que requiere para procesar la información. Las entradas pueden ser manuales o automáticas. Las entradas manuales son aquellas que se proporcionan en forma directa por el usuario, mientras que las entradas automáticas son datos o información que provienen o son tomados de otros sistemas o módulos.
- **Almacenamiento de información:** El almacenamiento es una de las actividades o capacidades más importantes que tiene una computadora, ya que a través de esta propiedad el sistema puede recordar la información guardada en la sección o proceso anterior. Esta información suele ser almacenada en estructuras de información denominadas archivos. La unidad típica de almacenamiento son los discos magnéticos o discos duros, los discos flexibles o diskettes y los discos compactos (CD-ROM).
- **Procesamiento de Información:** Es la capacidad del Sistema de Información para efectuar cálculos de acuerdo con una secuencia de operaciones preestablecida. Estos cálculos pueden efectuarse con datos introducidos recientemente en el sistema o bien con datos que están almacenados. Esta característica de los sistemas permite la transformación de datos fuente en información que puede ser utilizada para la toma de decisiones.
- **Salida de Información:** La salida es la capacidad de un Sistema de Información para sacar la información procesada o bien datos de entrada al exterior. Las unidades típicas de salida son las impresoras, terminales, diskettes, cintas magnéticas, la voz, los graficadores y los plotters, entre otros. Es importante aclarar que la salida de un Sistema de Información puede constituir la entrada a otro Sistema de Información o módulo. En este caso, también existe una interfase automática de salida.

Tipos y Usos de los Sistemas de Información.

Los Sistemas de Información cumplen tres objetivos básicos dentro de las organizaciones:

- Automatizar procesos operativos.
- Generar información que apoye al proceso de toma de decisiones.
- Lograr ventajas competitivas a través de su implantación y uso.

Los sistemas que se encargan de automatizar procesos operativos son los sistemas transaccionales, automatizan los procesos operativos en una organización y su función principal es procesar transacciones, así como: pagos, cobros, pólizas, entradas, salidas, etc. Otro sistema son los de soporte de toma de decisiones, los cuales apoyan el proceso de toma de decisiones porque son sistemas expertos a la toma de decisiones y sistemas de información ejecutivos. Existe un tercer tipo de sistema llamado el sistema estratégico, y se desarrolla en una organización, enfocado en lograr ventajas competitivas, de una organización, usando tecnologías de información. [13]

A continuación, se mencionan las principales características de estos tipos de sistemas de información:

a) Sistemas Transaccionales. Sus principales características son:

- A través de éstos suelen lograrse ahorros significativos de mano de obra, debido a que automatizan tareas operativas de la organización.
- Con frecuencia son el primer tipo de Sistemas de Información que se implanta en las organizaciones. Se empieza apoyando las tareas a nivel operativo de la organización.
- Son intensivos en entrada y salida de información; sus cálculos y procesos suelen ser simples y poco sofisticados.
- Tienen la propiedad de ser recolectores de información, es decir, a través de estos sistemas se cargan las grandes bases de información para su explotación posterior.
- Son fáciles de justificar ante la dirección general, ya que sus beneficios son visibles y palpables.

b) Sistemas de apoyo de las decisiones. Las principales características de estos son:

- Suelen introducirse después de haber implantado los Sistemas Transaccionales más relevantes de la empresa, ya que estos últimos constituyen su plataforma de información.
- La información que generan sirve de apoyo a los mandos intermedios y a la alta administración en el proceso de toma de decisiones.
- Suelen ser intensivos en cálculos y escasos en entradas y salidas de información.
- No suelen ahorrar mano de obra. Debido a ello, la justificación económica para el desarrollo de estos sistemas es difícil, ya que no se conocen los ingresos del proyecto de inversión.

- Suelen ser sistemas de información interactivos y amigables, con altos estándares de diseño gráfico y visual, ya que están dirigidos al usuario final.
 - Apoyan la toma de decisiones que, por su misma naturaleza son repetitivos y de decisiones no estructuradas que no suelen repetirse.
 - Estos sistemas pueden ser desarrollados directamente por el usuario final sin la participación operativa de los analistas y programadores del área de informática.
- c) Sistemas estratégicos. Sus principales características son:
- Suelen desarrollarse “*in house*”, es decir, dentro de la organización, por lo tanto, no pueden adaptarse fácilmente a paquetes disponibles en el mercado.
 - Típicamente su forma de desarrollo es a base de incrementos y a través de su evolución dentro de la organización. Se inicia con un proceso o función en particular y a partir de ahí se van agregando nuevas funciones o procesos.
 - Su función es lograr ventajas que los competidores no posean, tales como ventajas en costos y servicios diferenciados con clientes y proveedores. En este contexto, los Sistema Estratégicos son creadores de barreras de entrada al negocio.
 - Apoyan el proceso de innovación de productos y proceso dentro de la empresa debido a que buscan ventajas respecto a los competidores y una forma de hacerlo en innovando o creando productos y procesos

Inteligencia de negocios

El termino Inteligencia o business intelligence (BI) es utilizado por diferentes expertos y fabricantes de software para distinguir un amplio rango de tecnologías, plataformas de software, aplicaciones específicas y procesos”. Además, se puede utilizar el término desde tres diferentes perspectivas: [15]

- a) Tomar mejores decisiones rápidamente: El objetivo primario de BI es ayudar a la gente a tomar decisiones que mejoren el rendimiento de la compañía e impulsen su ventaja competitiva en el mercado. En forma breve BI faculta a las organizaciones a tomar las mejores decisiones rápidamente. BI ayuda a tomar mejores decisiones analizando si las acciones tomadas están de hecho dando resultados hacia los objetivos de la compañía. Decidir cuál es la mejor decisión para una organización, se lleva mejor a cabo con un definido y claro conjunto de objetivos y un plan para conseguirlos. Esta conexión entre el plan global de la compañía y BI no es una vía de

una sola dirección, con BI simplemente recibir el plan y utilizar éste como la escala de medición de la calidad de las decisiones.

- b) Convertir los datos en información: Para tomar mejores decisiones más rápidamente, los directivos y gerentes necesitan de información relevante y útil al alcance de la mano. Pero es común una larga brecha entre la información que los responsables en la toma de decisiones requieren y las grandes cantidades de datos que las organizaciones recopilan cada día. Nosotros llamamos a esto análisis gap (brecha del análisis). Para saltar esta brecha las organizaciones invierten en desarrollar sistemas de BI para convertir los datos de orígenes en información de utilidad. Los sistemas de BI más efectivos tienen acceso a inmensas cantidades de datos (medidas en gigabytes incluso en terabytes) para posteriormente entregar a los responsables en la toma de decisiones información expresada de una plataforma que ellos pueden asimilar fácilmente la misma.
- c) Utilizar un método razonable para la gestión empresarial: BI puede ser definido como un método para la gestión empresarial, una forma de pensamiento organizacional, una filosofía de gestión; en pocas palabras, un interés hacia la inteligencia de negocios. Tanto las personas como las organizaciones se interesan en el BI, porque creen que el uso de un enfoque racional y basado en hechos a la hora de tomar decisiones resulta positivo en la medida que sea posible. La actitud empresarial orientada hacia el BI de tomar decisiones basadas en hechos y basadas en análisis está influenciando la cultura corporativa en todas partes, principalmente porque vivimos en un mundo que está de forma creciente enriqueciéndose en información; por otro lado, la tecnología está disponible

Componentes de la inteligencia de negocios

Los componentes de inteligencia están descritos en la figura 1.

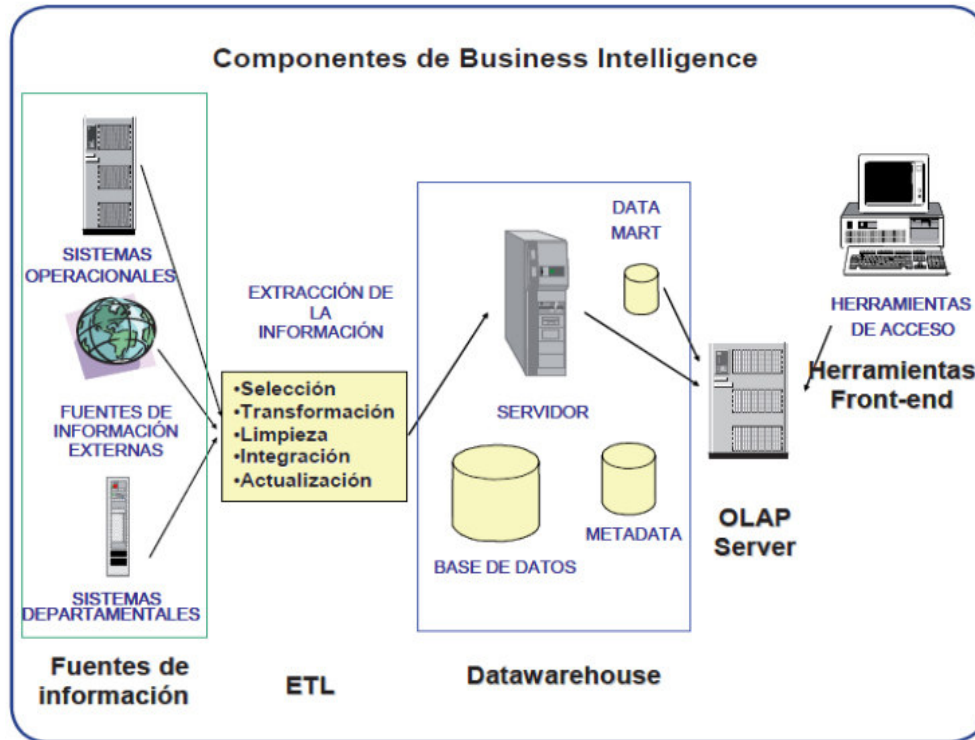


Figura 1 Componentes de la inteligencia de negocios

- Fuentes de Información: Generalmente son los datos generados por sistemas operacionales, o datos creados por aplicaciones como las hojas de cálculo o almacenados en archivos texto los cuales se utilizan para alimentar de información el DW.
- Proceso ETL: De extracción, transformación y carga de los datos en el DW. Los datos antes de almacenarse en un DW, pasan por procesos de filtrado, limpieza, transformación y redefinición.
- Almacén de datos: se almacenan los datos de una manera que optimice su flexibilidad, facilidad de acceso y administración, en donde los datos están estructurados para generar informes que ayuden a la toma de decisiones. En el DW se busca almacenar los datos de una forma que maximice su flexibilidad, facilidad de acceso y administración.
- El análisis dimensional: Es el que proporciona la capacidad de realizar cálculos, análisis, pronósticos, consultas en grandes volúmenes de datos.
- El motor OLAP, nos provee capacidad de cálculo, consultas, funciones de planeamiento, pronóstico y análisis de escenarios en grandes volúmenes de datos. En la actualidad, existen otras alternativas tecnológicas al OLAP.

Definición de Data Mart

Un data mart según Vitt, es un conjunto de datos que son estructurados de una forma que facilite su posterior análisis. [15]

Un data mart contiene la información referente a un área, un tema o una función en particular, con datos relevantes que provienen de diferentes aplicaciones operacionales. Los data marts pueden ser de diversa base de datos OLAP dependiendo del tipo de análisis que se quiera desarrollar. [15]

Herramientas

- Power Designer

Power Designer es una herramienta de modelado empresarial colaborativa producida por Sybase. [16]

- Spoon Pentaho

Transformaciones son Conjunto de pasos (steps) interconectados por medio de saltos (hops) consisten en mover y transformar filas de origen a destino. Flujo de información que se transforma. [17]

Trabajos son conjunto de transformaciones, trabajos o pasos que se ejecutan de manera secuencial. El trabajo (*JOB*) en sí no maneja registros sino secuencias de tarea. Flujo de procesos.

Los pasos, son pequeños funcionalidades lógicas o tareas específicas, que pueden admitir una entrada y/o salida. Se unen mediante los *hops*, y generan un flujo de datos, o en algunos casos lo reciben para devolver a la salida. Es donde reside la verdadera lógica de la ETL.

Saltos son las relaciones entre steps o direcciones que va a seguir el flujo de la información.

- Excel

Herramienta de Microsoft que nos permite revisar diferentes datos que se encuentran en documentos exportados con formato xls o clv.

- PgAdmin 4

PgAdmin es la herramienta *open source* de administración por excelencia para sus bases de datos PostgreSQL. Algunas de sus características son: el soporte completo para *unicode*, edición rápida de consultas y datos multihilo y soporte para todos los tipos de objetos de PostgreSQL. [18]

- Postgresql

PostgreSQL es un sistema de gestión de bases de datos relacional orientado a objetos y de código abierto, publicado bajo la licencia PostgreSQL [19]

- Tableau

Tableau es un software de análisis de datos con una excelente capa de visualización y presentación, considerado por muchos como uno de los mejores programas para la presentación visual de datos y con muy alta clasificación en la facilidad de uso. Especialmente utilizada para BI, simplifica los datos para presentarlos en un formato comprensible e intuitivo. [20]

1.6 Sistema de soporte de decisiones (DSS)

El sistema de soporte de decisiones es un sistema interactivo provisto de programas y herramientas, para ayudar a los responsables de la toma de decisiones a utilizar tecnologías de comunicaciones, datos, documentos, conocimiento y/o modelos para identificar y resolver problemas, para completar tareas del proceso de decisión, y para tomar decisiones. Es una amplia área de análisis que sirve para que la gente examine datos a fin de tomar decisiones, ya sean grandes o pequeños, sobre los negocios de sus compañías.

Data Warehouse (DW)

Almacenes de datos o data warehouse son repositorios de información extraída de otros sistemas corporativos, sean estos sistemas transaccionales, bases de datos departamentales, o Intranet de la compañía, a la que los hombres de negocios de la empresa pueden acceder. Data warehouse es un sistema, no un producto, en el que se almacenan datos. Data warehouse es una técnica para consolidar y administrar datos de variadas fuentes con el propósito de responder preguntas de negocios y tomar decisiones, de una forma rápida. Data warehouse se vale de una base de datos relacional diseñada para el acceso rápido y análisis y no proceso transaccional. El data warehouse separa la carga del análisis y normalmente contiene datos históricos derivados de datos transaccionales.

Datamart

El data mart es considerado data warehouse más pequeño; es decir, es un data warehouse orientada a algún tema específico. Los data mart son usados por un departamento o grupo de usuarios en una compañía, para un conjunto definido de tareas. Existen dos tipos de data mart:

- *Data marts* independientes: también conocimos como data mart aislados, son aquellos que extraen información directa de los sistemas transaccionales y no dependen de un data warehouse.
- *Data marts* dependientes: extrae la información de un data warehouse y almacena solo una parte de los datos del data warehouse.

Dimensiones

Las dimensiones son los diferentes puntos de vista por los que queremos analizar la información. Las dimensiones contienen los diversos atributos que queremos analizar, además se estructuran en forma jerárquica, conforme a diferentes niveles de detalle.

- Dimensiones normales: aquellas que agrupan diferentes atributos que están relacionados por el ámbito al que se refieren (todas las características de un cliente, los diferentes componentes de la dimensión tiempo, etc).
- Dimensiones causales: son atributos que pueden causar cambios en los procesos de negocio (por ejemplo, la dimensión promoción en el proceso de negocio de ventas)
- Dimensiones heterogéneas: agrupan conjuntos heterogéneos de atributos, que no están relacionados entre sí.
- Dimensiones roll-up: un subconjunto de otra, necesarias para el caso en que tenemos tablas de hechos con diferente granularidad (ver la entrada anterior del blog).
- Dimensiones *junk*: agrupan indicadores de baja cardinalidad como pueden ser *flags* o indicadores.
- Dimensiones *role-playing*: son dimensiones reutilizables cuando una misma dimensión interviene en una tabla de hechos varias veces (por ejemplo, la fecha en una tabla de hechos donde se registran varias fechas referidas a conceptos diferentes), es necesario reutilizar la misma dimensión, pues no tiene sentido crear tantas dimensiones como usos se hagan de ella.
- Mini dimensiones o dimensiones *outrigger*: son conjuntos de atributos de una dimensión que se extraen de la tabla de dimensión principal, pues se suelen analizar de forma diferente.

Métricas

Las medidas candidatas son los datos numéricos, pero no cada atributo numérico en una medida candidata.

Estas medidas controlan resultado de la actividad de negocio.

- Métricas de realización de actividad: miden la realización de una actividad. Por ejemplo, la participación de una persona en un evento.
- Métricas de resultado de una actividad: recogen los resultados de una actividad. Por ejemplo, la cantidad de unidades vendidas.

Indicadores

Los indicadores son una variable dimensional unitaria, expresada como un cociente, que correlaciona dos variables cualesquiera. A través de este concepto es posible relacionar diversas variables presentes en los procesos de una empresa. Ejemplo: N° Ciclos/Hora, LtsComb, /Maquina.

La importancia de un indicador radica en la particularidad de informar al usuario el estado actual del sistema. Revelar el estado del sistema, nos permitirá tomar decisiones preventivas o correctivas de acuerdo con los resultados de las comparaciones entre el valor esperado y el valor obtenido del sistema

Tabla de hechos

Los hechos son los indicadores de negocio que dan sentido al análisis de las dimensiones. Las tablas de hechos incluyen los indicadores asociados a un proceso de negocio en concreto y las claves de las dimensiones que intervienen en dicho proceso, en el mínimo nivel de granularidad o detalle. Podemos tener varios tipos de tablas de hechos:

- Tabla de hechos transaccionales: representan eventos que suceden en un determinado espacio-tiempo. Se caracterizan por analizar los datos con el máximo detalle. Reflejan las transacciones relacionadas con nuestros procesos de negocio (ventas, compras, inventario, contabilidad, etc.).
- Tabla de hechos sin medidas: no tienen medidas y representan la ocurrencia de un evento determinado. Por ejemplo, la asistencia a un curso puede ser una tabla de hechos sin métricas asociadas.
- Tabla de hechos de capturas periódicas: son tablas de hecho usadas para recoger información de forma periódica a intervalos de tiempo regulares sobre un hecho. Nos permiten tomar una foto de la situación en un momento establecido (por ejemplo, al final del día, de una semana o de un mes). Un ejemplo puede ser la foto del stock de materiales, al final de cada día.
- Tabla de hechos con enfoque acumulativo: representan el ciclo de vida completo de una actividad o proceso, que tiene un principio y final. Suelen representar valores acumulados.

- Tabla de hechos consolidada: tablas de hechos construidas como la acumulación, en un nivel de granularidad o detalle diferente, de las tablas de hechos de transacciones.

Enfoque “*bottom-up*”

El enfoque “*bottom-up*”, establece que el desarrollo de una solución de inteligencia de negocios debe partir por el desarrollo de los data marts y que el conjunto de estos se constituye en la data warehouse.

Enfoque “*top-down*”

El enfoque “*top-down*”, establece que el desarrollo de una solución de inteligencia de negocios debe partir por el desarrollo del data warehouse que deriva en la data marts.

Dashboard

Los Dashboard son tableros o paneles en donde se muestran información fácil de abstraer, por lo general se encuentran métricas necesarias que permitan conocer acerca de la efectividad de alguna campaña o para la toma de decisiones basados en objetivos.
[21]

2 METODOLOGÍA

La Metodología Kimball, es aplicada para la construcción de un almacén de datos (*data warehouse*, DW) que es la colección de datos encaminada a un determinado ámbito ya sea organización, empresa, etc., que ayuda principalmente a la resolución y toma de decisiones en la entidad en la que sea utilizada.

La metodología se desarrolla en base a lo que Kimball denomina Business Dimensional Lifecycle (Ciclo de Vida Dimensional del Negocio) que es descrita en el libro titulado “*The Data Warehouse Lifecycle Toolkit*”. [22]

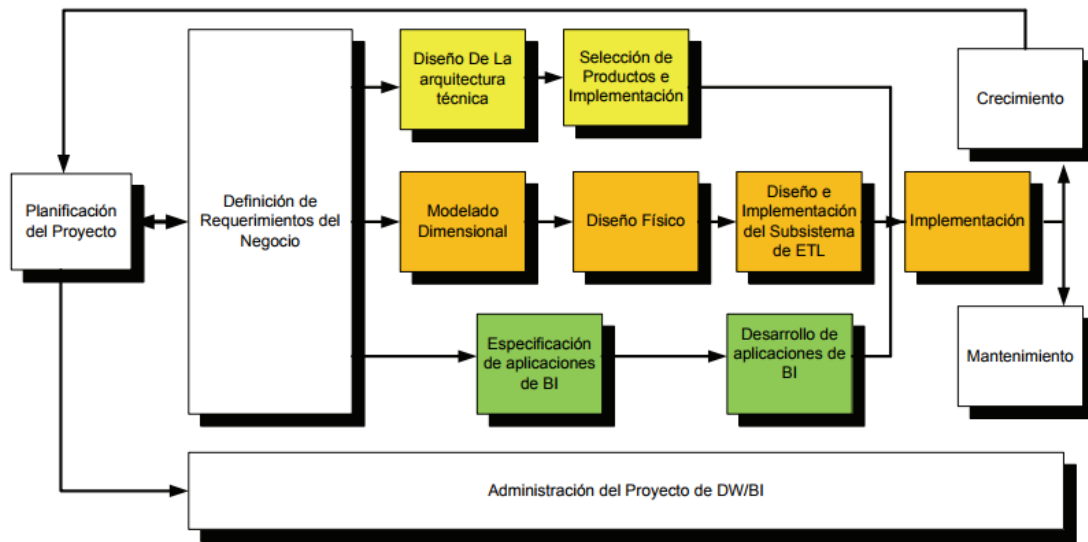


Figura 2 Tareas de la metodología de Kimball, denominada Business Dimensional Lifecycle.

Este ciclo de vida está planteado en cuatro principios básicos:

- Centrarse en el negocio.
- Construir adecuadamente una infraestructura de información.
- Realizar entregas en incrementos significativos (se crean almacenes de datos en plazos de 6 a 12 meses).
- Ofrecer una solución de manera integral, para este principio se deben proporcionar todos los componentes necesarios para de esta manera proveer valor a los usuarios de negocios, para lo cual debe tener bien establecido y diseñado un almacén de datos, entregar herramientas de consulta ad hoc, análisis avanzados, soporte, aplicaciones para informes, documentación, sitio web y capacitación.

La construcción de una solución de *data warehouse* (DW) es compleja, y lo que hace la metodología Kimball es ayudar a simplificar esa complejidad. Las tareas de esta metodología (ciclo de vida) se especifican a continuación:

2.1 Planificación del proyecto.

En este proceso se establece el propósito del proyecto de DW/BI, El alcance y sus objetivos específicos, principales riesgos y una aproximación inicial a las necesidades de información. Este proceso incluye acciones típicas de un plan de proyecto:

- Definir el alcance (entender los requerimientos del negocio).
- Identificar las tareas
- Programar las tareas
- Planificar el uso de los recursos.
- Asignar la carga de trabajo a los recursos
- Elaboración de un documento final que representa un plan del proyecto.
- Además, en esta parte definimos cómo realizar la administración o gestión de esta subfase que es todo un proyecto en sí mismo, con las siguientes actividades:
- Monitoreo del estado de los procesos y actividades.
- Rastreo de problemas
- Desarrollo de un plan de comunicación comprensiva que dirija la empresa y las áreas de TI

2.2 Definición de requerimientos del negocio

La definición de requerimientos es un proceso de entrevistar al personal de negocio y técnico. En este punto, se debe conocer sobre el negocio, industria, competidores y los clientes. Es importante desarrollar una revisión a todos los informes posibles de la organización; entrevistar a empleados, rastrear los documentos de estrategia interna, analizar lo que se dice en la prensa acerca de la organización y la competencia y a su vez conocer términos y la terminología del negocio.

Lo recomendable es entrevistar al personal que se encuentra en los cuatro grupos a continuación:

- El directivo responsable de tomar las decisiones estratégicas.
- Los administradores intermedios y de negocio responsables de explorar alternativas estratégicas y aplicar decisiones

- El personal de sistemas, si existe (estas son las personas que realmente saben qué tipos de problemas informáticos y de datos existen en la organización)
- El personal que se entrevista por razones políticas.

Entre las tareas antes mencionadas, existe una flecha bidireccional, es decir que los requerimientos del negocio son el soporte inicial de las tareas subsiguientes, influenciando en el plan de proyecto.

Si avanzamos por el camino central del diagrama, se encontrarán tareas relacionadas al área de Datos, se implementaremos y diseñara el modelo dimensional, y se desarrollara el subsistema de extracción, transformación y carga (*extract, transformation, and load* - ETL) para cargar el DW. Las tareas que pertenecen al área se describen a continuación:

2.3 Modelado dimensional

El modelo dimensional es un proceso dinámico e iterativo. Inicia con un modelo dimensional de alto nivel que se lo obtiene a partir de los procesos priorizados y descritos en la tarea anterior, y el proceso iterativo consiste en cuatro pasos:

- a) Elegir el proceso de negocio: consiste en seleccionar el área a modelizar. Esta es una decisión de la dirección, y depende principalmente del análisis de requerimientos y temas analíticos.
- b) Establecer el nivel de granularidad: Significa especificar el nivel de detalle. La elección de la granularidad depende de requerimientos del negocio y lo que es posible a partir de los datos que actualmente se tienen. Lo que se sugiere es empezar a diseñar el DW al mayor nivel de detalle porque se podrían realizar agrupamientos posteriores, al nivel que uno desea.
- c) Elegir las dimensiones: Las dimensiones aparecen de las discusiones del equipo, y facilitadas por la elección del nivel de granularidad y de la matriz de procesos/dimensiones. Las tablas de dimensiones tienen una agrupación de atributos que brindan una perspectiva o forma de análisis sobre una medida en una tabla hechos. Una manera de identificar las tablas de dimensiones es que sus atributos son posibles candidatos para encabezar en los informes, cubos, tablas *pivot*, o cualquier forma de visualización, unidimensional o multidimensional.
- d) Identificar medidas y las tablas de hechos: Consiste en establecer las medidas que empiezan en los procesos de negocios. Una medida es un atributo de una tabla que se quiere analizar, sumando o reuniendo datos y usando los criterios de corte

conocidos como dimensiones. Las medidas por lo general se vinculan con el nivel de granularidad y se encuentran en tablas que denominamos tablas de hechos (*fact* en inglés). Cada tabla de hechos tiene una o más medidas de un proceso organizacional, de acuerdo con los requerimientos. Un registro contiene una medida expresada en números, en la cual se desea realizar una operación de agregación (promedio, conteo, suma, etc.) en función de una o más dimensiones.

2.4 Diseño físico

En esta tarea, se contestan las siguientes preguntas:

¿Cómo puede determinar cuán grande será el sistema de DW/BI?

¿Cuáles son los factores de uso que llevarán a una configuración más grande y compleja?

¿Cómo se debe configurar el sistema?

¿Cuánta memoria y servidores se necesitan? ¿Qué tipo de almacenamiento y procesadores?

¿Cómo instalar el software en los servidores de desarrollo, prueba y producción?

¿Qué necesitan instalar los diferentes miembros del equipo de DW/BI en sus estaciones de trabajo?

¿Cómo convertir el modelo de datos lógico en un modelo de datos físicos en la base de datos relacional?

¿Cómo conseguir un plan de indexación inicial?

¿Debe usarse la partición en las tablas relacionales?

2.5 Diseño e implementación del subsistema de extracción, transformación y carga (ETL)

El subsistema de extracción, transformación y carga (ETL) es la base del data warehouse. Si se diseña de una manera adecuada, puede extraer los datos de los sistemas de origen, aplicando diferentes reglas con el objetivo de incrementar la calidad y consistencia de estos, consolidando la información proveniente de distintos sistemas, y finalmente grabar (cargar) la información en el DW en un formato que este acorde a la utilización por parte de las herramientas de análisis.

2.6 Implementación

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio. Existen varios factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación y las estrategias de feedback.

2.7 Mantenimiento y crecimiento del *data warehouse*

El mantenimiento y crecimiento del data warehouse sirve para administrar el entorno del DW existente es importante enfocarse en los usuarios de negocio, los cuales son el motivo de su existencia, además de gestionar adecuadamente las operaciones del DW, medir y proyectar su éxito y comunicarse constantemente con los usuarios para establecer un flujo de retroalimentación, en esto consiste el mantenimiento. Finalmente, es importante sentar las bases para el crecimiento y evolución del DW en donde el aspecto clave es manejar el crecimiento y evolución de forma iterativa utilizando el ciclo de vida propuesto, y establecer las oportunidades de crecimiento y evolución en orden por nivel prioridad.

Si avanzamos por el camino inferior del diagrama, encontramos las tareas asociadas al área aplicaciones de inteligencia de negocios, en esta ruta se encuentran tareas en las que diseñamos y desarrollamos las aplicaciones de negocios para los usuarios finales. Las tareas pertenecientes al área se describen a continuación:

2.8 Especificación de aplicaciones de BI

En el proceso especificación de aplicaciones de BI, esta tarea se proporciona, a una gran comunidad de usuarios una forma más estructurada y, por lo tanto, más fácil de acceder al almacén de datos. Se proporciona este acceso estructurado a través de lo que llamamos, aplicaciones de inteligencia de negocios (*business intelligence applications*). Las aplicaciones de BI son la cara visible de la inteligencia de negocios: los informes y aplicaciones de análisis proporcionan información útil a los usuarios. Las aplicaciones de BI incluyen un amplio espectro de tipos de informes y herramientas de análisis, que van desde informes simples de formato fijo, a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio. Kimball divide a estas aplicaciones en dos categorías basadas en el nivel de sofisticación, y les llama:

- a) Informes estándar: son informes relativamente simples, de formato predefinido, y parámetros de consulta fijos, proporcionan a los usuarios un conjunto básico de información acerca de lo que está sucediendo en un área determinada de la empresa y se utilizan día a día.
- b) Aplicaciones analíticas: Son más complejas que los informes estándar. Estas aplicaciones pueden incluir algoritmos y modelos de minería de datos, que ayudan a identificar oportunidades o cuestiones subyacentes en los datos, y el usuario puede pedir cambios en los sistemas transaccionales basándose en los conocimientos obtenidos del uso de la aplicación de BI. Algunas aplicaciones analíticas comunes incluyen:
- Análisis de la eficacia de las promociones
 - Análisis de rutas de acceso en un sitio Web
 - Análisis de afinidad de programas
 - Planificación del espacio en espacios comerciales
 - Detección de fraudes
 - Administración y manejo de categorías de productos
 - Por último, en el camino superior, encontramos las tareas asociadas al área Tecnología en esta ruta, se encuentran las tareas relacionadas con software específico.

2.9 Diseño de la arquitectura técnica

El área de arquitectura técnica cubre los procesos y herramientas que se aplican a los datos. En el área técnica existen dos conjuntos que tienen distintos requerimientos, brindan sus propios servicios y componentes de almacenaje de datos, por lo que se consideran cada uno aparte: El *back room* y el *front room*. El *back room* es el responsable de la obtención y preparación de los datos, por lo que también se conoce como adquisición de datos y el *front room* es responsable de entregar los datos a la comunidad de usuario y también se le conoce como acceso de datos.

3 IMPLEMENTACIÓN

La metodología Kimball, descrita en el capítulo 2 es usada para la implementación del sistema de toma de decisiones en moras tributarias.

3.1 Planificación

En la etapa de planificación se ha creado un documento, en donde se detalla los objetivos generales y específicos, así como el plan de trabajo y el cronograma detallado.

3.2 Definición de requerimientos del negocio

Definición de requerimientos del negocio.

La definición de los requerimientos se obtuvo a partir de conversaciones entre la Dra. María Hallo, mi persona y en base a requerimientos e información del trabajo de titulación “Modelo para predecir las deudas de los contribuyentes con alta probabilidad de no pago en la administración tributaria” siendo el autor José Ordoñez. [23] Además, fue necesario investigar y revisar información y guías del Servicio de Rentas Internas del Ecuador. [24].

Para la creación del sistema de apoyo de decisiones se requiere resolver u obtener los siguientes requerimientos presentados en dashboards en la herramienta de visualización: ver tabla 1.

Tabla 1 Requerimientos

Identificador:	Nombre:
R-01	El valor total de las deudas por provincias
R-02	Porcentaje de obligaciones en remisión
R-03	Porcentaje de obligaciones impagas
R-04	Valor de impuesto generados por provincia
R-05	Valor en porcentaje de obligaciones con dificultad del recaudación
R-06	Número de contribuyentes por provincia
R-07	Número de contribuyentes activos por actividad económica.
R-08	Número de pagos realizados en intervalos de 30 días
R-09	Número de obligaciones en remisión por año
R-10	Valor de deuda por actividad económica
R-11	Valor de deuda por contribuyente
R-12	Valor de deuda por zonas
R-13	Número de obligaciones que se encuentran en remisión
R-14	Los Días transcurridos entre la fecha de generación de deuda hasta la fecha de pago

R-15	Los Días transcurridos entre la fecha de generación de deuda y el día de la consulta de información.
R-16	La fecha que se ha generado una deuda
R-17	La fecha que se ha pagado una deuda
R-18	Las provincias con mayor número obligaciones impagas
R-19	Pagos en remisión por tipo de impuesto
R-20	Número de obligaciones que se encuentra en revisión
R-21	Valor total de las deudas impagas

Fuente de datos

Las fuentes de datos usadas en el desarrollo del sistema de apoyo en la toma de decisiones según el documento “Modelo para predecir las deudas de los contribuyentes con alta probabilidad de no pago en la administración tributaria” son cuatro bases de datos de la institución tributaria [23]. Ver figura N° 3.

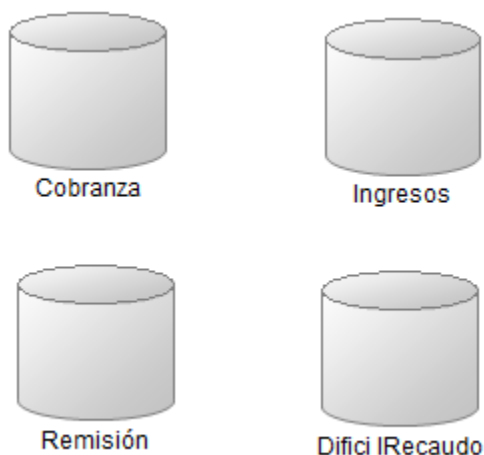


Figura 3 Fuente de datos

En la base de datos con nombre “Ingresos” se encuentra la información de ingresos y gastos de los contribuyentes.

La base de datos con nombre “Remisión” contiene todas las deudas que se acogieron al beneficio de condonación de multas en base a una normativa, aprobada mediante la ley. [23]

Además, la base de datos difícil recaudo que fue parte del análisis para este problema, contiene obligaciones donde la institución tributaria, a pesar de diferentes instancias legales y procedimiento de coactivas no logro recuperar los valores adeudados por un contribuyente. [23]

Otra base de datos necesaria para la generación del data mart es la de “Cobranza”, en donde se encuentra la información del contribuyente, valor de impuesto, valor de la multa, fechas de generación de deuda y pago entre otros campos más que serán descritos en la tabla N° 2

Tabla 2 Atributos de las fuentes de datos

Base de datos	Atributo	Tipo de Variable
Cobranza	Año de la deuda	Categórica
	Tipo de contribuyente	Categórica
	Clave de obligación	Clave primaria
	Departamento que genera la deuda	Categórica
	Clase de contribuyente	Categórica
	Estado del contribuyente	Categórica
	Actividad	Categórica
	Tipo de impuesto	Categórica
	Fecha de emisión de la deuda	Tiempo
	Fecha de pago	Tiempo
	Gran contribuyente	Categórica
	Mes de generación de la deuda	Categórica
	Valor del impuesto	Continua
	Valor de la multa	Continua
	Región a la que pertenece el contribuyente	Categórica
	Revisión del proceso	Categórica
	Registro del contribuyente	Clave foránea
Tipo de documento	Categórica	
Ingresos	Ingreso	Continua
	Registro del contribuyente	Clave primaria
Remisión	Pago en remisión	Categórica
	Código en remisión	Clave primaria
	Clave de obligación	Clave foránea

Difícil recaudo	Difícil recaudo	Categorica
	Código de difícil recaudo	Clave primaria
	Clave de obligación	Clave foránea

Además, es necesario informar que los datos para el “*data mart*” han sido obtenidos de la página del SRI y otros han sido simulados para el presente proyecto.

3.3 Dirección de datos preprocesamiento.

La dirección de datos preprocesamiento, describimos los datos a usar de las diferentes fuentes de datos antes de realizar algún proceso de extracción de datos. En la tabla 3 describe la base de datos cobranza, la tabla 4 la base de datos ingreso, la tabla 5 base de datos remisión, la tabla 6 la base de datos difícil recaudo. El detalle de los requerimientos están en el anexo 2. En cada tabla “PK” hace referencia a clave primaria y “FK “a clave foránea.

Tabla 3 Base de datos Cobranza

\	Atributo	Tipo de dato	Rango	PK	FK
Cobranza	Año de la deuda	Numeric	1 al 7	-	-
	Tipo de contribuyente	Varchar	Tipo de contribuyente	-	-
	Clave de obligación	Numeric	no aplica	PK	-
	Departamento que genera la deuda	Varchar	0 a 12	-	-
	Clase de contribuyente	Varchar	1 al 4	-	-
	Estado del contribuyente	Varchar	1 al 4	-	-
	Actividad	Varchar	1 al 26	-	-
	Tipo de impuesto	Varchar	1 al 4	-	-
	Fecha de emisión de la deuda	Date	desde 01/01/2014	-	-
	Fecha de pago	Date	desde 01/01/2014	-	-
	Gran contribuyente	Varchar	dos posibles valores	-	-
	Mes de generación de la deuda	Numeric	0 al 12	-	-
	Valor del impuesto	Decimal	Continuo	-	-

Valor de la multa	Decimal	Continuo	-	-
Región a la que pertenece el contribuyente	Varchar	1 al 9	-	-
Revisión del proceso	Varchar	dos posibles valores	-	-
Registro del contribuyente	Varchar	no aplica	-	FK
Tipo de documento	Varchar	no aplica	-	-

Tabla 4 Base de datos ingresos

Atributo	Tipo de dato	Rango	PK	FK
Ingreso	decimal	Continuo	-	-
Registro del contribuyente	varchar	no aplica	PK	-

Tabla 5 Base de datos remisión

Atributo	Tipo de dato	Rango	PK	FK
Pago en remisión	Varchar	dos posibles valores	-	-
Código en remisión	Varchar	no aplica	-	-
Clave de obligación	Numeric	no aplica	-	FK

Tabla 6 Base de datos difícil recaudo

Atributo	Tipo de dato	Rango	PK	FK
Difícil recaudo	varchar	dos posibles valores	-	-
Código de difícil recaudo	varchar	no aplica	-	-
Clave de obligación	numeric	no aplica	-	FK

3.4 Modelo dimensional

Luego de realizar la definición de requerimientos se procede analizar las medidas y dimensiones.

Elección de dimensiones

Para determinar las dimensiones del *data mart*, se identifica que variables son las que se necesita para el análisis y se las agrupa, cada variable es un atributo.

- Año de la deuda
- Tipo de contribuyente
- Clave de obligación
- Clase de contribuyente
- Estado del contribuyente
- Actividad
- Tipo de impuesto
- Fecha de emisión de la deuda
- Fecha de pago
- Gran contribuyente
- Mes de generación de la deuda
- Valor del impuesto
- Valor de la multa
- Región a la que pertenece el contribuyente
- Revisión del proceso
- Registro del contribuyente
- Ingreso
- Pago en remisión
- Díficil recaudo
- Clave de obligación
- Valor pago

Agrupamos las variables que tenga relación y se determina las dimensiones. Ver tabla 7

Tabla 7 Dimensiones

Dimensiones	Atributos
Fecha	Año de la deuda
	Fecha de emisión de la deuda
	Fecha de pago
	Mes de generación de la deuda
Contribuyente	Estado del contribuyente
	Registro del contribuyente
	Gran contribuyente
	Clase de contribuyente
	Tipo de contribuyente
Actividad económica	Actividad
Ubicación	Región a la que pertenece el contribuyente
Revisión de proceso	Revisión del proceso
Impuesto	Tipo de impuesto
Dificultad de recaudación	Difícil recaudo
Obligación	Clave de obligación
Pago en remisión	Pago en remisión

Medidas encontradas

El análisis realizado permite determinar las siguientes medidas:

- Valor de deuda
- Valor de multa
- Días fin
- Ingreso
- Valor Pago

3.5 Descripción de dimensiones

Dimensión fecha

La jerarquía de la dimensión fecha inicia con año, mes y por último día, siendo nuestro nivel más bajo y granularidad día.

La dimensión de fecha es un calendario a partir del 1 de enero del 2014, esta dimensión describe la fecha que se generó una deuda y la fecha en la que se ha pagado una deuda.

La dimensión fecha contribuye con el id para fecha de pago y el id fecha generación de deuda.

Dimensión contribuyente

En la dimensión contribuyente describe la información del contribuyente.

Dimensión actividad económica

La dimensión actividad describe las diferentes actividades económicas que existen.

Dimensión ubicación

La dimensión ubicación tiene jerarquía, empezando por la descripción de zona y luego por provincia.

Dimensión revisión de proceso

La dimensión de revisión de proceso describe si una obligación está en revisión

Dimensión impuesto

La dimensión impuesto nos describe el tipo de impuesto que pertenece una deuda.

Dimensión dificultad de recaudación.

La dimensión de dificultad de recaudación nos describe si una deuda es difícil o no de recaudar.

Dimensión obligación.

La dimensión obligación describe el tipo de obligación de una deuda.

Dimensión Pago en remisión

La dimensión pago en remisión describe si un pago está en estado de remisión o no, al estar en remisión significa que el contribuyente no debe cancelar el valor de esa deuda.

Tabla de hechos moras tributarias.

En esta tabla se encuentran las medidas de nuestro *data mart*:

- Valor deuda
- Valor impuesto
- Valor ingreso
- Días Fin
- Valor pago
- Valor multa

Gráfico dimensional de alto nivel

El gráfico dimensional de alto nivel nos permite describir las dimensiones y la tabla de hechos de nuestro *data mart*. Ver figura 3.

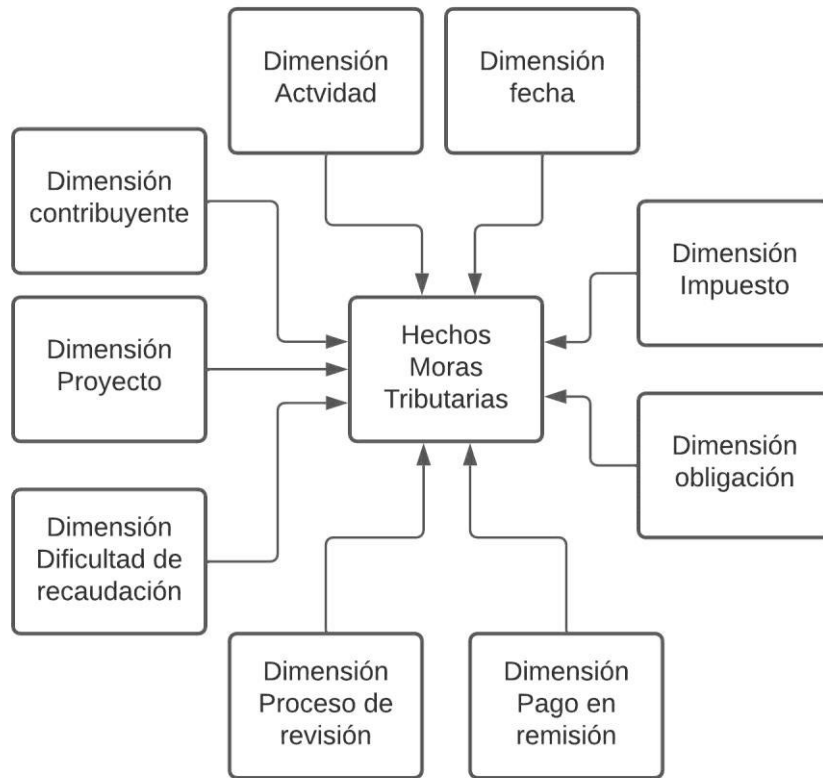


Figura 4 Gráfico dimensional de alto nivel

3.6 Arquitectura

La arquitectura para el desarrollo del sistema inicia con el origen de datos que son archivos exportados de base de datos con formato CSV, luego se procede a la extracción, transformación y limpieza de datos, para lo cual se usa la herramienta “Spoon” de “Pentaho”, dicha herramienta nos permite cargar los datos la base de datos realizada en PostgreSQL, y finalmente para la visualización de los datos de parte de los usuarios la herramienta a usar es Tableau. En la imagen 4 se observa la arquitectura a utiliza:



Figura 5 Arquitectura técnica

Las fuentes de datos dada por la institución tributaria son 4: cobranzas, ingreso, difícil recaudación y remisión.

Para la extracción de datos hemos usado la herramienta "Spoon" de la suite de Pentaho y se ha creado en la base de datos "MORAS TRIBUTARIAS" un esquema llamado DSA que es abreviatura de "data staging area" o área de análisis, se encuentran las siguientes tablas creadas cobranzas, ingreso, difícil recaudación y remisión, éstas tablas contienen datos temporales que serán luego ingresados a las diferentes dimensiones del data mart "Moras tributarias".

Para la creación dimensiones se ha generado un modelo lógico basado en los requerimientos del negocio.

La arquitectura utilizada para el *data mart* Moras Tributarias es en estrella.

La granularidad es diaria.

Las dimensiones creadas son las siguientes:

- La dimensión contribuyente "dim_contribuyente".
- La dimensión actividad económica "dim_actividad".
- La dimensión Fecha "dim_fecha".
- La dimensión ubicación "dim_ubicacion"

- La dimensión Dificultad de recaudación “dim_dificultad_recaudacion”
- La dimensión Proceso en revisión “dim_proceso_revision”
- La dimensión Impuesto “dim_impuesto”
- La dimensión Impuesto “dim_obligacion”
- La dimensión Impuesto “dim_pago_en_remision”
- Y la tabla de hechos: “fact_moras_tributarias”

Modelo lógico

El modelo lógico lo creamos después de conocer nuestras dimensiones y los campos de cada una de ellas. A continuación, el modelo lógico: Ver figura 6.

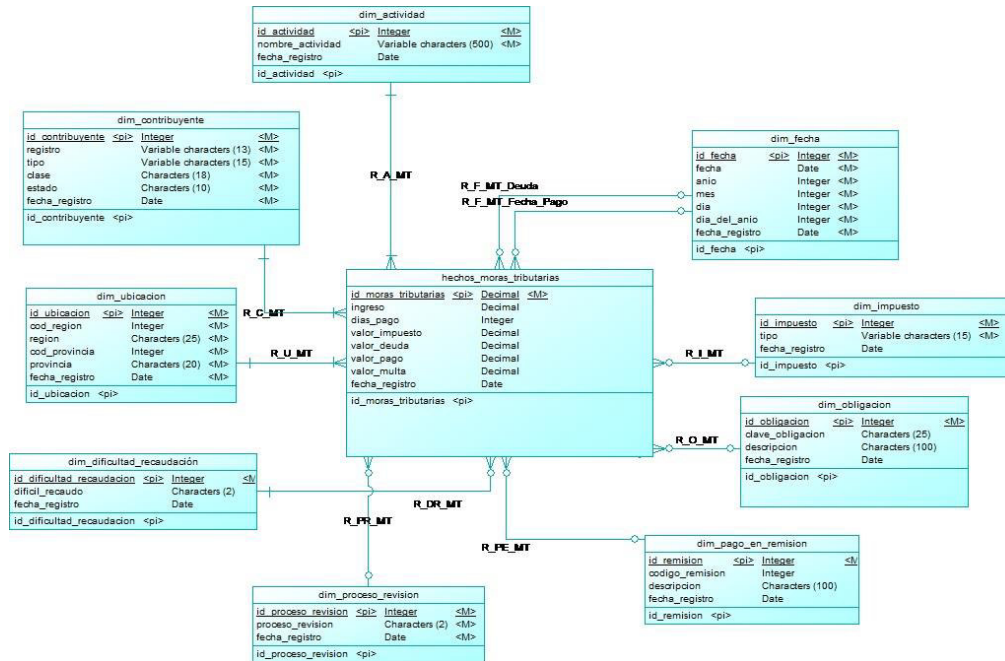


Figura 6 Modelo lógico

Modelo físico

El modelo físico se lo genera luego de tener un modelo conceptual completo. Ver figura 7.

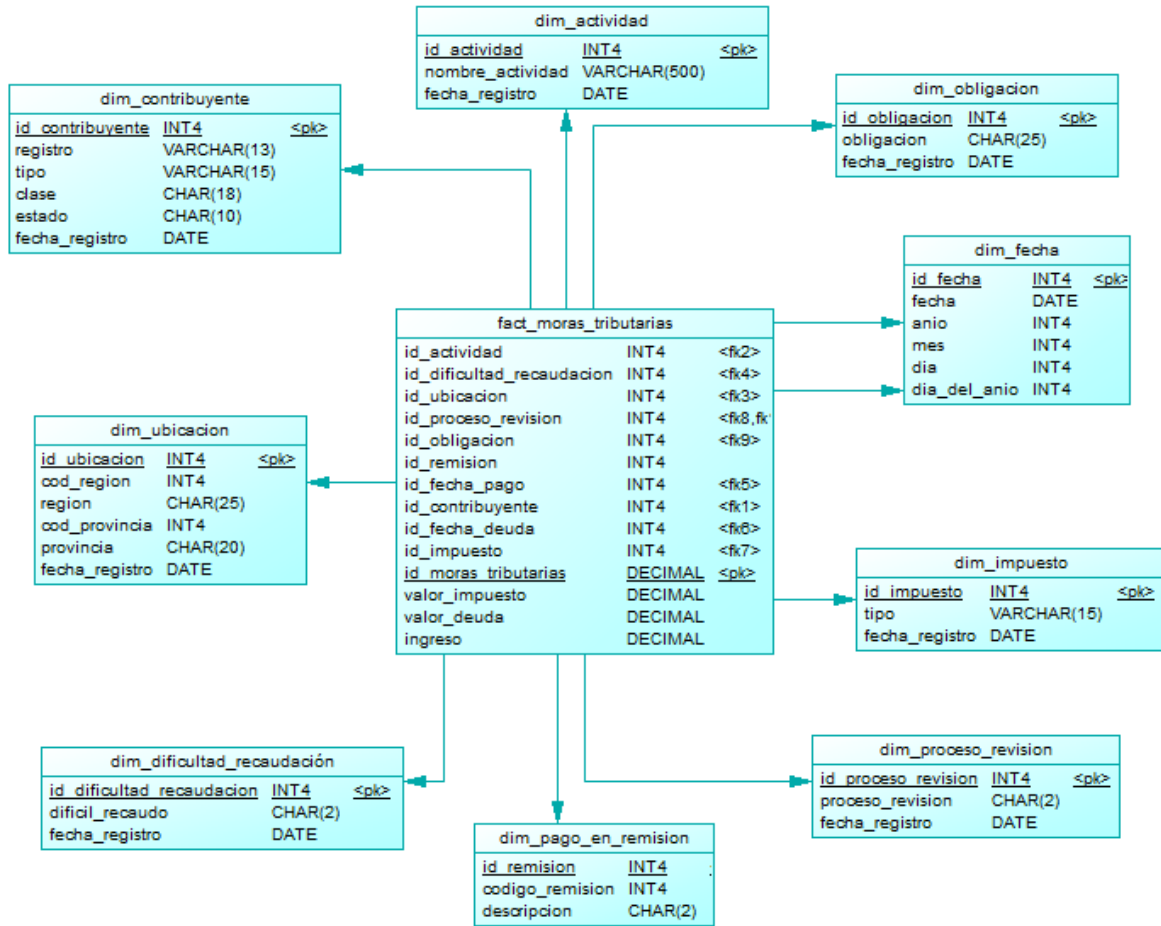


Figura 7 Modelo Físico

En la figura 8 describimos la dimensión contribuyente con sus campos, el formato y la clave primara.

dim_contribuyente		
<u>id_contribuyente</u>	INT4	<pk>
registro	VARCHAR(13)	
tipo	VARCHAR(15)	
clase	CHAR(18)	
estado	CHAR(10)	
fecha_carga	DATE	

Figura 8 Dimensión contribuyente

En la figura 9 describimos la dimensión actividad con sus campos, el formato y la clave primara.

dim_actividad		
<u>id_actividad</u>	INT4	<pk>
nombre_actividad	VARCHAR(500)	
fecha_carga	DATE	

Figura 9 Dimensión actividad

En la figura 10 describimos la dimensión fecha con sus campos, el formato y la clave primara.

dim_fecha		
<u>id_fecha</u>	INT4	<pk>
fecha	DATE	
año	INT4	
mes	INT4	
día	INT4	
fecha_carga	DATE	

Figura 10 Dimensión fecha

En la figura 11 describimos la dimensión impuesta con sus campos, el formato y la clave primara.

dim_impuesto		
<u>id_impuesto</u>	INT4	<pk>
tipo	VARCHAR(15)	
fecha_carga	DATE	

Figura 11 Dimensión impuesto

En la figura 12 describimos la dimensión proceso revisión con sus campos, el formato y la clave primara.

dim_proceso_revision		
<u>id_proceso_revision</u>	INT4	<pk>
proceso_revision	CHAR(2)	
fecha_carga	DATE	

Figura 12 Dimensión proceso revisión

En la figura 13 describimos la dimensión dificultad recaudación con sus campos, el formato y la clave primara.

dim_dificultad_recaudación		
<u>id_dificultad_recaudacion</u>	INT4	<pk>
dificil_recaudo	CHAR(2)	
fecha_carga	DATE	

Figura 13 Dimensión dificultad recaudación

En la figura 14 describimos la dimensión ubicación con sus campos, el formato y la clave primara.

dim_ubicacion		
<u>id_ubicacion</u>	INT4	<pk>
cod_region	INT4	
region	CHAR(25)	
cod_provincia	INT4	
provincia	CHAR(20)	
fecha_carga	DATE	

Figura 14 Dimensión ubicación

En la figura 15 describimos la dimensión ubicación con sus campos, el formato y la clave primara.

dim_obligacion			
<u>id_obligacion</u>	<pi>	Integer	<M>
clave_obligacion		Characters (25)	
descripcion		Characters (100)	
fecha_registro		Date	
<u>id_obligacion</u> <pi>			

Figura 15 Dimensión Obligación

En la figura 16 describimos la dimensión ubicación con sus campos, el formato y la clave primara.

dim_pago_en_remision			
<u>id_remision</u>	<pi>	Integer	<M>
codigo_remision		Integer	
descripcion		Characters (2)	
fecha_registro		Date	
<u>id_remision</u> <pi>			

Figura 16 Dimensión pago en remisión

En la tabla de hechos podemos observar en la figura 17.

hechos_moras_tributarias		
<u>id_moras_tributarias</u>	DECIMAL	<pk>
id_contribuyente	INT4	<fk1>
id_actividad	INT4	<fk2>
id_dificultad_recaudacion	INT4	<fk4>
id_ubicacion	INT4	<fk3>
id_proceso_revision	INT4	<fk8>
id_obligacion	INT4	<fk9>
id_remision	INT4	<fk10>
id_fecha_generacion_deuda	INT4	<fk5>
id_fecha_pago	INT4	<fk6>
id_impuesto	INT4	<fk7>
ingreso	DECIMAL	
dias_pago	INT4	
valor_impuesto	DECIMAL	
valor_deuda	DECIMAL	
valor_pago	DECIMAL	
valor_multa	DECIMAL	
fecha_registro	DATE	

Figura 17 Tabla de hechos

3.7 Diseño e implementación del subsistema del ETL.

La creación el esquema “DSA” se realiza para la extracción y transformación de datos que permanecen en 4 tablas “Cobranza”, “Ingreso”, “Remisión” y “Difícil Recaudo”. Aquellos datos serán borrados una vez que han sido cargados al data mart “Moras Tributarias”. Para cada tabla se ha creado una transformación en la cual ingresa los datos en formato CVS se usa pasos para modificar según los requerimientos se crea una tabla dentro del esquema DSA en la Base de datos Postgresql, la cual es administrada por PgAdmin 4.0. Ver figura 18.

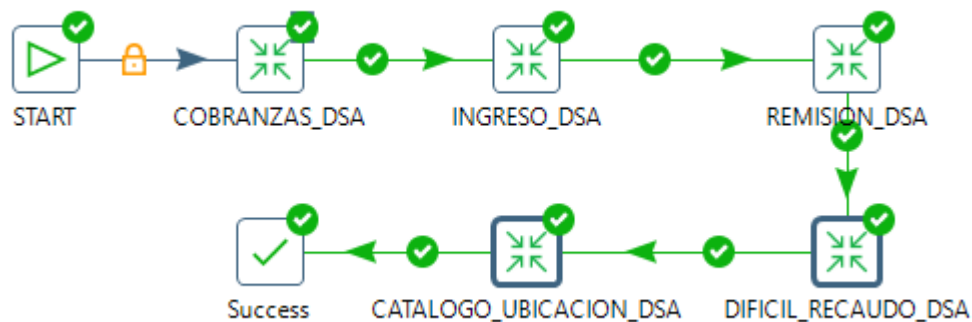


Figura 18 ETL DSA

En la figura 19 se observa los pasos realizados en Pentaho para cargar en nuestra área de validación la base de datos cobranza.

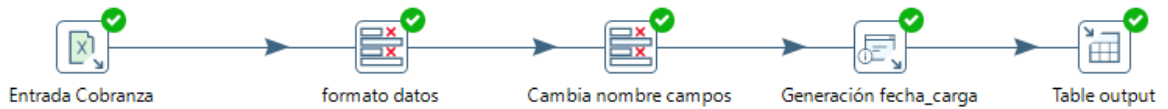


Figura 19 ETL DSA Cobranza

En la figura 20 se observa los pasos realizados en Pentaho para cargar en nuestra área de validación la base de datos Ingreso.

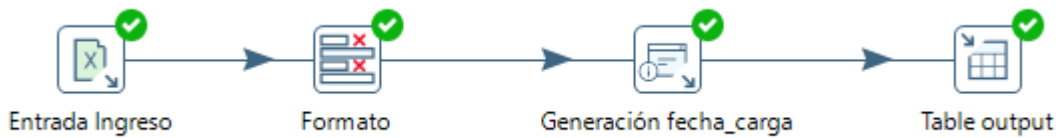


Figura 20 ELT DSA Ingreso

En la figura 21 se observa los pasos realizados en Pentaho para cargar en nuestra área de validación la base de datos remisión.

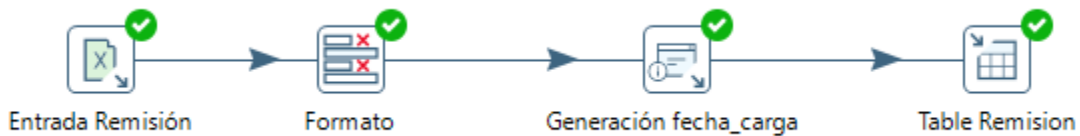


Figura 21 ETL DSA Remisión

En la figura 22 se observa los pasos realizados en Pentaho para cargar en nuestra área de validación la base de datos difícil recaudó.

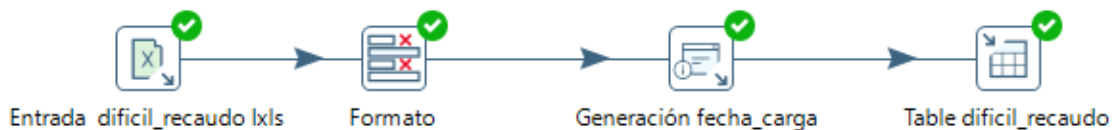


Figura 22 ETL DSA difícil recaudo

Los pasos para continuar con la carga de datos del DWH son los siguientes:

Los pasos para cargar todos los datos en el DWH realizado en PostgreSQL lo describimos en la figura 23.

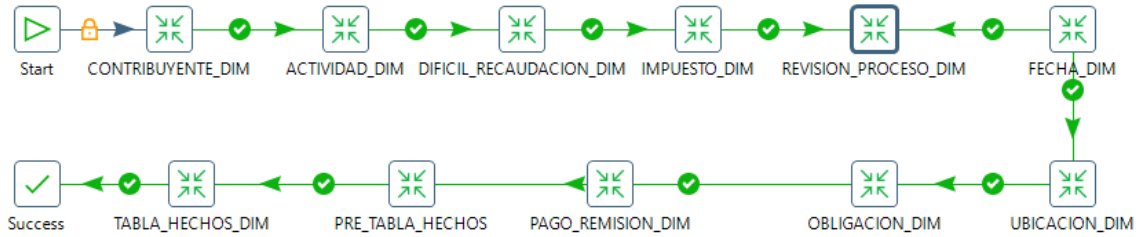


Figura 23 ETL DWH

Los pasos para cargar la dimensión contribuyente en el DWH se describen en la figura 24.

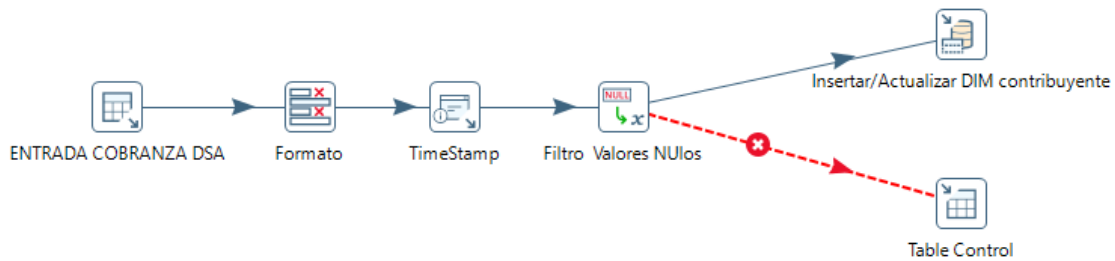


Figura 24 ETL DWH Dimensión contribuyente

Los pasos para cargar la dimensión actividad en el DWH se describe en la figura 25.

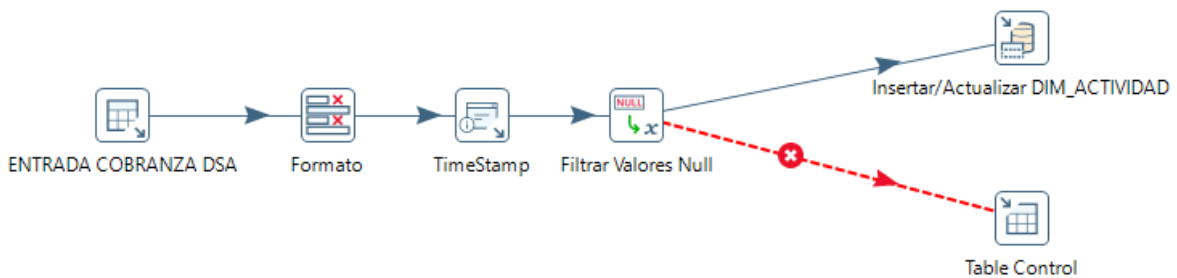


Figura 25 ETL DWH Dimensión actividad

Los pasos para cargar la dimensión difícil recaudo en el DWH se describen en la figura 26.

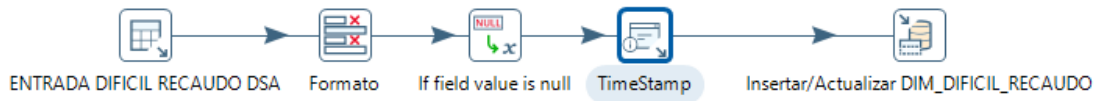


Figura 26 ETL DWH difícil recaudo

Los pasos para cargar la dimensión impuesto en el DWH se describen en la figura 27.



Figura 27 ETL DWH Dimensión impuesto

Los pasos para cargar la dimensión revisión proceso en el DWH se describen en la figura 28.



Figura 28 ETL DWH Dimensión revisión de proceso

Los pasos para cargar la dimensión fecha en el DWH se describen en la figura 29.

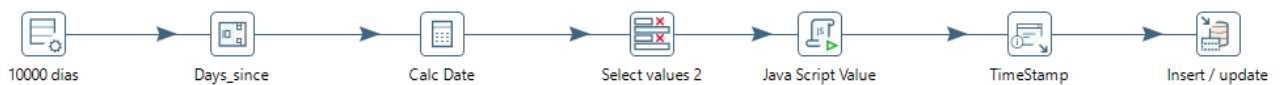


Figura 29 ETL DWH Dimensión fecha

Los pasos para cargar la dimensión ubicación en el DWH se describen en la figura 30.



Figura 30 ETL DWH Dimensión ubicación.

Los pasos para cargar la dimensión obligación en el DWH se describen en la figura 31.

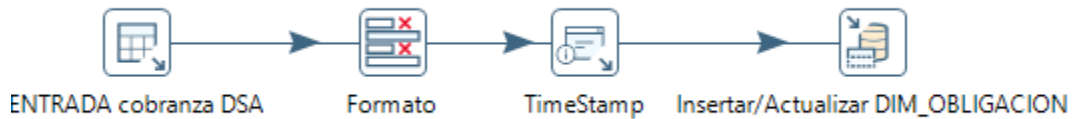


Figura 31 ETL DWH Dimensión obligación.

Los pasos para cargar la dimensión Pago Remisión en el DWH se describen en la figura 32.



Figura 32 ETL DWH Dimensión Pago remisión.

AL realizar la carga en la tabla de hechos se realiza una carga con los datos a usar en la tabla de hechos para luego comparar con los datos de cada dimensión para obtener los id. En la figura 33 describimos este proceso.



Figura 33 ETL DWH Pre-tabla hechos

Para la creación de la table de Hechos “FACT_MORAS_TRIBUTARIAS” se necesita crear una transformación en la herramienta SPOON de la suite de

Pentaho al cual se la nombra como “T_FAC_MORAS_TRIBUTARIAS”. Tal como se observa en la figura 34.

El id de cada dimensión del esquema DWH de la base de datos “MORAS TRIBUTARIAS”:

- Id_actividad
- Id_dificultad_recaudación
- Id_region
- Id_proceso_revision
- Id_fecha_deuda
- Id_contribuyente
- Id_fecha_pago
- Id_impuesto
- Id_obigacion

Los siguientes valores se encuentra en la tabla “DSA_COBRANZA” que se encuentra en el esquema DSA en la base de datos “MORAS TRIBUTARIAS”:

- valor_impuesto
- valor_deuda
- ingreso
- El valor de Pago en remisión se lo obtiene de la tabla “DSA_REMISION”

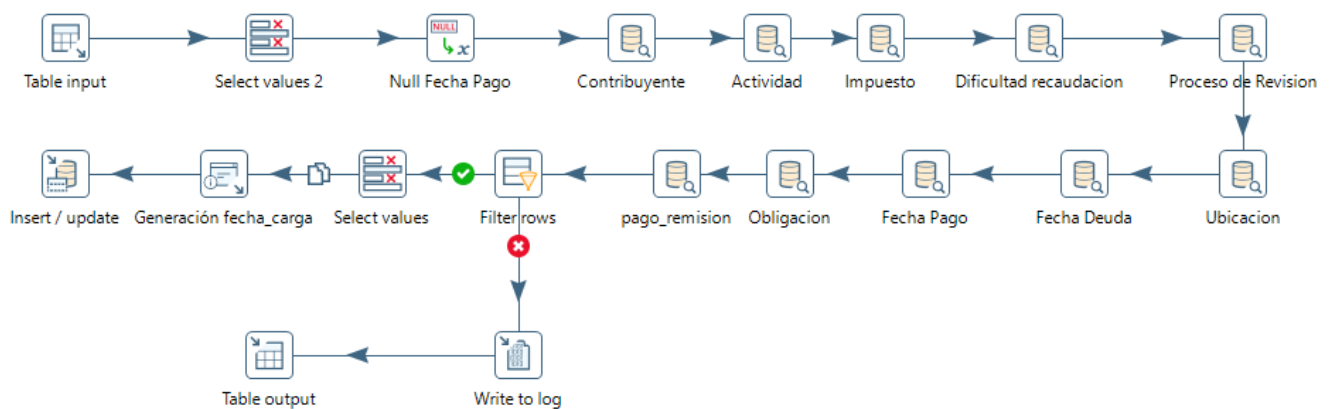


Figura 34 ETL DWH Tabla de hechos

Hemos realizado un trabajo en el cual se encuentra automatizado el proceso de ETL. Ver figura 35.



Figura 35 ETL DWH.

4 RESULTADO OBTENIDOS Y DISCUSIÓN

Los resultados obtenidos a partir de los datos simulados se pueden evidenciar utilizando la herramienta Tableau, en donde visualizamos graficas que aportan a la toma de decisiones. Además, se pueden graficar y visualizar los requerimientos dados en los anexos, dos ejemplos de los gráficos que podemos obtener se observan en la figura 36 y figura 37.

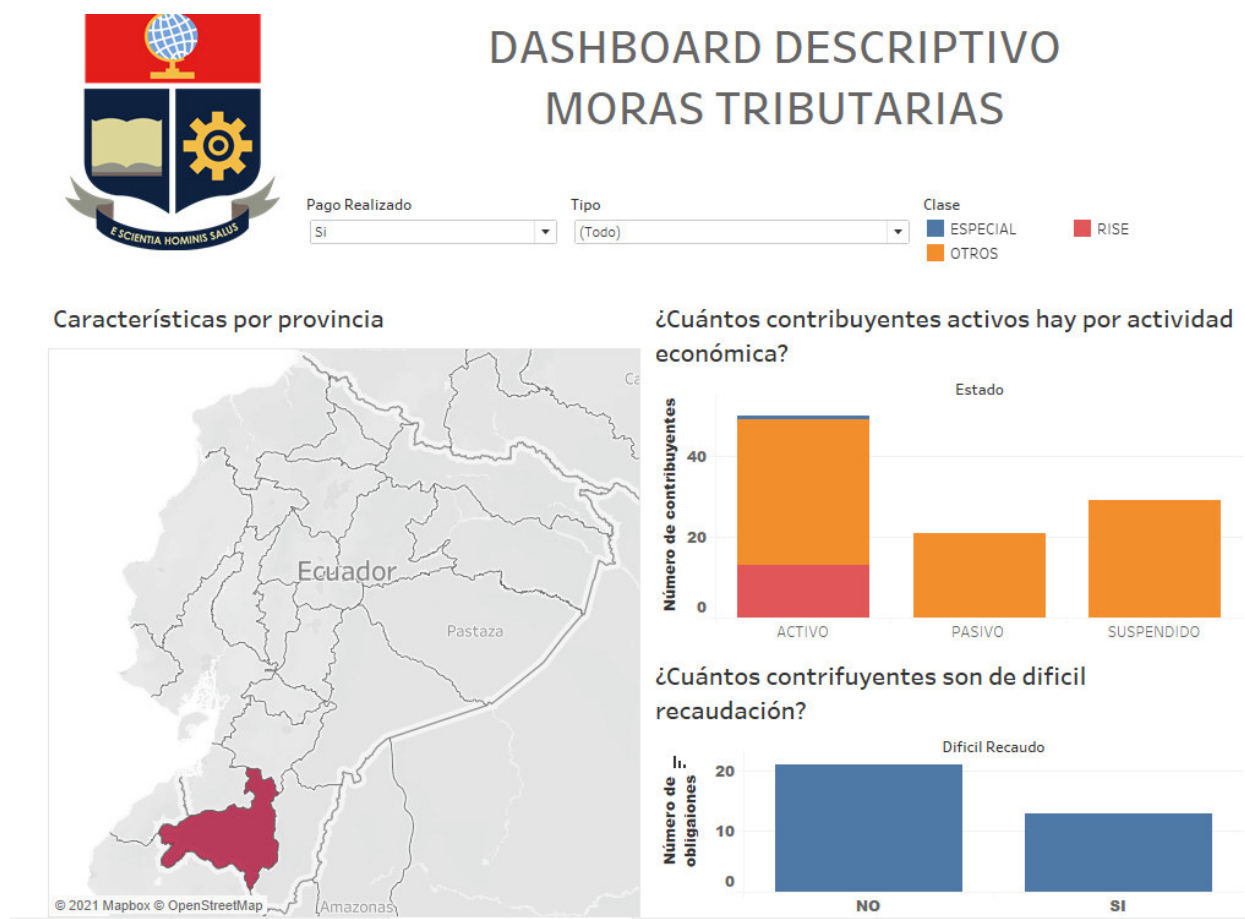


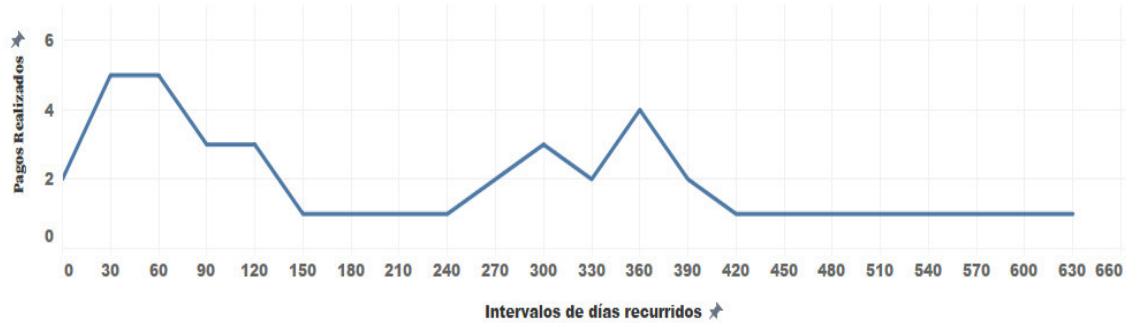
Figura 36 Dashboard descriptivo



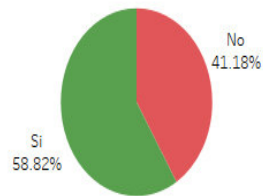
DASHBOARD PREDICTIVO MORAS TRIBUTARIAS

Tipo (Dim Impuesto) Pago Realizado Pago Realizado No Si No Si Pago Remisino Discreto Año de Fecha

Pagos realizados en diferentes intervalos de días



Porcentaje de Obligaciones pagadas



Porcentaje de obligaciones están en remisión?

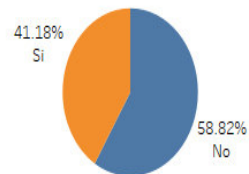


Figura
37 Dashboard predictivo

5 CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

El sistema de apoyo en la toma de decisiones, desarrollado con la metodología Kimball y con datos simulados nos permitió visualizar datos a través de *dashboards* realizados en la aplicación Tableau, obteniendo información relevante como ver que provincias tienen mayor deuda y que provincias el valor de recaudación es mayor. Además, en qué provincia se encuentra la mayor cantidad de obligaciones con dificultad de recaudo, contribuyendo a la toma de decisiones estratégicas como la intensificación de cobro de impuesto en una provincia.

En el proyecto se identificó cuatro bases de datos transaccionales como origen de datos para el desarrollo del sistema: contribuyente, ingreso, impuesto. Y, para la creación del *data mart* es necesario el proceso de extracción, carga y limpieza de los datos para al final lograr la visualización de información.

5.2 Recomendaciones

A futuro, si existen otros *data mart* en la institución, se podrían integrar y construir un *data warehouse* permitiendo que el alcance del sistema sea mayor.

Además, con los datos brindados por el *data mart* se puede encontrar correlaciones entre las variables para poder determinar patrones entre las deudas impagas con otras dimensiones, así predecir futuros morosos.

Se puede implementar el *Data mart* en servidores públicos de Tableau para que sean visualizados desde cualquier parte que se tenga acceso a internet.

6 BIBLIOGRAFÍA

- [1] R. K. a. J. Caserta, *The Data Warehouse*, Indianapolis: Wiley Publishing, 2004.
- [2] J. G. a. D. Morán, *Política tributaria en América Latina: agenda para una segunda generación*, Santiago de Chile: Naciones Unidas, 2013.
- [3] P. G. a. H. Romo, «Tributación en el contexto económico de América Latina,» *Revista Publicando*, vol. III, pp. 467 - 478, 2017.
- [4] Comisión Económica para América Latina y el Caribe (CEPAL), *Panorama Social de América Latina*, Santiago de Chile: Naciones Unidas, 2019.
- [5] R. P. a. U. Troiano, «Shaming tax delinquents,» *Journal of Public conomics*, vol. 167, pp. 120-137, 2018.
- [6] M. Ruiz, «Using stretegic tools for the detection of tax fraud,» *Investigación, Ciencia y Universidad (ICU)*, vol. 1, nº 1, pp. 57-63, 2016.
- [7] G. K. & G. J. B. G., «Decision Making and reduction techniques for tax revenue using data warehouse,» *International Journal of ENgineering trends and Technology (IJETT)*, vol. 3, nº 1, pp. 24-31, 2012.
- [8] R. Kimball, «The Future Is Bright,» Kimball Group, [En línea]. Available: <http://www.kimballgroup.com/2015/12/design-tip-180-the-future-is-bright/>. [Último acceso: 05 Junio 2020].
- [9] A. TÖLE, «The Importance of Data Warehouses in the Development of Computerized Decision Support Solutions. A Comparison between Data Warehouses and Data Marts,» *Database Systems Journal*, vol. VI, nº 4, pp. 19-26, 2015.
- [10] A. B. a. N. S. L. Brynne, «Effective visualization of integrated knowledge and data to enable informed decisions in drug development and translational medicine,» *Journal of translational medicine*, vol. XI, nº 1, p. 250, 2013.
- [11] T. Bucher, «Process-centric buusiness intelligence,» *Business Process Managment Journal*, vol. 15, nº 3, pp. 409-429, 2009.
- [12] R. Rodríguez, «Propuesta del nuevo Sistema de Información de Planes de Estudio de la Universidad de,» Costa Rica, 2009.
- [13] L. C. & J. Abreu, «El impacto de las herramientas de inteligencia de negocios en la toma de decisiones de los ejecutivos,» *Daema: International Journal of Good Conscience*, vol. IV, nº 2, pp. 16-52, 2009.
- [14] D. Cohen, *Sistema de información para los negocios*. 3ª edición, México : McGraw-Hill, 2000.
- [15] E. L. M. & M. S. Vitt, *Business Intelligence.Técnicas de análisis*, España: McGraw-Hill, 2003.
- [16] Novalys, «powerdesigner,» Novalys, 2019. [En línea]. Available: <https://www.powerdesigner.biz/ES/>. [Último acceso: 06 06 2020].
- [17] Hitachi, «Hitachivantara,» www.hitachivantara.com, 2020. [En línea]. Available: <https://www.hitachivantara.com/en-us/home.html>. [Último acceso: 06 06 2020].
- [18] PostgreSQL Community Association of Canada,, «pgadmin.org,» PostgreSQL Community Association of Canada,, 2020. [En línea]. Available: <https://www.pgadmin.org/>. [Último acceso: 06 06 2020].
- [19] The PostgreSQL Global Development Group, «postgresql.org,» The PostgreSQL Global Development Group, 2020. [En línea]. Available: <https://www.postgresql.org/>. [Último acceso: 06 06 2020].

- [20] TABLEAU SOFTWARE, LLC, A SALESFORCE COMPANY., «tableau.com,» 2020. [En línea]. Available: <https://www.tableau.com/>. [Último acceso: 20 06 2020].
- [21] Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., ... Wiesel, T., «Dashboards as a Service: Why, What, How, and What Research Is Needed?,» *Journal of Service Research*, p. 175–189, 2009.
- [22] R. Kimball, *The Data Warehouse Lifecycle Toolkit*, Indianapolis: Wiley Publishing, Inc., 2009.
- [23] J. Ordoñez, «Modelo para predecir las deudas de los contribuyentes con alta probabilidad de no pago en la administración tributaria,» *Trabajo de Titulación previo a la obtención del grado de magister en ciencias de computación*, 2020.
- [24] Gobierno de la República del Ecuador, «Guías para el contribuyente,» [En línea]. Available: <https://www.sri.gob.ec/guias-tributarias>. [Último acceso: 06 2021].

7 ANEXOS

Anexo 1 Cronograma

Actividades	Mes 1		Mes 2		Mes 3		Mes 4		Mes 5		Mes 6		Horas						
	Semana				Semana				Semana										
	1	2	3	4	1	2	3	4	1	2	3	4		1	2	3	4		
Planteamiento del problema.		x	x														20		
Planificación del proyecto.				x	x													20	
Recopilación de información					x	x											20		
Definición de requerimientos para el sistema.						x											10		
Diseño del modelo dimensional.							x										10		
Definición de la arquitectura.							x										10		
Inducción de las herramientas a utilizar.								x									10		
Instalación y configuración de herramientas.									x								10		
Construcción del <i>data mart</i>										x	x	x	x	x			45		
Implementación.													x	x	x	x	x	x	60
Pruebas																	x	x	20
Elaboración de la documentación,		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	180
Entrega de la documentación.																		x	10
Horas Totales: 425																			

Anexo 2

Tabla 8 Requerimiento N°1

Identificador:	R-01	Nombre:	El valor total de las deudas por provincias
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el monto de deudas generadas por cada provincia.		

Tabla 9 Requerimiento N°2

Identificador:	R-02	Nombre:	Porcentaje de obligaciones en remisión
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el porcentaje sobre el total de obligaciones generadas por contribuyentes están en estado de remisión.		

Tabla 10 Requerimiento N°3

Identificador:	R-03	Nombre:	Porcentaje de obligaciones en pagadas
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el porcentaje sobre el total de obligaciones generadas han sido pagadas por los contribuyentes.		

Tabla 11 Requerimiento N°4

Identificador:	R-04	Nombre:	Valor de impuesto de por provincia
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor del monto de impuestos generados por provincia.		

Tabla 12 Requerimiento 5

Identificador:	R-05	Nombre:	Valor en porcentaje de obligaciones por difícil recaudación
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el porcentaje de obligaciones que tienen dificultad recaudatoria.		

Tabla 13 Requerimiento N°6

Identificador:	R-06	Nombre:	Número de contribuyentes por provincia
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el número total de contribuyentes que se han registrado en cada provincia		

Tabla 14 Requerimiento N°7

Identificador:	R-07	Nombre:	Número de contribuyentes activos por actividad económica.
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el número total de contribuyentes según su actividad económica registrada y que actualmente se encuentren en estado activo		

Tabla 15 Requerimiento N°8

Identificador:	R-08	Nombre:	Número de pagos realizados en intervalos de 30 días
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el número total de obligaciones pagadas según el tiempo que se han demorado en pagar en intervalos de 30 días		

Tabla 16 Requerimiento N°9

Identificador:	R-09	Nombre:	Número de obligaciones en remisión por año
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el número total de deudas remitidas por cada año.		

Tabla 17 Requerimiento N°10

Identificador:	R-10	Nombre:	Valor de deuda por actividad económica
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor de deudas por actividad económica		

Tabla 18 Requerimiento N°11

Identificador:	R-11	Nombre:	Valor de deuda por contribuyente
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor de deudas por contribuyente		

Tabla 19 Requerimiento N°12

Identificador:	R-12	Nombre:	Valor de deuda por zonas
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor total de deudas por zonas		

Tabla 20 Requerimiento N°13

Identificador:	R-13	Nombre:	Número de obligaciones que se encuentran en remisión
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor total de deudas por zonas		

Tabla 21 Requerimiento N°14

Identificador:	R-14	Nombre:	Los Días transcurridos entre la fecha de generación de deuda hasta la fecha de pago
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer cuántos días han transcurrido desde el día de generación de deuda hasta el día de pago en deudas que han sido pagadas		

Tabla 22 Requerimiento N°15

Identificador:	R-15	Nombre:	Los Días transcurridos entre la fecha de generación de deuda y el día de la consulta de información.
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer cuántos días han transcurrido desde el día de generación hasta el día de generación de reporte para deudas impagas.		

Tabla 23 Requerimiento N°16

Identificador:	R-16	Nombre:	La fecha que se ha generado una deuda
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el día exacto que se ha generado una deuda		

Tabla 24 Requerimiento N°17

Identificador:	R-17	Nombre:	La fecha que se ha pagado una deuda
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el día exacto que ha sido pagada una deuda		

Tabla 25 Requerimiento N°18

Identificador:	R-18	Nombre:	Las provincias con mayor número obligaciones impagas
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer que provincias tienen un mayor número de obligaciones pendientes por pagar		

Tabla 26 Requerimiento N°19

Identificador:	R-19	Nombre:	Valor de deuda por tipo de impuesto
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor total de deudas por impuesto.		

Tabla 27 Requerimiento 20

Identificador:	R-20	Nombre:	Obligaciones en remisión por tipo de impuesto
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor total de deudas por impuesto.		

Tabla 28 Requerimiento N°21

Identificador:	R-21	Nombre:	Pagos en remisión por tipo de impuesto
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor total de deudas por impuesto.		

Tabla 29 Requerimiento N°22

Identificador:	R-22	Nombre:	Número de obligaciones que se encuentra en revisión
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer cuántas obligaciones se encuentran en proceso de revisión		

Tabla 30 Requerimiento N°23

Identificador:	R-23	Nombre:	El valor deuda por tipo de contribuyente
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor total de deuda que se ha generado por cada tipo de contribuyente.		

Tabla 31 Requerimiento N°24

Identificador:	R-24	Nombre:	El valor total de las obligaciones impagas
Tipo:	Funcional	Fecha:	20-01-2020
Prioridad:	Alta	Necesidad:	Si
Descripción:	La solución de inteligencia de negocios permitirá conocer el valor total de obligaciones que están impagas		