

**UNIVERSIDAD NACIONAL DE CHIMBORAZO**



**FACULTAD DE INGENIERÍA**

**CARRERA DE INGENIERÍA SISTEMAS Y COMPUTACIÓN**

Proyecto de Investigación previo a la obtención del título de Ingeniero en Sistemas y Computación

**TRABAJO DE TITULACIÓN**

**DISEÑO Y GENERACIÓN DEL ALMACÉN DE DATOS APLICANDO ETL PARA INSTITUCIONES DE SALUD.**

**Autor:**

Jorge Luis Ortiz Paguay

**Tutor:**

PhD. Fernando Molina

**Riobamba - Ecuador:**

Año 2021

## PÁGINA DE ACEPTACIÓN

Los miembros del tribunal de Graduación del proyecto de investigación de título:

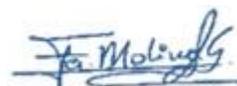
**“DISEÑO Y GENERACIÓN DEL ALMACÉN DE DATOS APLICANDO ETL PARA INSTITUCIONES DE SALUD.”**, presentado por el estudiante el Sr. Jorge Luis Ortiz Paguay, dirigido por el PhD. Fernando Molina.

Una vez escuchada la defensa oral y revisado el informe final del proyecto de investigación escrita, con fines de graduación en el cual se ha constatado el cumplimiento de las observaciones realizadas, remite la presente para uso y custodia en la biblioteca de la Facultad de Ingeniería de la UNACH.

Para constancia de lo expuesto firman:

PhD. Fernando Molina

**Tutor del Proyecto**



.....  
**Firma**

Mgs. Wayner Bustamante

**Miembro del Tribunal**



.....  
**Firma**

PhD. Ximena Quintana

**Miembro del Tribunal**



.....  
**Firma**

## **DERECHO DE AUTORÍA**

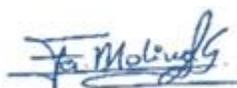
La responsabilidad del contenido de este proyecto de graduación corresponde exclusivamente a: Jorge Luis Ortiz Paguay bajo la dirección del Mgs. Fernando Molina, y al patrimonio intelectual de la Universidad Nacional de Chimborazo.

### **Autor**



.....  
Jorge Luis Ortiz Paguay

### **Director del Proyecto**



.....  
PhD. Fernando Molina

## **DEDICATORIA**

Esta tesis está dedicada a mi Madre  
quien me ilustró con el mejor conocimiento  
de todos “aprender de nuestros errores”,  
frase que me motivó a siempre buscar la mejor  
versión de mí, que con su apoyo incondicional  
y gran esfuerzo me enseñó que todo es posible  
sin importar las trabas, y mi hija que fue el motor principal  
para cada día a día seguir esforzándome.

## **AGRADECIMIENTO**

Principalmente agradezco a Dios por todas las

bendiciones que he recibido.

Un agradecimiento especial a mi tutor

PhD Fernando Molina y colaboradores

MsC Wayner Bustamante, PhD Ximena Quintana

por el apoyo brindado a lo largo de este proyecto de investigación.

A los docentes de la carrera que siempre

fueron una guía y un ejemplo de superación.

A mi familia por estar siempre apoyándome

incondicionalmente en la etapa de mi formación profesional.

## ÍNDICE

<b>RESUMEN</b> .....	<b>XII</b>
1. INTRODUCCIÓN .....	1
1. Capítulo I .....	3
1.1. Problema y justificación.....	3
1.2. Justificación de la investigación.....	4
1.3. Viabilidad de la investigación .....	5
1.4. Objetivos .....	5
1.4.1. Objetivo General.....	5
1.4.2. Objetivos específicos .....	5
2. Capítulo II.....	6
2.1. Almacén de datos .....	6
2.1.1. Características.....	6
2.2. Data mart.....	7
2.2.1. Data mart OLAP.....	7
2.2.2. Data mart OLTP.....	7
2.3. Proceso ETL.....	8
2.4. Data Mining o Minería de Datos.....	9
2.4.1. Beneficios .....	10
2.5. Inteligencia de negocios.....	10
2.6. Big Data.....	11

2.7.	Analítica de Big Data .....	11
2.8.	Big Data en el sector de la salud .....	12
2.8.1.	The Next Frontier for Innovation, Competition and Productivity .....	13
2.8.2.	Healthcare Hype and Hope .....	13
2.9.	Big data en nuestro sistema sanitario (Ecuador) .....	14
2.10.	Aporte del Big Data .....	15
3.	Capítulo III.....	17
3.1.	Tipo y Diseño de Investigación.....	17
3.1.1.	Enfoque de la Investigación.....	17
3.2.	Unidad de análisis .....	17
3.2.1.	Variables Secundarias o dependientes .....	18
3.2.2.	Variable principal o independiente .....	18
3.2.3.	Técnicas de recolección de Datos.....	18
3.2.4.	Técnicas de Análisis e interpretación de la información .....	18
3.3.	Población y Muestra.....	18
3.4.	Operacionalización de variables.....	19
3.5.	MÉTODO APLICADO.....	20
3.6.	Procedimientos .....	21
3.6.1.	Empresa analizada .....	21
3.6.2.	Análisis de Requerimientos .....	21

3.6.3.	Análisis de los OLTP .....	23
3.6.4.	Modelo Lógico del DW .....	37
3.6.5.	Integración de Datos .....	41
3.6.6.	Análisis de Información.....	52
3.6.7.	Validación.....	56
3.7.	Comprobación de hipótesis .....	60
4.	Capítulo IV RESULTADOS.....	63
4.1.	Resultados .....	63
4.2.	Conclusiones .....	67
4.3.	Recomendaciones.....	69
5.	REFERENCIAS BIBLIOGRAFÍA .....	70
6.	ANEXOS .....	73

## ÍNDICE DE TABLAS

<b>Tabla 1</b> Comparación entre dos autores más reconocidos en Data Warehouse .....	16
<b>Tabla 2</b> Operacionalización de variables .....	19
<b>Tabla 3</b> Correspondencia entre los modelos de variables .....	28
<b>Tabla 4</b> Descripción de figuras para el modelo direccional .....	29
<b>Tabla 5</b> Descripción de dimensiones del DW .....	31
<b>Tabla 6</b> Descripción de variables de la dimensión Paciente. ....	32
<b>Tabla 7</b> Descripción de variables de la dimensión Médico. ....	33
<b>Tabla 8</b> Descripción de variables de la dimensión Establecimiento. ....	34
<b>Tabla 9</b> Descripción de variables de la dimensión Diagnóstico. ....	35
<b>Tabla 10</b> Descripción de variables de la dimensión Tiempo. ....	36
<b>Tabla 11</b> Descripción de variables de la fact table Atención Médica. ....	36
<b>Tabla 12</b> Descripción de valores de indicadores .....	61
<b>Tabla 13</b> Análisis de varianza de un factor .....	61
<b>Tabla 14</b> Análisis de varianza .....	62

## ÍNDICE DE FIGURAS

<b>Figura 1</b> Esquema tres fases de un ETL .....	9
<b>Figura 2</b> Esquema de la Metodología Hefesto.....	20
<b>Figura 3</b> Estructura Orgánica Zonal.....	21
<b>Figura 4</b> Esquema del modelo conceptual. ....	23
<b>Figura 5</b> Variables Fuente de datos Rdacaa V 1.6. ....	24
<b>Figura 6</b> Variables Fuente de datos Rdacca V 2.0.....	25
<b>Figura 7</b> Variables Fuente de datos Pras.....	26
<b>Figura 8</b> Variables Fuente de datos Pras.....	27
<b>Figura 9</b> Dimensiones y tabla de hechos. ....	30
<b>Figura 10</b> Modelo conceptual Ampliado. ....	37
<b>Figura 11</b> Dimensión Paciente.....	38
<b>Figura 12</b> Dimensión Médico. ....	38
<b>Figura 13</b> Dimensión Establecimiento.....	39
<b>Figura 14</b> Dimensión Diagnóstico. ....	39
<b>Figura 15</b> Dimensión Tiempo.....	40
<b>Figura 16</b> Fact-table Atención Médica .....	40
<b>Figura 17</b> Uniones de dimensiones.....	41
<b>Figura 18</b> Data Preparation.....	42
<b>Figura 19</b> Limpieza de datos.....	43
<b>Figura 20</b> Sistema de gestión de bases de datos Postgresql 9.6.....	44
<b>Figura 21</b> Conexión de PostgreSQL 9.6. y Pentaho Data Integration. ....	44
<b>Figura 22</b> ETL Paciente. ....	45

<b>Figura 23</b> ETL Médico. ....	45
<b>Figura 24</b> ETL Establecimiento.....	45
<b>Figura 25</b> ETL Diagnóstico. ....	46
<b>Figura 26</b> ETL Tiempo. ....	46
<b>Figura 27</b> ETL Atención.....	46
<b>Figura 28</b> Fórmulas.....	47
<b>Figura 29</b> Proceso ETL.....	48
<b>Figura 30</b> Datos poblados en el DW.....	48
<b>Figura 31</b> Datos Pacientes.....	49
<b>Figura 32</b> Datos Médicos.....	49
<b>Figura 33</b> Datos Establecimientos. ....	50
<b>Figura 34</b> Datos Diagnóstico. ....	51
<b>Figura 35</b> Datos Tiempo. ....	51
<b>Figura 36</b> Porcentaje de satisfacción de atributos de la dimensión Paciente.....	52
<b>Figura 37</b> Porcentaje de satisfacción de atributos de la dimensión Médico. ....	53
<b>Figura 38</b> Porcentaje de satisfacción de atributos de la dimensión Diagnóstico. ....	54
<b>Figura 39</b> Porcentaje de satisfacción de atributos de la dimensión Establecimiento.....	55
<b>Figura 40</b> Porcentaje de satisfacción de las medidas del DW. ....	56
<b>Figura 41</b> Porcentaje de satisfacción de variables para el análisis de información. ....	57
<b>Figura 42</b> Porcentaje de satisfacción de la estructura del DW. ....	58
<b>Figura 43</b> Porcentaje de satisfacción de acceso a datos y análisis de información. ....	59
<b>Figura 44</b> Reducción de tiempo en consultas de información.....	60
<b>Figura 45</b> Gestor de consultas.....	64

<b>Figura 46</b> Total, atenciones año 2020 - 2021.....	65
<b>Figura 47</b> Total, atenciones año 2020.....	66
<b>Figura 48</b> Total, atenciones año 2021.....	67
<b>Figura 49</b> Descripción de variables RDACAA.....	73
<b>Figura 50</b> Descripción de variables PRASS .....	78

## RESUMEN

En el sector público las instituciones de salud tienen el trabajo de tomar decisiones enfocadas a satisfacer la demanda de servicios de los pacientes que acuden a los centros de salud y es por ello muy importante incorporar mejoras en los sistemas de información que están ligados a estos procesos de decisión. En el presente proyecto de investigación se diseñó un modelo de un Almacén de datos apoyado de procesos ETL para aportar con información en el proceso de toma de decisiones al directorio de la coordinación zonal de salud 3, el cual, decidirá en base a datos históricos. Los datos fueron sometidos a un proceso de limpieza ETL (extracción, transformación y carga) de datos provenientes de dos fuentes de datos PRAS Y RDACAA, con la ayuda de herramientas de analítica de datos Talend Data Preparation, Spoon, Power BI para poder garantizar su confiabilidad. Se aplicó un diseño con intervención por lo tanto es una investigación experimental es de nivel explicativo, en el proceso de confección del DW se aplicó la metodología HEFESTO, la cual se ajusta a las necesidades de nuestra investigación.

Se generó un Almacén de Datos en el cual reposa información de atención integral de salud atención que brinda el Ministerio de Salud Pública de establecimientos de salud del Primer Nivel de Atención de Consultas y Atenciones Ambulatorias misma información que hará posible el cruce de variables para análisis estadístico y la generación de reportes y permitir un análisis de información.

**Palabras clave:** Almacén de datos, ETL, análisis de datos, minería de datos, Inteligencia de negocios.

## ABSTRACT

In the public sector, health institutions make decisions focused on satisfying patients' demands. Therefore, it is crucial to incorporate improvements in information systems associated with these decision-making processes. In this research project, a data warehouse model supported by ETL processes was designed to contribute information in the decision-making process to the board of directors of the health zonal coordination 3, which will decide based on historical data. The data were subjected to an ETL cleaning process (extraction, transformation, and loading) from two data sources PRAS and RDACAA, with the help of data analytics tools to ensure reliability like Talend Data Preparation, Spoon, Power BI. The research design was the intervention; it is experimental research with an explanatory level. In making the DW, the HEFESTO methodology was applied, which fit the needs of our study.

A Data Warehouse was generated in which there is information of critical health care provided by the Ministry of Public Health of health facilities of the First Level of Care of Consultations and Outpatient Care. The data mentioned would make possible the crossing of variables for statistical analysis and the generation of reports while allows an analysis of information.

**Keywords:** Data warehousing, ETL, data analysis, data mining, business intelligence.

Reviewed by: MsC. Adriana Cundar Ruano, Ph.D.

**ENGLISH PROFESSOR**

c.c. 1709268534

## INTRODUCCIÓN

El adelanto tecnológico y la incorporación de las (TIC) al sistema de salud se han originado grandes cantidades de información que se originan de distintas fuentes con ello se ha logrado obtener bases de datos con un volumen grande con datos de tipo estructurado y no estructurado dicha información requiere de un tratamiento apropiado para que los responsables de tomar decisiones puedan sacar ventajas de ella es por esto que resulta decisivo implementar una nueva informática de opción para obtener un mejor análisis del valor de las informaciones disponibles, delimitar indicadores de negocio pertinentes para simplificar la toma de decisiones y conservar la memoria de la corporación sobre esta información se puede aplicar técnicas de analítica de datos permitiendo extraer conocimiento útil desde los datos almacenados.

Este problema es de gran importancia porque con los grandes volúmenes de datos almacenados en sistemas tradicionales denominados (datos históricos) no se obtiene ningún valor, pero con una correcta aplicación de técnicas de analítica de datos se puede obtener indicadores que proporcionen información importante para la toma de decisiones en las entidades de salud es por ello por lo que surge una la solución para este problema lo que se denominada almacenes de datos o Data Warehouse, DW. Empezando los años noventa, Inmon definió el término “almacén de datos “como: “Un almacén de datos es un conjunto de datos orientados por temas, integrados, variables en el tiempo y no volátiles para el apoyo de la toma de decisiones”. Un almacén de datos es “integrado” porque los datos que se integran en el almacén de datos se obtienen de distintas fuentes de datos (sistemas heredados, bases de datos relacionales, ficheros etc.). Los procesos de extracción, transformación y carga (ETL) son los encargados de la extracción de los datos a partir de las distintas fuentes de datos heterogéneas, su transformación (conversión, limpieza, etc.) y su carga en el almacén de datos (Mora, 2014) .

Kimball la define como “una réplica de los datos transaccionales estructurados específicamente para consultas y análisis” (Rivadera, 2019).

La importancia de gestionar conocimiento en empresas radica en que las implementaciones de gestión de conocimiento han generado una gran variedad de acciones, es así que Rubikey (2011) indica que el 76 por ciento ha generado una estrategia de conocimiento, el 64 por ciento ha adoptado el entrenamiento, el 58 por ciento ha establecido compartir mejores prácticas, el 57 por ciento ha instaurado políticas de conocimiento y el 50 por ciento ha establecido redes formales de gestores de conocimiento. Este proyecto de investigación propone el diseño y la generación de un Data Warehouse para entidades de salud con la identificación de técnicas analíticas avanzadas para aportar a la toma de decisiones enmarcadas a satisfacer la demanda de servicios de los pacientes que acuden a las entidades de salud. El Almacén de datos se alimentará de información que proviene de sistemas operacionales ya existentes conocidos como sistemas legados esta información que será procesada con técnicas y herramientas (ETL, Data Warehouse, Data Mining).

## Capítulo I

### PLANTEAMIENTO DEL PROBLEMA

#### 1.1. Problema y justificación

¿El diseño de un almacén de datos permitirá el análisis eficiente de la información en instituciones de salud para la toma de decisiones? De forma desigual las TIC se van insertado al Sistema Nacional de Salud la receta electrónica es una realidad en la mayoría de las entidades administrativas, en un porcentaje superior al 90 por ciento de los médicos de atención primaria hacen uso de dicha receta electrónica, pero eso depende del nivel de cobertura del servicio. La agregación de la historia clínica electrónica en las entidades de salud se ha realizado de manera desigual. En las entidades de salud nuevas si poseen una implementación completa de las TIC, pero en entidades más antiguas este contexto ha sido un problema de implementarlo. Las organizaciones autónomas no hacen conocer avances o documentación de los proyectos y las mejoras que conlleva la implementación de las TIC siendo ellas una herramienta que permiten realizar un tratamiento organizado de datos, información y conocimiento en lo que se refiere a el ámbito sanitario y aportan considerablemente (González A. A., 2016).

Las metodologías avanzadas de procesamiento de la información junto con las (TIC) se han desarrollado significativamente, influyendo activamente en nuestra sociedad dentro de la salud el objetivo de este desarrollo es optimizar procesos y aumentar la productividad y también gestionar las finanzas a todo esto se lo denomina sistemas de información hospitalaria estos sistemas evolucionaron hacia “sistemas de información en salud” (o sanitarios) al comprender los aspectos operacionales y contexto clínico (capa clínica) como una lista de tareas o procesos difíciles donde los pacientes interactúan con el sistema de salud y siendo este el centro en este contexto, en el cual el intercambio, captura, almacenamiento, accesibilidad y gestión de la información clínica poseen un rol importante con la meta de poder aportar eficiencia en la atención y mejorar el acceso a los servicios es por esto

que aparece la informática en salud, con el objetivo de desarrollar gestionar y estratégicamente estos nuevos sistemas de información sanitaria (Plazzotta, 2015).

La cantidad de información que existe en los sistemas transaccionales utilizados en salud pública contiene un volumen grande y tratar de analizarla es un problema, pero con la aparición de técnicas de explotación de datos o data Mining que permiten organizar o limpiar los datos almacenados se puede mitigar este problema. Las nuevas oportunidades de la era de la información han acortado las distancias y han aumentado las ventajas para quienes tienen acceso al gran caudal de datos, pero, lo que conforma un valioso recurso para la sociedad, se ha convertido en el gran problema desde principios de siglo: manejar de forma eficaz inmensos volúmenes de información (Fontana, 2019).

Grandes cantidades de datos de diferentes áreas como negocios, sociedades, ciencia, ingeniería, medicina y ámbitos generales de la vida, se alojan y están siendo procesados en las redes de computadoras y en los dispositivos de almacenamiento este resultado se debe a la informatización en la sociedad dando lugar a buscar una nueva manera de estudiar grandes cantidades de datos con el objetivo de obtener información importante para ser empleada en la toma de decisiones esta informatización en las empresas tiene como objetivo apoyar a los procesos de negocio con la aparición de nuevas metas de negocio (Montserrat, 2016).

## **1.2. Justificación de la investigación**

Podemos mencionar que con el surgimiento y el desarrollo de los sistemas informáticos en la actualidad la capacidad de almacenar y compartir datos e información es inmensa, pero eso no es un problema. El problema es cómo responder a las preguntas: ¿Cuáles datos?, ¿Por qué?, ¿De qué forma? Mediante avanza el tiempo aumenta el tamaño de datos e información en el ámbito de la salud y de la mano los problemas a la hora de manejar dicha información De manera que se puede emplear acciones problemas que se presentan de diferentes formas selección, jerarquización, interpretación que con el desarrollo informático no se obtiene una solución se definen al concepto “una abundancia de

información de difícil utilización, tanto por su volumen como por su calidad “como precariedades del exceso (Alazraqui, 2016).

### **1.3. Viabilidad de la investigación**

El proyecto de investigación es viable, porque al diseñar y generar un almacén de datos se contribuirá en la limpieza de toda la información que venga de distintas fuentes dentro de las entidades de salud y mediante ello se podrá explotar la información que es fundamental para apoyar en el aspecto de toma de decisiones en este campo, como la detección oportuna de riesgos en cirugías, afecciones en la seguridad del paciente, organización de atención al paciente, Mejoramiento en aspectos de gestión de infraestructura y tecnología, inclusive en aspectos de mantenimiento de equipos médicos. No existen estrategias para su gestión. Se obtendrán nuevos indicadores que aportarán conocimiento a las gerencias de medicina, los mismos que podrá obtener información confiable y oportuna en el tiempo. El desarrollo del proyecto se realizará mediante procedimientos de extracción, transformación y carga (ETL) de información desde el sistema transaccional, hacia la colección de datos.

### **1.4. Objetivos**

#### ***1.4.1. Objetivo General***

Diseñar el almacén de datos para el análisis eficiente de la información en instituciones de salud.

#### ***1.4.2. Objetivos específicos***

Identificación de técnicas analíticas avanzadas de Análisis de Datos.

Establecer procesos, relaciones, y reglas de normalización para extraer, transformar y cargar datos.

Generación y validación del almacén de datos en instituciones de salud.

## Capítulo II

### MARCO TEÓRICO

#### 2.1. Almacén de datos

Es una combinación de conceptos y tecnologías destinadas a satisfacer los requerimientos de una organización o empresa, en términos de mejorar la gestión con eficiencia y facilidad de acceso. Es una base de datos relacional diseñada para la consulta y análisis en lugar del procesamiento de transacciones. Por lo general, contiene información histórica derivada de una transacción, pero puede incluir datos de otras fuentes. Separa el análisis de carga de trabajo de las transacciones y permite a una organización consolidar datos de varias fuentes. William Inmon considerado el padre del Data Warehouse en 1992. Un DW es una colección de datos orientados a temas, integrados, no volátiles y variantes en el tiempo, organizados para soportar necesidades empresariales. Ralph Kimball (considerado el principal promotor del enfoque dimensional para el diseño de almacenes de datos), “un Data Warehouse es una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis.” También en 1993, Susan Osterfeldt publicó una definición que sin duda acierta en la clave del DW: "Yo considero al DW como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. DW elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico" (Silva, 2016).

##### 2.1.1. Características

**Temático:** Los Almacenes de Datos están diseñados para ayudar a analizar los datos de un determinado tema o significado.

**Integrado:** La integración está muy vinculada con el punto “temático” los Almacenes de Datos deben juntar datos de fuentes dispares de una forma consistente. Deben dar solución a problemas tales como el nombre de los campos, conflictos de inconsistencia en unidades y medidas antes de ser almacenados.

**No volátil:** No volátil significa que, una vez introducidos los datos en el Data Warehouse, estos datos no deben ser cambiados.

**Variante en el tiempo:** El foco del Data Warehouse se centra en los cambios realizados a lo largo del tiempo, es decir, analiza cómo el tiempo hace cambiar los diferentes elementos. Para esto se requiere una gran cantidad de datos almacenados a lo largo de mucho tiempo. Esta es la diferencia con un sistema transaccional, donde los datos históricos son archivados y poco accedidos (Garay, 2016).

## **2.2. Data mart**

Es una base de datos departamental, especializada en almacenar información de un negocio específico. Se caracteriza por disponer una estructura óptima de datos para analizar información desde varias perspectivas que afecten los procesos de dicho departamento. Un Data mart puede ser alimentado desde los datos de un Data Warehouse o integrar por sí mismo un compendio de distintas fuentes de información (García, 2015).

### **2.2.1. Data mart OLAP**

Se basan en los cubos OLAP populares, que se construyen según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es heterogéneo, en función de la herramienta final que se utilice (Lemahieu, 2015).

### **2.2.2. Data mart OLTP**

Las bases de datos OLTP pueden basarse en un simple extracto de la data Warehouse, no obstante, lo común es introducir mejoras en su rendimiento (los agregados y filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras comunes son las tablas report, que vienen a ser fact-tables reducidas (que agregan dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores,

aun con el objetivo de explotar la reescritura de queries (aunque solo es posible en algunos SGBD avanzados, como Oracle) (Euclides, 2019).

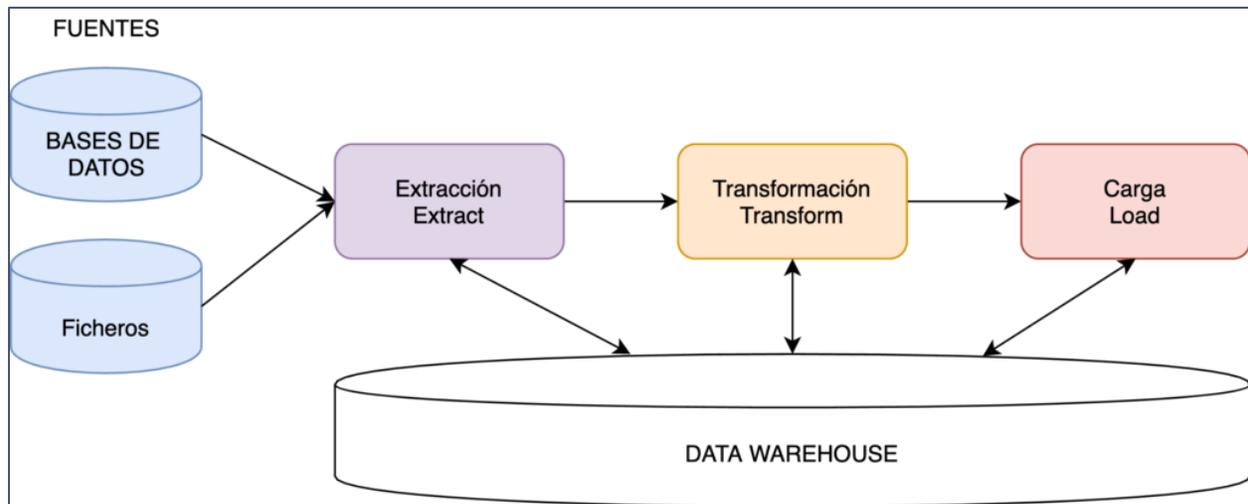
### **2.3. Proceso ETL**

Proceso de Extracción Transformación y Carga de datos desde los Sistemas de Origen a las bases de datos del Data Warehouse. La responsabilidad de mantener la consistencia de los datos, en un Data Warehouse, está en la programación correcta de los procesos ETL. Un proceso ETL puede ocupar una base de datos (BD) auxiliar, siendo sus funciones principales limpiar, depurar y homologar los datos, sobre todo, cuando provienen de diferentes fuentes. Las diversas plataformas, para el desarrollo de soluciones de BI, otorgan herramientas y servicios integrados para las funciones de ETL (Morales, 2017).

Los procesos ETL cargan normalmente a los almacenes de datos los datos que están en las fuentes, teniendo en cuenta siempre si los datos sufrieron alguna variación a lo largo del tiempo es por esto que los almacenes de datos tienen la característica de ser utilizados para la lectura de la información lo cual la diferencia de las bases de datos ya que en ellas si se puede modificar los datos. En este punto de vista estos procesos son trascendentales porque cumplen un rol importante de trabajo el cual es recolectar datos y posicionar los mismos, en el lugar que servirá como la fuente principal de conocimiento para los interesados en realizar análisis sobre este conocimiento. Los individuos que creen estos procesos y la presenten mediante herramientas OLAP serán quienes lleven la responsabilidad de generar toda la información para su respectivo análisis, por ello se debe tener en consideración el diseño y mantenimiento de estos procesos ETL son las causas importantes para el éxito de proyectos en el cual intervengan los data Warehouse (March, 2007).

**Figura 1**

*Esquema tres fases de un ETL*



**Fuente:** Oscar, F. (2021). Comparativa herramientas ETL más usadas en la empresa. Tomado de <https://aprenderbigdata.com/>.

#### **2.4. Data Mining o Minería de Datos.**

Según Molina (2001) menciona que la Data Mining se refiere al proceso de extraer conocimiento de bases de datos. Su objetivo es descubrir situaciones anómalas e interesantes, tendencias, patrones y secuencias en los datos. La Data Mining es una etapa dentro del proceso completo del descubrimiento del conocimiento, este intenta obtener patrones o modelos a partir de los datos recopilados. Decidir si los modelos obtenidos son útiles o no suele requerir una valoración subjetiva por parte del usuario. Los algoritmos de Data Mining suelen tener tres componentes: El modelo, que contiene parámetros que han de fijarse a partir de los datos de entrada. El criterio de preferencia, que sirve para comparar modelos alternativos. El algoritmo de búsqueda, que viene a ser como cualquier otro programa de inteligencia artificial (IA) (Villada, 2019). Es el proceso de Seleccionar, Explorar, Modificar, Modelar y valorar grandes cantidades de datos con el objetivo de descubrir conocimiento (González, 2019).

La minería de datos nos permite la administración en tiempo real de manera eficaz, es una herramienta aplicable a cualquier tipo de empresa. Una amplia gama de compañías puede tener aplicaciones exitosas con ella (Angeles,2010). El proceso debe ser automático o semiautomático. Los modelos encontrados deben ser significativos demostrando cierto patrón o regla de comportamiento. Las aplicaciones que más se utilizan son las que necesitan algún tipo de predicción (Aular, 2017).

#### **2.4.1. Beneficios**

Incremento de los resultados como consecuencia del aumento de la cuota de mercado; Clientela Fidelizada una mejor respuesta a sus requerimientos; Mejora del rendimiento; Reducción del factor riesgo; Optimización de las estrategias y toma de decisiones y Optimización de la gestión, maximizando rentabilidades (López 2004).

La Minería de Datos permite simular el comportamiento humano en procesos forenses establece los diferentes escenarios de ocurrencia de accidentes como también en procesos biométricos con el reconocimiento de emociones faciales y en la medicina con la determinación de la dosificación ideal de medicamentos en el tratamiento de trastornos del habla (Rosado, 2014).

#### **2.5. Inteligencia de negocios**

La definición de inteligencia de negocios es la habilidad corporativa para tomar decisiones. Esto se logra mediante el uso de metodologías, aplicaciones y tecnologías que proporcionan reunir, depurar, transformar datos, y aplicar en ellos técnicas analíticas de extracción de conocimiento (Parr, 2000) los datos pueden ser estructurados para que indiquen las características de un área de interés (Stackowiak, 2007), aportando el conocimiento sobre los problemas y oportunidades del negocio para que puedan ser corregidos y aprovechados (Ballard, 2006).

La inteligencia de negocios apoya a la toma de decisiones del negocio: estratégicas, tácticas y operativas; según varios autores: “Inteligencia Empresarial es el proceso de transformar datos en conocimiento, y el conocimiento en acción, para la toma de decisiones” (Bernabéu, 2010). Vinculado

con las tecnologías de la información y desde un punto de vista pragmático se define a Business Intelligence como “el grupo de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa de la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP/OLAP, alertas) o para su análisis y conversión en conocimiento aportando así soporte a la toma de decisiones sobre el negocio”

## **2.6. Big Data**

Cuando hablamos de analítica de Big Data ocurre como con el propio concepto de Big Data, donde no hay una única definición, por lo que en aras de elegir aquella que goza de mayor reconocimiento hemos optado por la propuesta por la asociación Isaca en 2011:

“Analítica de datos (Data Analytics) engloba aquellos procesos y actividades diseñadas para obtener y evaluar los datos, obteniendo de esta información útil” (Joyanes, 2011).

- Analítica de datos más propiamente de organizaciones

Es el proceso del análisis de los datos obtenidos de forma tradicional.

- Analítica web

Nos referimos al análisis de los datos obtenidos del tráfico web de una página.

- Analítica social

Aquellos datos provenientes de las redes sociales, blogs o la misma Wikipedia entre otros.

- Analítica de Big Data

Centrada en grandes volúmenes de datos. Por tanto, vemos que la analítica de Big Data es una categoría del análisis de datos.

## **2.7. Analítica de Big Data**

La analítica de Big Data sirve a las empresas para analizar los cambios que se producen, de forma que esta herramienta provea a la organización de medios para afrontar los retos surgidos por este inmenso

crecimiento de datos. Además, la analítica es un proceso que permite a las empresas descubrir mejores proveedores, nuevos clientes, observar la estacionalidad en su negocio. Para Joyanes (243, 2013): Analítica de Big Data es el proceso de examinar grandes cantidades de datos de una variedad de tipos para descubrir patrones ocultos, correlaciones desconocidas y otra información útil. Dicha información puede proporcionar ventajas competitivas, brindar beneficios en los negocios y un aumento en los ingresos (Russo, 2016).

La analítica avanzada se compone de una serie de técnicas orientadas a obtener ventajas para la empresa que las implementa. Dentro de las diferentes técnicas y herramientas que engloba, hay que destacar la analítica predictiva, minería de datos, análisis estadísticos de los datos y programación utilizando lenguaje SQL. No obstante, el concepto de analítica avanzada está dejando paso al de analítica de descubrimiento o analítica exploratoria, ya que el usuario, es decir, el analista de la empresa, a partir de la existencia de un gran volumen de datos nuevos, está tratando de descubrir oportunidades y hechos existentes previamente en la empresa pero que no eran conocidos por esta. El análisis de Big Data suele llevarse a cabo siempre a partir de las mismas herramientas de software, de entre las cuales podemos destacar: análisis estadístico de datos, minería de datos, de textos, web y social, análisis y modelado predictivo, visualización de datos y consultas en lenguaje SQL (Russo, 2016).

## **2.8. Big Data en el sector de la salud**

En el siguiente capítulo se presenta el Big Data en el ámbito de la salud consta de tres partes: Primero se analizan las posibilidades de la analítica de Big Data en el sector sanitario a nivel internacional, segundo, se presenta el estado en que se encuentra actualmente su adopción en nuestro entorno sanitario. Finalmente, se presentan algunos casos de éxito. Posibilidades de la analítica de Big Data en el ámbito sanitario. En este apartado se incluyen algunos de los trabajos que más contribuyeron en su momento a dar a conocer las posibilidades que la tecnología Big Data ya había comenzado a ofrecer en el sector de la salud. Así lo pusieron de relieve de modo especial dos estudios de referencia, como

son el extenso informe que el McKinsey Global Institute (MGI) publicó en junio de 2011 bajo el título *Big Data: The Next Frontier for Innovation, Competition and Productivity* y el informe *Big Data Healthcare Hype and Hope*, que la doctora y consultora Bonnie Feldman, en colaboración con Ellen Martin y Toby Skotness, publicó en octubre de 2012. En los subapartados siguientes se revisan las principales aportaciones de cada uno de ellos (Vallejo, 2016).

### ***2.8.1. The Next Frontier for Innovation, Competition and Productivity***

En este documento se destaca el potencial transformador de Big Data en cinco grandes sectores, que los expertos de la consultora MGI han estudiado en profundidad: la atención sanitaria y el comercio minorista en EE. UU., la administración del sector público en la Unión Europea, y, a nivel global, la fabricación y los datos de geolocalización de personas y entidades. Los citados ámbitos representaban en 2010 cerca del 40 por ciento del PIB mundial. Para cada uno de ellos se identifican “palancas” o recursos mediante los cuales Big Data facilitará la creación de valor (Vallejo, 2016).

En el caso de la atención sanitaria en EE. UU., se estima que en el plazo de diez años el valor creado mediante el uso de Big Data podría alcanzar los 300.000 millones de dólares al año, lo que supondría un ahorro anual de más de 200.000 millones en el gasto nacional en el sector. En el informe se distinguen quince recursos Big Data relacionados con el sector sanitario, que se distribuyen en diferentes categorías según ejerzan su impacto sobre la operativa clínica, la investigación y desarrollo, el sistema de pagos y fijación de precios, el impulso de nuevos modelos de negocio o la salud pública (Vallejo, 2016).

### ***2.8.2. Healthcare Hype and Hope***

Este estudio constituye una valiosa aportación. Se basa en el análisis de proyectos y herramientas que ya están en funcionamiento, siguiendo diferentes estrategias. En palabras de Paniagua (2012), el trabajo de Feldman “explora con bastante precisión cómo Big Data se está convirtiendo en una creciente fuerza de cambio en el panorama sanitario”. Dentro de su detallada exposición, se pueden considerar de

especial interés los apartados que dedica a las cuatro principales dimensiones que caracterizan a Big Data aplicadas al sector sanitario y a las formas que propone para mejorar el cuidado de la salud. Dimensiones Big Data aplicadas a los datos del sector sanitario Volumen, Variedad, Velocidad, Veracidad. Formas de utilizar Big Data para mejorar el estado de salud Apoyo a la investigación genómica y otras transformar datos en información (y la información en datos), Apoyo al autocuidado y colaboración ciudadana (Peñañiel, 2019).

## **2.9. Big data en nuestro sistema sanitario (Ecuador)**

En Ecuador se realizó un informe denominado aplicación de Big Data en detección de concentración poblacional, para evitar brotes epidemiológicos por covid-19 en el cual resume de la siguiente manera el Big data con datos de georreferencia con la emergencia sanitaria por COVID-19 decretada por el Gobierno del Ecuador ante la pandemia declarada por la OMS, trajo consigo una serie de actividades y planes que generan información importante y de interés general para la población. El objetivo de la investigación es analizar cómo los sistemas de georreferencia y herramientas que identifican el volumen de estos datos informativos, permiten tomar medidas de aislamiento a los organismos de salud para ciertos sectores del país. La metodología utilizada está enfocada en un análisis cuantitativo sobre los datos estadísticos presentados por los sistemas georeferenciales del MSP y cuán importante es esta información digital y en tiempo real para las autoridades de salud pública. Los resultados pretenden describir la realidad nacional sobre la dimensión de aglomeración de personas en las principales ciudades y la detección de brotes epidemiológicos por Covid-19 a nivel nacional desde la percepción de la tecnología digital Big Data. Concluyendo en un análisis de las acciones de prevención que tomaron los COE provinciales para enfatizar si fueron o no eficaces para evitar mayores contagios de brotes epidemiológicos por coronavirus. El objetivo de la investigación se centra en identificar la aplicación de herramientas digitales con tecnología Big Data en la salud pública y sus efectos positivos en la toma de decisiones en el sector sanitario frente a los brotes epidemiológicos. (Vallejo, 2016)

## **2.10. Aporte del Big Data**

Uno de los instrumentos tecnológicos que se habilitaron para la emergencia sanitaria fueron los mapas de calor. De esta manera los análisis de datos permiten tomar decisiones sobre las provincias con más aglomeraciones y mediante las consultas web a los sitios se pueden consultar las opciones para conocer las poblaciones más activas (Ortega, 2020). Todo esto gracias a la 24 tecnología que proporcionan sitios especializados como la herramienta OpenStreetMap. Según las mediciones de mapas corporales Web, la ciudad de Guayaquil resultó encontrarse entre las ciudades más afectadas por aglomeración durante el mes de abril hasta mayo de 2020, sobre todo en parroquias como Tarqui (4,740 casos confirmados), Ximena (1,781 casos confirmados), Febres Cordero (1,449 casos confirmados), donde los contagios fueron masivos. En estos lugares fue muy importante haber establecido un cerco epidemiológico. Otras localidades cercanas a la ciudad de Guayaquil como es Durán en la parroquia Eloy Alfaro se presentaron 516 casos confirmados y en Milagro 562. La provincia de Guayas fue la primera en utilizar herramientas de medición basadas en Big Data que emplea el ECU 911, la tecnología del servicio de emergencias ecuatoriano incluye cámaras térmicas para monitoreo de la actividad volcánica, lectoras de placas de vehículos, monitoreo de eventos de concentración masiva, reconocimiento facial, drones, herramientas que se utilizan para realizar labores de monitoreo, control de incendios, videovigilancia de lugares poco accesibles, rescate, prevención y acciones de respuesta oportuna ante incidentes. Estos sistemas tecnológicos son de última generación (Ortega, 2020).

Los sistemas del ECU 911 permitieron medir la concentración de aplicaciones utilizadas por celulares y el desplazamiento de estos y almacenarlas en sus bases de datos como Mobile Locato y APP ECU 911, también los sistemas interactúan con aplicaciones de georreferencia como QGIS, y Arcmap, y herramientas basadas en Heatmaps como Crazyegg y SumoMe. Definitivamente estas herramientas contribuyeron a las acciones positivas tomadas por el COE Nacional y COE cantonal de Guayaquil, al

implementar restricciones de movilidad en sus diferentes sectores de la ciudad y evitar contagios probables por coronavirus (Ortega, 2020).

**Tabla 1**

*Comparación entre dos autores más reconocidos en Data Warehouse*

<b>Autor</b>	<b>Definición</b>	<b>Arquitectura</b>
<b>Ralph Kimball</b>	“Un Data Warehouse es un duplicado de los datos transaccionales específicamente estructurada para la consulta y el análisis”	Data Warehouse no es más que la unión de los Data mart en una sola entidad y a partir de ello se realiza la explotación de datos. Defiende su metodología ascendente para cuando empezamos a diseñar el almacén de datos.
<b>William Inmon</b>	“Un Data Warehouse es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales”	Existe un proyecto inicial, es decir un Enterprise Data Warehouse y posteriormente de él se producen Data mart Inicia con un proceso que es directamente tomar la información y empezar a almacenarla en el Enterprise Data Warehouse y después dividía la parte del Data Mart Podemos recalcar que Willian Inmon trabaja con el modelo copo de nieve.

**Fuente:** Elaborado por el autor.

## **Capítulo III**

### **METODOLOGÍA**

#### **3.1. Tipo y Diseño de Investigación**

Dado que el objetivo de estudio será diseñar un almacén de datos para producir información aplicando ETL el investigador recurrió a un diseño con intervención por lo tanto es una investigación experimental es de nivel explicativo por qué se va a obtener indicadores que ayuden a la toma de decisiones en la entidad de salud. Según el autor (Fidias G. Arias (2012)), define: La investigación experimental es un proceso que consiste en someter a un objeto o grupo de individuos, a determinadas condiciones, estímulos o tratamiento (variable independiente), para observar los efectos o reacciones que se producen (variable dependiente). Si de originalidad se trata, las normas APA tienen un doctorado. Es por eso que se han mantenido vigentes con el pasar de los años y es un criterio casi obligatorio a la hora de redactar un informe, documento, artículo, etc. de carácter científico.

##### ***3.1.1. Enfoque de la Investigación***

El presente proyecto será diseñado bajo el planteamiento metodológico del enfoque cuantitativo puesto que este es el que mejor se adapta a las necesidades de la investigación de acuerdo con Tamayo (2007), consiste en el contraste de teorías ya existentes a partir de una serie de hipótesis surgidas de la misma, siendo necesario obtener una muestra, ya sea en forma aleatoria o discriminada, pero representativa de una población o fenómeno objeto de estudio.

#### **3.2. Unidad de análisis**

Volúmenes de datos sanitarios dispersos en diferentes fuentes de datos en la coordinación zonal 3. Donde ocurre el problema: Coordinación zonal 3. Sobre quien o quienes ocurre el problema: Sobre los datos sanitarios Coordinación zonal 3. Qué problema tienen: Volúmenes de datos sanitarios dispersos en diferentes fuentes.

**Causa:** Volúmenes de datos dispersos en diferentes fuentes en la coordinación zonal 3. **Efecto:** Datos históricos sin producir valor. **Aporte:** Diseñar un almacén de datos para tener información centralizada y estandarizada con la aplicación de procesos ETL.

### ***3.2.1. Variables Secundarias o dependientes***

Diseñar un almacén de datos para tener información centralizada y estandarizada con la aplicación de procesos ETL.

Datos históricos sin producir valor.

### ***3.2.2. Variable principal o independiente***

Volúmenes de datos sanitarios dispersos en diferentes fuentes, DW.

### ***3.2.3. Técnicas de recolección de Datos***

La recogida de datos tal como Tejada (1.997) expresa es una “las fases más trascendentales en el proceso de investigación científica” (p. 95). Dadas las condiciones en la que se realizó la investigación los datos fueron recogidos por el propio investigador. Por cuanto la investigación se realiza en nombre de una institución de salud.

### ***3.2.4. Técnicas de Análisis e interpretación de la información***

Una vez analizados los datos se obtuvieron los resultados y se procedió a elaborar las conclusiones y recomendaciones pertinentes al diagnóstico que originó el diseño de la propuesta de un almacén de datos orientado a un análisis eficiente de la información.

## **3.3. Población y Muestra**

En la investigación no se dispone de población y muestra porque se va a trabajar en un ambiente de modelado de un Almacén de datos.

### 3.4. Operacionalización de variables

**Tabla 2**

*Operacionalización de variables*

<b>Problema</b>	<b>Tema</b>	<b>Objetivos</b>	<b>Variables</b>	<b>Indicadores</b>
Volúmenes de datos sanitarios dispersos en diferentes fuentes de información.	Diseño y generación del Almacén de datos aplicando ETL para instituciones de salud.	<p>Diseñar el almacén de datos para el análisis eficiente de la información en instituciones de salud.</p> <p>Identificación de técnicas analíticas avanzadas de Análisis de Datos.</p> <p>Establecer procesos, relaciones, y reglas de normalización para extraer, transformar y cargar datos. Generación y validación del almacén de datos en instituciones de salud.</p>	<p><b>Independiente:</b> Almacén de datos.</p> <p><b>Dependiente:</b> permitirá el Análisis eficiente de la información en instituciones de salud.</p>	<p><b>Independiente:</b> Toma de decisiones. Acceso a datos.</p> <p><b>Dependiente:</b> Tiempo de consultas. Calidad de Información. Tiempo de análisis.</p>

**Fuente:** Elaborado por el autor.

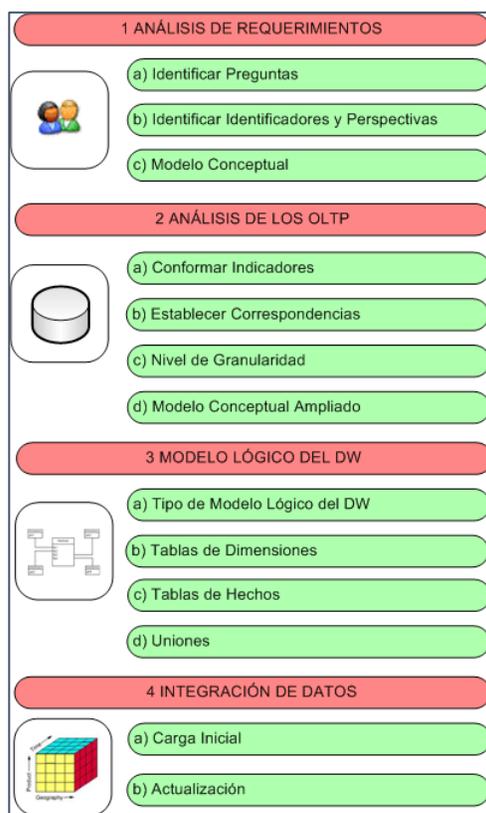
### 3.5. MÉTODO APLICADO

La metodología que se utilizó en el desarrollo del proyecto se denomina Hefestos por las características de los datos de Salud que se tiene, las cuales se ajustan a las necesidades de nuestra investigación. HEFESTO es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos.

La metodología HEFESTO puede resumirse mediante el siguiente gráfico:

**Figura 2**

*Esquema de la Metodología Hefesto.*



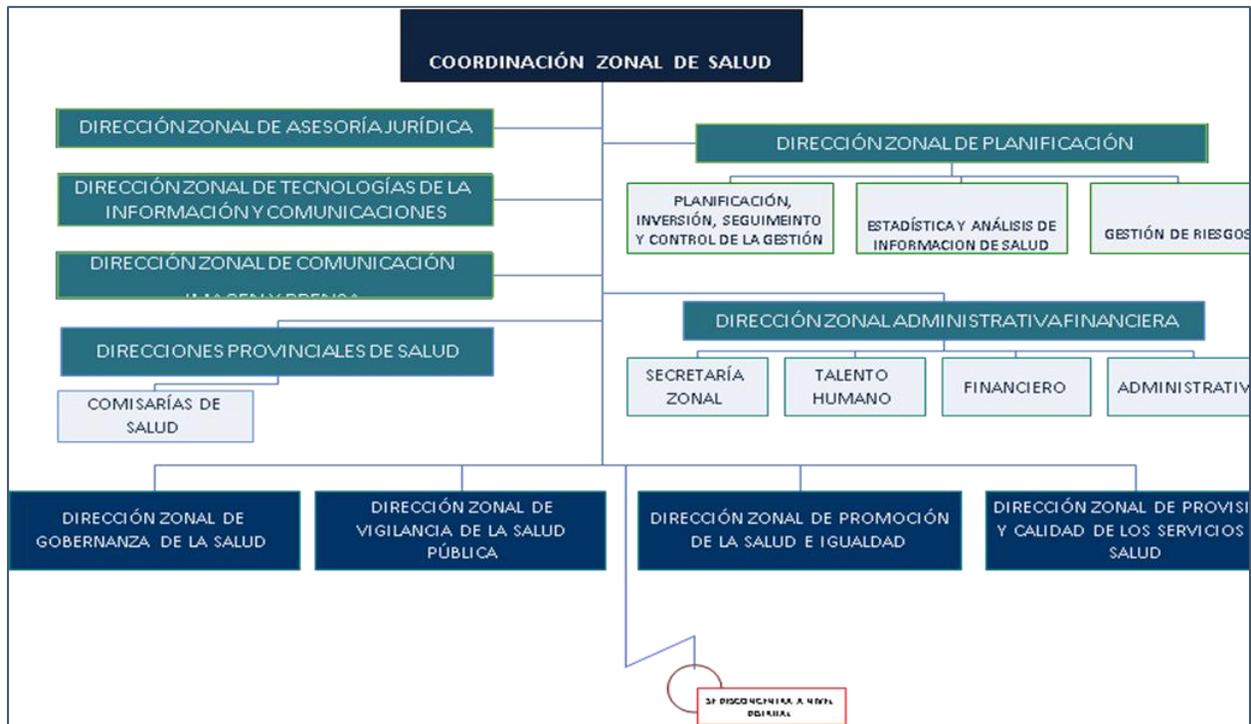
**Fuente:** Bernabéu. (2009). Esquema de la Metodología Hefesto con sus fases. Tomado de <https://modelohefesto.com/>.

### 3.6. Procedimientos

#### 3.6.1. Empresa analizada

Figura 3

Estructura Orgánica Zonal



**Fuente:** Msp. (2015). Estructura orgánica de las Coordinaciones Zonales. Tomado de <https://http://www.saludzonal.gob.ec/>.

#### 3.6.2. Análisis de Requerimientos

##### Identificar preguntas

Se preguntó a los representantes de la coordinación de salud en una encuesta sus necesidades de información, con preguntas referidas a cuál es el proceso más importante en salud y los indicadores que representan este proceso y qué sería exactamente lo que se desea analizar del mismo.

El proceso elegido fue el de Atención médica con las siguientes preguntas.

1. Cantidad de pacientes atendidos en un determinado tiempo, agrupado por tipo de establecimiento.
2. Cantidad de pacientes atendidos por profesional de salud y por sexo.
3. Consultar pacientes manejo de atenciones: por tipo de atención, diagnóstico, por grupo etario, por especialidad, por cédula, localidad.
4. Total, pacientes atendidos.

### **Identificar indicadores y perspectivas**

- Indicadores

Cantidad de pacientes atendidos

Total, atenciones

- Perspectivas

Pacientes

Médicos

Establecimiento

Diagnóstico

Tiempo

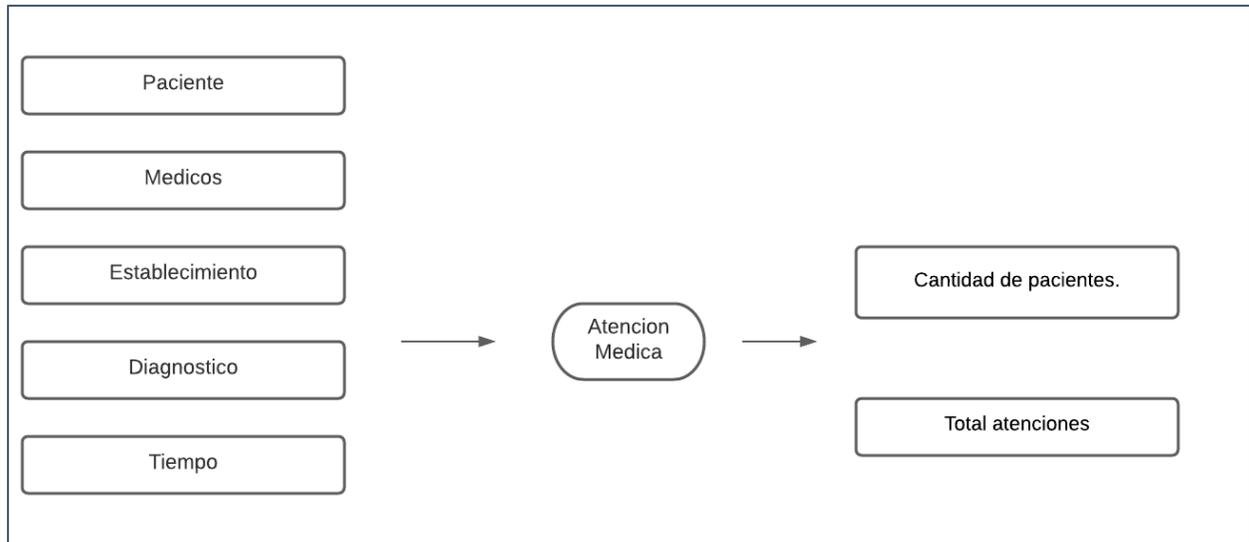
### **Modelo Conceptual**

Se construyó a través de los indicadores y perspectivas establecidos en el paso anterior.

El siguiente gráfico visualiza el Modelo Conceptual.

## Figura 4

*Esquema del modelo conceptual.*



**Fuente:** Elaborado por el autor.

### 3.6.3. Análisis de los OLTP

#### a) Conformar Indicadores

En esta fase se determinó, las funciones para calcular cada uno de indicadores.

Cantidad Paciente

**Hechos:** Cantidad de pacientes atendidos

Función de Porcentaje: COUNT

**Aclaración:** el indicador “Cantidad pacientes.” representa la sumatoria de los pacientes que se han atendido agrupado por tipo de establecimiento.

Cantidad Paciente

**Hechos:** Cantidad de pacientes atendidos

Función de Porcentaje: COUNT

**Aclaración:** el indicador “Cantidad pacientes.” representa la sumatoria de pacientes atendidos por profesional de salud y por sexo.

b) Establecer correspondencias

En la coordinación de salud disponen de variables del proceso de atención médica que proviene de un OLTP denominado RDACAA, pero contiene dos versiones 1.6 y 2.0 y la segunda fuente de datos PRASS a continuación se presenta la información a través de entidades, atributos y jerarquías de generalización.

**Figura 5**

*Variables Fuente de datos Rdacaa V 1.6.*



N°	NOMBRES DE VARIABLES BASE DE DATOS	NOMBRES DE VARIABLES
1	N_ID_UNIDAD_OPERATIVA	cod_esta
2	F_FECHA_ATENCION	fecha_aten
3	N_ID_UNIDAD_OPERATIVA_des	nom_esta
4	N_LUGAR_ATENCION	lugar_aten
5	C_ATC_BLK_A_6_PROVINCIA	prov_esta
6	C_ATC_BLK_A_7_CANTON	can_esta
7	C_ATC_BLK_A_8_PARROQUIA	parr_esta
8	C_ATC_BLK_A_9_ZONA	zona_esta
9	C_ATC_BLK_A_10_DISTRITO	distrito_esta
10	C_ATC_BLK_A_11_CIRCUITO	circuito_esta
11	N_ATC_BLK_B_5_PROF_FORMA	forma_prof
12	N_ATC_BLK_B_8_PROF_SEXO	sexo_prof
13	F_ATC_BLK_B_9_PROF_FEC_NAC	fecha_nac_prof
14	N_ATC_BLK_B_10_PROF_PAIS	pais_prof
15	N_ATC_BLK_B_11_PROF_AUTOID	etnia_prof
16	N_ATC_BLK_C_5_SEXO	sexo_pac
17	F_ATC_BLK_C_6_FECHA_NACIMI	fecha_nac_pac
18	N_ATC_BLK_C_6_1_ANIO	anio_pac
19	N_ATC_BLK_C_6_2_MES	mes_pac
20	N_ATC_BLK_C_6_3_DIA	dia_pac
21	N_ATC_BLK_C_7_NACIONALIDAD	nacionalidad_pac
22	N_ATC_BLK_C_8_AUTOID_ETNICA	etnia_pac
23	N_ATC_BLK_C_9_NACIONALIDAD	nacio_pac
24	C_ATC_BLK_C_9_1_PUEBLO	pueblo_pac
25	N_ATC_BLK_C_10_GRP_PRIORI_1	gp1
26	N_ATC_BLK_C_11_GRP_PRIORI_2	gp2
27	N_ATC_BLK_C_12_GRP_PRIORI_3	gp3
28	C_ATC_BLK_C_13_PROVINCIA	prov_res
29	C_ATC_BLK_C_14_CANTON	cant_res
30	C_ATC_BLK_C_15_PARROQUIA	parr_res
31	C_BARRIO_SECTOR_RECINTO	recinto_res
32	N_ATC_BLK_C_17_SEGURO_SALUD	ssalud
33	C_ATC_BLK_D_1_DIAGN_CIE10_1	cie_1
34	C_ATC_BLK_D_1_DIAGN_CIE10_1_des	dcie_1
35	C_ATC_BLK_D_1_1_TIP_ATC_PRE	tap1
36	C_ATC_BLK_D_1_2_TIP_ATC_MOR	tam1
37	C_ATC_BLK_D_1_3_TIPO_DIAGNO	td1
38	C_ATC_BLK_D_2_DIAGN_CIE10_2	cie_2
39	C_ATC_BLK_D_2_DIAGN_CIE10_2_des	dcie_2
40	N_ATC_BLK_D_2_1_TIP_ATC_PRE	tap2
41	N_ATC_BLK_D_2_2_TIP_ATC_MOR	tam2
42	N_ATC_BLK_D_2_3_TIPO_DIAGNO	td2
43	C_ATC_BLK_D_3_DIAGN_CIE10_3	cie_3
44	C_ATC_BLK_D_3_DIAGN_CIE10_3_des	dcie_3
45	N_ATC_BLK_D_3_1_TIP_ATC_PRE	tap3
46	N_ATC_BLK_D_3_2_TIP_ATC_MOR	tam3
47	N_ATC_BLK_D_3_3_TIPO_DIAGNO	td3
48	PAR_URBANO_RURAL_INEC	area_res
49	NUN_DESCRIPCION	nivel_esta
50	TUN_DESCRIPCION	tipo_esta
51	TUN_SIMBOLO	simbolo_tipo_esta

**Fuente:** Elaborado por el autor.

**Figura 6**

*Variables Fuente de datos Rdacca V 2.0.*

No.	CODIGO VARIABLE				
1	COD_EST_SALUD	43	COD_DAT_ANT_TIP_TOMA_TALLA	90	COD_VACUNA_2
2	COD_LIUD	44	DAT_ANT_PUNTAJE_2_TE	91	COD_VAC_2_DOSIS
3	COD_PROFESIONAL	45	DAT_ANT_PUNTAJE_2_PE	92	COD_VAC_2_ESQUEMA
4	COD_ESPECIALIDAD	46	DAT_ANT_PUNTAJE_2_PT	93	COD_VACUNA_3
5	COD ESTRATEGIA	47	DAT_ANT_PUNTAJE_2_IMC_E	94	COD_VAC_3DOSIS
6	FEC_ATENCION	48	DAT_ANT_PUNTAJE_2_PC_E	95	COD_VAC_3_ESQUEMA
7	PER_NUM_ARCHIVO	49	DAT_ANT_CATEGORIA_TE	96	COD_VACUNA_4
8	COD_PER_TIP_IDENTIFICACION	50	DAT_ANT_CATEGORIA_PE	97	COD_VAC_4_DOSIS
9	COD_PER_IDENTIFICACION	51	DAT_ANT_CATEGORIA_PT	98	COD_VAC_4_ESQUEMA
10	PER FEC NACIMIENTO	52	DAT_ANT_CATEGORIA_IMC_E	99	COD_VACUNA_5
11	PER_PRI_NOMBRE	53	DAT_ANT_CATEGORIA_PC_E	100	COD_VAC_5_DOSIS
12	PER_SEG_NOMBRE	54	DAT_ANT_HB	101	COD_VAC_5_ESQUEMA
13	PER_PRI_APELLIDO	55	DAT_ANT_HB_CORREGIDO	102	COD_PROCEDIMIENTO_1
14	PER_SEG_APELLIDO	56	COD_IND_ANEMIA	103	PROC_CANTIDAD_1
15	COD_PER_SEXO	57	COD_VIH_MOT_PRUEBA	104	COD_PROCEDIMIENTO_2
16	COD_PER_ORI_SEXUAL	58	COD_VIH_PRI_PRUEBA	105	PROC_CANTIDAD_2
17	COD_PER_IDE_GENERO	59	COD_VIH_PRUEBA_PRI_CONF	106	COD_PROCEDIMIENTO_3
18	COD_PER_NACIONALIDAD	60	COD_VIH_SEG_PRUEBA	107	PROC_CANTIDAD_3
19	COD_PER_AUT_ETNICA	61	COD_VIH_PRUEBA_SEG_CONF	108	COD_PROCEDIMIENTO_4
20	COD_PER_NAC_ETNICA	62	COD_VIH_VIA_TRANSMISION	109	PROC_CANTIDAD_4
21	COD_PER_PUEBLOS	63	COD_DIAG_C_E_1	110	COD_PROCEDIMIENTO_5
22	COD_PER_SEGURO	64	COD_DIAG_TIP_DIAGNOSTICO_1	111	PROC_CANTIDAD_5
23	COD_PER_TIPO_BONO	65	COD_DIAG_TIP_ATENCION_1	112	ODON_DIE_CARIADOS
24	COD_PER_PARROQUIA	66	COD_DIAG_CON_DIAGNOSTICO_1	113	ODON_DIE_RXTRAIDO
25	PER_TEL_PACIENTE	67	DIAG_NUM_CER_MEDICO_UNICO	114	ODON_DIE_OBTURADO
26	PER_TEL_FAMILIAR	68	COD_DIAG_C_E_2	115	ODON_DIE_PERDIDO
27	PER_DIR_DOMICILIO	69	COD_DIAG_TIP_DIAGNOSTICO_2	116	PRE_SUP_HIE_MULTIVITAMINAS
28	COD_PER_TIP_IDE_REPRESENTANTE	70	COD_DIAG_TIP_ATENCION_2	117	PRE_SUP_VITAMINA_A
29	PER_IDE_REPRESENTANTE	71	COD_DIAG_CON_DIAGNOSTICO_2	118	PRE_SUP_HIE_JARABE
30	COD_EST_LUG_ATENCION	72	COD_DIAG_C_E_3	119	PRE_SUP_HIE_ACL_FOLICO
31	COD_EST_TIP_ATENCION	73	COD_DIAG_TIP_DIAGNOSTICO_3	120	REF_CON_SUBSISTEMA
32	COD_GPR_PRIORITARIO_1	74	COD_DIAG_TIP_ATENCION_3	121	REF_CON_INTERCONSULTA
33	COD_GPR_PRIORITARIO_2	75	COD_DIAG_CON_DIAGNOSTICO_3	122	REF_CON_LUG_REFERIDO
34	COD_GPR_PRIORITARIO_3	76	COD_DIAG_C_E_4	123	VIOL_NUM_SERIE
35	COD_GPR_PRIORITARIO_4	77	COD_DIAG_TIP_DIAGNOSTICO_4	124	COD_VIOL_IDE_AGRESOR
36	COD_GPR_PRIORITARIO_5	78	COD_DIAG_TIP_ATENCION_4	125	COD_VIOL_LESIONES
37	DAT_ANT_TALLA	79	COD_DIAG_CON_DIAGNOSTICO_4	126	COD_VIOL_PARENTESCO
38	DAT_ANT_TALLA_CORREGIDA	80	COD_DIAG_C_E_5	127	COD_GRP_VULNERABLE_1
39	DAT_ANT_PESO	81	COD_DIAG_TIP_DIAGNOSTICO_5	128	COD_GRP_VULNERABLE_2
40	DAT_ANT_PER_CEFALICO	82	COD_DIAG_TIP_ATENCION_5	129	COD_GRP_VULNERABLE_3
41	DAT_ANT_IMC	83	COD_DIAG_CON_DIAGNOSTICO_5	130	COD_GRP_VULNERABLE_4
42	DAT_ANT_IMC_CLASIFICACION	84	SIV_MAD_PER_LACTANCIA	131	COD_GRP_VULNERABLE_5
		85	SIV_24H_LECHE	132	COD_OBS_RIES Obstetrico
		86	SIV_24H_ALIMENTOS	133	COD_OBS_PLA_PARTO
		87	COD_VACUNA_1	134	COD_OBS_PLA_TRANSPORTE
		88	COD_VAC_1_DOSIS	135	COD_OBS_EMB_PLANIFICADO
		89	COD_VAC_1_ESQUEMA		
				136	COD_OBS_MET_DET_SEMANAS_GESTACION
				137	OBS_SEM_GESTACION
				138	OBS_FEC_LUL_MENSTRUACION
				139	COD_EXA_LAB_SIFILIS_NO_TREPONEMICA
				140	COD_EXA_LAB_SIFILIS_TREPONEMICA
				141	COD_EXA_LAB_SIFILIS_TRATAMIENTO
				142	COD_EXA_LAB_SIFILIS_TRATAMIENTO_PAREJA
				143	COD_RES_BACTERIURIA
				144	COD_CARGA_ARCH_ESTADO
				145	COD_VERSION_SOFTWARE
				146	COD_DPA_EST_PARROQUIA

**Fuente:** Elaborado por el autor.

## Figura 7

*Variables Fuente de datos Prs.*

ITEM	EJEMPLO	NOMBRES DE NOMBRE DE	DESCRIPCION TABLA IDENTIFICADA
Nro.	75538		
ENT_ID	397	ENTIDAD	Identificación de la entidad
ENT_RUC	6.61E+16	ENTIDAD	RUC de la entidad
ENT_NOM	CHAMBO	ENTIDAD	Nombre de la Entidad
ENT_SIM_TIP_CS-B		ENTIDAD	Tipología de la Entidad
ENT_DES_TIP_CENTRO DE SALUD TIPO B		ENTIDAD	Descripción del tipo de la Entidad
ENT_COD_PRI	6	ENTIDAD	Código de la provincia de la Entidad
ENT_DES_PRC CHIMBORAZO		ENTIDAD	Descripción de la provincia de la Entidad
ENT_COD_CA	604	ENTIDAD	Código del cantón de la Entidad
ENT_DES_CANTON CHAMBO		ENTIDAD	Descripción del cantón de la Entidad
ENT_COD_PA	60450	ENTIDAD	Código de la parroquia de la Entidad
ENT_DES_PAF CHAMBO		ENTIDAD	Descripción de la parroquia de la Entidad
ENT_DES_TIP_Urbano		ENTIDAD	Descripción del tipo de parroquia de la Entidad
ENT_INT	MSP	ENTIDAD	Institución a la que pertenece de la Entidad
ENT_NIV	NIVEL 1	ENTIDAD	Nivel de atención de la Entidad
ENT_COD_ZO	Z03	ENTIDAD	Código de la zona de la Entidad
ENT_DES_ZON ZONA 3		ENTIDAD	Descripción de la zona de la Entidad
ENT_COD_DIS	06D01	ENTIDAD	Código del distrito de la Entidad
ENT_DIST_DIS CHAMBO, RIOBAMBA		ENTIDAD	Descripción del distrito de la Entidad
ENT_COD_CIR	06D01C20	ENTIDAD	Código del circuito de la Entidad
PROF_NOMBRE ERAZO ESTRELLA XAVIER DA		PROFESIONAL	Nombres completos del Profesional
PROF_NOMBRE XAVIER DANILO		PROFESIONAL	Nombres del Profesional
PROF_APELLIDO ERAZO ESTRELLA		PROFESIONAL	Apellidos del Profesional
PROF_FEC_NA	30298	PROFESIONAL	Fecha de nacimiento del Profesional
PROF_SEXO	Hombre	PROFESIONAL	Sexo del Profesional
PROF_TIP_IDENTIFICACION ?ula de Identidad		PROFESIONAL	Tipo de identificación del Profesional
PROF_IDEN	602468860	PROFESIONAL	Identificación del Profesional
PROF_ESPECIALIDAD AT Medicina Familiar Y Comunitaria		PROFESIONAL	Especialidad en atención del Profesional

**Fuente:** Elaborado por el autor.

## Figura 8

*Variables Fuente de datos Pras.*

PROF_REG_NI	602468860	PROFESIONAL	Numero de registro del Profesional
PROF_CORREO	xerazoestrella@gmail.com	PROFESIONAL	Correo del Profesional
PCTE_NOM	REINO QUISHPI AYISETH NO	PACIENTE	Nombres completos del Paciente
PCTE_NOMBRA	AYISETH NOEMI	PACIENTE	Nombres del Paciente
PCTE_APELLID	REINO QUISHPI	PACIENTE	Apellidos del Paciente
PCTE_TIP_IDE	C?ula de Identidad	PACIENTE	Tipo de identificación del Paciente
PCTE_IDE	651019085	PACIENTE	Identificación del Paciente
PCTE_SEXO	Mujer	PACIENTE	Sexo del Paciente
PCTE_ORI_SEX	NA	PACIENTE	Orientació sexual del Paciente
PCTE_IDE_GE	NA	PACIENTE	Identificación del género del Paciente
PCTE_FEC_NA	44061	PACIENTE	Fecha de nacimiento del Paciente
PCTE_ANIOS	0	PACIENTE	Años del Paciente
PCTE_MESES	5	PACIENTE	Meses del Paciente
PCTE_DIAS	4	PACIENTE	Días del Paciente
PCTE_EDAD_(	0a 5m 4d	PACIENTE	Edad compuesta del Paciente
PCTE_ANIOS_	5	PACIENTE	Edad en años y meses del Paciente
PCTE_IDE_RE	NA	PACIENTE	Identificación del representante del Pacien
PCTE_NACION	ECUATORIANO/A	PACIENTE	Nacionalidad del Paciente
PCTE_AUTID_	NA	PACIENTE	Autoidentificación étnica del Paciente
PCTE_NAC_ET	NA	PACIENTE	Nacionalidad del Paciente
PCTE_PUEBL	CNA	PACIENTE	Pueblo del Paciente
PCTE_TEL_C	EI099-435-9502	PACIENTE	Número de celular del Paciente

**Fuente:** Elaborado por el autor.

A continuación, se expondrá la correspondencia entre los dos modelos:

Las relaciones identificadas fueron las siguientes:

### **Tabla 3**

#### *Correspondencia entre los modelos de variables*

---

La tabla “Paciente” se relaciona con la perspectiva “Paciente”.
La tabla “profesional” con la perspectiva “Médicos”.
La tabla “Establecimiento” con la perspectiva “Establecimiento”.
El campo “IMC” con la perspectiva “Diagnóstico”.
El campo “Fecha Atención” se vinculó con la perspectiva “Tiempo”.
La variable” Id Registro” de la tabla” PACIENTE” representa el indicador” Cantidad de pacientes”.

---

**Fuente:** Elaborado por el autor.

#### c) Nivel de granularidad

De acuerdo con las correspondencias establecidas, se analizaron los campos residentes en cada tabla a la que se hacía referencia, a través de dos métodos diferentes. Primero se examinó la base de datos para intuir los significados de cada campo, y luego los representantes de la coordinación explicaron cada una de las variables en reuniones mantenidas. Los nombres de los campos son bastante explícitos y se deducen con facilidad, pero aun así fue necesario investigarlos para evitar cualquier tipo de inconvenientes.

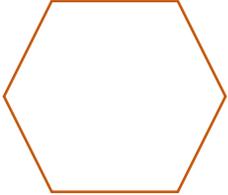
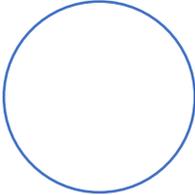
#### d) Modelo Conceptual ampliado

Una vez que se recolectó toda la información pertinente y se consultó con los usuarios cuáles eran los datos que consideraban de interés para analizar los indicadores ya expuestos, los resultados obtenidos fueron los siguientes:

Modelo Dimensional

**Tabla 4**

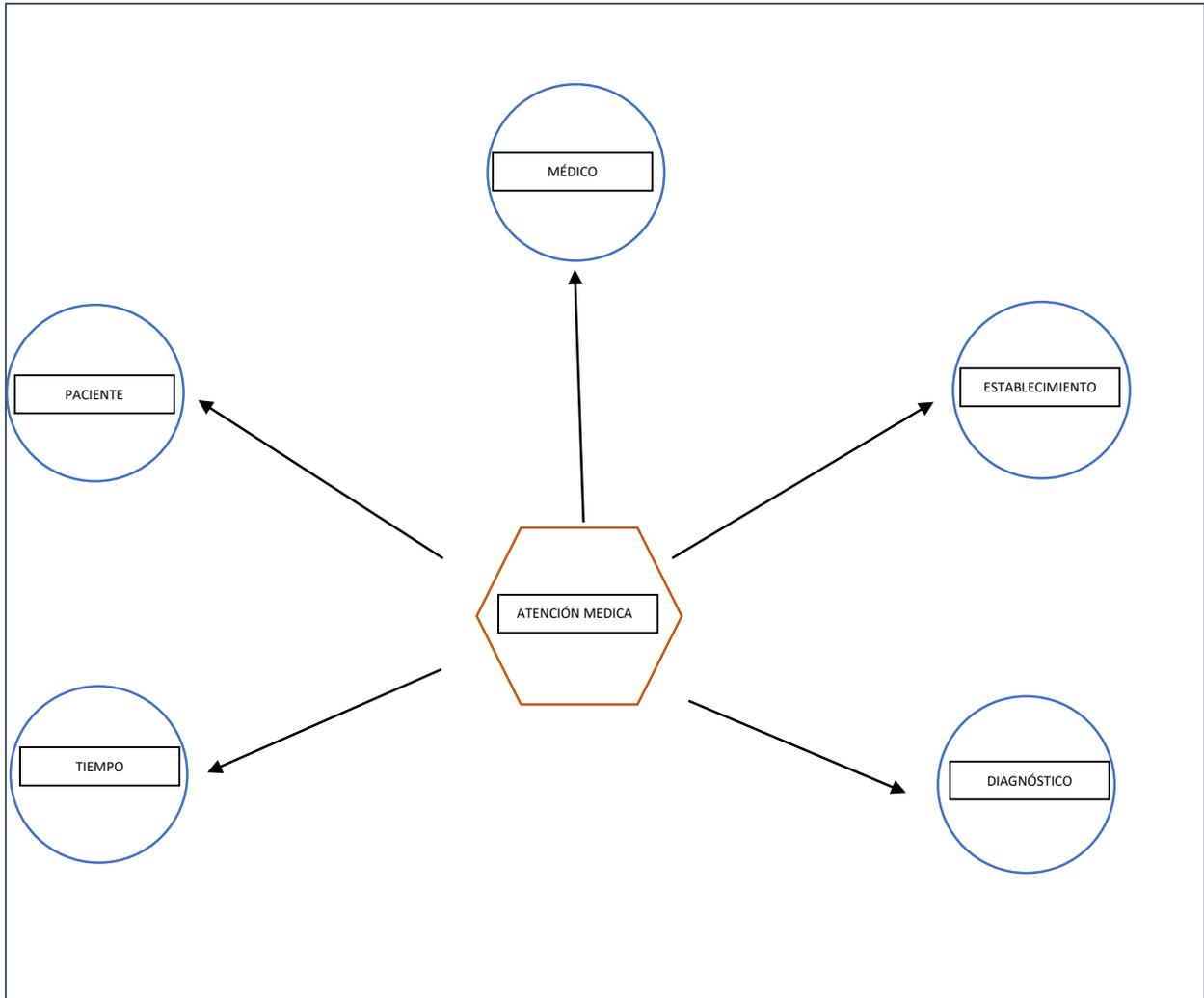
*Descripción de figuras para el modelo direccional*

<b>Figura</b>	<b>Descripción</b>
 A regular hexagon with an orange outline.	Representa un hecho o un data mart.
 A circle with a blue outline.	Representa las dimensiones que están involucradas en el hecho.
 A rounded rectangle with a light blue fill and a green outline.	Representan los atributos que juntos forman una dimensión.
 A black arrow pointing to the right.	Representa la relación entre la dimensión y el hecho o data mart.

**Fuente:** Elaborado por el autor.

**Figura 9**

*Dimensiones y tabla de hechos.*



**Fuente:** Elaborado por el autor.

**Tabla 5***Descripción de dimensiones del DW.*

<b>Dimensiones</b>	<b>Descripción</b>
Paciente	Se refiere a la perspectiva del paciente desde la cual se puede filtrar y agrupar los hechos y sus medidas por medio de sus atributos: <b>sexo, edad, bono, seguro social, nacionalidad, grupo, provincia, cantón, parroquia.</b>
Médico	Hace referencia a la perspectiva sobre los médicos del Ministerio de salud mediante esta dimensión se puede filtrar agrupar las consultas y tratamientos realizados por cada médico.
Establecimiento	Se refiere a la perspectiva sobre los establecimientos las entidades de salud del Ministerio de salud por esta dimensión se puede filtrar agrupar las consultas de que pacientes fueron atendidos en qué <b>lugar y qué tipo de establecimiento.</b>
Diagnóstico	Se refiere a la perspectiva sobre los diagnósticos realizados por el profesional de salud con esta entidad con esta dimensión se puede filtrar agrupar las consultas con la condición del paciente, <b>talla peso anemia código CIE 10.</b>
Tiempo	Hace referencia a la perspectiva de la fecha cuando se realizó la atención para agrupar y promediar las medidas de los hechos por medio de una <b>jerarquía día mes trimestre y año</b>

**Fuente:** Elaborado por el autor.

**Tabla 6***Descripción de variables de la dimensión Paciente.*

<b>NOMBRE:</b>		<b>PACIENTE</b>		<b>DIMENSIÓN:</b>	<b>dim_paciente</b>
<b>TIPO:</b>		<b>Dimensión</b>			
<b>DESCRIPCIÓN:</b>		Tabla que contiene los datos importantes del paciente para generar los hechos.			
<b>NOMBRE</b>		<b>NOMBRE VARIABLE</b>		<b>DESCRIPCION</b>	
		<b>PRAS</b>		<b>RDACA</b>	
<b>1</b>	<b>Pk_paciente</b>				Clave principal
<b>2</b>	<b>nombres</b>	PCTE_NOMBRES			Nombres del paciente
<b>3</b>	<b>apellidos</b>	PCTE_APELLIDOS			Apellidos del paciente
<b>4</b>	<b>identificacion</b>		COD_PER_IDENTIFICACION		Identificación del paciente
<b>5</b>	<b>edad_a</b>		anio_pac		Año de nacimiento del paciente
<b>6</b>	<b>edad_m</b>		mes_pac		Mes de nacimiento del paciente
<b>7</b>	<b>edad_d</b>		dia_pac		Día de nacimiento del paciente
<b>8</b>	<b>provincia</b>		prov_res		Provincia de residencia habitual del paciente
<b>9</b>	<b>cantón</b>		can_res		Cantón de residencia habitual del paciente
<b>10</b>	<b>parroquia</b>		parr_res		Parroquia de residencia habitual del paciente
<b>11</b>	<b>discapacidad</b>	PCTE_DISC			Paciente, tipo de discapacidad
<b>12</b>	<b>sexo</b>		sexo_pac		Sexo del paciente
<b>13</b>	<b>seguro</b>	PCTE_SEG			Seguro social que posea el paciente
<b>14</b>	<b>bono</b>	PCTE_TIP_BON			Tipo de bono del paciente
<b>15</b>	<b>nacionalidad</b>	PCTE_NACIONALIDAD			Nacionalidad del paciente
<b>16</b>	<b>grupo_prioritario</b>	PCTE_GRP_PRI			Grupo prioritario del paciente

**Fuente:** Elaborado por el autor.

**Tabla 7***Descripción de variables de la dimensión Médico.*

<b>NOMBRE:</b>	<b>MEDICO</b>	<b>DIMENSIÓN:</b>	<b>dim_medico</b>
<b>TIPO:</b>	<b>Dimensión</b>		
<b>DESCRIPCIÓN:</b>	Tabla que contiene los datos importantes del médico para generar los hechos.		
<b>NOMBRE</b>	<b>NOMBRE VARIABLE</b>		<b>DESCRIPCION</b>
	<b>PRAS</b>	<b>RDACA</b>	
<b>1</b>	<b>Pk_medico</b>		clave principal
<b>2</b>	<b>nombres</b>	PCTE_NOMBRES	Nombres del Profesional
<b>3</b>	<b>apellidos</b>	PCTE_APELLIDOS	Apellidos del Profesional
<b>4</b>	<b>identificacion</b>	PROF_IDEN	Identificación del Profesional
<b>5</b>	<b>fecha_nacimiento</b>		fecha_nac_prof Fecha de nacimiento del Profesional
<b>6</b>	<b>especialidad</b>	PROF_ESP_ATE	Especialidad en atención del Profesional
<b>7</b>	<b>sexo</b>		sexo_prof Sexo del Profesional
<b>8</b>	<b>nacionalidad</b>		pais_prof Nacionalidad del profesional

**Fuente:** Elaborado por el autor.

**Tabla 8***Descripción de variables de la dimensión Establecimiento.*

<b>NOMBRE:</b>		<b>ESTABLECIMIENTO</b>		<b>DIMENSIÓN</b>	<b>dim_establecimiento</b>
<b>TIPO:</b>		<b>Dimensión</b>			
<b>DESCRIPCIÓN:</b>		Tabla que contiene los datos importantes del establecimiento para generar los hechos.			
<b>NOMBRE</b>		<b>NOMBRE VARIABLE</b>		<b>DESCRIPCION</b>	
		<b>PRAS</b>	<b>RDACA</b>		
<b>1</b>	<b>pk_establecimiento</b>			Clave principal	
<b>2</b>	<b>distrito</b>	ENT_DIST_DIS		Descripción del distrito de la Entidad	
<b>3</b>	<b>zona</b>	ENT_DES_ZON		Descripción de la zona de la Entidad	
<b>4</b>	<b>circuito</b>		circuito_esta	Circuito al cual pertenece el establecimiento	
<b>5</b>	<b>nivel</b>	ENT_NIV		Nivel de atención de la Entidad	
<b>6</b>	<b>tipo</b>	ENT_DES_TIP_EST		Descripción del tipo de la Entidad	
<b>7</b>	<b>nombre</b>		nom_esta	Nombre de establecimiento que se registra la atención	
<b>8</b>	<b>lugar</b>		lugar_aten	Lugar donde el paciente se realiza la atención	
<b>9</b>	<b>parroquia</b>	ENT_DES_TIP_PARR		Descripción del tipo de parroquia de la Entidad	
<b>10</b>	<b>tipo_atencion</b>	COD_EST_TIP_ATTENCION		tipo de atención intra o extra mura	

**Fuente:** Elaborado por el autor.

**Tabla 9**

*Descripción de variables de la dimensión Diagnóstico.*

<b>NOMBRE:</b>		<b>DIAGNÓSTICO</b>		<b>DIMENSIÓN:</b>	<b>dim_diagnostico</b>
<b>TIPO:</b>		<b>Dimensión</b>			
<b>DESCRIPCIÓN:</b>		Tabla que contiene los datos importantes del diagnóstico para generar los hechos.			
<b>NOMBRE</b>		<b>NOMBRE VARIABLE</b>		<b>DESCRIPCION</b>	
		<b>PRAS</b>	<b>RDACA</b>		
<b>1</b>	<b>pk_diagnostico</b>			Clave principal	
<b>2</b>	<b>imc</b>		DAT_ANT_IMC	Índice de masa corporal del paciente	
<b>3</b>	<b>imcClaficacion</b>		DAT_ANT_IMC_CLASIFICACION	Clasificación de datos antropométricos del índice de masa corporal del paciente	
<b>4</b>	<b>anemia</b>		COD_IND_ANEMIA	Datos antropométricos de indicador de anemia del paciente	
<b>5</b>	<b>peso</b>	PCTE_PESO		Peso del Paciente	
<b>6</b>	<b>talla</b>	PCTE_TALLA		Talla del Paciente	
<b>7</b>	<b>codigoCIE</b>		COD_DIAG_CIE_1	Registro de diagnóstico de la codificación cie 10	
<b>8</b>	<b>tipo</b>		COD_DIAG_TIP_DIAGNOSTICO_1	Tipo de diagnóstico cie 10	
<b>9</b>	<b>condicion</b>		COD_DIAG_CON_DIAGNOSTICO_1	Condicion de diagnostico	
<b>10</b>	<b>morbilidad</b>		MORBILIDAD		

**Fuente:** Elaborado por el autor.

**Tabla 10***Descripción de variables de la dimensión Tiempo.*

<b>NOMBRE:</b>	<b>TIEMPO</b>	<b>DIMENSIÓN:</b>	<b>dim_tiempo</b>
<b>TIPO:</b>	<b>Dimensión</b>		
<b>DESCRIPCION:</b>	Tabla que contiene los datos importantes del tiempo para generar los hechos.		
<b>NOMBRE</b>	<b>NOMBRE VARIABLE</b>		<b>DESCRIPCION</b>
	<b>PRAS</b>	<b>RDACA</b>	
<b>1</b>	<b>Pk_tiempo</b>		Clave principal
<b>2</b>	<b>fecha</b>	fecha_aten	FEC_ATENCION
<b>3</b>	<b>anio</b>	fecha_aten	
<b>4</b>	<b>trimestre</b>	fecha_aten	
<b>5</b>	<b>mes</b>	fecha_aten	
<b>6</b>	<b>dia</b>	fecha_aten	

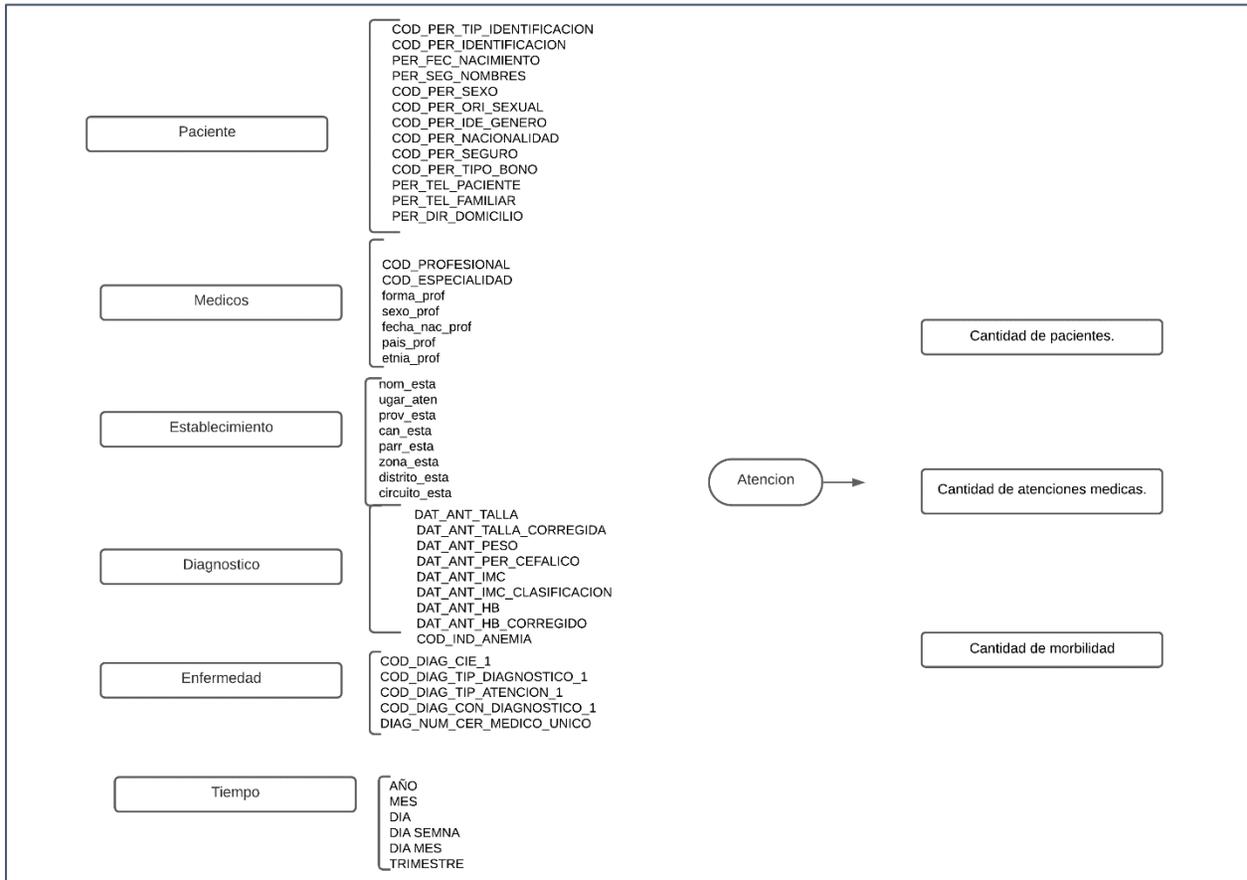
**Fuente:** Elaborado por el autor.**Tabla 11***Descripción de variables de la fact table Atención Médica.*

<b>NOMBRE:</b>	<b>ATENCIÓN</b>	<b>DIMENSIÓN:</b>	<b>H_atencion</b>
<b>TIPO:</b>	<b>Hecho</b>		
<b>DESCRIPCIÓN</b>	Tabla que contiene los datos importantes de la atención médica para consultar información de las dimensiones.		
<b>NOMBRE</b>	<b>NOMBRE VARIABLE</b>		<b>DESCRIPCION</b>
	<b>PRAS</b>	<b>RDACA</b>	
<b>1</b>	<b>Pk_atencion</b>		Clave principal
<b>2</b>	<b>Pk_paciente</b>		Clave foránea
<b>3</b>	<b>Pk_medico</b>		Clave foránea
	<b>Pk_establecimiento</b>		Clave foránea
<b>4</b>	<b>Pk_diagnostico</b>		Clave foránea
<b>5</b>	<b>Pk_tiempo</b>		Clave foránea
<b>6</b>	<b>Cpacientes</b>	identificacion	
<b>7</b>	<b>Catenciones</b>		FEC_ATENCION

Fuente: Elaborado por el autor.

Figura 10

Modelo conceptual Ampliado.



Fuente: Elaborado por el autor.

### 3.6.4. Modelo Lógico del DW

A continuación, se construyó el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que ya había sido creado. Para ello, primero se definirá el tipo de modelo que se utilizará y luego se llevarán a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizan las uniones pertinentes entre estas tablas.

#### a. Tipo de Modelo Lógico del DW

El esquema que se utilizará será en estrella, debido a sus características, ventajas y diferencias

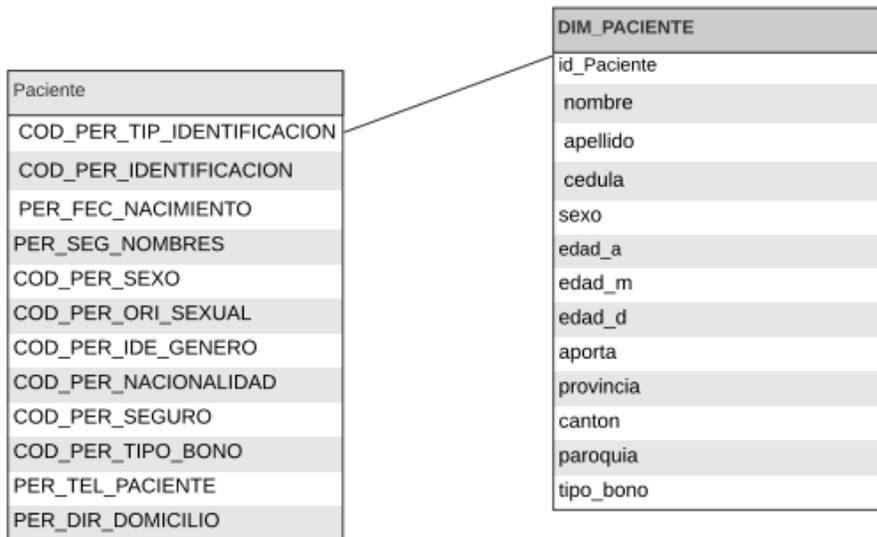
con los otros esquemas.

b. Tablas de dimensiones

A continuación, se diseñarán las tablas de dimensiones.

**Figura 11**

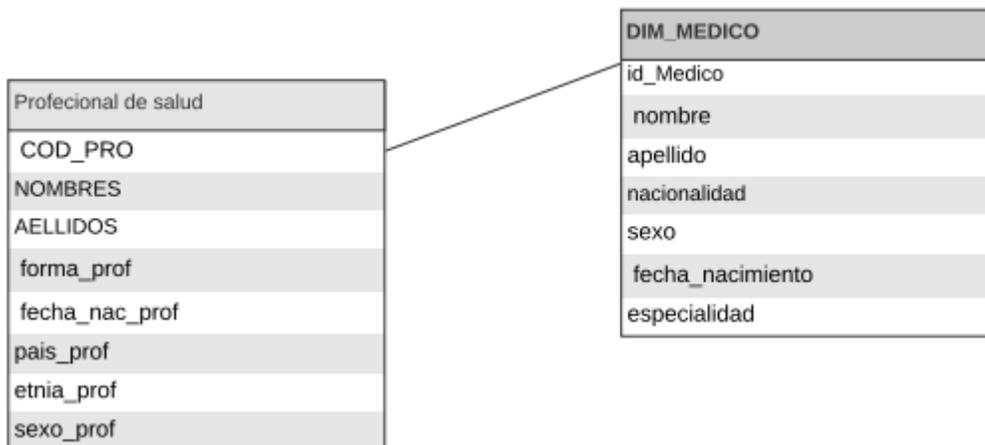
*Dimensión Paciente*



**Fuente:** Elaborado por el autor.

**Figura 12**

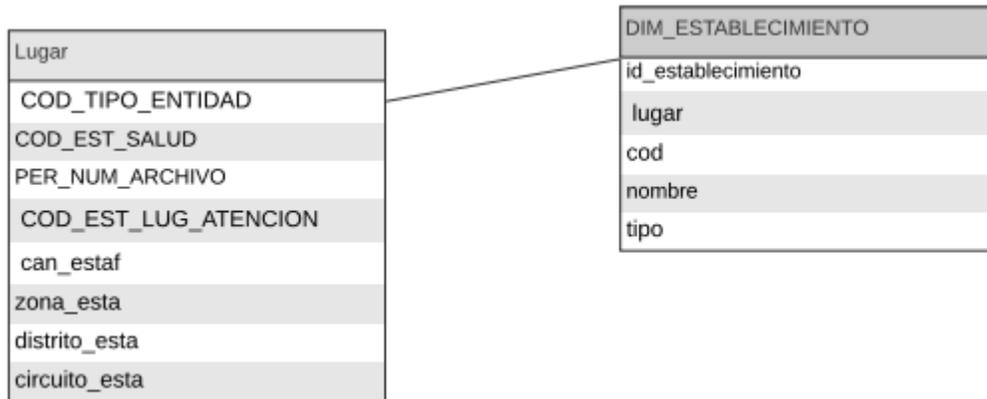
*Dimensión Médico.*



**Fuente:** Elaborado por el autor.

**Figura 13**

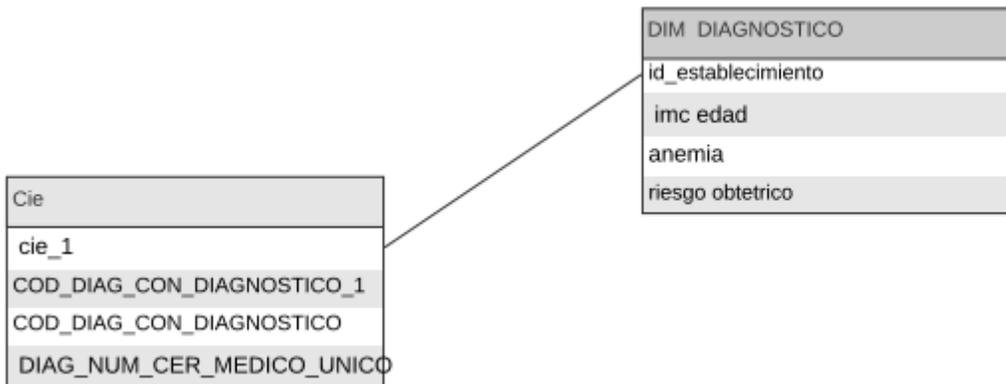
*Dimensión Establecimiento.*



**Fuente:** Elaborado por el autor.

**Figura 14**

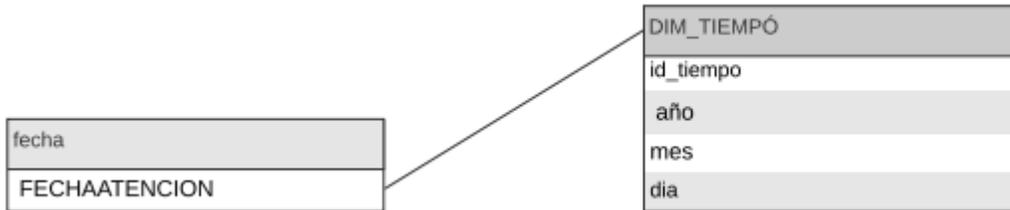
*Dimensión Diagnóstico.*



**Fuente:** Elaborado por el autor.

**Figura 15**

*Dimensión Tiempo*



**Fuente:** Elaborado por el autor.

c. Tablas de hechos

A continuación, se elabora la tabla de hechos:

La tabla de hechos tendrá el nombre “ATENCIÓN”.

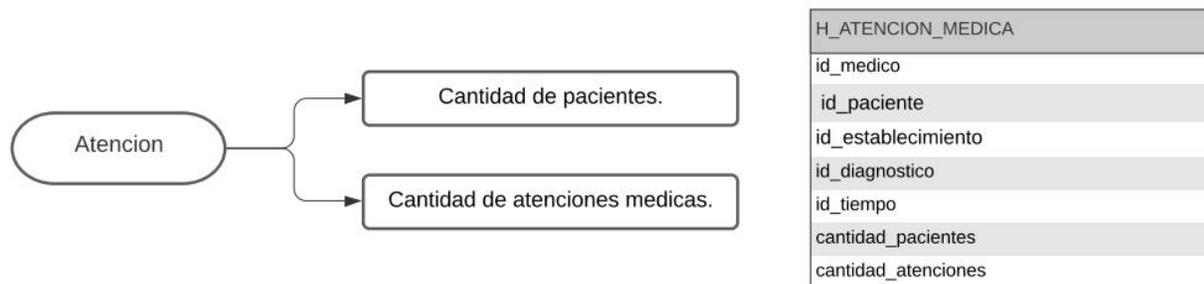
Su clave principal será la combinación de las claves principales de las tablas de dimensiones antes definidas: “id paciente”, “id médico”, “id establecimiento”, “id diagnostico”, “id tiempo”.

Se crearán dos indicadores y serán renombrados, “Cantidad pacientes” y “Cantidad de atenciones”

En el gráfico siguiente se puede apreciar mejor este paso:

**Figura 16**

*Fact-table Atención Médica*



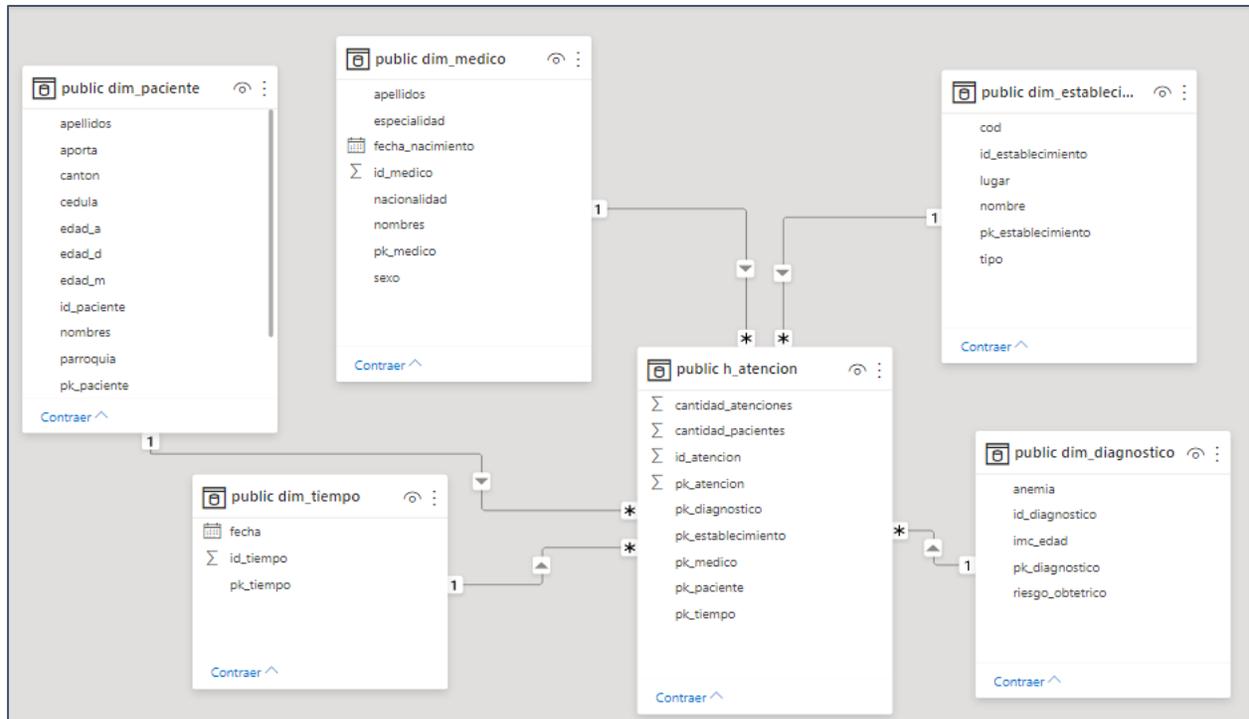
**Fuente:** Elaborado por el autor.

#### d. Uniones

Se realizaron las uniones pertinentes, de acuerdo las dimensiones y la tabla de hechos.

**Figura 17**

*Uniones de dimensiones*



**Fuente:** Elaborado por el autor.

### 3.6.5. Integración de Datos

Una vez construido el modelo lógico, se procedió a poblarlo con datos, no sin antes utilizar técnicas de limpieza y calidad de datos, denominados procesos ETL. Para el ETL se llevó adelante una serie de tareas básicas como son extracción transformación y carga con dos herramientas denominadas PDI Y Talend Data Preparation.

## Extracción

Para realizar de manera correcta el proceso de extracción se realizó los siguientes pasos:

Extracción de los datos desde los sistemas de origen PRASS y RDACAA.

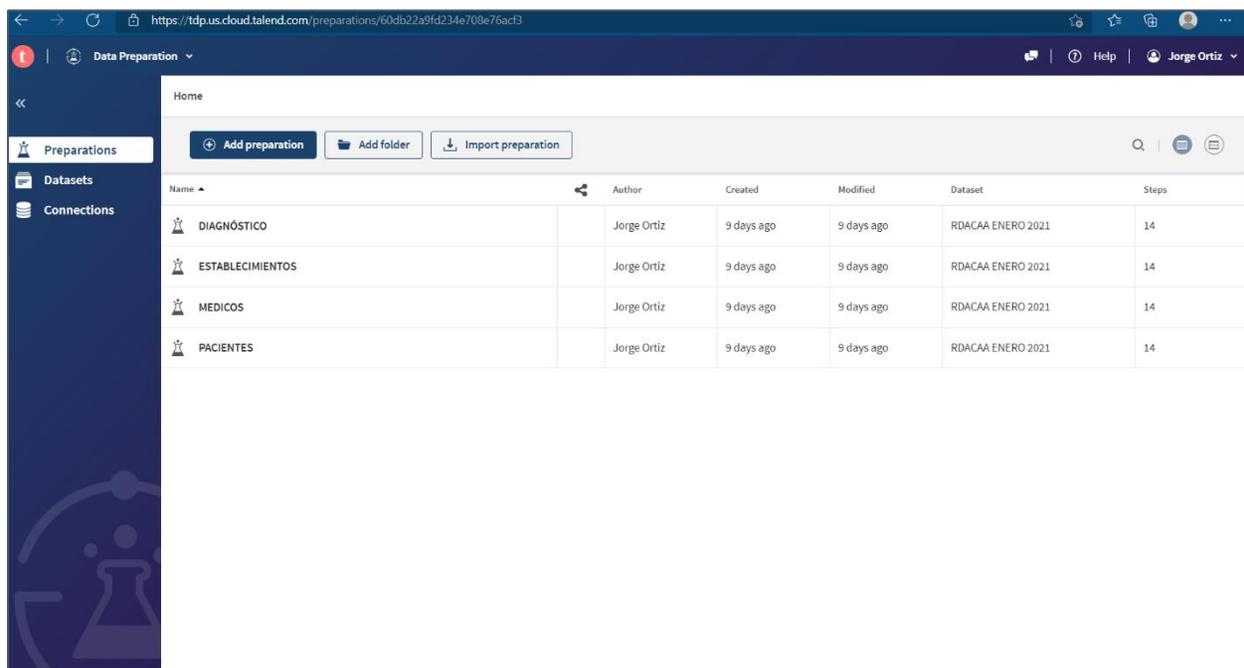
Se analizaron los datos extraídos aplicando un chequeo mediante la herramienta Talend Data **Preparación.**

Se verificó que los datos extraídos cumplan alguna estructura de organización de datos.

Aplicamos un formato estándar a los datos para iniciar el proceso de transformación con información fiable.

## Figura 18

### *Data Preparation*



The screenshot shows the Talend Data Preparation web interface. The browser address bar displays the URL: <https://tdp.us.cloud.talend.com/preparations/60db22a9fd234e708e76acf3>. The interface includes a navigation menu on the left with 'Preparations', 'Datasets', and 'Connections'. The main content area shows a 'Home' header with buttons for 'Add preparation', 'Add folder', and 'Import preparation'. Below this is a table listing four data preparations:

Name	Author	Created	Modified	Dataset	Steps
DIAGNÓSTICO	Jorge Ortiz	9 days ago	9 days ago	RDACAA ENERO 2021	14
ESTABLECIMIENTOS	Jorge Ortiz	9 days ago	9 days ago	RDACAA ENERO 2021	14
MEDICOS	Jorge Ortiz	9 days ago	9 days ago	RDACAA ENERO 2021	14
PACIENTES	Jorge Ortiz	9 days ago	9 days ago	RDACAA ENERO 2021	14

**Fuente:** Elaborado por el autor.

## Transformación

Se aplicaron una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados al DW como son:

Limpie los datos irrelevantes de los conjuntos de datos PRAS y RDACAA, borrar filas duplicadas, borrar columnas duplicadas, dividir campos ejemplo nombres = nombre apellido,

Cambiar formatos de tipos de datos ejemplo fechas dd-mm-aa edad.

En la figura se puede observar la serie de reglas aplicadas.

## Figura 19

*Limpieza de datos.*

The screenshot shows the 'Data Preparation' interface for a dataset named 'RDACAA'. On the left, there is a list of 8 data cleaning rules:

- Delete column on column Archivo
- Delete column on column Hierro + Acido fólico (Tabletas)
- Delete column on column Hierro jarabe (24 a 59 meses de edad)
- Delete column on column Vitamina A cápsulas (6 a 59 meses d...)
- Delete column on column Hierro, multivitaminas y minerales e...
- Delete column on column Dientes perdidos
- Delete column on column Dientes obturados
- Delete column on column Dientes extraídos

On the right, the 'Filters' table is displayed with the following columns: ULA, SEXO (text), ORIENTACION SEXUAL (text), FECHA NACIMIEN... (date), and EDAD AÑOS (integer). The table contains 11 rows of data:

	ULA	SEXO	ORIENTACION SEXUAL	FECHA NACIMIEN...	EDAD AÑOS
1		Hombre	Heterosexual	1960-07-30	
2		Mujer	No aplica	2011-09-24	
3		Hombre	Heterosexual	2009-12-20	
4		Mujer	No aplica	2017-08-27	
5		Mujer	Heterosexual	1998-02-09	
6		Mujer	Heterosexual	1982-03-25	
7		Mujer	Heterosexual	2002-06-04	
8		Hombre	Heterosexual	1958-11-15	
9		Hombre	Heterosexual	1969-01-24	
10		Mujer	Heterosexual	1940-09-15	
11		Mujer	Heterosexual	1946-02-02	

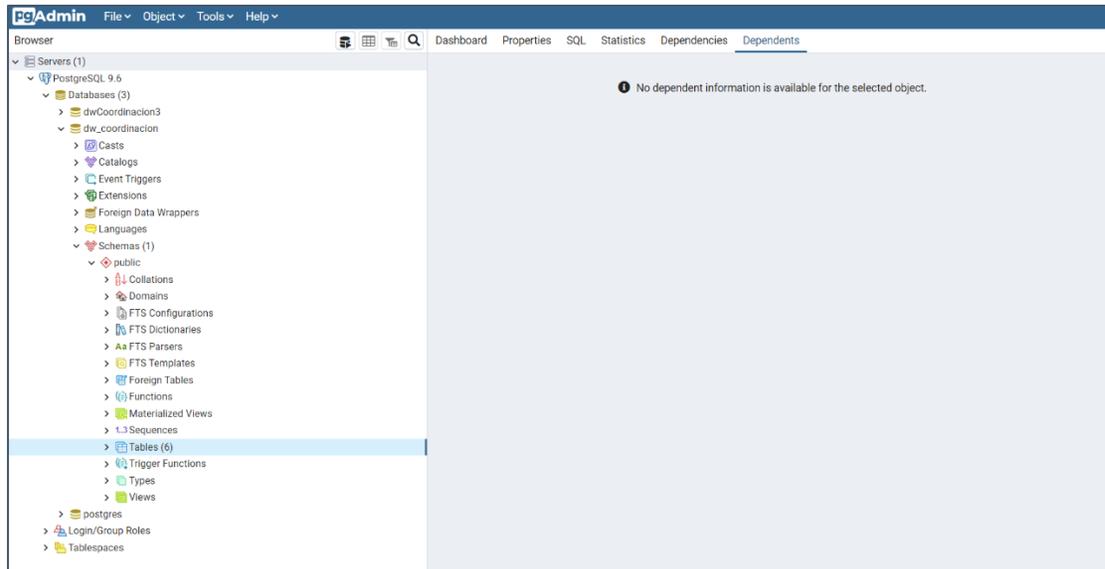
**Fuente:** Elaborado por el autor.

### a) Carga Inicial

Se utilizó PostgreSQL 9.6 y pgAdmin 4 para la creación de la base de datos y en la cual se alojarán los datos de las dimensiones establecidas luego se realizó la conexión con herramienta PDI Pentaho Data Integration para que el proceso ETL pueda agregar los datos en el DW.

**Figura 20**

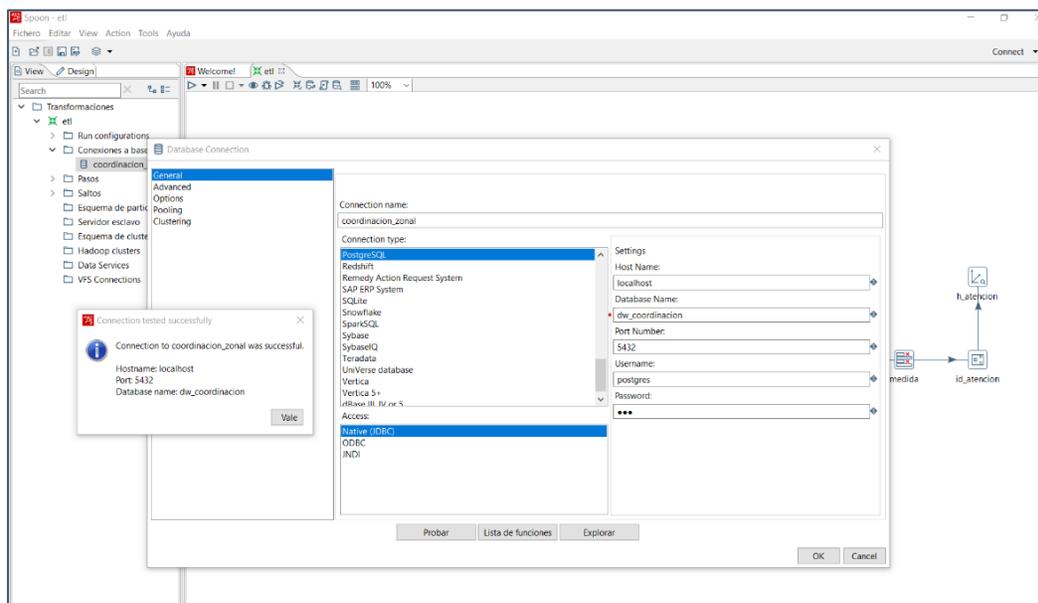
*Sistema de gestión de bases de datos Postgresql 9.6*



**Fuente:** Elaborado por el autor.

**Figura 21**

*Conexión de PostgreSQL 9.6. y Pentaho Data Integration.*



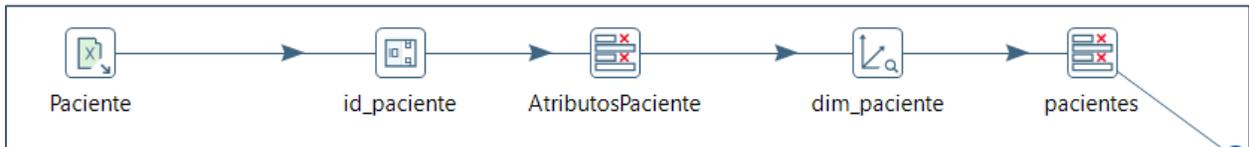
**Fuente:** Elaborado por el autor.

El proceso ETL planteado para la Carga Inicial es el siguiente las tareas que lleva a cabo este proceso son:

Crear y poblar la dimensión PACIENTE:

**Figura 22**

*ETL Paciente.*



**Fuente:** Elaborado por el autor.

Crear y poblar la dimensión MÉDICO:

**Figura 23**

*ETL Médico.*

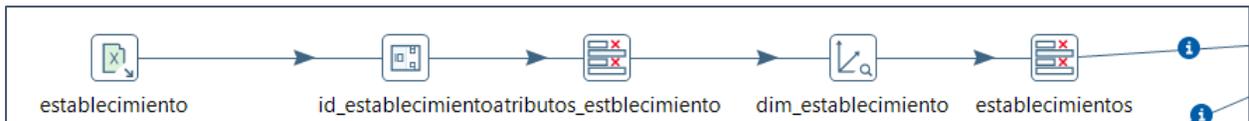


**Fuente:** Elaborado por el autor.

Crear y poblar la dimensión ESTABLECIMIENTO:

**Figura 24**

*ETL Establecimiento.*

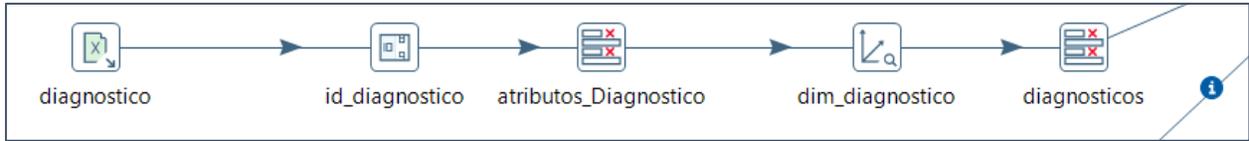


**Fuente:** Elaborado por el autor.

Crear y poblar la dimensión DIAGNÓSTICO:

**Figura 25**

*ETL Diagnóstico.*

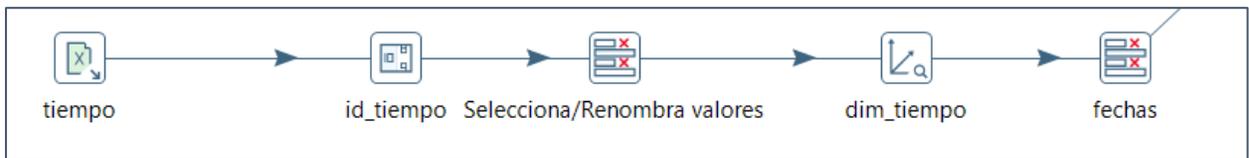


**Fuente:** Elaborado por el autor.

Crear y poblar la dimensión TIEMPO:

**Figura 26**

*ETL Tiempo.*

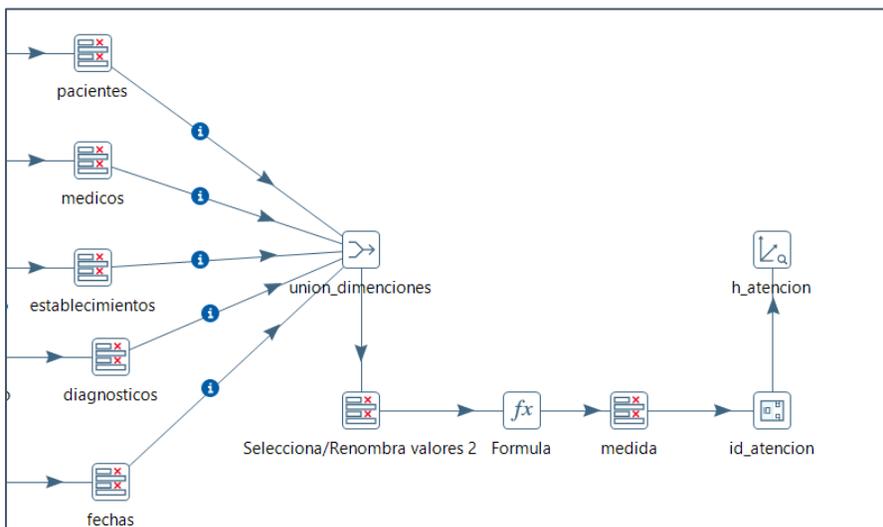


**Fuente:** Elaborado por el autor.

Carga de Hecho ATENCIÓN:

**Figura 27**

*ETL Atención.*

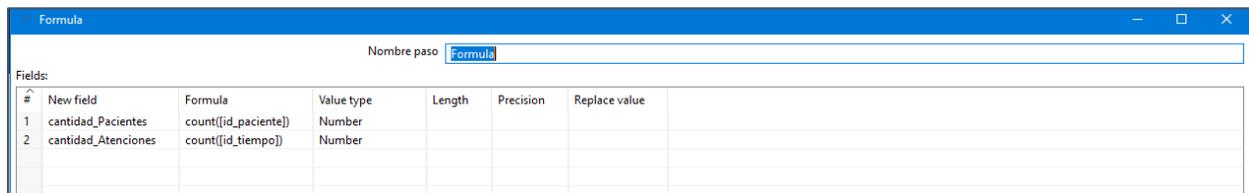


**Fuente:** Elaborado por el autor.

El proceso ETL que se realizó empieza por extraer los datos de las fuentes, crea un id para cada dimensión del DW se somete a una selección de atributos y crea la dimensión con un SQL, luego se unen todas las dimensiones para llegar a identificar todos los Primary keys (pks) de las mismas una vez obtenidas se crean las medidas o indicadores para el análisis de la información finalmente se crea la tabla de hechos con los Primary keys (pks) y las medidas.

## Figura 28

*Fórmulas.*



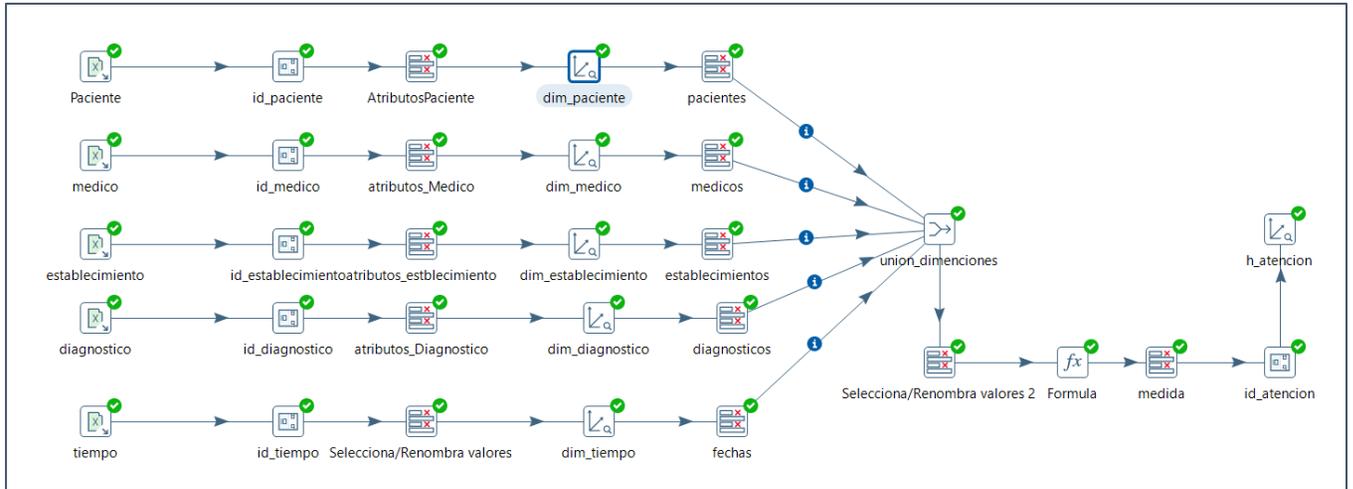
#	New field	Formula	Value type	Length	Precision	Replace value
1	cantidad_Pacientes	count([id_paciente])	Number			
2	cantidad_Atenciones	count([id_tiempo])	Number			

**Fuente:** Elaborado por el autor.

Continúa el proceso con el ítem formula el mismo que se encarga de crear las métricas o también llamadas medidas para medir o analizar a los pacientes, luego con el ítem medida se selecciona los Primary keys (pks) junto con las métricas creadas para posteriormente crear la Tabla de hechos finalizando en ETL.

**Figura 29**

*Proceso ETL.*



**Fuente:** Elaborado por el autor.

Una vez concluido el proceso ETL se muestran los resultados con los datos poblados en el almacén de datos.

**Figura 30**

*Datos poblados en el DW.*

#	Nombre paso	Numero Copia	Leído	Escrito	Entrada	Salida	Actualizado	Rejected	Errores	Activo	Tiempo	Velocidad (r/s)	Pt/E/S
1	medico	0	0	16455	16455	0	0	0	0	Finalizado	18.8s	875	-
2	Paciente	0	0	16455	16455	0	0	0	0	Finalizado	21.7s	759	-
3	establecimiento	0	0	16455	16455	0	0	0	0	Finalizado	22.3s	738	-
4	id_establecimiento	0	16455	16455	0	0	0	0	0	Finalizado	22.3s	738	-
5	diagnostico	0	0	16455	16455	0	0	0	0	Finalizado	19.9s	826	-
6	tiempo	0	0	16455	16455	0	0	0	0	Finalizado	21.8s	756	-
7	id_paciente	0	16455	16455	0	0	0	0	0	Finalizado	21.7s	757	-
8	id_tiempo	0	16455	16455	0	0	0	0	0	Finalizado	22.3s	738	-
9	id_diagnostico	0	16455	16455	0	0	0	0	0	Finalizado	21.7s	759	-
1.	atributos_Diagnostico	0	16455	16455	0	0	0	0	0	Finalizado	25.6s	642	-
1.	atributos_estblecimiento	0	16455	16455	0	0	0	0	0	Finalizado	25.9s	636	-
1.	Selecciona/Renombra valores	0	16455	16455	0	0	0	0	0	Finalizado	24.7s	666	-
1.	AtributosPaciente	0	16455	16455	0	0	0	0	0	Finalizado	26.5s	620	-
1.	id_medico	0	16455	16455	0	0	0	0	0	Finalizado	18.8s	874	-
1.	dim_diagnostico	0	16455	16455	16455	0	0	0	0	Finalizado	29.2s	564	-
1.	atributos_Medico	0	16455	16455	0	0	0	0	0	Finalizado	26.2s	628	-
1.	dim_medico	0	16455	16455	16455	0	0	0	0	Finalizado	29.9s	551	-
1.	diagnosticos	0	16455	16455	0	0	0	0	0	Finalizado	29.2s	563	-
1.	medicos	0	16455	16455	0	0	0	0	0	Finalizado	29.9s	551	-
2.	dim_establecimiento	0	16455	16455	16455	0	0	0	0	Finalizado	29.4s	560	-
2.	establecimientos	0	16455	16455	0	0	0	0	0	Finalizado	29.4s	559	-
2.	dim_tiempo	0	16455	16455	16455	0	0	0	0	Finalizado	27.2s	604	-
2.	fechas	0	16455	16455	0	0	0	0	0	Finalizado	27.2s	604	-
2.	dim_paciente	0	16455	16455	16455	0	0	0	0	Finalizado	30.6s	537	-
2.	pacientes	0	16455	16455	0	0	0	0	0	Finalizado	30.6s	537	-
2.	union dimensiones	0	82275	16455	0	0	0	0	0	Finalizado	32.3s	2,550	-
2.	Selecciona/Renombra valores 2	0	16455	16455	0	0	0	0	0	Finalizado	32.3s	510	-
2.	Formula	0	16455	16455	0	0	0	0	0	Finalizado	32.3s	509	-
2.	medida	0	16455	16455	0	0	0	0	0	Finalizado	32.3s	509	-
3.	id_atencion	0	16455	16455	0	0	0	0	0	Finalizado	32.3s	509	-
3.	h_atencion	0	16455	16455	16455	0	0	0	0	Finalizado	32.4s	508	-

Fuente: Elaborado por el autor.

Figura 31

Datos Pacientes.

pk_paciente	version	date_from	date_to	id_paciente	nombre	apellido	cedula	sexo
1	0	1 [null]	[null]	[null]	[null]	[null]	[null]	[null]
2	1	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	1	[null]	[null]	[null]	Homb
3	2	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	2	[null]	[null]	[null]	Mujer
4	3	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	3	[null]	[null]	[null]	Homb
5	4	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	4	[null]	[null]	[null]	Mujer
6	5	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	5	[null]	[null]	[null]	Mujer
7	6	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	6	[null]	[null]	[null]	Mujer
8	7	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	7	[null]	[null]	[null]	Mujer
9	8	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	8	[null]	[null]	[null]	Homb
10	9	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	9	[null]	[null]	[null]	Homb
11	10	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	10	[null]	[null]	[null]	Mujer
12	11	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	11	[null]	[null]	[null]	Mujer
13	12	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	12	[null]	[null]	[null]	Mujer
14	13	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	13	[null]	[null]	[null]	Mujer
15	14	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	14	[null]	[null]	[null]	Mujer
16	15	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	15	[null]	[null]	[null]	Homb
17	16	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	16	[null]	[null]	[null]	Mujer

Fuente: Elaborado por el autor.

Figura 32

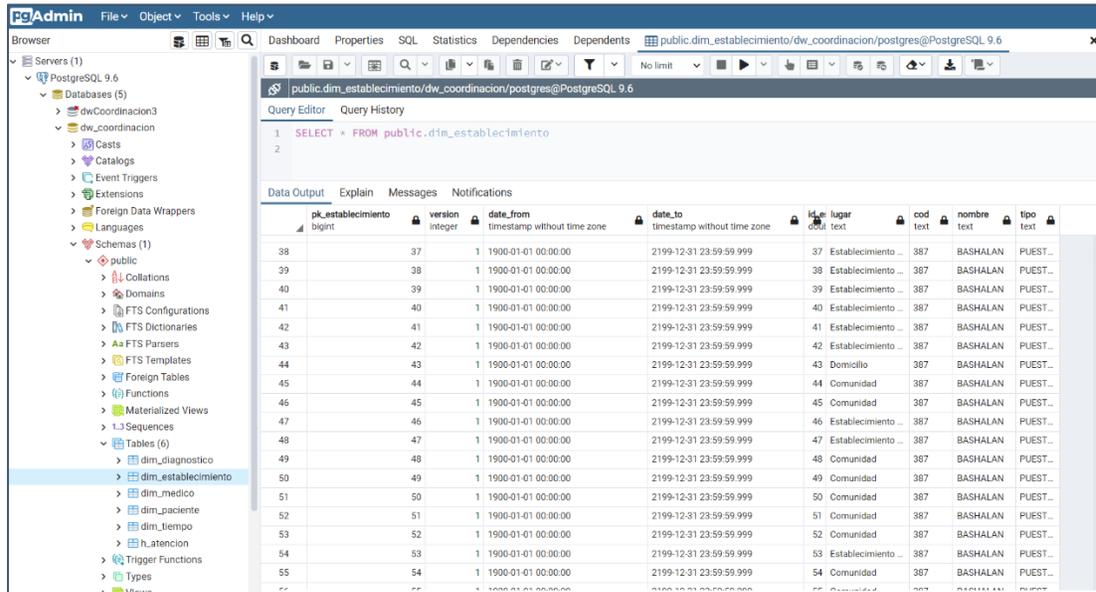
Datos Médicos.

pk_medico	version	date_from	date_to	id_medico	nombre	apellido	nacionalidad	sexo	fecha_nacimiento	especialidad
1	0	1 [null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]	[null]
2	1	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	1	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
3	2	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	2	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
4	3	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	3	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
5	4	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	4	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
6	5	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	5	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
7	6	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	6	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
8	7	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	7	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
9	8	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	8	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
10	9	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	9	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
11	10	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	10	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
12	11	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	11	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
13	12	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	12	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
14	13	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	13	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
15	14	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	14	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
16	15	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	15	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi
17	16	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	16	FAUSTO	VAQUILEMA	EC	Hombre	1979-06-26 00:00:00	Medi

Fuente: Elaborado por el autor.

### Figura 33

Datos Establecimientos.



The screenshot shows the PostgreSQL Admin interface. The left sidebar displays a tree view of the database structure, with 'public.dim\_establecimiento' selected. The main window shows a query editor with the following SQL query:

```
1 SELECT * FROM public.dim_establecimiento
2
```

Below the query editor, the 'Data Output' tab is active, displaying a table with the following columns and data:

pk_establecimiento	version	date_from	date_to	lugar	cod	nombre	tipo
38	37	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
39	38	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
40	39	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
41	40	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
42	41	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
43	42	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
44	43	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Domicilio	387	BASHALAN	PUEST...
45	44	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Comunidad	387	BASHALAN	PUEST...
46	45	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Comunidad	387	BASHALAN	PUEST...
47	46	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
48	47	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
49	48	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Comunidad	387	BASHALAN	PUEST...
50	49	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Comunidad	387	BASHALAN	PUEST...
51	50	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Comunidad	387	BASHALAN	PUEST...
52	51	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Comunidad	387	BASHALAN	PUEST...
53	52	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Comunidad	387	BASHALAN	PUEST...
54	53	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Establecimiento ...	387	BASHALAN	PUEST...
55	54	1 1900-01-01 00:00:00	2199-12-31 23:59:59.999	Comunidad	387	BASHALAN	PUEST...

Fuente: Elaborado por el autor.

Figura 34

Datos Diagnóstico.

The screenshot shows the PgAdmin interface with a query executed against the 'public.dim\_diagnostico' table. The query is 'SELECT \* FROM public.dim\_diagnostico'. The results are displayed in a table with the following columns: pk\_diagnostico, version, date\_from, date\_to, id\_diagnostico, ime\_adad, anemia, and riesgo\_obterico. The data shows 18 rows of records, each with a unique primary key and a timestamp range from 1900-01-01 00:00:00 to 2199-12-31 23:59:59.999. The 'ime\_adad' and 'anemia' columns contain values like 'NO APLICACION' or '380 Normal/S...'. The 'riesgo\_obterico' column contains values like 'NO APLICACION' or '2574 Bajo Riesgo'. A status message at the bottom indicates 'Successfully run. Total query runtime: 191 msec. 16456 rows affected.'

pk_diagnostico	version	date_from	date_to	id_diagnostico	ime_adad	anemia	riesgo_obterico
1	0	[null]	[null]				
2	1	1900-01-01 00:00:00	2199-12-31 23:59:59.999		1 NO APLICACION	NO APLICACION	NO APLICACION
3	2	1900-01-01 00:00:00	2199-12-31 23:59:59.999		2 380 Normal/S...	NO APLICACION	NO APLICACION
4	3	1900-01-01 00:00:00	2199-12-31 23:59:59.999		3 379 Sobrepos...	NO APLICACION	NO APLICACION
5	4	1900-01-01 00:00:00	2199-12-31 23:59:59.999		4 373 Normal...	NO APLICACION	NO APLICACION
6	5	1900-01-01 00:00:00	2199-12-31 23:59:59.999		5 NO APLICACION	NO APLICACION	NO APLICACION
7	6	1900-01-01 00:00:00	2199-12-31 23:59:59.999		6 NO APLICACION	NO APLICACION	NO APLICACION
8	7	1900-01-01 00:00:00	2199-12-31 23:59:59.999		7 NO APLICACION	NO APLICACION	NO APLICACION
9	8	1900-01-01 00:00:00	2199-12-31 23:59:59.999		8 NO APLICACION	NO APLICACION	NO APLICACION
10	9	1900-01-01 00:00:00	2199-12-31 23:59:59.999		9 NO APLICACION	NO APLICACION	NO APLICACION
11	10	1900-01-01 00:00:00	2199-12-31 23:59:59.999		10 NO APLICACION	NO APLICACION	NO APLICACION
12	11	1900-01-01 00:00:00	2199-12-31 23:59:59.999		11 NO APLICACION	NO APLICACION	NO APLICACION
13	12	1900-01-01 00:00:00	2199-12-31 23:59:59.999		12 NO APLICACION	NO APLICACION	NO APLICACION
14	13	1900-01-01 00:00:00	2199-12-31 23:59:59.999		13 380 Normal/S...	NO APLICACION	NO APLICACION
15	14	1900-01-01 00:00:00	2199-12-31 23:59:59.999		14 NO APLICACION	NO APLICACION	NO APLICACION
16	15	1900-01-01 00:00:00	2199-12-31 23:59:59.999		15 NO APLICACION	NO APLICACION	NO APLICACION
17	16	1900-01-01 00:00:00	2199-12-31 23:59:59.999		16 NO APLICACION	NO APLICACION	NO APLICACION
18	17	1900-01-01 00:00:00	2199-12-31 23:59:59.999		17 NO APLICACION	NO APLICACION	2574 Bajo Riesgo

Fuente: Elaborado por el autor.

Figura 35

Datos Tiempo.

The screenshot shows the PgAdmin interface with a query executed against the 'public.dim\_tiempo' table. The query is 'SELECT \* FROM public.dim\_tiempo'. The results are displayed in a table with the following columns: pk\_tiempo, version, date\_from, date\_to, id\_tiempo, and fecha. The data shows 113 rows of records, each with a unique primary key and a timestamp range from 1900-01-01 00:00:00 to 2199-12-31 23:59:59.999. The 'fecha' column contains values like '2021-01-05 00:00:00' or '2021-01-28 00:00:00'. A status message at the bottom indicates 'Successfully run. Total query runtime: 191 msec. 16456 rows affected.'

pk_tiempo	version	date_from	date_to	id_tiempo	fecha
95	94	1900-01-01 00:00:00	2199-12-31 23:59:59.999		94 2021-01-05 00:00:00
96	95	1900-01-01 00:00:00	2199-12-31 23:59:59.999		95 2021-01-07 00:00:00
97	96	1900-01-01 00:00:00	2199-12-31 23:59:59.999		96 2021-01-07 00:00:00
98	97	1900-01-01 00:00:00	2199-12-31 23:59:59.999		97 2021-01-07 00:00:00
99	98	1900-01-01 00:00:00	2199-12-31 23:59:59.999		98 2021-01-07 00:00:00
100	99	1900-01-01 00:00:00	2199-12-31 23:59:59.999		99 2021-01-07 00:00:00
101	100	1900-01-01 00:00:00	2199-12-31 23:59:59.999		100 2021-01-25 00:00:00
102	101	1900-01-01 00:00:00	2199-12-31 23:59:59.999		101 2021-01-25 00:00:00
103	102	1900-01-01 00:00:00	2199-12-31 23:59:59.999		102 2021-01-25 00:00:00
104	103	1900-01-01 00:00:00	2199-12-31 23:59:59.999		103 2021-01-25 00:00:00
105	104	1900-01-01 00:00:00	2199-12-31 23:59:59.999		104 2021-01-25 00:00:00
106	105	1900-01-01 00:00:00	2199-12-31 23:59:59.999		105 2021-01-25 00:00:00
107	106	1900-01-01 00:00:00	2199-12-31 23:59:59.999		106 2021-01-26 00:00:00
108	107	1900-01-01 00:00:00	2199-12-31 23:59:59.999		107 2021-01-26 00:00:00
109	108	1900-01-01 00:00:00	2199-12-31 23:59:59.999		108 2021-01-26 00:00:00
110	109	1900-01-01 00:00:00	2199-12-31 23:59:59.999		109 2021-01-26 00:00:00
111	110	1900-01-01 00:00:00	2199-12-31 23:59:59.999		110 2021-01-26 00:00:00
112	111	1900-01-01 00:00:00	2199-12-31 23:59:59.999		111 2021-01-28 00:00:00
113	112	1900-01-01 00:00:00	2199-12-31 23:59:59.999		112 2021-01-28 00:00:00

Fuente: Elaborado por el autor.

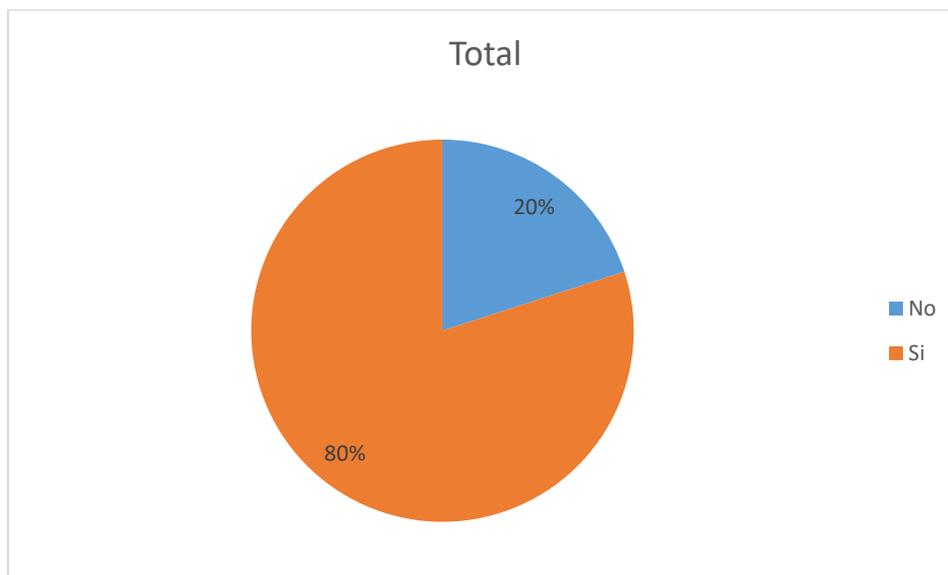
### 3.6.6. *Análisis de Información*

Para la verificación del análisis de información y validación del DW se aplicó una encuesta basada en preguntas cerradas, con dos partes la primera validación de atributos y luego validación de análisis, las encuestas fueron aplicadas a 10 usuarios expertos en DW (6 usuarios expertos externos a la institución y 4 usuarios expertos pertenecientes a la Carrera de Tecnologías de la información) obteniendo los siguientes resultados:

**Pregunta 1.-** ¿Considera que la dimensión PACIENTE está adecuadamente incorporada en la estructura del almacén teniendo en cuenta que desde la cual se puede filtrar y agrupar los hechos y sus medidas por medio de sus atributos: apellidos, nombres, sexo, edad, bono, seguro social, ¿nacionalidad, grupo, provincia, cantón, parroquia en el almacén de datos?

**Figura 36**

*Porcentaje de satisfacción de atributos de la dimensión Paciente.*

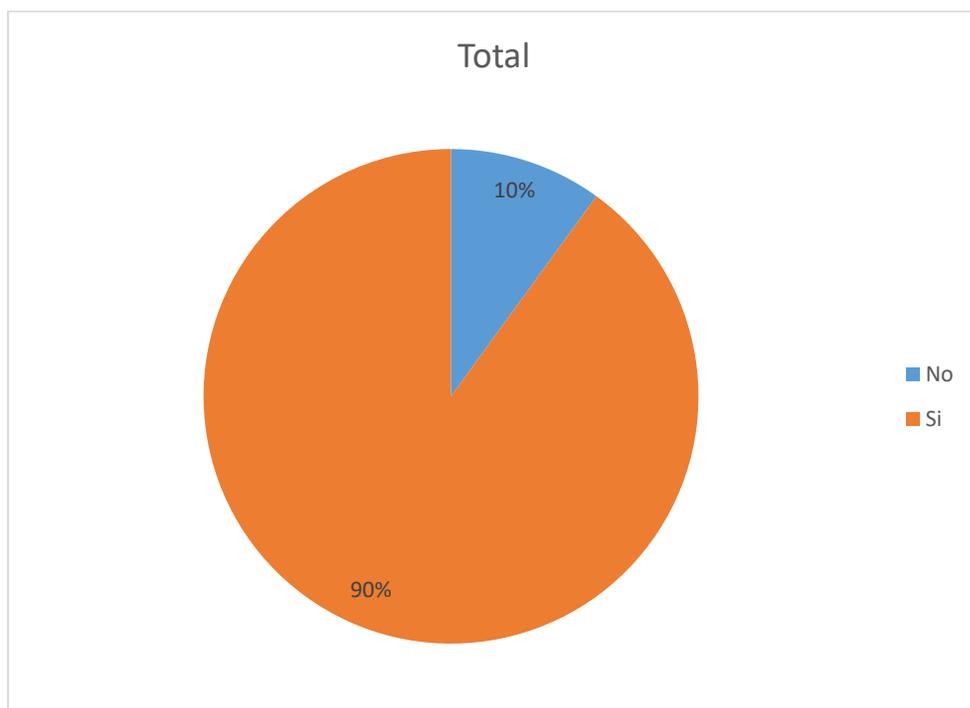


**Fuente:** Elaborado por el autor.

**Pregunta 2.-** ¿Considera que la dimensión MÉDICO está adecuadamente incorporada en la estructura del almacén teniendo en cuenta que mediante esta dimensión se puede filtrar agrupar las consultas y tratamientos realizados por cada médico por sus atributos: nombres, apellidos, identificación, fecha nacimiento, especialidad, sexo, nacionalidad en el almacén de datos?

**Figura 37**

*Porcentaje de satisfacción de atributos de la dimensión Médico.*

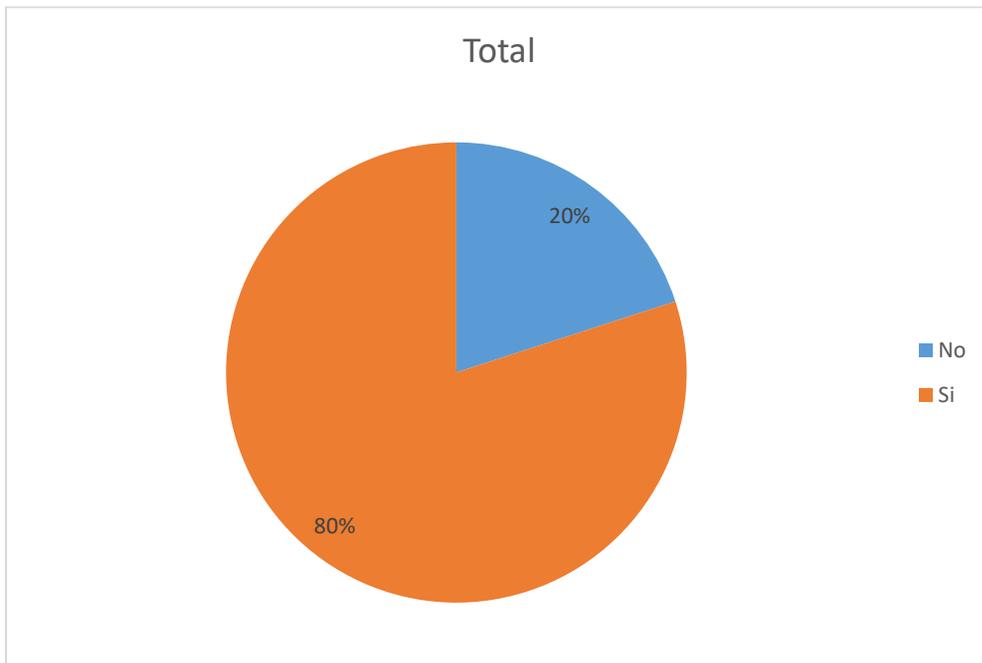


**Fuente:** Elaborado por el autor.

**Pregunta 3.-** ¿Considera que la dimensión DIAGNÓSTICO está bien incorporada en la estructura del almacén teniendo en cuenta que mediante esta dimensión se puede filtrar agrupar las consultas por sus atributos: imcClaficacion, anemia, peso, talla, codigoCIE, tipo, ¿condición en el almacén de datos?

**Figura 38**

*Porcentaje de satisfacción de atributos de la dimensión Diagnóstico.*

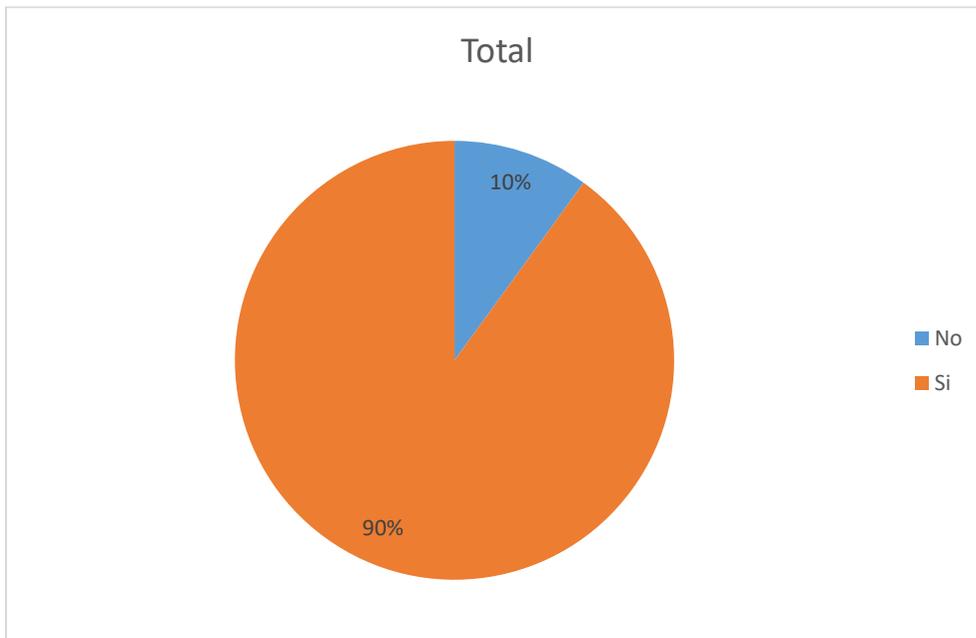


**Fuente:** Elaborado por el autor.

**Pregunta 4.-** ¿Considera que la dimensión ESTABLECIMIENTO está adecuadamente incorporada en la estructura del almacén teniendo en cuenta que mediante esta dimensión se puede filtrar agrupar las consultas por zona, circuito, nivel, tipo, nombre, lugar, parroquia, tipo atención?

**Figura 39**

*Porcentaje de satisfacción de atributos de la dimensión Establecimiento.*



**Fuente:** Elaborado por el autor.

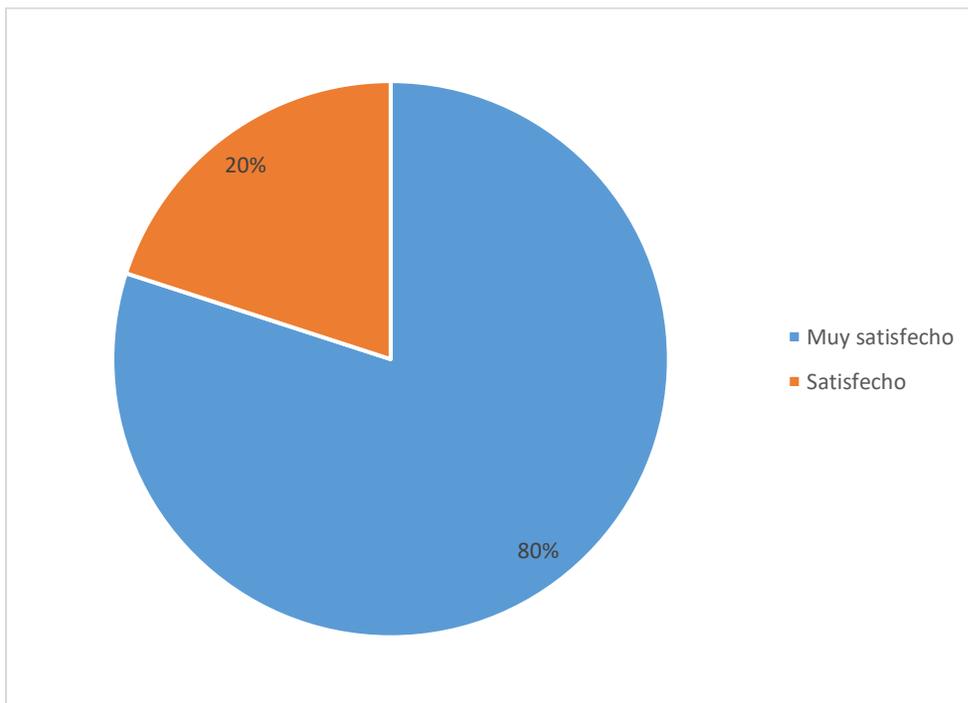
Se puede argumentar que los atributos que fueron incluidos en el DW son válidos, mediante estos resultados, siendo estos de gran importancia, porque mediante ellos se realiza el análisis de información.

### 3.6.7. Validación

**Pregunta 1.-** ¿En la estructura del almacén de datos en su tabla principal contiene métricas las cuales son total de atenciones y cantidad de pacientes, estas medidas son adecuadas?

**Figura 40**

*Porcentaje de satisfacción de las medidas del DW.*

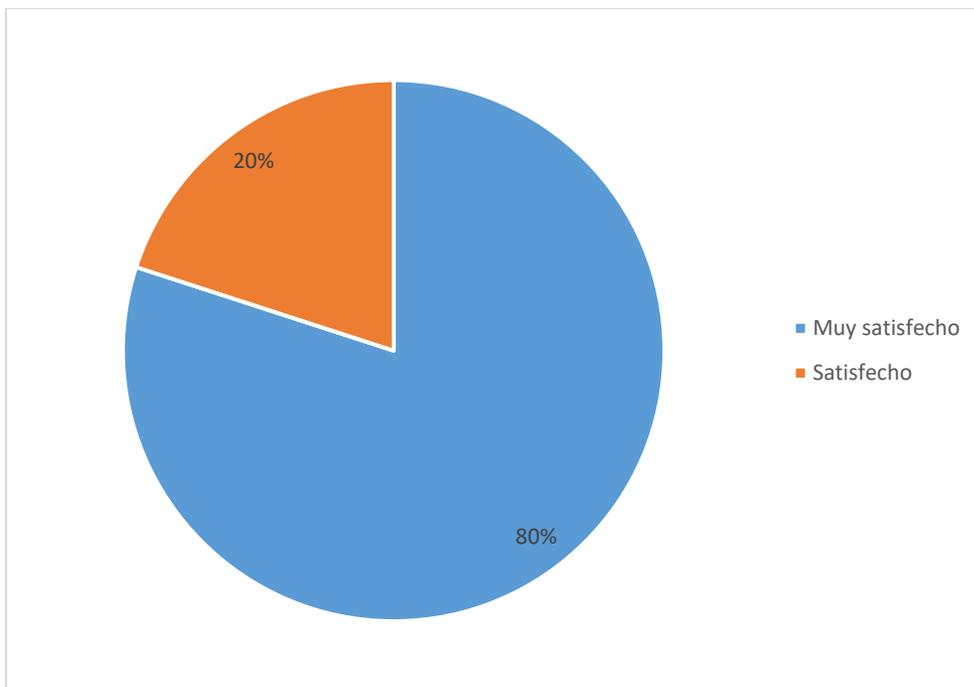


**Fuente:** Elaborado por el autor.

**Pregunta 2.-** ¿Considera que se agregaron todas las variables necesarias para filtrar y agrupar los hechos (Atención médica) y sus medidas por medio de sus atributos en el almacén de datos?

**Figura 41**

*Porcentaje de satisfacción de variables para el análisis de información.*

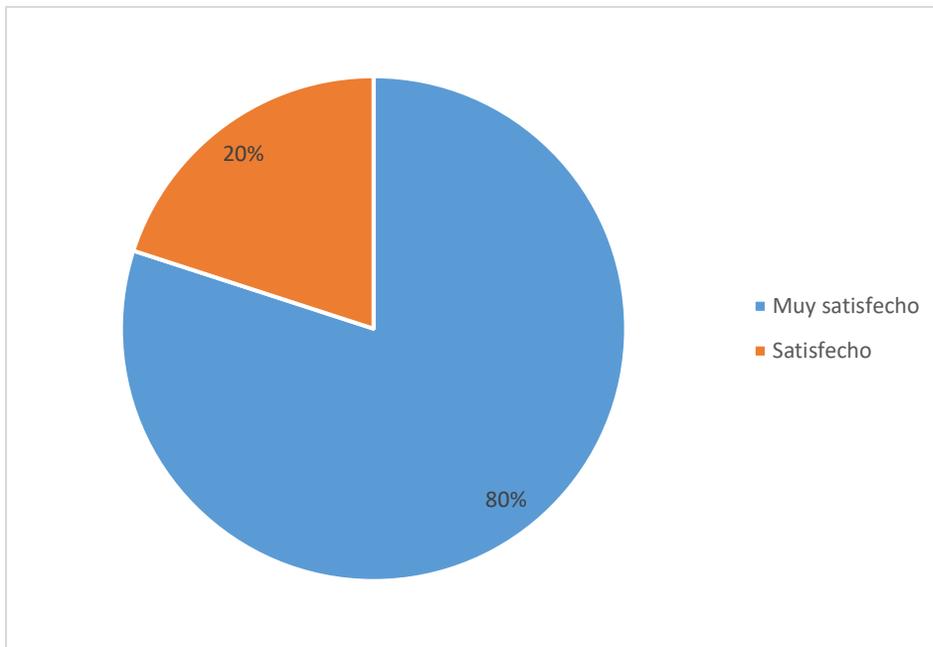


**Fuente:** Elaborado por el autor.

**Pregunta 3.-** ¿Considera que el diseño de la estructura, organización de variables como también de sus medidas del almacén de datos son adecuados?

**Figura 42**

*Porcentaje de satisfacción de la estructura del DW.*

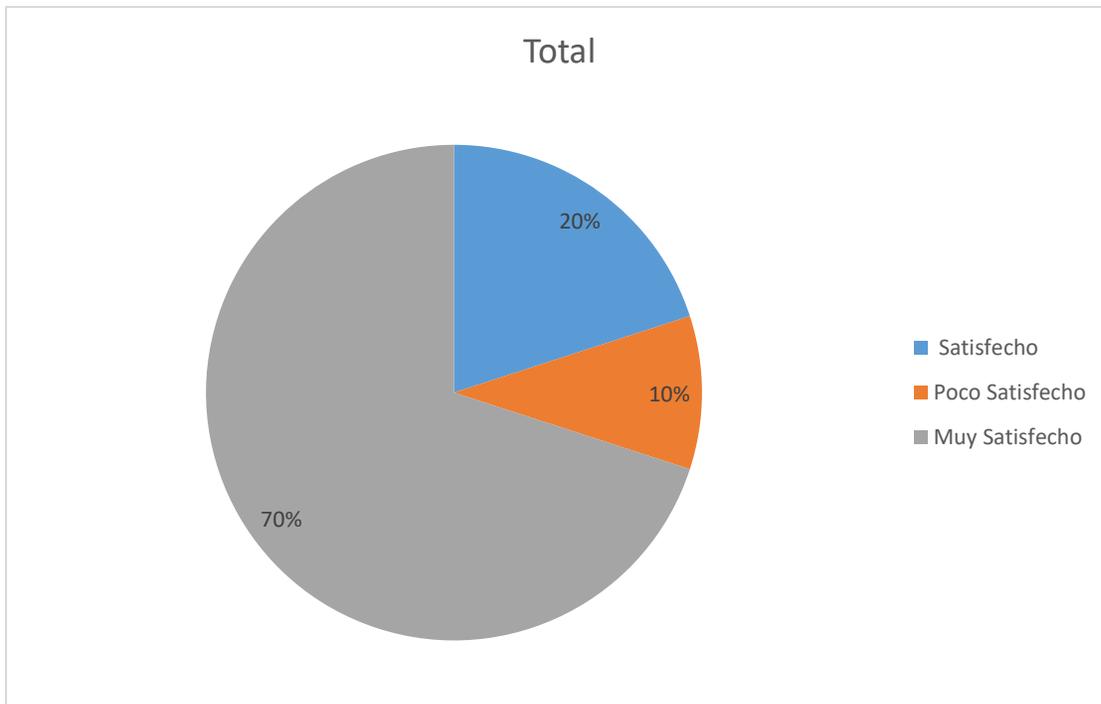


**Fuente:** Elaborado por el autor.

**Pregunta 4.-** ¿Considera que la estructura del almacén facilita el acceso a una gran variedad de datos y permite un análisis de información?

**Figura 43**

*Porcentaje de satisfacción de acceso a datos y análisis de información.*

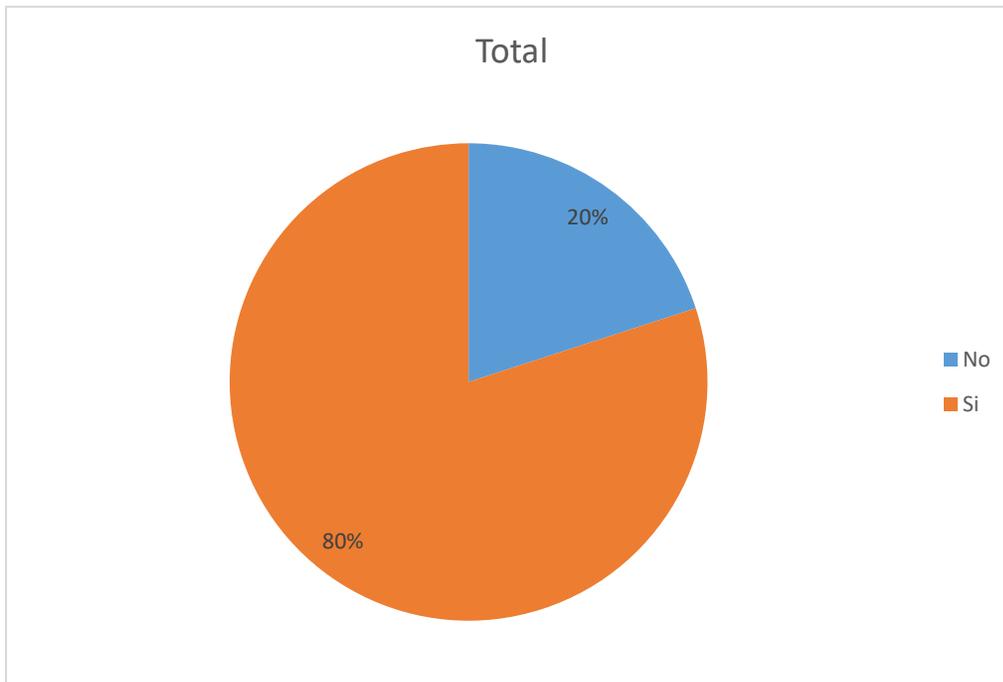


**Fuente:** Elaborado por el autor.

**Pregunta 5.-** ¿Considera que la estructura del almacén minimiza el tiempo de búsqueda de información en relación con los sistemas tradicionales?

**Figura 44**

*Reducción de tiempo en consultas de información.*



**Fuente:** Elaborado por el autor.

Según los resultados se puede argumentar que el DW con su estructura y variables, permite un análisis de información minimizando tiempo de búsqueda de información y aportando la toma de decisiones.

### **3.7. Comprobación de hipótesis**

Se realizó la prueba ANOVA o análisis de varianza la cual es un método estadístico que sirve para descubrir si los resultados de una prueba son significativos, es decir, permiten determinar si es necesario rechazar la hipótesis nula o aceptar la hipótesis alternativa.

**H0**=El diseño de un almacén de datos no permitirá el análisis eficiente de la información en instituciones de salud para la toma de decisiones

**HA**=El diseño de un almacén de datos permitirá el análisis eficiente de la información en instituciones de salud para la toma de decisiones

**Tabla 12**

*Descripción de valores de indicadores*

	DW	PRASS	RDACAA
Tiempo Consulta	4.5	7	6
Tiempo Análisis	4	6	6.2
Calidad información	8.5	7	7

**Fuente:** Elaborado por el autor.

**Tabla 13**

*Análisis de varianza de un factor*

RESUMEN					
<i>Grupos</i>	<i>Cuenta</i>	<i>Suma</i>	<i>Promedio</i>	<i>Varianza</i>	
Tiempo Consulta	3	17,5	5,83333333	1,58333333	
Tiempo Análisis	3	16,2	5,4	1,48	
Calidad información	3	22,5	7,5	0,75	

**Fuente:** Elaborado por el autor.

**Tabla 14***Análisis de varianza*

<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
Entre grupos	7,37555556	2	3,68777778	2,90122378	<b>0,131382498</b>	5,14325285
Dentro de los grupos	7,62666667	6	1,27111111			
Total	15,0022222	8				

**Fuente:** Elaborado por el autor.

Podemos afirmar que se rechaza la hipótesis nula  $H_0$  teniendo en cuenta que la probabilidad es **0,131382498** siendo menor al nivel de significación el cual es **0.05** esto con lleva a comprobar la hipótesis alternativa.

## Capítulo IV

### RESULTADOS

#### 4.1. Resultados

Se diseñó el almacén de datos para el análisis eficiente de información en instituciones de salud obteniendo un DW con información centralizada de dos fuentes de datos que contenían un número de 82 variables las cuales fueron depuradas para el diseño del DW se logró identificar técnicas de analítica de datos como es el análisis descriptivo el cual describe los hechos del problema también denominado inteligencia de negocios. En el cual se logró capturar una gran masa de datos sanitarios que fueron depurados a través de un ETL para una mayor utilidad buscando entender el nivel de correlación de estos datos para responder, explicar la situación de los datos y dar respuestas a preguntas que se conocen como análisis de diagnóstico que vendría a ser la segunda técnica de análisis de datos.

Se aplicó procesos, relaciones, y reglas de normalización para extraer, transformar y cargar datos, los cuales ayudaron a producir datos limpios con una arquitectura basada en metadatos, aumentando la productividad, otorgando agilidad en la entrega de informes siendo accesibles para ser utilizados en analíticas de datos.

Por otra parte, las relaciones entre variables o el diseño del DW se realizó mediante la técnica de estructuración de datos optimizada, más conocida como el modelo dimensional el cual sirve para leer analizar resumir información numérica con valores como cantidad de pacientes y total de atenciones en el Almacén de datos siendo estas medidas aquello que se quiere medir o analizar.

Se logró generar reportes de cantidad de pacientes y total de atenciones segmentados por año, médico, diagnóstico, riesgo obstétrico y morbilidad representando la primera técnica de analítica siendo esta soportada por el DW que es la columna vertebral de este tipo de analítica obteniendo información centralizada para luego poder explotar esta información.

## Figura 45

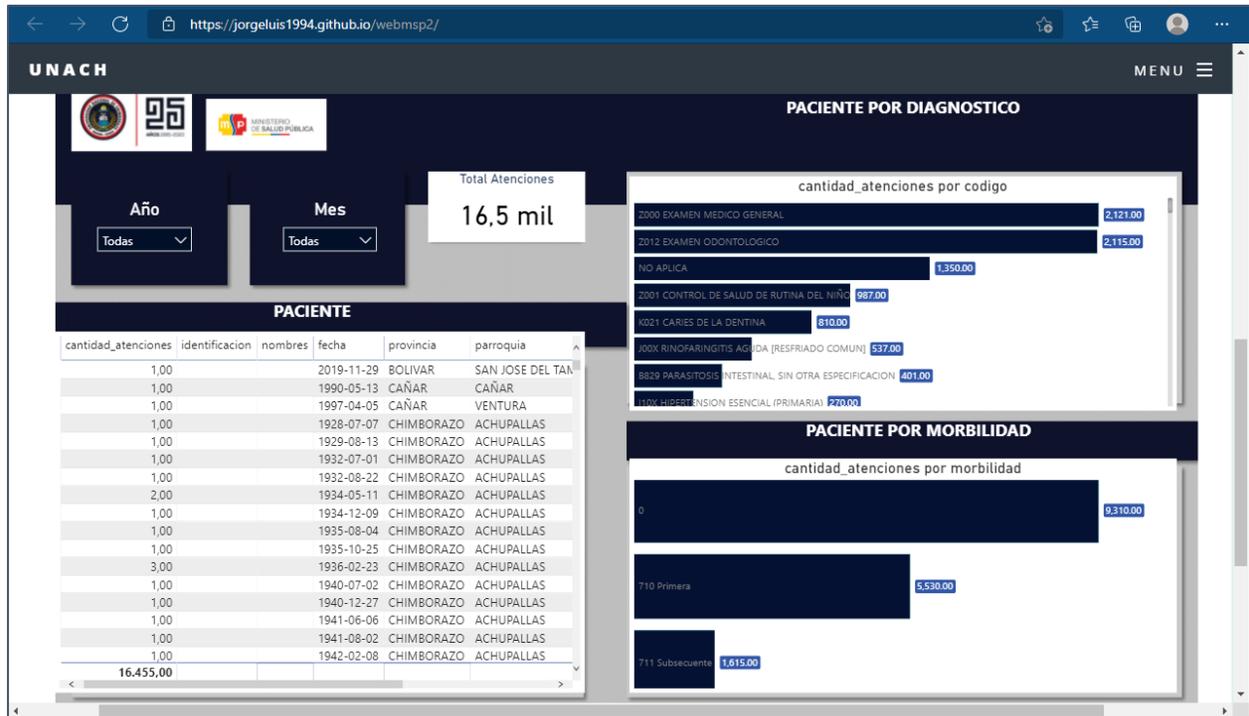
*Gestor de consultas*



**Fuente:** Elaborado por el autor.

**Figura 46**

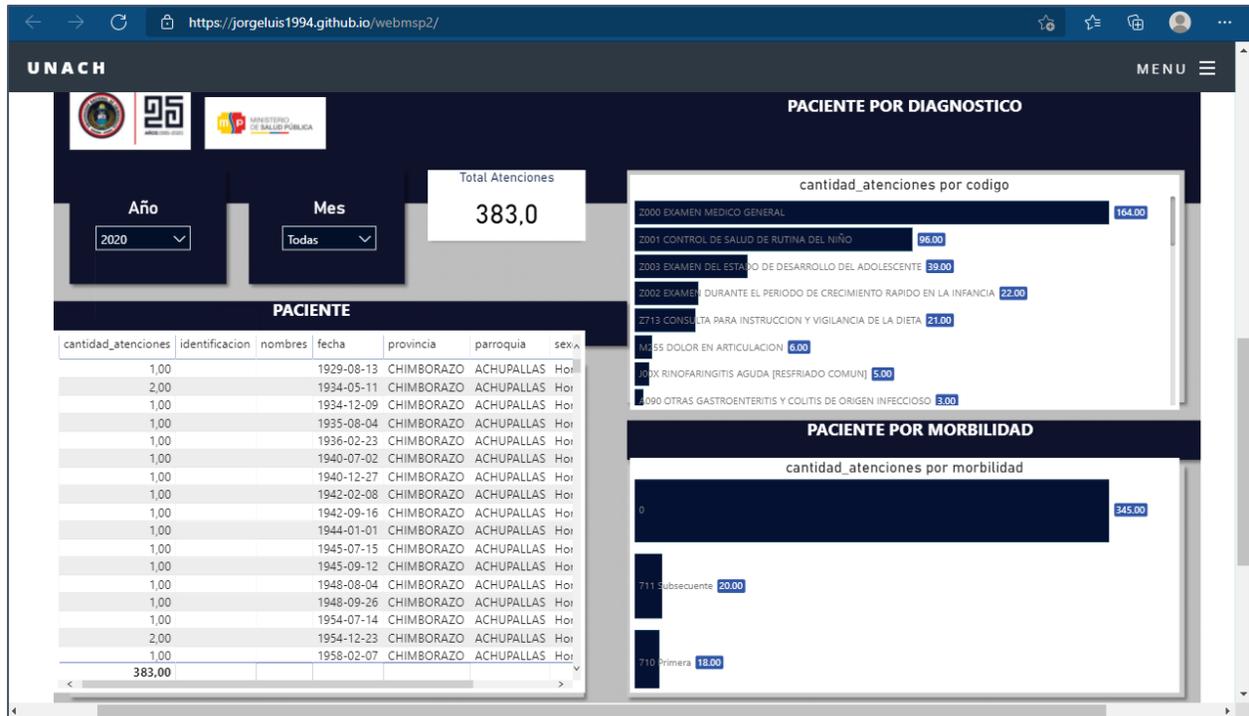
Total, atenciones año 2020 - 2021



Fuente: Elaborado por el autor.

**Figura 47**

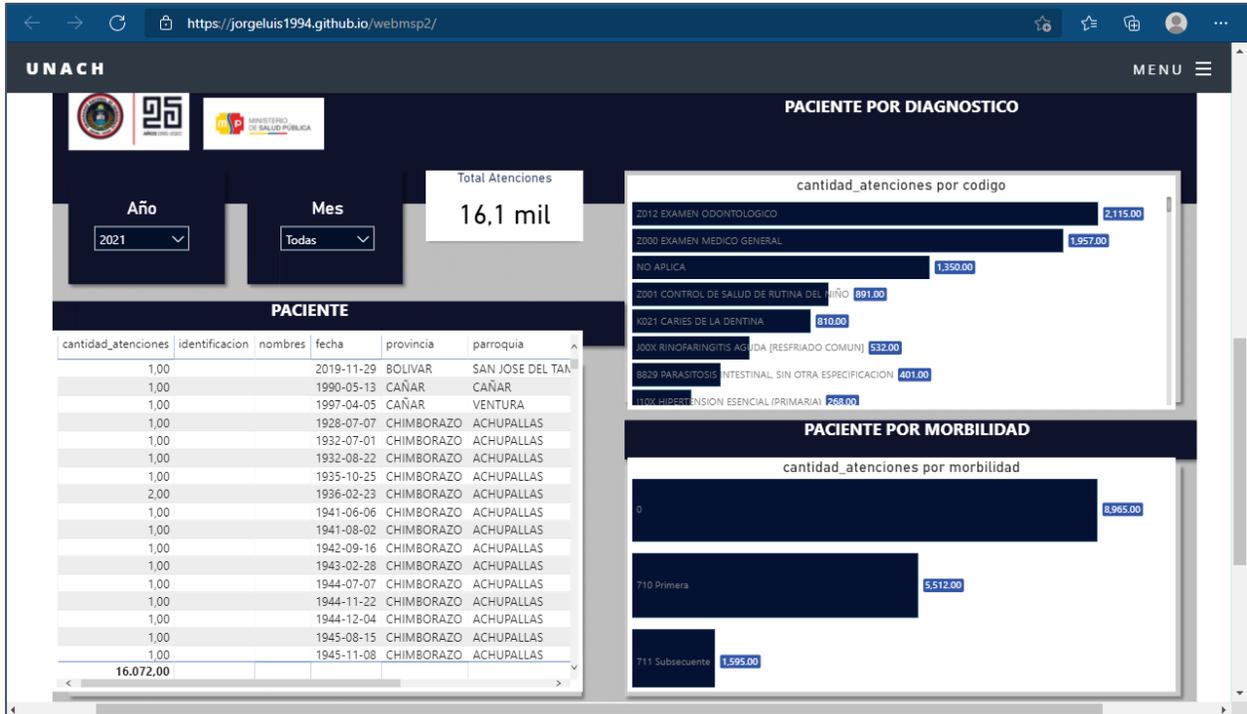
*Total, atenciones año 2020*



**Fuente:** Elaborado por el autor.

**Figura 48**

*Total, atenciones año 2021*



**Fuente:** Elaborado por el autor.

## 4.2. Conclusiones

En esta tesis se diseñó el almacén de datos para el análisis eficiente de la información en instituciones de salud el cual consta de cinco dimensiones, las cuales representan los factores de análisis de la empresa y están desnormalizadas las mismas contienen atributos o campos que nos permiten filtrar, agrupar o seccionar la información, por otra parte su tabla de hechos contiene los campos claves que unen las dimensiones, con sus métricas o medidas para realizar el análisis descriptivo de la información, se identificó técnicas analíticas avanzadas de análisis de datos, siendo este Almacén de Datos la columna vertebral para este tipo de analítica de información, también denominado Inteligencia

de negocios que viene hacer la primera técnica en la análisis avanzado de datos para posteriormente aplicar diferentes estudios predictivos de datos.

Se estableció procesos, relaciones, y reglas de normalización para extraer, transformar y cargar datos, la propuesta se centra en los pasos del proceso ETL que es el encargado de extraer los datos de dos fuentes, transformarlos y luego poblar el DW se aplicó reglas de ETL en su primera fase denominada de Extracción se realizó una verificación de calidad de datos de las dos fuentes que alimentan el DW si los datos estaban estructurados o no estructurados en la segunda fase de Transformación se aplicó reglas de negocio las cuales son, limpieza de columnas ejemplo la tabla paciente y admisión de las fuentes de datos PRAS y RDACAA que contenía 40 columnas se limpió y dejó 15 para posteriormente construir la dimensión paciente, de las tablas profesional y funcionarios de la fuentes de datos PRAS y RDACAA que contenía 16 columnas se limpió y dejó 7 para posteriormente construir la dimensión médico, de las tablas unidad operativa y atención de la fuentes de datos PRAS y RDACAA que contenía 14 columnas se limpió y dejó 9 para posteriormente construir la dimensión establecimiento, de igual forma para la construcción de la dimensión diagnóstico de las tablas tratamiento enfermería y atención de la fuente de datos PRAS que contenía 32 columnas se limpió y dejó 9, dividimos columnas paciente : nombre = nombres apellidos aplicamos INNER para unir múltiples flujos provenientes de las dimensiones seleccionando todas las Primary Keys, luego generamos nuevos valores count ([id paciente]), count ([id tiempo]) siendo estas el objeto de los análisis.

Una vez generado el DW mediante una encuesta realizada a expertos se logró validar la estructura de este, con sus variables y métricas las mismas que permiten el análisis de información, como también se verificó que el tiempo de consultas y acceso a información de pacientes disminuyó, siendo este uno de los aspectos que marcan el grado de satisfacción de los profesionales de salud.

### **4.3. Recomendaciones**

Se recomienda que al inicio de la construcción de un almacén de datos se tenga muy claro los requerimientos (medidas y dimensiones) para posteriormente no tener que realizar cambios.

Definir un modelo dimensional correcto y no hacerlo muy complejo para no tener un bajo rendimiento en las consultas de información para el usuario final.

Respecto a los datos con la herramienta PDI Spoon la configuración que viene por defecto es de 256 Mb de RAM cuando se maneja una gran cantidad esta cantidad no es suficiente se recomienda elevar dependiendo los recursos del hardware ya que puede llegar a presentar un error.

## REFERENCIAS BIBLIOGRAFÍA

- Alazraqui. (2016). Sistemas de Información en Salud: de sistemas cerrados a la ciudadanía social. Un desafío en la reducción de desigualdades en la gestión local. *Scielo*.
- Aular, Y. J. (2017). Minería de Datos como soporte a la toma de decisiones empresariales. *Scientific Electronic Library Online*.
- Dirven , B. B., Pérez, R., Cáceres, R. J., Tito, A. T., Gómez , R. K., & Ticona, A. (2018). *El desarrollo rural establecido en las áreas Vulnerables*. Lima: Colección Racso.
- Euclides. (10 de 09 de 2019). *link.springer.com*. Obtenido de [https://link.springer.com/chapter/10.1007/3-540-46146-9\\_21](https://link.springer.com/chapter/10.1007/3-540-46146-9_21)
- Fontana. (2019). APLICACIÓN DE TÉCNICAS DE BUSINESS INTELLIGENCE EN EL ÁMBITO DE LA SALUD PÚBLICA. *Rumbos tecnológicos*.
- Garay. (13 de 12 de 2016). *SciELO*. Obtenido de <http://www.ijopm.org/index.php/IJOPM/article/view/245>
- García. (29 de 08 de 2015). *Metodologia\_etl*. Obtenido de <https://silo.tips/download/metodologia-etl-para-el-procesamiento-de-datos-en-repositorios-de-proyectos-de-s>
- González. (06 de 09 de 2019). *Sociedad Española*. Obtenido de <https://seis.es/wp-content/uploads/2018/02/LA-EXPLORACION-DE-DATOS-DE-SALUD.pdf>
- González, A. A. (2016). *LA EXPLORACIÓN DE DATOS DE SALUD*. ISBN.
- Lemahieu. (16 de 10 de 2015). Obtenido de SciELO: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2227-18992015000400007](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992015000400007)

- Monserrat, S. (2016). Minería de Datos en Base de Datos de Servicios de Salud. *Minería de Datos en Base de Datos de Servicios de Salud*.
- Mora. (2014). Diseño de Almacenes de Datos con UML. *ResearchGate*.
- Morales. (14 de 03 de 2017). *cienciadigital.org*. Obtenido de <https://www.reci.org.mx/index.php/reci/article/view/40>
- Peñañiel. (10 de 10 de 2019). *Ciencia Digitalk*. Obtenido de <http://cienciadigital.org/revistacienciadigital2/index.php/CienciaDigital/article/view/922>
- Plazzotta, D. L. (2015). SISTEMAS DE INFORMACIÓN EN SALUD: INTEGRANDO. *Scientific Electronic Library Online*.
- Rivadera, G. R. (24 de 06 de 2019). *La metodología de Kimball para el diseño de almacenes de*. Obtenido de <https://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>
- Rosado, G. (2014). INTELIGENCIA DE NEGOCIOS. *Scientia Et Technica*,.
- Russo, C. C. (2016). Aplicación de minería de datos espacial en el área de salud en la zona de influencia de la UNNOBA. *SEDECI*.
- Silva. (05 de 06 de 2016). *Grupo csi*. Obtenido de <https://www.fing.edu.uy/inco/grupos/csi/esp/Publicaciones/2001/jaiio2001-ag.pdf>
- Tovar, G. L. (1986). *El asentamiento y la segregación de los Blancos y Mestizos*. Bogotá: Cengage.
- Vallejo, G. L. (11 de 01 de 2016). “*La analítica de Big Data en el sector sanitario*. Obtenido de <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/45611/7/glinaresvTFG0116memoria.pdf>

Villada. (01 de 06 de 2019). *scipedia*. Obtenido de

[https://www.scipedia.com/public/Botero\\_Villada\\_2019a](https://www.scipedia.com/public/Botero_Villada_2019a)

## ANEXOS

**Figura 49**

*Descripción de variables RDACAA*

TABLA IDENTIFICADA	NOMBRES DE VARIABLES RDACAA USUARIO FINAL	DESCRIPCION DE VARIABLES RDACAA
ARCHIVO	(en blanco)	ESTADO DEL ARCHIVO DE CARGA DE DATOS DE ATENCIONES MEDICAS VERSION DE SOFTWARE DEL ARCHIVO GENERADO (en blanco)
ATENCION	cie_1 cie_2 cie_3 dcie_1 dcie_2 dcie_3 fecha_aten lugar_aten nivel_esta td1 td2 td3 (en blanco)	BLOQUE D: CODIFICACION DEL PRIMER DIAGNOSTICO CIE10 BLOQUE D: CODIFICACION DEL SEGUNDO DIAGNOSTICO CIE10 BLOQUE D: CODIFICACION DEL TERCER DIAGNOSTICO CIE10 BLOQUE D: DESCRIPCION DE LA CODIFICACION DEL PRIMER DIAGNOSTICO CIE10 BLOQUE D: DESCRIPCION DE LA CODIFICACION DEL SEGUNDO DIAGNOSTICO CIE10 BLOQUE D: DESCRIPCION DE LA CODIFICACION DEL TERCER DIAGNOSTICO CIE10 FECHA DE ATENCION DEL PACIENTE LUGAR DONDE EL PACIENTE SE REALIZA LA ATENCION (ESTABLECIMIENTO-CENTRO EDUCATIVO-ETC) BLOQUE D: NIVEL DE ATENCION AL CUAL CORRESPONDE EL ESTABLECIMIENTO BLOQUE D: CONDICION PRIMER DIAGNOSTICO (PRESUNTIVO-DEFINITIVO INICIAL-DEFINITIVO INICIAL CONFIRMADO POR LABORATORIO-DEFINITIVO CONTROL) BLOQUE D: CONDICION SEGUNDO DIAGNOSTICO (PRESUNTIVO-DEFINITIVO INICIAL-DEFINITIVO INICIAL CONFIRMADO POR LABORATORIO-DEFINITIVO CONTROL) BLOQUE D: CONDICION TERCER DIAGNOSTICO (PRESUNTIVO-DEFINITIVO INICIAL-DEFINITIVO INICIAL CONFIRMADO POR LABORATORIO-DEFINITIVO CONTROL) BLOQUE A: FECHA DE ATENCION PACIENTE BLOQUE A: IDENTIFICADOR DE LA ESTRATEGIA BLOQUE A: LUGAR DONDE SE REALIZA LA ATENCION MEDICA BLOQUE A: TIPO DE ATENCION INTRA O EXTRA MURA BLOQUE C: CLASIFICACION DE DATOS ANTROPOMETRICOS DEL INDICE DE MASA CORPORAL DEL PACIENTE BLOQUE C: DATOS ANTROPOMETRICOS DE HEMOGLOBINA CORREGIDO DEL PACIENTE BLOQUE C: DATOS ANTROPOMETRICOS DE HEMOGLOBINA DEL PACIENTE BLOQUE C: DATOS ANTROPOMETRICOS DE INDICADOR DE ANEMIA DEL PACIENTE BLOQUE C: DATOS ANTROPOMETRICOS DE LA TALLA CORREGIDA DEL PACIENTE

td1	INICIAL-DEFINITIVO INICIAL CONFIRMADO POR LABORATORIO-DEFINITIVO CONTROL)
td2	BLOQUE D: CONDICION SEGUNDO DIAGNOSITCO (PRESUNTIVO-DEFINITIVO INICIAL-DEFINITIVO INICIAL CONFIRMADO POR LABORATORIO-DEFINITIVO CONTROL)
td3	BLOQUE D: CONDICION TERCER DIAGNOSITCO (PRESUNTIVO-DEFINITIVO INICIAL-DEFINITIVO INICIAL CONFIRMADO POR LABORATORIO-DEFINITIVO CONTROL)
(en blanco)	<p>BLOQUE A: FECHA DE ATENCION PACIENTE</p> <p>BLOQUE A: IDENTIFICADOR DE LA ESTRATEGIA</p> <p>BLOQUE A: LUGAR DONDE SE REALIZA LA ATENCION MEDICA</p> <p>BLOQUE A: TIPO DE ATENCION INTRA O EXTRA MURA</p> <p>BLOQUE C: CLASIFICACION DE DATOS ANTROPOMETRICOS DEL INDICE DE MASA CORPORAL DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DE HEMOGLOBINA CORREGIDO DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DE HEMOGLOBINA DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DE INDICADOR DE ANEMIA DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DE LA TALLA CORREGIDA DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DE LA TALLA DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DEL INDICE DE MASA CORPORAL DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DEL PERIMETRO CEFALICO</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DEL PESO DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS DEL TIPO DE TOMA DE MEDIDA DE LA TALLA DEL PACIENTE</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS PUNTAJE Z IMC PARA EDAD</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS PUNTAJE Z PERIMETRO CEFALICO PARA EDAD</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS PUNTAJE Z PESO PARA TALLA</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS PUNTAJE Z PESO PARA TALLA</p> <p>BLOQUE C: DATOS ANTROPOMETRICOS PUNTAJE Z TALLA PARA EDAD</p> <p>BLOQUE C: GRUPO PRIORITARIO DE ATENCION</p> <p>BLOQUE C: LESIONES DE VIOLENCIA</p> <p>BLOQUE D: CATEGORIZACION RIESGO OBSTETRICO</p> <p>BLOQUE D: PLAN DE PARTO</p> <p>BLOQUE D: PLAN DE TRANSPORTE</p> <p>BLOQUE D: CATALOGO DE PRIMERA PRUEBA DE CONFIRMACION</p> <p>BLOQUE D: CATALOGO DE SEGUNDA PRUEBA DE CONFIRMACION</p> <p>BLOQUE D: DATOS DE VIH MOTIVOS DE PRUEBA</p> <p>BLOQUE D: EMBARAZO PLANIFICADO</p> <p>BLOQUE D: FECHA ULTIMA MENSTRUACION</p> <p>BLOQUE D: METODO PARA DETERMINAR SEMANAS DE GESTACION</p> <p>BLOQUE D: PRIMERA PRUEBA DE VIH</p> <p>BLOQUE D: PRUEBA NO TRPONEMICA</p>

		BLOQUE D: SEGUNDA PRUEBA DE VIH
		BLOQUE D: SEMANAS DE GESTACION
		BLOQUE D: SIFILIS TRATAMIENTO
		BLOQUE D: SIFILIS TRATAMIENTO PAREJA
		BLOQUE D: TIPO DE ATENCION MORBILIDAD
		BLOQUE D: TIPO DE ATENCION PREVENTIVA
		BLOQUE D: VIAS DE TRANSMISION DE VIH
		BLOQUE E: CONDICION DE DIAGNOSTICO
		BLOQUE E: REGISTRO DE DIAGNOSTICO DE LA CODIFICACION CIE 10
		BLOQUE E: TIPO DE ATENCION DEL DIAGNOSTICO
		BLOQUE E_ NUMERO CERTIFICADO MEDICO UNICO
		BLOQUE F: 24 HORAS RECIBIO ALIMENTOS SOLIDOS
		BLOQUE F: DIENTES EXTRAIDOS
		BLOQUE F: DIENTES PERDIDOS
		BLOQUE F: DOSIS DE LA VACUNA
		BLOQUE F: HIERRO JARBE
		BLOQUE F: HIERRO MULTIVITAMINAS Y MINERALES
		BLOQUE F: HIERRO Y ACIDO FOLICO
		BLOQUE F: IDENTIFICA AL PRESUNTO AGRESOR/A
		BLOQUE F: INTERCONSULTA SOLICITADA/RECIBIDA
		BLOQUE F: LUGAR REFERIDO
		BLOQUE F: MADRE EN PERIODO DE LACTANCIA
		BLOQUE F: NOMBRE DEL ESQUEMA DE VACUNAS
		BLOQUE F: NUMERO DE SERIE
		BLOQUE F: SUBSISTEMA DE REFERENCIA/CONTRAREFERENCIA
		BLOQUE F: VACUNA
		BLOQUE F: VITAMINA A
		BLOQUE F:24 HORA LECHE MATERNA
		BLOQUE F:DIENTES OBTURADOS
		BLOQUE F_DIENTES CARIADOS
		BLOQUE G: CANTIDAD DE PROCEDIMIENTO
		BLOQUE G: GRUPO VULNERABLE
		BLOQUE G: PROCEDIMIENTO
		NUEVO
		(en blanco)
ESTABLECIMIENTO	can_esta	BLOQUE A: CANTON DEL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	circuito_esta	BLOQUE A: CIRCUITO AL CUAL PERTENECE EL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	cod_esta	CODIGO DE ESTABLECIMIENTO
	distrito_esta	BLOQUE A: DISTRITO AL CUAL PERTENECE EL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	nom_esta	NOMBRE DE ESTABLECIMIENTO QUE SE REGISTRA LA ATENCION
		BLOQUE A: PARROQUIA DEL ESTABLECIMIENTO QUE SE BRINDA LA

		BLOQUE G: GRUPO VULNERABLE BLOQUE G: PROCEDIMIENTO NUEVO (en blanco)
ESTABLECIMIENTO	can_esta	BLOQUE A: CANTON DEL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	circuito_esta	BLOQUE A: CIRCUITO AL CUAL PERTENECE EL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	cod_esta	CODIGO DE ESTABLECIMIENTO
	distrito_esta	BLOQUE A: DISTRITO AL CUAL PERTENECE EL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	nom_esta	NOMBRE DE ESTABLECIMIENTO QUE SE REGISTRA LA ATENCION
	parr_esta	BLOQUE A: PARROQUIA DEL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	prov_esta	BLOQUE A: PROVINCIA DEL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	simbolo_tipo_esta	SIMBOLOGÍA DEL TIPO DE ESTABLECIMIENTO EN LA QUE SE REALIZA LA ATENCION
	tipo_esta	DESCRIPCION DEL TIPO DE ESTABLECIMIENTO EN EL QUE SE REALIZA LA ATENCION
	zona_esta	BLOQUE A: ZONA AL CUAL PERTENECE EL ESTABLECIMIENTO QUE SE BRINDA LA ATENCION
	(en blanco)	BLOQUE A: CODIGO DEL ESTABLECIMIENTO DE SALUD PARROQUIA DEL LUGAR DE ATENCION DEL PROFESSIONAL DE LA SALUD (en blanco)
PACIENTE	anio_pac	BLOQUE C: AÑO DE NACIMIENTO DEL PACIENTE
	area_res	BLOQUE D: AREA DE RESIDENCIA DEL PACIENTE
	cant_res	BLOQUE C: CANTON DE RESIDENCIA HABITUAL DEL PACIENTE
	dia_pac	BLOQUE C: DIA DE NACIMIENTO DEL PACIENTE
	etnia_pac	BLOQUE C: AUTO IDENTIFICACION ÉTNICA DEL PACIENTE
	fecha_nac_pac	BLOQUE C: FECHA DE NACIMIENTO DEL PACIENTE
	mes_pac	BLOQUE C: MES DE NACIMIENTO DEL PACIENTE
	nacio_pac	BLOQUE C: NACIONALIDADES INDIGENAS A LAS CUALES CORRESPONDE EL PACIENTE
	nacionalidad_pac	BLOQUE C: NACIONALIDAD DEL PACIENTE
	parr_res	BLOQUE C: PARROQUIA DE RESIDENCIA HABITUAL DEL PACIENTE
	prov_res	BLOQUE C: PROVINCIA DE RESIDENCIA HABITUAL DEL PACIENTE
	pueblo_pac	BLOQUE C: PUEBLO INDÍGENA AL CUAL CORRESPONDE EL PACIENTE
recinto_res	BLOQUE C: BARRIO-SECTOR-RECINTO DE RESIDENCIA HABITUAL DEL PACIENTE	
sexo_pac	BLOQUE C: SEXO DEL PACIENTE	

	ssalud (en blanco)	<p>BLOQUE C: SEGURO SOCIAL QUE POSEA EL PACIENTE</p> <p>BLOQUE A: IDENTIFICACION DE ATENCION DEL PACIENTE</p> <p>BLOQUE C: AUTO IDENTIFICACION ÉTNICA DEL PACIENTE</p> <p>BLOQUE C: DIRECCION DEL DOMICILIO DEL PACIENTE</p> <p>BLOQUE C: FECHA DE NACIMIENTO DEL PACIENTE</p> <p>BLOQUE C: IDENTIDAD DE GÉNERO DEL PACIENTE</p> <p>BLOQUE C: IDENTIFICACION DEL REPRESENTANTE</p> <p>BLOQUE C: LUGAR DE RESIDENCIA DEL PACIENTE</p> <p>BLOQUE C: NACIONALIDAD DEL PACIENTE</p> <p>BLOQUE C: NACIONALIDADES A LAS CUALES CORRESPONDE EL PACIENTE</p> <p>BLOQUE C: NUMERO DE ARCHIVO DEL ESTABLECIMIENTO DE SALUD</p> <p>BLOQUE C: NUMERO DE IDENTIFICACION DEL PACIENTE A REALIZAR UNA ATENCION MEDICA</p> <p>BLOQUE C: ORIENTACIÓN SEXUAL DEL PACIENTE</p> <p>BLOQUE C: PARENTESCO CON EL AGRESOR</p> <p>BLOQUE C: PRIMER APELLIDO DEL PACIENTE</p> <p>BLOQUE C: PRIMER NOMBRE DEL PACIENTE</p> <p>BLOQUE C: PUEBLO INDÍGENA AL CUAL CORRESPONDE EL PACIENTE</p> <p>BLOQUE C: SEGUNDO APELLIDO DEL PACIENTE</p> <p>BLOQUE C: SEGUNDO NOMBRE DEL PACIENTE</p> <p>BLOQUE C: SEGURO SOCIAL QUE POSEA DEL PACIENTE</p> <p>BLOQUE C: SEXO DEL PACIENTE</p> <p>BLOQUE C: TELEFONO DEL FAMILIAR</p> <p>BLOQUE C: TELEFONO DEL PACIENTE</p> <p>BLOQUE C: TIPO DE BONO QUE POSEA DEL PACIENTE</p> <p>BLOQUE C: TIPO DE IDENTIFICACION DEL PACIENTE</p> <p>BLOQUE C: TIPO DE IDENTIFICAION DEL PACIENTE NUEVO</p> <p>(en blanco)</p>
PROFESIONAL DE SALUD	(en blanco)	(en blanco)
PROFESIONAL SALUD	etnia_prof fecha_nac_prof forma_prof pais_prof sexo_prof (en blanco)	<p>BLOQUE B: AUTO IDENTIFICACION DEL PROFESIONAL</p> <p>BLOQUE B: FECHA DE NACIMIENTO DEL PROFESIONAL</p> <p>FORMACION PROFESIONAL DEL QUE REALIZA LA ATENCION</p> <p>BLOQUE B: PAIS DE NACIMIENTO DEL PROFESIONAL</p> <p>BLOQUE B: SEXO DEL PROFESIONAL QUE REALIZA LA ATENCION</p> <p>BLOQUE A: IDENTIFICACION DE ESPECIALIDAD DE SALUD DEL PROFESIONAL DE LA SALUD</p> <p>BLOQUE A: IDENTIFICACION DE PROFESIONAL DE SALUD</p>
Total general		

## Figura 50

### Descripción de variables PRASS

ITEM	EJEMPLO	NOMBRES DE NOMBRE DE	DESCRIPCION TABLA IDENTIFICADA
Nro.	75538		
ENT_ID	397	ENTIDAD	Identificación de la entidad
ENT_RUC	6.61E+16	ENTIDAD	RUC de la entidad
ENT_NOM	CHAMBO	ENTIDAD	Nombre de la Entidad
ENT_SIM_TIP_CS-B		ENTIDAD	Tipología de la Entidad
ENT_DES_TIP_CENTRO DE SALUD TIPO B		ENTIDAD	Descripción del tipo de la Entidad
ENT_COD_PRI	6	ENTIDAD	Código de la provincia de la Entidad
ENT_DES_PRC CHIMBORAZO		ENTIDAD	Descripción de la provincia de la Entidad
ENT_COD_CA	604	ENTIDAD	Código del cantón de la Entidad
ENT_DES_CA CHAMBO		ENTIDAD	Descripción del cantón de la Entidad
ENT_COD_PA	60450	ENTIDAD	Código de la parroquia de la Entidad
ENT_DES_PAF CHAMBO		ENTIDAD	Descripción de la parroquia de la Entidad
ENT_DES_TIP Urbano		ENTIDAD	Descripción del tipo de parroquia de la Entidad
ENT_INT	MSP	ENTIDAD	Institución a la que pertenece de la Entidad
ENT_NIV	NIVEL 1	ENTIDAD	Nivel de atención de la Entidad
ENT_COD_ZO	Z03	ENTIDAD	Código de la zona de la Entidad
ENT_DES_ZO	ZONA 3	ENTIDAD	Descripción de la zona de la Entidad
ENT_COD_DIS	06D01	ENTIDAD	Código del distrito de la Entidad
ENT_DIST_DIS	CHAMBO, RIOBAMBA	ENTIDAD	Descripción del distrito de la Entidad
ENT_COD_CIR	06D01C20	ENTIDAD	Código del circuito de la Entidad
PROF_NOMBI ERAZO ESTRELLA XAVIER DA		PROFESIONAL	Nombres completos del Profesional
PROF_NOMBI XAVIER DANILO		PROFESIONAL	Nombres del Profesional
PROF_APELLI ERAZO ESTRELLA		PROFESIONAL	Apellidos del Profesional
PROF_FEC_N/	30298	PROFESIONAL	Fecha de nacimiento del Profesional
PROF_SEXO	Hombre	PROFESIONAL	Sexo del Profesional
PROF_TIP_IDE C?ula de Identidad		PROFESIONAL	Tipo de identificación del Profesional
PROF_IDEN	602468860	PROFESIONAL	Identificación del Profesional
PROF_ESP_AT	Medicina Familiar Y Comunitaria	PROFESIONAL	Especialidad en atención del Profesional

PROF_REG_NI	602468860	PROFESIONAL	Numero de registro del Profesional
PROF_CORREO	xerazoestrella@gmail.com	PROFESIONAL	Correo del Profesional
PCTE_NOM	REINO QUISHPI AYISETH NO	PACIENTE	Nombres completos del Paciente
PCTE_NOMB	AYISETH NOEMI	PACIENTE	Nombres del Paciente
PCTE_APELLI	REINO QUISHPI	PACIENTE	Apellidos del Paciente
PCTE_TIP_IDE	C?ula de Identidad	PACIENTE	Tipo de identificación del Paciente
PCTE_IDE	651019085	PACIENTE	Identificación del Paciente
PCTE_SEXO	Mujer	PACIENTE	Sexo del Paciente
PCTE_ORI_SE	NA	PACIENTE	Orientació sexual del Paciente
PCTE_IDE_GE	NA	PACIENTE	Identificación del género del Paciente
PCTE_FEC_NA	44061	PACIENTE	Fecha de nacimiento del Paciente
PCTE_ANIOS	0	PACIENTE	Años del Paciente
PCTE_MESES	5	PACIENTE	Meses del Paciente
PCTE_DIAS	4	PACIENTE	Días del Paciente
PCTE_EDAD_(	0a 5m 4d	PACIENTE	Edad compuesta del Paciente
PCTE_ANIOS_	5	PACIENTE	Edad en años y meses del Paciente
PCTE_IDE_RE	NA	PACIENTE	Identificación del representante del Pacien
PCTE_NACION	ECUATORIANO/A	PACIENTE	Nacionalidad del Paciente
PCTE_AUTID_	NA	PACIENTE	Autoidentificación étnica del Paciente
PCTE_NAC_ET	NA	PACIENTE	Nacionalidad del Paciente
PCTE_PUEBL	CNA	PACIENTE	Pueblo del Paciente
PCTE_TEL_CE	099-435-9502	PACIENTE	Número de celular del Paciente