



UNIVERSIDAD TECNOLÓGICA ISRAEL

**MAESTRÍA EN TELEMÁTICA, MENCIÓN CALIDAD EN
EL SERVICIO**

TRABAJO FINAL DE GRADUACIÓN

**TÍTULO: APLICACIÓN DE LA MINERÍA DE DATOS
PARA ANÁLISIS DE TENDENCIAS POLÍTICAS EN REDES
SOCIALES**

AUTOR: ESMERALDA PROAÑO A.

TUTOR: Ing. Fidel Parra M. Ph. D.

QUITO, 2019

DEDICATORIA

El presente trabajo lo dedico principalmente a Dios, por ser mi inspirador y darme la fuerza todos los días para continuar en este proceso de obtener uno de mis sueños más deseados.

A mi padre, por su amor, paciencia, trabajo y sacrificio en todo este tiempo, gracias a su apoyo he logrado llegar hasta aquí y convertirme en lo que soy. Tengo el orgullo y el privilegio de ser su hija.

A mis hijas por estar siempre presentes, acompañándome y por el apoyo moral, que me brindaron a lo largo de esta etapa de mi vida.

A todas las personas por su apoyo, eso hizo posible que el trabajo se realice con éxito en especial a aquellos que me abrieron las puertas y compartieron sus conocimientos.

AGRADECIMIENTO

Mi profundo agradecimiento a las autoridades y personal que hacen la Universidad Tecnológica Israel, por confiar en mí, abrirme sus puertas, que con sus enseñanzas y sus valiosos conocimientos hicieron que pueda crecer cada día como profesional, gracias a todos mis maestros por su paciencia, dedicación, apoyo incondicional y amistad.

Pensamiento

Hoy he hablado de tecnología, no solo se debe tener en cuenta el (hardware) el instrumento, el aparato...) y el software (las reglas que permiten usar el hardware), si no también lo que se llama el branwaire (knoware), el (porque, como, cuando, donde... usar el hardware y software.

(Agustí Chalaux)

ÍNDICE DE CONTENIDOS

CERTIFICADO DE RESPONSABILIDAD.....	ii
CERTIFICADO DE AUDITORIA.....	iii
..... DEDICATORIA	
.....	iv
AGRADECIMIENTOS	vi
PENSAMIENTOS.....	vi
ÍNDICE DE CONTENIDOS.....	vii-x
ÍNDICE DE FIGURAS	xi-xiii
RESUMEN.....	xiv
ABSTRACT	xv
INTRODUCCIÓN.....	1
Situación Problemática	1
Formulación del Proyecto	2
Objetivos	2
Objetivo General	2
Objetivos Específicos	2
Hipótesis o idea a defender	2
Justificación	3

Capítulo I. Fundamentación teórica	4
1.1 Contextualización espacio temporal del problema	4
1.2 Cuerpo teórico conceptual.....	4
1.2.1 Aproximación al concepto de minería de datos.....	4
1.2.1.1 EIS: Executive Information System	5
1.2.1.2 OLAP: On-Line Analysis Planning.....	5
1.2.1.3 Consolidación	7
1.2.1.4 Dril – Down	7
1.2.1.5 Slicing and dicing.....	7
1.2.1.6 Herramientas de minería de datos	8
1.2.2 Almacén de datos (Data Warehouse).....	8
1.2.2.1 Diseño de una Data Warehouse	9
1.2.3 Proceso de extracción del conocimiento a partir de datos (KDD).....	10
1.2.3.1 Etapas del proceso KDD.....	10
1.2.4 Las redes sociales.....	11
1.2.4.1 Proceso de conexión.....	13
1.2.4.2 Las redes como mapas de información	14
1.2.4.3 El fenómeno de las redes sociales	15
1.2.4.4 El rol de las redes sociales en la proyección corporativa	16
1.3 Revisión de las investigaciones previas del objeto de estudio	17
CAPÍTULO II. Marco Metodológico	19
2.1 Enfoque metodológico de la investigación.....	19
2.2 Modalidad de la investigación	19
2.3 Población y muestra	20
2.4 Métodos empíricos y técnicas empleadas para la recolección de información	20
2.5 Formas de procesamiento de la información	21

2.5.1	Data Warehouse	21
2.5.2	Análisis de requerimientos	22
2.5.2.1	Preguntas del Negocio	22
2.5.2.2	Indicadores y perspectivas	22
2.5.2.3	Modelo Conceptual.....	22
2.5.3	Análisis de Data Sources	23
2.5.3.1	Mapeo	23
2.5.3.2	Granularidad.....	23
2.5.4	Modelo Lógico del Data Warehouse.....	23
2.5.4.1	Tipología	23
2.5.4.2	Tablas de Dimensiones	23
2.5.4.3	Tablas de Hechos.....	23
2.5.4.4	Uniones	24
2.5.5	Integración de datos	24
2.6	Metodología seleccionada	24
Capítulo III. Propuesta		25
3.1	Fundamentos de la propuesta.....	25
3.2	Misión.....	25
3.3	Visión	25
3.4	Justificación	25
3.5	Objetivos de la propuesta	26
3.5.1	Objetivo general.....	26
3.5.2	Objetivos Específicos.....	26
3.6	Presentación del proyecto	26
3.6.1	Presentación de la Propuesta.....	26
3.6.2	Desarrollo del Data Mining de tendencias políticas en redes sociales.....	27

3.6.2.1	Modelado del Data Warehouse	27
3.6.2.2	Análisis de requerimientos	27
3.6.2.3	Análisis de data sources	30
3.6.2.4	Modelo lógico del Data Warehouse	35
3.6.2.5	Integración de datos	38
3.6.2.6	Creación del proceso Data mining – ETL (Extraer, transformar y cargar).....	43
3.6.2.7	Creación de Cubos de información.....	56
3.6.2.8	Validación del procesamiento de la información	66
CONCLUSIONES.....		75
RECOMENDACIONES		74
Bibliografía.....		76
Artículo Científico.....		81

ÍNDICE DE FIGURAS

Figura 1. Vista de los datos en forma de cubo.....	6
Figura 2. Operación Slice and Dice	7
Figura 3. Herramientas de Minería de Datos usadas habitualmente	8
Figura 4. Etapas del proceso KDD	11
Figura 5. Línea del tiempo de las redes sociales.....	12
Figura 6. Libros comprados por las mismas personas en la campaña presidencial de 2008.	14
Figura 7. Modelo conceptual.....	29
Figura 8. Aplicación del modelo conceptual	30
Figura 9. Modelo conceptual Ampliado.....	34
Figura 10. Diseño de diagrama conceptual	35
Figura 11. Tablas de Dimensiones.....	37
Figura 12. Tabla de hechos factor Tendencia Política	37
Figura 13. Unión Tabla de dimensión con tabla de hechos.....	38
Figura 14. Diagrama físico del modelado.	39
Figura 15. Exportación del Diagrama Físico.....	40
Figura 16. Código script a generar.....	40
Figura 17. Tabla script generada.....	41
Figura 18. Creación de la base de datos	41
Figura 19. Base de datos creada	42
Figura 20. Carga Satisfactoria	42
Figura 21. Carga inicial al Data Warehouse.....	43
Figura 22. Gráfica del proceso ETL.....	44
Figura 23. Estructuración de la información a extraer	45
Figura 24. Transformaciones y trabajos del sistema de ETL	46
Figura 25. Apertura de Microsoft Excel Input.....	46

Figura 26. Carga de libros de información	47
Figura 27. Selección de la información.....	47
Figura 28. Selección - Dim Persona.....	48
Figura 29. Extracción de la información Dim persona	48
Figura 30. Información Dim persona	49
Figura 31. Identificación de datos.....	49
Figura 32. Table output	50
Figura 33. Configuración de parámetros	50
Figura 34. Parámetros requeridos para establecer una conexión	51
Figura 35. Mensaje de configuración correcta.....	51
Figura 36. Direccionamiento de la tabla para carga de información.....	52
Figura 37. Database fields	52
Figura 38. Carga dimensión dim_persona a la base de datos.	53
Figura 39. Apertura de la base de datos	53
Figura 40. Visualización de la base de datos	54
Figura 41. Carga de dimensiones.....	54
Figura 42. DWH_Tendencia	55
Figura 43. DWH_Tendencia 1-5	55
Figura 44. DWH_Tendencia 1-6	56
Figura 45. Interfaz gráfica de SpagoBiMeta	57
Figura 46. Conexión con SpagoBI Server.....	57
Figura 47. Agregar los parámetros necesarios para establecer una nueva conexión.....	58
Figura 48. Visualización de la conexión al servidor creada.	58
Figura 49. Conexión con la base de datos	59
Figura 50. Configuración de parámetros	59
Figura 51. Descarga los drivers de la página oficial de PostgreSQL	60
Figura 52. Configuración de parámetros necesarios para establecer conexión con la BD	60
Figura 53. Creación de un nuevo modelo.....	61

Figura 54. Extracción de las dimensiones y la tabla.....	61
Figura 55. Definir a las tablas como dimensiones y a la tabla de hechos como Cubo.....	62
Figura 56. Editar las jerarquías de cada una de las dimensiones	62
Figura 57. Jerarquía establecida para Dim_actividad_red.....	63
Figura 58. Generación de la relación	63
Figura 59. Dim partido político	64
Figura 60. Relación Fact tendencia política – Id persona.....	64
Figura 61. Selección de la ubicación del OLAP	65
Figura 62. Corrección del archivo y errores	65
Figura 63. Esquema Olap final	66
Figura 64. Interfaz de trabajo de SpagoBi Server.....	67
Figura 65. Visualización del cubo	67
Figura 66. Ingreso al cubo y visualización de barras de herramientas.....	68
Figura 67. Interpretación gráfica agrupada de Dim personas.....	68
Figura 68. Interpretación gráfica agrupada de Dim. Actividad red	69
Figura 69. Interpretación gráfica agrupada horizontal de Dim actividad red	70
Figura 70. Interpretación gráfica apilada de Dim personas.....	71
Figura 71. Tabulación de datos Dim personas.....	71
Figura 72. Tabulación de datos Dim actividad red	71
Figura 73. Gráfica del Dim tiempo red (masculino y femenino).....	72
Figura 74. Gráfica del Dim persona (masculino y femenino)	72
Figura 75. Análisis de personas que apoyan a un partido político.....	73
Figura 76. Personas que apoyan a un partido político desde otra posición	73

RESUMEN

Actualmente las Redes Sociales se encuentran inmersas en todos los campos del diario vivir, en áreas como ingeniería, arquitectura, salud, economía, política, etc., por lo que se han constituido en un factor de gran peso en la decisión del usuario, por tal motivo, el presente trabajo se desarrolla con la finalidad de utilizar la Minería de Datos como herramienta para transformar la información recopilada de las redes sociales sobre las tendencias políticas.

La investigación parte del estudio teórico de la minería de datos, con la definición de conceptos necesarios para el proceso, posteriormente se diseñó la herramienta para la recopilación de información, la cual fue aplicada a los usuarios de redes sociales como Facebook, Twitter, WhatsApp e Instagram, debido a que se estudió acerca de las tendencias políticas en dichas redes. Del total de la población, mediante un muestreo no probabilístico por conveniencia, se seleccionó un total de 1016 personas para la aplicación de la encuesta, con la cual se pudo obtener la información necesaria para realizar el análisis requerido con la técnica de minería de datos.

Posteriormente para el procesamiento de información se utilizó la metodología Data Warehouse que es una base de datos corporativos encargada de integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista, con lo cual, se pudo determinar que los usuarios si utilizan las redes sociales para compartir contenido político, siendo evidente que existe mayor afinidad por los partidos como Movimiento Centro Democrático, Partido Sociedad Patriótica, Movimiento Alianza País, entre otras.

ABSTRACT

Currently Social Networks are immersed in all fields of daily life, in areas such as engineering, architecture, health, economics, politics, etc., which is why they have become a major factor in the decision of the user, for this reason, this work is developed with the purpose of using Data Mining as a tool to transform the information collected from social networks on political trends.

The research is based on the theoretical study of data mining, with the definition of necessary concepts for the process, then the tool was designed for the collection of information, which was applied to users of social networks such as Facebook, Twitter, WhatsApp and Instagram, because it was studied about the political trends in these networks. Of the total population, through a non-probabilistic convenience sampling, a total of 1016 people were selected for the application of the survey, with which it was possible to obtain the necessary information to perform the required analysis with the data mining technique.

Subsequently, for the processing of information, the Data Warehouse methodology was used, which is a corporate database in charge of integrating and debugging the information and then processing it, allowing the analysis of the information from different points of view, with which it was possible to determine that users if they use social networks to share political content, being evident that there is greater affinity for parties like Movimiento Centro Democrático, Partido Sociedad Patriótica, Movimiento Alianza País, among others.

INTRODUCCIÓN

En los últimos años, la industria de las comunicaciones ha centrado su atención en sistemas que transportan datos a largas distancia. La industria de la informática y de las redes WAN (Wide Area Network) es un sector estable en constante desarrollo abarcando cada vez más y más actividades personales, profesionales y empresariales.

Tradicionalmente las personas disponían de su propia computadora, con la cual ejecutaban sus propias aplicaciones y administraban sus actividades e información de forma independiente conectada a una red de recursos, pronto surgió la necesidad de compartir esos recursos con el entorno social, acoplándose a la tendencia de las comunicaciones globales para lograr los objetivos establecidos.

El factor tecnológico disponible solo al entorno profesional pronto llegó al sector personal cubriendo un entorno social muy diverso con la aplicación de las TIC (Tecnologías de Información y Comunicación), por lo que engloba un amplio rango de las Redes Sociales (RRSS) disponibles que están siendo utilizadas por personas sin diferenciar nivel social, género o edad.

Situación Problemática

Las Redes Sociales (RRSS) han copado todos los campos del diario vivir, en áreas como ingeniería, arquitectura, salud, economía, política, etc., llegando a ser un factor de gran peso en la decisión del usuario y que no necesariamente debe tener el conocimiento y razonamiento de la veracidad o falsedad de la información, según, Corbella (2018), indica que las noticias falsas en redes sociales son el 70% más que las ciertas.

Las redes sociales se convierten en caminos directos para comunicarse entre los candidatos y los votantes según indica, NN (2015), estos intensifican sus movimientos y propuestas en períodos electorales estando la mayor parte de tiempo compartiendo con los internautas, además siendo un canal para el debate y la conversación.

Al momento de interactuar los políticos con los internautas se pueden crear grescas, esto debido a la existencia de personas aliadas a otro partido político, los cuales publican información falsa que va en contra de la integridad del político, realizando campañas sucias. En ese momento es en el que muchos de los internautas comparten tal información creando desinformación a las personas que lo leen.

Por tal motivo se buscó un medio accesible y confiable para la recopilación y transformación de la información por medios electrónicos, donde la aplicación de Minería de Datos usando el Dataware se encargará de proporcionar los datos sobre las tendencias políticas que tiene o existe en el país.

Formulación del Proyecto

Los antecedentes antes mencionados dirigen hacia el planteo del problema científico:

¿De qué forma se puede validar los datos de las Redes Sociales para conocer la tendencia política que tienen las personas?

Al aplicar la minería de datos se puede conocer la tendencia política que tiene cada una de las personas y al mismo tiempo conocer sobre la veracidad de tal información.

Objetivos

Objetivo General

Crear un Data mining - ETL a través de la Minería de Datos como herramienta de transformación de la información recopilada de las redes sociales sobre las tendencias políticas.

Objetivos Específicos

- Analizar los modelos de datos, para la determinación de la arquitectura y el proceso de selección de información.
- Diseñar un modelo de extracción de datos para el descubrimiento de información desconocida.
- Crear un Data mining - ETL a través de la Minería de Datos, utilizando información recopilada de las redes sociales sobre las tendencias políticas.
- Validar los patrones del procesamiento de datos para la evaluación y difusión del conocimiento.

Hipótesis o idea a defender

En caso de no ser la minería de datos un medio viable en las redes sociales para la transformación de datos, misma que permitirá conocer la tendencia política que tienen las personas, seguido de los problemas al no tener seguridad en la información recolectada,

además tales personas dedicadas a la política tendrían que buscar otro medio físico o realizar encuestas.

Con lo planteado, se menciona que la minería de datos ayudará en la recolección de información y permitirá procesarla con la herramienta Datawarehouse, información que se encuentra en la web y sobre todo en las redes sociales como Facebook, Twiter e Instagram.

Con la ayuda de estas herramientas se podrá conocer la tendencia política que tienen las personas en el país, de esta manera el programa arrojará datos viables y ordenados, los cuales deberán ser claros a cualquier persona.

Justificación

En el Ecuador se realizan elecciones para presidentes y asambleístas cada 4 años, al igual que para prefecturas, concejales, alcaldes, etc., siendo la segunda en la mitad de la primera, esto quiere decir que cada 2 años existen elecciones y los medios utilizados para conocer la tendencia política no siempre es confiable, ya que en muchas ocasiones dan a un candidato como ganador y en las elecciones el triunfador es otro.

Por tal motivo la presente investigación científica es de gran importancia, debido a que por medio de la minería de datos se podrá conocer la tendencia de las personas que utilizan las redes sociales en el país, ya que, con solo enviar un mensaje, compartir una publicación, hacer un comentario en cualquier medio web, la información quedará guardada, almacenada y procesada y de esa forma conocer quien o quienes serán los políticos o listas ganadoras en cada período.

En el Ecuador esta herramienta puede ser de mucha utilidad para conocer las tendencias políticas de las personas, pero a su vez ayudan en otras áreas como la educación, el deporte, entre otras. Partiendo de la minería de datos se puede encontrar información de valiosa cuantía para lo que se desee.

Con el desarrollo de esta investigación se aspira conocer información de gran importancia y tener claro lo que las personas quieren y aspiran.

Capítulo I. Fundamentación teórica

1.1 Contextualización espacio temporal del problema

Desde hace años atrás se ha evidenciado la necesidad de administrar la información que generan todo tipo de organizaciones, lo que exige llevar de manera ordenada datos que faciliten la toma de decisiones, por tal motivo se promueve la generación masiva de información que creen históricos de los movimientos que realizan las organizaciones, sin embargo actualmente en la mayoría de los casos este control se lleva de forma no muy desarrollada, por lo que los resultados no son los que se desean, (Benalcázar, 2017).

Dentro de las organizaciones mundiales se encuentran las redes sociales, las cuales indiscutiblemente juegan un papel importante en la vida social, económica y política de las personas, debido a que han permitido transformar la dinámica para la comunicación y adquisición de información, llegando a ser un factor de gran peso en la decisión del usuario despreciando la veracidad o falsedad de la información.

De acuerdo a ello, en la actualidad la técnica de minería de datos está siendo aplicada de forma muy notoria en diversas áreas, tales como el marketing, las finanzas, la medicina, salud pública, educación, entre otras, debido a que la aplicación de esta técnica sobre los datos provenientes de fuentes como las redes sociales, mismas que pueden revelar patrones sobre individuos inmersos en el ambiente compartido y producir conocimiento que años atrás no era factible encontrar, pues la información era muy variada y compleja, (Jaramillo, Cardona, & Fernández, 2015).

Por lo expuesto anteriormente, el presente trabajo de investigación tiene como finalidad analizar las redes sociales de mayor impacto en conjunto con los diferentes algoritmos de minería de datos que puedan utilizarse para generar conocimiento a partir de la información de dichas redes.

1.2 Cuerpo teórico conceptual

1.2.1 Aproximación al concepto de minería de datos

Inicialmente la minería de datos se puede definir como un proceso en el cual se analiza una gran cantidad de datos, con la finalidad de descubrir nuevas relaciones, tendencias y patrones significativos, (Pérez C. , 2007).

El análisis de datos toma el nombre de Data Mining (Minería de datos) gracias a la disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas cuyo objetivo es, descubrir de manera automática el conocimiento contenido en una extensa base de datos.

Las técnicas que se utilizan en esta metodología tienen como meta descubrir patrones, perfiles y tendencias mediante el análisis de los datos haciendo uso de tecnologías de reconocimiento de patrones, lógica difusa, redes neuronales, algoritmos genéticos y otras técnicas especiales de análisis de datos.

La minería de datos es un concepto muy extenso que necesita ser disgregado conceptualmente por etapas, gracias a la informatización de las organizaciones y la aparición de programas operacionales. Sobre el sistema de información es posible utilizar esta herramienta para dar soporte a los procesos fundamentales de la organización (ventas, producción, personal, inventarios, etc). Una vez delimitado la necesidad de tener un soporte informático para los procesos básicos organizacionales (Sistemas de Información para la Gestión), es necesario pasar a la siguiente fase de los sistemas de información (Sistemas de Información para la Toma de Decisiones).

Las herramientas de negocio que han aparecido para la toma de decisiones (DSS o Decision Support Systems) que se interrelacionan son: Executive Information System (EIS), Executive Information System (OLAP), informes, consultas y las propias herramientas de minería de datos.

1.2.1.1 EIS: Executive Information System

Es un sistema que permite a los ejecutivos tomar decisiones en tiempo real, el cual no solo consta de software y hardware, sino también de toda la infraestructura que permite tener la información completa al alcance de su mano representada de manera gráfica al estilo de hojas de cálculo (ventas semanales, inventarios, balances, etc), (Nieto, 2016).

1.2.1.2 OLAP: On-Line Analysis Planning

Se denominan Sistemas de Procesamientos Analíticos en Línea (OLAP), y son aquellos que deben analizar datos desde diferentes perspectivas y soportar requerimientos complejos de análisis en un volumen gigantesco de datos.

Este sistema se caracteriza por ser una técnica de estudio multidimensional de datos, en el cual el usuario puede navegar por los mismos de modo asistido, (Molina & García, 2006).

En este sistema, la información es vista como cubos, los cuales están formados por categorías descriptivas (dimensiones) y valores cuantitativos (medidas) como se puede ver en la figura 1.

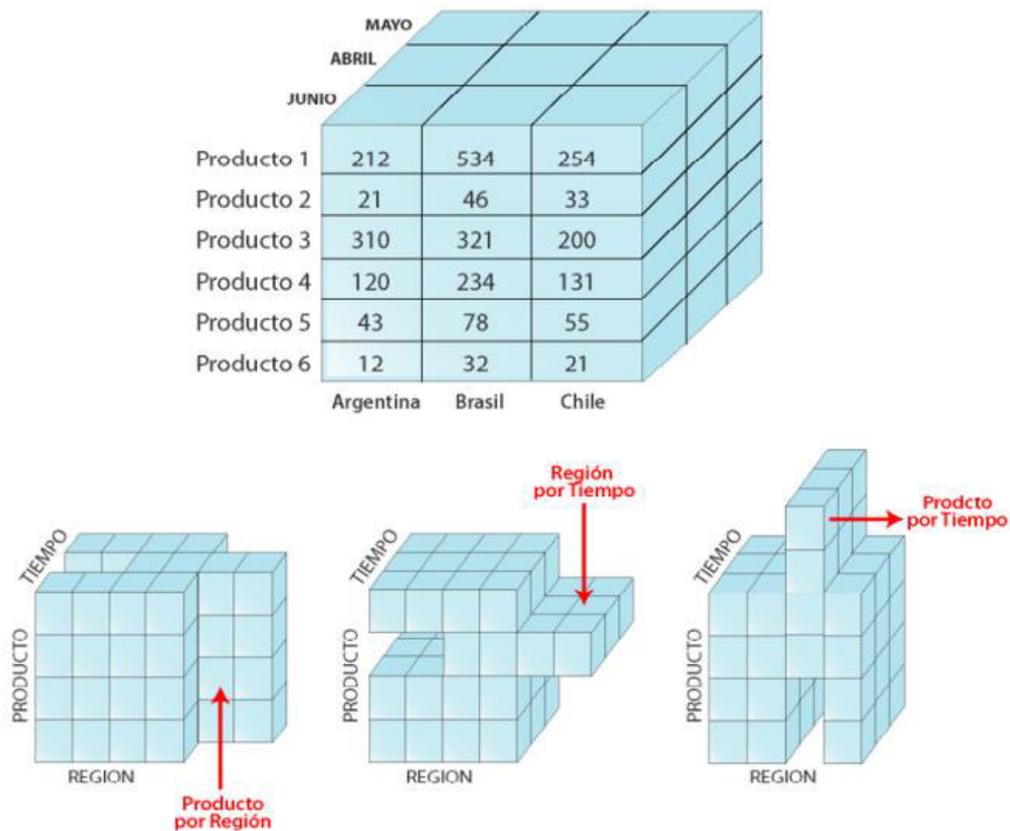


Figura 1. Vista de los datos en forma de cubo

Fuente: (Ibarra, 2006)

Las ventajas de presentar de esta manera la información es que facilita a los usuarios diseñar consultas complejas, ordenar datos en un reporte, filtrar o disgregar datos en subconjuntos significativos y cambiar de datos resumidos a datos detallados.

Las dimensiones típicas de un cubo que contenga información de ventas, por ejemplo, incluirían: tiempo, región, producto, canal, escenario y organización (planeado - real). Las medidas tradicionales incluirían: ventas en tipo de moneda, unidades de artículos vendidos, cantidad de personas, ingresos y egresos.

OLAP está conformada de varias operaciones analíticas fundamentales como: consolidación, drill-down y slicing and dicing.

1.2.1.3 Consolidación

Está formada por el conjunto de datos que pueden involucrar agrupaciones simples o complejas que incluyen datos enlazados. Un claro ejemplo puede ser el agrupar a las oficinas de ventas con respecto a las provincias y las provincias respecto a las regiones.

1.2.1.4 Dril – Down

OLAP puede dirigirse en sentido contrario y presentar automáticamente datos detallados que forman datos consolidados. Un ejemplo es que se puede consultar fácilmente las ventas por tipo de producto o representante que forman el total de ventas de una región.

1.2.1.5 Slicing and dicing

Es la capacidad para observar diferentes perspectivas a las bases de datos. Una agrupación de la base de datos de ventas podría exponer todas las ventas de la clase de artículo dentro de las regiones, mientras que otro grupo podría mostrar las ventas por clase dentro de cada tipo de artículo. El análisis del Slicing and Dicing se lo realiza generalmente a lo largo del eje del tiempo, con la finalidad de encontrar patrones y generar pronósticos. La operación se puede observar en la figura 2.

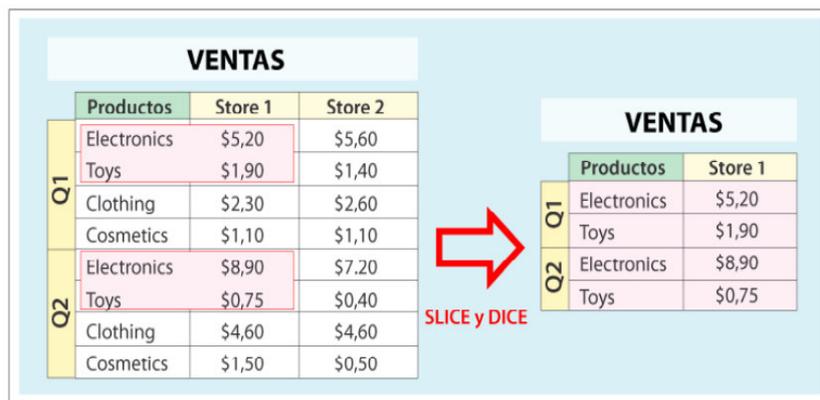


Figura 2. Operación Slice and Dice

Fuente: (Rodríguez & Díaz, 2009)

1.2.1.6 Herramientas de minería de datos

Las herramientas de la minería de datos utilizadas para la extracción de conocimientos se pueden clasificar en dos principales grupos:

1. Técnica de verificación en donde el usuario entrega una hipótesis y el sistema lo comprueba.
2. Método de descubrimiento en el cual se van encontrando patrones potencialmente importantes de manera automática, (Rodríguez & Díaz, 2009).

Las herramientas más populares que se utilizan habitualmente en la minería de datos se enlistan en la figura 3.

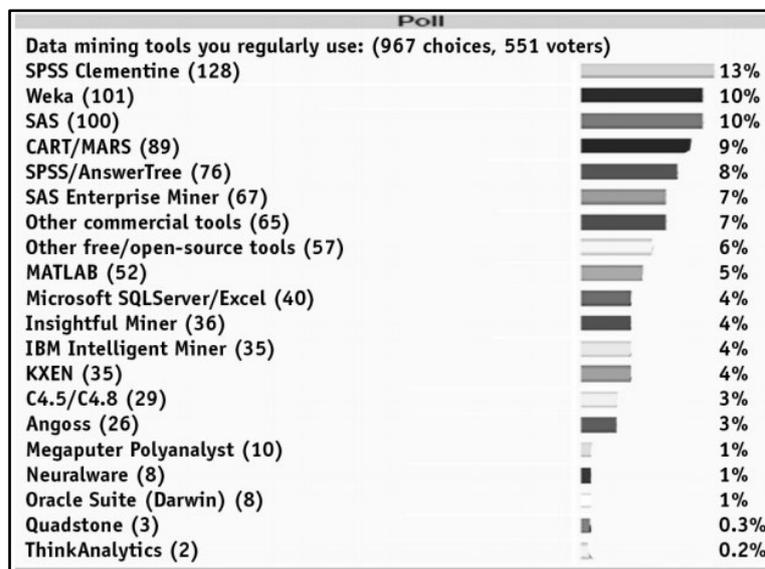


Figura 3. Herramientas de Minería de Datos usadas habitualmente

Fuente: (Rodríguez & Díaz, 2009)

Sin embargo, las herramientas mencionadas con anterioridad, habitualmente suelen necesitar de la existencia previa de una base de datos (Data Warehouse) cuyo objetivo es ser el sistema de información central de todo este proceso.

1.2.2 Almacén de datos (Data Warehouse)

Este tipo de almacén se originó por dos razones principales:

1. La necesidad de contar con una sola fuente de datos limpia y consistente con el objetivo de ser utilizada para la toma de decisiones.

2. La necesidad de realizarlo sin invadir los sistemas de operación.

Es por ello que la Data Warehouse se denomina como “un almacén de datos orientado a un tema, integrado, no volátil y variante en el tiempo, que soporta decisiones de administración”. El término no volátil significa que los datos agregados no pueden ser cambiados, pero si borrados, (Molina & García, 2006).

1.2.2.1 Diseño de una Data Warehouse

Se han propuesto diferentes metodologías para el diseño y construcción de Data Warehouse (DW) las cuales tienen en común las siguientes etapas:

1. **Analizar los requerimientos del usuario final:** es necesario que las herramientas y aplicaciones entreguen información útil de acuerdo a la decisión estratégica solicitada por el cliente.
2. **Seleccionar las fuentes de información:** luego de haber definido el requerimiento final del cliente se buscan las fuentes de información que permitan cumplir con el objetivo planteado.
3. **Elaborar un modelo lógico de datos:** los más utilizados son el modelo cubo y el modelo estrella.
4. **Generar un prototipo para el cliente:** la meta es indicar al usuario una idea preliminar de DW a desarrollar para que este apruebe y asevere que el modelo se ajusta a sus requerimientos.
5. **Seleccionar el sistema administrador de base de datos (SABD):** este sistema se encarga de la velocidad de respuesta en las consultas, por lo cual se convierte en una de las variables más importantes dentro del desarrollo del DW.
6. **Diseño físico de la BDD:** es la implementación física del modelo lógico de datos en base a la estructura que entregue el SABD (relaciones entre el modelo de datos y las tablas).
7. **Almacenamiento:** guardar la información a través de un trabajo de extracción, conversión y carga de datos.

8. **Mejora continua del desempeño obtenido:** es necesario realizar modificaciones continuas a la estructura interna de datos para mejorar el tiempo de respuesta ante las consultas, (Martínez C. , 2012).

1.2.3 Proceso de extracción del conocimiento a partir de datos (KDD)

El Knowledge Discovery in Databases es un proceso automático en el que se combinan análisis y descubrimiento. La tarea consiste en extraer patrones en forma de reglas o funciones para que el cliente los revise.

Las etapas básicas de este proceso son, reprocesar los datos, hacer minería de datos y presentar resultados.

1.2.3.1 Etapas del proceso KDD

Las etapas del proceso KDD se describen a continuación (Ver figura 4):

1. **Selección:** De acuerdo a los requerimientos del usuario y metas del proceso, se crea un grupo de datos objetivo, escogiendo todo el conjunto de datos o una muestra de este sobre el cual se efectuará el proceso de análisis.
2. **Procesamiento / Clasificación:** En esta etapa se disgrega los datos de menor calidad aplicando la técnica de remoción de datos ruidosos, datos nulos, datos duplicados. En esta fase es de suma importancia el trabajo en conjunto entre el usuario y el analista.
3. **Conversión / Reducción:** Se utilizan técnicas de reducción de dimensiones para disminuir el número de variables simplificando una tabla o base de datos vertical u horizontalmente. Las más utilizadas son agregaciones, compresión de datos, segmentación, histogramas, muestreo, etc.
4. **Data mining (Minería de datos):** El objetivo de esta fase es descubrir patrones inimaginables e importantes aplicando las técnicas anteriormente expuestas.
5. **Análisis de resultados / Evaluación:** Interpretación de patrones descubiertos y generalmente se repite el ciclo en busca de nuevos hallazgos.

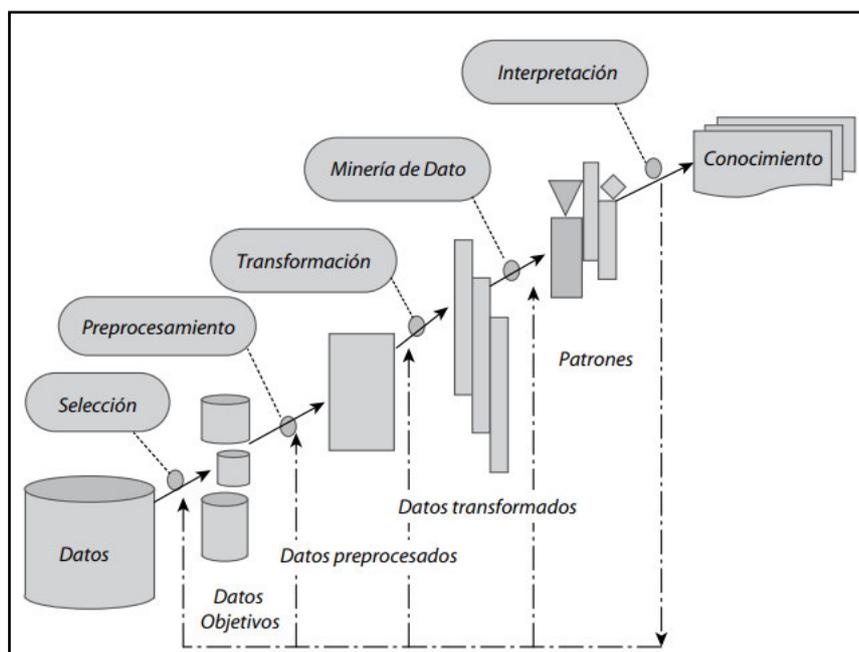


Figura 4. Etapas del proceso KDD

Fuente: (Timarán, Hernández, Caicedo, Hidalgo, & Alvarado, 2016)

1.2.4 Las redes sociales

El término networking se ha convertido en una palabra común, todas las personas van a un lugar para gestionar su red de contactos, para un gran porcentaje de población la *World Wide Web* tiene como principal objetivo establecer contactos, (Kadushin, 2012).

Las redes sociales han sido parte de la vida desde la época de los recolectores y cazadores, los individuos se generaban vínculos entre sí a través de sus relaciones unas con otras y su dependencia entre ellos. Las relaciones familiares y el parentesco, al igual que el barrio, las ciudades, amigos, etc., son redes sociales; y fuera del parentesco familiar las personas de la sociedad moderna dependen unas de otras para el servicio de taxis, la venta del periódico el dentista, etc. Es decir, el parentesco en comunidades se forma a través de las necesidades y en torno a las redes sociales geográficamente dispersas, es por ellos que se dice que se lleva milenios siendo networkers.

Networking es el uso de una de red de contactos para alcanzar los objetivos personales, por lo tanto, una red es simplemente un grupo de relaciones entre objetos que pueden ser organizaciones, personas o naciones en las cuales se comparte ideas de amistad, dinero, poder e incluso la enfermedad, (Flores, Morán, & Rodríguez, 2015).

Hace solo algunos años (2003) se ha venido manejando el concepto de las redes sociales, pero como se describe en los párrafos anteriores esto no es del todo nuevo, sin embargo, lo que es relativamente nuevo es la manera de referirse a ellas, describiéndolas, analizándolas y relacionándolas con instancias sociales más formales como organizaciones y gobiernos.

La red social viéndolo desde el lado informático, se denomina como un servicio que brinda la oportunidad a los individuos de construir un perfil semipúblico o público dentro de un sistema delimitado. También permite socializar con otros usuarios con los que comparten una conexión, ver y recorrer su historial de conexiones y consultar información de los demás usuarios del sistema, (Flores, Morán, & Rodríguez, 2015).

A continuación, se puede apreciar el desarrollo de las redes sociales a través del tiempo (Ver figura 5).

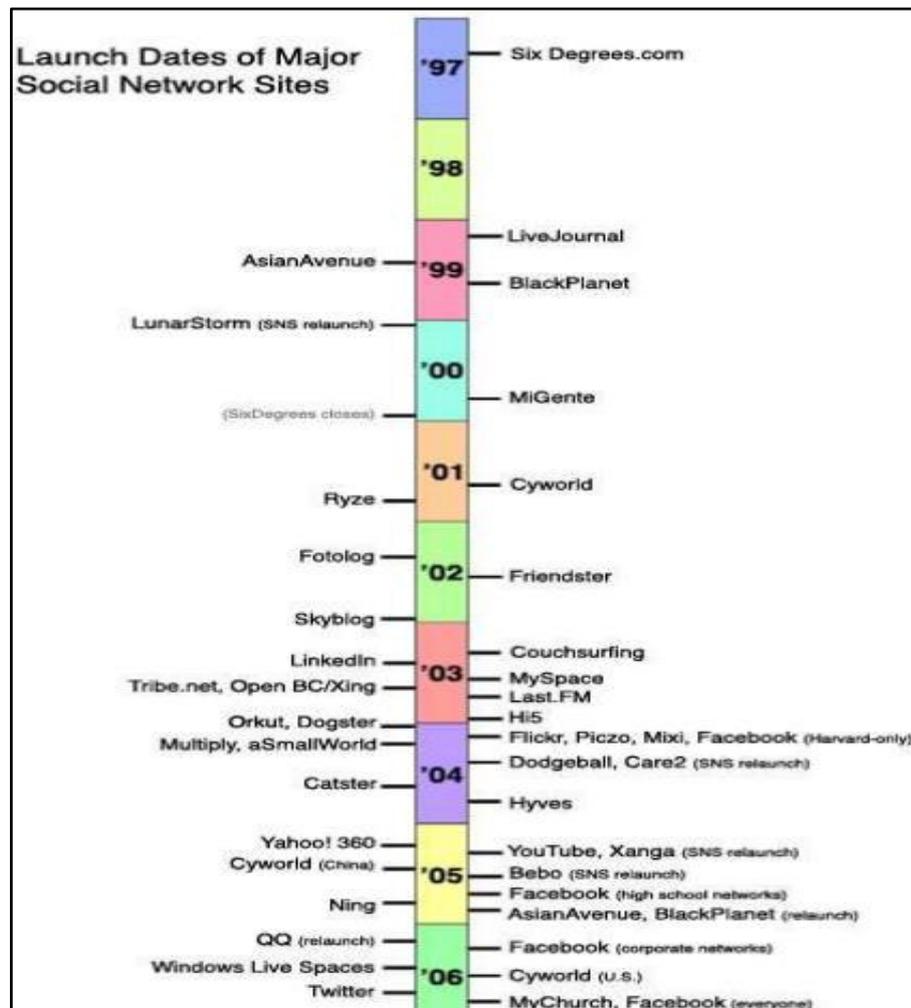


Figura 5. Línea del tiempo de las redes sociales

Fuente: (Flores, Morán, & Rodríguez, 2015)

1.2.4.1 Proceso de conexión

Hoy por hoy el sistema de conexión networking a través de las redes sociales está en pleno auge, ya que este no se limita a los adolescentes en búsqueda de amigos. Según Kadushin (2012), en diciembre de 2008, el 75% de la población adulta de Estados Unidos utilizaba internet, de estos el 35% mantienen un perfil en un sitio networking social, el 75% de los adultos entre 18 y 24 años de edad tienen un perfil, y casi el 70% de dichos adolescentes estudian. Estas cifras sugieren que conforme estas generaciones vayan creciendo, en Estados Unidos el total de usuarios de los sitios networking social irá creciendo paulatinamente.

Los lugares networking online tienen profundas implicaciones, tomando como ejemplo un perfil con 100 amigos este puede tener contacto con 1.000.000 de personas y exponencialmente esta cifra puede ir creciendo y en poco tiempo el mundo entero es un potencial amigo, para bien o para mal.

Esperando que la situación sea para bien la conexión en mención tiene el potencial de permitir a una persona tener acceso a información valiosa como: referencias para encontrar trabajo, ayuda con problemas personales, capacidad para ascender en la escala social de ocupaciones, recomendaciones de nuevos libros, películas, restaurantes, etc.

La red social (RRSS), continúan avanzando a pasos agigantados como una estructura social compuesta por un grupo de personas, las cuales están conectadas por uno o varios tipos de relaciones, tales como amistad, parentesco, interés común o que comparten conocimiento.

Las áreas de mayor desarrollo son:

- Entretenimiento: Facebook, WhatsApp, Instagram, Multicines, entre otras.
- Informativas: Twitter, entre otras.
- Negocios: Mercadolibre, Computrabajo, entre otras.
- Grupo de lectura.
- Grupo de oración.
- Otras.

1.2.4.2 Las redes como mapas de información

El análisis de redes sociales revela lo que a simple vista no se puede observar, por ejemplo, cuando se compra un libro en Amazon, la página web te informa que otros individuos escogieron el mismo libro y de igual manera cuáles son sus preferencias literarias. Aprovechando este análisis y la información disponible en la base de datos es posible crear un mapa de libros relacionada con cualquier tema seleccionado.

En el año 2008, Valdis Krebs un consultor de organizaciones especializado en generación de aplicaciones para redes sociales creó un mapa de libros relacionados con la campaña presidencial de Estados Unidos en ese año, (Ver figura 6).

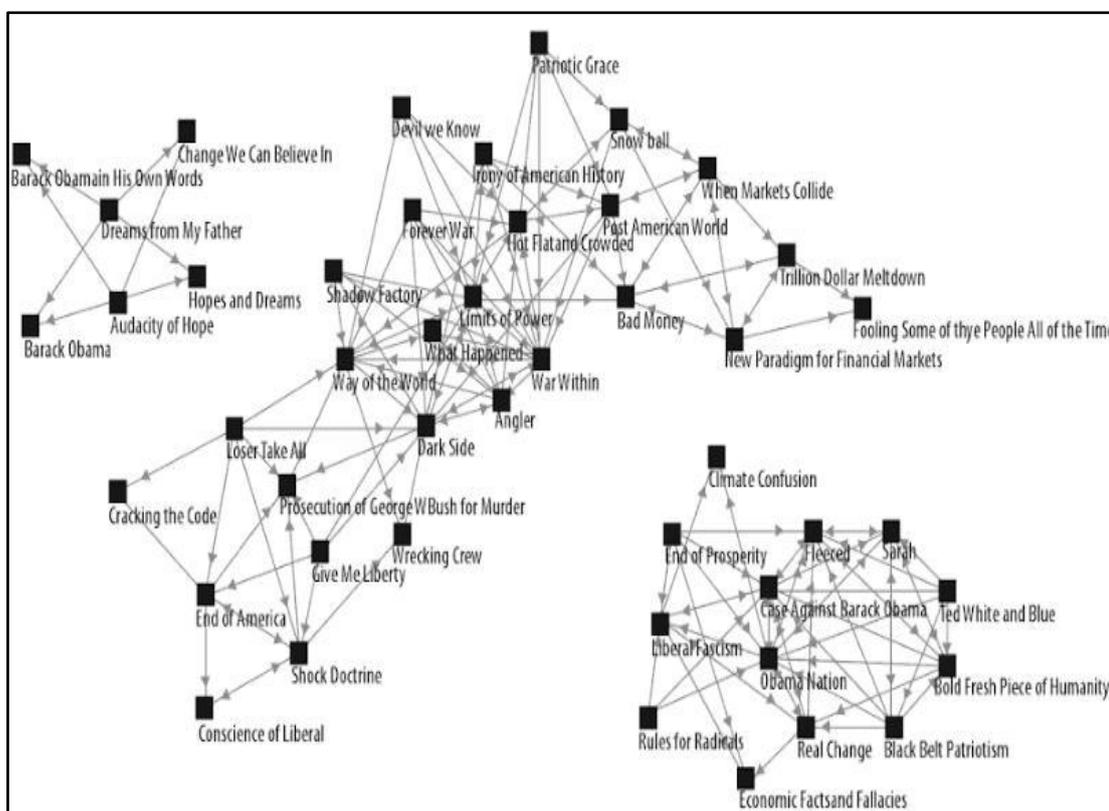


Figura 6. Libros comprados por las mismas personas en la campaña presidencial de 2008.

Fuente: (Kadushin, 2012)

En la figura anterior se nota claramente que en el 2008 la población lectora estaba polarizada, ya que existe un grupo sobre Obama en la parte superior izquierda; un grupo sobre la campaña demócrata en el medio y un grupo republicano a la derecha, los grupos no se mezclan entre sí.

Las aplicaciones relativas a las redes son beneficiosas para mostrar información como quien compró un libro, pero se puede ir más allá y encontrar sentido a una noticia que envuelve conexiones, como: ¿Quién estuvo implicado en qué operación bancaria?, ¿Cómo interactuar con un grupo específico de acuerdo a sus tendencias políticas? ¿Quién pertenecía a la red de los secuestradores el 11 de septiembre?, etc.

1.2.4.3 El fenómeno de las redes sociales

En el año 1997, la historia de las redes sociales se da a conocer en un formato similar como se lo conoce en la actualidad, en donde muchos usuarios podían crear su perfil y agregar a ellos a las amistades que deseaban, y en el año 1998 se les permitió navegar en las mismas; desde ese momento los usuarios en las redes sociales se han ido incrementando de manera incontrolable.

Facebook nace en el año 2004 y es una de las redes sociales más utilizadas en la actualidad, a pesar que desde ese año hasta el 2006 no estaba permitido ingresar ningún usuario sin una cuenta de correo electrónico de una Universidad de Norteamérica, y en el año 2008, en solo en 5 semanas ya habían sobrepasado los 150 millones de usuarios en registrarse en dicha red.

Las redes sociales han sido utilizadas para diferentes ámbitos educativos y es por eso que han vuelto a resurgir como uno de los medios principales de comunicación a nivel mundial, permitiendo que las personas estén en contacto y compartiendo información de una manera inmediata.

Las teorías matemáticas tienen mucha relación con las redes sociales, ya que el estudio de las estructuras y sus conexiones fueron utilizados por algunos autores, en donde los llevaban al campo psicológico social y en base a eso aplicaban dichas teorías para investigar la vida en la sociedad, (Pérez & López, 2015).

A mediados de los años 90, el psicólogo estadounidense Stanley Milgram aplicó el experimento del mundo pequeño, el cual consistía en enviar una postal a una persona conocida para que a través de su red de contactos llegara a su destinatario del cual sabía solo sus nombres y su localización, y es ahí donde constató que se necesita 5 o 7 intermediarios para llegar a esa persona desconocida pasando por pocos puntos de conexión.

Otra fuente de origen de las redes sociales también surge de la antropología, ya que al realizar pruebas con los trabajadores con el objetivo de aumentar la productividad en las empresas se determinó que las redes sociales se trataban más de un aspecto motivacional, ya que ayudaba a tener una mejor relación interpersonal, (Pérez & López, 2015).

Tras realizar pocos estudios de las redes sociales en las décadas 60 y 70, en adelante nacen investigaciones científicas y otros numerosos tipos de investigaciones para ir descifrando la importancia del uso de las redes.

En la actualidad se puede utilizar las redes sociales virtuales, en donde todos pueden hablar con otro usuario y la comunicación sea más real (cara a cara) gracias al avance tecnológico de telecomunicaciones que hoy en día existen.

Las redes sociales inician en el año 2003 justo en la crisis informática que atravesaban en ese tiempo y por ende las empresas que tenían páginas web tenían que cerrar sus cuentas debido a que nadie los visitaban; es ahí donde 3 personas, Marc Pincus, Reid Hoffman y Jonathan Abrams crean empresas online en donde se puede interaccionar entre personas y ya no solo entre empresas permitiendo darle importancia al usuario en general.

Es aquí en donde el desarrollo del internet y las redes sociales se convierten en protagonistas como medios de conexión entre usuarios de cualquier parte del mundo, el desarrollo tecnológico y un mercado globalizado facilitan la creación de múltiples páginas de sitio web de uso social, ayudando a comunicarse de nuevas formas en periodos más cortos con usuarios a grandes distancias, (Pérez & López, 2015).

1.2.4.4 El rol de las redes sociales en la proyección corporativa

En la actualidad nadie duda del éxito de las redes sociales al transformar el modo en que las empresas interactúan con sus clientes y la base misma del como concretar un sinnúmero de negocios.

Las empresas han sabido aprovechar su incorporación en las redes sociales, las cuales se utilizan como canal de comunicación con sus clientes, como herramienta para crear una marca y mejorar su imagen o como medio para ofrecer sus productos o servicios.

La forma que interactúan las empresas entre sus integrantes y el entorno debe estar dentro de un esquema direccionado y estratégico, los resultados irán relacionados a los objetivos

específicos planteados inicialmente y procurarán el envío de ideas claras, consistentes y asertivas, garantizando así la fluidez y eficiencia de sus mensajes.

Lo anteriormente descrito, ha implicado un desarrollo importante en el proceso de comunicación debido al impacto que ha generado en la sociedad y permitiendo de esta manera un intercambio de mensajes de manera sincrónica y asincrónica, pero sobre todo interactiva que se enmarcan en un ámbito de alcance global, que además ofrece una variedad de opciones para que el cliente pueda expresarse, emitir mensajes escritos, audio o video, emitir comentarios, realizar consultas, recopilar información, entre otros.

Toda esta información se irá guardando de manera automática en una base de datos, la cual servirá como punto de partida para la toma de decisiones corporativas de mejora continua y para la elaboración de planes estratégicos, utilizando a la minería de datos como herramienta para el descubrimiento de patrones, tendencias que darán la dirección de hacia dónde apuntalar los esfuerzos y planes de acción, (Hüt Herrera, 2012).

1.3 Revisión de las investigaciones previas del objeto de estudio

Una vez realizada la revisión bibliográfica del tema de investigación, se ha encontrado investigaciones previas que se toma como antecedentes para el desarrollo del presente proyecto, las mismas que se presentan a continuación:

Contreras & Rosales (2016), en su investigación denominada: “Análisis del comportamiento de los clientes en las redes sociales mediante técnicas de Minería de datos”, describieron los resultados de la utilización de técnicas de minería de datos para analizar el comportamiento de los clientes de una empresa de moda en la red social Instagram, para lo cual se utilizó la metodología CRISP-DM con lo que se evaluó los modelos descriptivos, utilizando las técnicas de reglas de agrupación y asociación, concluyendo que el aprovechamiento de la información generada por redes sociales, en conjunto con técnicas informáticas modernas como la minería de datos, permite a las empresas conocer las preferencias y el comportamiento de sus clientes, sin la necesidad de realizar encuestas ni cualquier otro trabajo de campo.

De la misma manera, Roales (2014), realizó una investigación con el tema: “Detección de tendencias en twitter utilizando minería de datos adaptativa” con el objetivo extraer mensajes que los usuarios comparten en Twitter y tratarlos mediante diferentes técnicas de

minería de datos, para de esa forma poder distinguir grupos de tendencia de opiniones sobre un tema específico, en este caso Eurovisión, para lo cual se realizó el filtrado de datos con la extracción de tweets, logrando la recopilación de un total de 155422 mensajes. Con la información obtenida se concluye que al tomar como objetivo de análisis únicamente la frecuencia de las menciones es muy difícil predecir resultados, ya que se tienen en consideración tanto las buenas como las malas opiniones de los usuarios.

Cortés, et al. (2015), en su trabajo investigativo denominado “Estudios de los hábitos de conexión en redes sociales virtuales, por medio de la minería de datos” tuvo como finalidad presentar una visión general de los hábitos de conexión en las redes sociales virtuales entre los estudiantes de pregrado, esta información se obtuvo con la aplicación de una encuesta a 458 estudiantes, cuyo análisis e interpretación de resultados se realizó utilizando la técnica de minería de datos, mediante las cuales se obtuvieron grupos de estudiantes definidos por el género, la frecuencia de conexión y las redes sociales que consultan, concluyendo que existe correlación de uso entre las redes sociales Facebook y Youtube, así como el servicio de mensajería instantánea WhatsApp.

Finalmente, Díaz (2017), en su investigación denominada “Aplicación de técnicas de minería de datos para el análisis de sentimientos en las redes sociales sobre productos de la marca fideos Cayambe”, utiliza la metodología CRISP-DM para analizar información recolectada de la red social Facebook, con la finalidad de evaluar y determinar las opiniones de las personas, dejando como resultado una percepción positiva en base a los comentarios publicados por los seguidores de la marca a nivel nacional.

CAPÍTULO II. Marco Metodológico

2.1 Enfoque metodológico de la investigación

La presente investigación se basó en el enfoque cuali-cuantitativo, por lo que es cualitativa debido a que se centra en determinar las cualidades y características del fenómeno y la población de estudio, pues de acuerdo a lo que plantea, Martínez (2006), la investigación cualitativa trata de identificar la naturaleza de las realidades, su estructura dinámica, aquella que es la causal de su comportamiento y manifestaciones para generar resultados y sugerencias que permitan solucionar el problema de estudio; bajo este contexto, la presente investigación se basó en el descripción de modelo de datos con bodega y minería de datos, para determinar en entorno general y específico el comportamiento de la población involucrada.

Además, la investigación es cuantitativa debido a que trata de identificar la naturaleza de sus realidades, asociación entre las variables a través de datos numéricos, pues como lo manifiesta, Del Canto & Silva (2013), la investigación cuantitativa parte de datos evidenciales, es decir usa la recolección de los datos de manera estructurada y sistemática para comprobar hipótesis planteadas previamente, con base a la medición numérica y el análisis estadístico, para con ello establecer patrones de comportamiento y probar teorías. Por su parte, Palacios (2006), menciona que para que exista esta metodología cuantitativa, debe existir claridad entre los elementos de investigación desde el momento que inicia hasta donde termina, abordando los datos estáticamente y asignándoles un significado numérico por medio de la estadística, con el objetivo de hacer inferencia.

De acuerdo a lo manifestado, para el desarrollo de la investigación, se utilizó el enfoque cuali-cuantitativo considerando que lo cualitativo no se opone a lo cuantitativa, sino que lo integra, especialmente donde sea importante.

2.2 Modalidad de la investigación

La investigación tuvo una modalidad bibliográfica documental, debido a que se recolectó información de libros, revistas, artículos científicos, folletos y demás documentos que permitieron fundamentar conceptos y teorías acerca de las herramientas de minería de datos y las redes sociales, pues como lo menciona, Morales (2004), la investigación documental se caracteriza por utilizar el documento escrito en sus diferentes formas: documentos

impresos, electrónicos y audiovisuales, los cuales son el resultado de otras investigaciones de reflexiones teóricas que representan la base teórica del área de objeto de investigación; lo que permitirá construir conocimiento a través de su lectura, análisis, síntesis, reflexión e interpretación.

De igual forma fue una investigación de campo debido a que se recolectó información de la fuente primaria, es decir directamente de las personas involucradas en la investigación, quienes proporcionaron los datos necesarios acerca de las variables planteadas para alcanzar los objetivos establecidos previamente. Por todo lo dicho se puede describir que, La Secretaría Marina de México (2016), menciona que la investigación de campo permite recabar información a partir del contacto directo con el objeto de investigación, para la obtención de información empírica, y que las técnicas más comunes para dicha recolección son la observación, diario de campo, entrevista y cuestionario.

2.3 Población y muestra

En el presente proyecto de investigación se tomó como población a usuarios de la ciudad de Quito que utilizan redes sociales como Facebook, Twitter, WhatsApp e Instagram, por lo que el estudio está basado en las tendencias políticas en dichas redes sociales. Del total de la población, mediante un muestreo no probabilístico por conveniencia, que de acuerdo a, Otzen & Manterola (2017), permite seleccionar aquellos casos accesibles que acepten ser incluidos en el proceso o se encuentren disponibles en el tiempo o periodo de investigación, se seleccionó un total de 1016 personas para la aplicación de la encuesta, con la cual se pudo obtener la información necesaria para realizar el análisis requerido con la técnica de minería de datos.

2.4 Métodos empíricos y técnicas empleadas para la recolección de información

Los métodos empíricos aportan al proceso de investigación al proporcionar resultados que se fundamentan en la experiencia. Estos métodos posibilitan revelar las relaciones esenciales y las características fundamentales del objeto de investigación, accesibles a la detección sensorial, por medio de procedimientos prácticos con el objeto y diversos medios de estudio.

Entre los métodos empíricos utilizados en la presente investigación se encuentran los siguientes:

Método de Observación

La observación trata de registrar de manera precisa y sistemática, las características esenciales que explican o hacen comprender lo observado (sean hechos, acciones, situaciones, grupos o individuos, colectividades, acontecimientos, etc.), como un problema digno de investigar, (Muñoz, 2002, p. 28).

Método de Medición

Es el método empírico que se desarrolla con la finalidad de obtener información numérica acerca de una propiedad o cualidad del objeto, proceso o problema de estudio, donde se pueden comparar magnitudes medibles conocidas, es decir es la asignación de datos numéricos a determinadas propiedades del objeto, así como relaciones para evaluarlas y representarlas de forma adecuada, (Martínez & Rodríguez, 2016).

2.5 Formas de procesamiento de la información

2.5.1 Data Warehouse

Traducido como almacén de datos, se trata de una base de datos corporativos que se caracteriza por integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista. La creación de un Data Warehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

Para el desarrollo del análisis y construcción del Data Warehouse, se utilizó la Metodología Hefesto, detallada a continuación en 4 puntos importantes que son:

1. Análisis de requerimientos

- Preguntas del Negocio
- Indicadores y Perspectivas
- Modelo Conceptual

2. Análisis de data sources

- Hechos e Indicadores
- Mapeo
- Granularidad
- Modelo Conceptual Ampliado

3. Modelo lógico del DW

- Tipología
- Tablas de Dimensiones
- Tablas de Hechos
- Uniones

4. Integración de datos

- Carga Inicial

2.5.2 Análisis de requerimientos

Primero se tiene que recolectar la información del tema en análisis y de allí plantear las preguntas claves para su investigación, posterior obtener los diferentes indicadores para su análisis y de allí se fabricará en modelo Data Warehouse.

2.5.2.1 Preguntas del Negocio

Es el primer paso a dar, ya que ésta va a ser la guía de la investigación sobre el tema deseado, donde se obtendrá las necesidades de información clave de alto nivel, siendo punto esencial para determinar las metas y estrategias de la investigación, facilitando así la toma de decisiones.

2.5.2.2 Indicadores y perspectivas

Los indicadores que se utilizarán se van a extraer de las preguntas planteadas, estos generalmente son valores numéricos donde indicarán lo que se desea analizar.

Las perspectivas son los objetos que se necesitan para analizar los indicadores con la finalidad de responder a las preguntas planteadas.

2.5.2.3 Modelo Conceptual

Al obtener los indicadores y perspectivas, se puede graficar un modelo conceptual del tema planteado, donde se describirán la estructura de la base de datos, representadas a través de objetos relaciones y atributos.

2.5.3 Análisis de Data Sources

En este punto se va analizar los Data Sources para decretar la forma de construir los indicadores, indicando el mapeo y eligiendo los campos en estudio de cada perspectiva.

2.5.3.1 Mapeo

Se debe analizar los Data Source y reconocer sus características y de esa forma saber si los Data Sources contiene los datos requeridos y de esta forma establecer una comunicación directa entre elementos del Modelo Conceptual y el Data Source.

2.5.3.2 Granularidad

En este apartado se determinarán los ítems que debe tener cada perspectiva para analizar los indicadores.

2.5.4 Modelo Lógico del Data Warehouse

Primero se determinará el tipo de Modelo Lógico a emplear y posterior delinear las tablas de Dimensiones y de Hechos con las relaciones respectivas

2.5.4.1 Tipología

En esta parte seleccionamos el tipo ideal de esquema a usarse en nuestra investigación. De esto depende mucho la elaboración del Modelo Lógico.

2.5.4.2 Tablas de Dimensiones

En este punto serán realizadas las Tabla de Dimensiones que va a ser parte del Data Warehouse.

2.5.4.3 Tablas de Hechos

La tabla de hechos es una forma de representar la información, donde se le debe establecer un nombre, su clave primaria y esto debe realizarse para cada uno de los indicadores establecidos.

2.5.4.4 Uniones

Se debe realizar uniones para los esquemas tipo estrella para sus tablas de Dimensiones y de Hechos.

2.5.5 Integración de datos

En esta parte se elaborará la carga inicial de la información al Data Warehouse empleando métodos de limpieza y calidad de datos.

2.6 Metodología seleccionada

Para la elaboración del presente documento de investigación se efectuó varios pasos en su inicio, desarrollo y culminación, detallados a continuación:

1. Elección del tema a investigar.
2. Aprobación del proyecto de investigación.
3. Estructuración de a información en base a los lineamientos de investigación de la Maestría en Telemática.
4. Boceto de la problemática y realización de los objetivos, hipótesis y justificación del tema investigativo.
5. Enunciación del marco teórico.
6. Elaboramiento del marco metodológico.
7. Ejecución de la herramienta a las personas seleccionadas.
8. Análisis y discusión de los resultados.
9. Redacción de la propuesta.
10. Elaboramiento de las respectivas conclusiones y recomendaciones del tema de investigación.

Capítulo III. Propuesta

3.1 Fundamentos de la propuesta

En este capítulo se expone la propuesta planteada en razón de los lineamientos técnicos y teóricos analizados en los capítulos anteriores para el análisis de las redes sociales de mayor impacto en conjunto con los diferentes algoritmos de minería de datos, con la finalidad de ordenar la información para generar conocimiento y posteriormente tomar decisiones efectivas y precisas con lo obtenido.

Es por ende que, la presente propuesta se convierte en una herramienta de pautas y procedimientos del como realizar la recolección de información de las principales redes sociales que día a día está produciendo información y que posteriormente servirá para convertirlas en ideas y toma de decisiones.

3.2 Misión

La misión de la propuesta va encaminada al desarrollo del uso de la Minería de Datos con las personas usuarias de las principales redes sociales, partiendo desde allí administrar la información para que sea de gran ayuda para la toma de decisiones.

3.3 Visión

La visión de la propuesta es obtener la mayor cantidad de información de todas las personas de Quito por medio de las redes sociales, para saber sus necesidades, costumbres, alimentación, comportamiento, entre otros, logrando así, generar históricos de los movimientos que realizan las organizaciones de forma desarrollada y por ende obtener excelentes resultados para el mejoramiento del vivir de los seres humanos.

3.4 Justificación

Debido a que en el mundo existen diferentes tipos de personas, teniendo cada una de ellas diferentes necesidades para su buen vivir, y que todavía no hay algo inventado por la ciencia para cubrir tal necesidad, entonces se puede mencionar que, la información que obtiene la ciencia o las organizaciones no está siendo correctamente organizada, por tal motivo no se están tomando las mejores decisiones. También se debe aclarar que la mayoría de estas

personas manejan las redes sociales y se comunican con personas de otras ciudades o de otros países del mundo.

Ante esto, la presente propuesta es de gran ayuda para las organizaciones que innovan día tras día al obtener información de las necesidades de las personas, utilizando la Minería de Datos, ya que dicha información se la pueden obtener de las principales redes sociales y al analizar la información obtenida de la mejor manera se puede obtener mejor control y por supuesto se obtendrán mejores resultados, llegando a ser óptimos para las personas que la requieran.

Al plantear el uso de esta técnica es con la finalidad de manejar de manera eficiente la información obtenida de las personas que tienen mayor necesidad que otras y que se comunican mediante las redes sociales, para posterior obtener resultados satisfactorios.

3.5 Objetivos de la propuesta

3.5.1 Objetivo general

Generar conocimiento con la información obtenida de la minería de datos que se consigue de las principales redes sociales y que son usadas por organizaciones para la toma de decisiones.

3.5.2 Objetivos Específicos

- Identificar a las organizaciones que utilizan la minería de datos para la generación de ideas innovadoras para el bien común.
- Analizar las principales redes sociales utilizadas por las personas para comunicarse y que están generando información.
- Establecer el uso del DataWarehouse para el manejo y procesamiento de la información.

3.6 Presentación del proyecto

3.6.1 Presentación de la Propuesta

En este apartado, se presenta el desarrollo de la propuesta que debe ser usado por organizaciones para el procesamiento de la información obtenida desde las principales redes sociales, mismas que permitirán generar conocimiento y tomar las mejores decisiones.

Es importante describir que en dicha propuesta se considera a las redes sociales como una herramienta esencial en este tema de estudio, debido a que la mayoría de personas en el mundo la utilizan y es una forma de interactuar y comunicarse entre sí, sin importar la distancia que los separa, su cultura o su nacionalidad.

En cuanto a la minería de datos, se usará para el procesamiento de la investigación ayudada con el DataWarehouse para integrar y depurar la información y ayudará a que la toma de decisiones sea de una forma eficiente y de esta forma minimizar errores que puedan afectar en el futuro.

3.6.2 Desarrollo del Data Mining de tendencias políticas en redes sociales

3.6.2.1 Modelado del Data Warehouse

Data Warehouse traducido como almacén de datos, se trata de una base de datos corporativos que se caracteriza por integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista. La creación de un Data Warehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

Para el desarrollo y construcción de la propuesta se usó el Data Warehouse, basado en la Metodología Hefesto, la cual se detalla a continuación:

3.6.2.2 Análisis de requerimientos

En este punto, se comenzó recolectando información sobre el tema que se desea analizar y se obtienen las preguntas claves sobre la investigación. Después se requiere identificar los indicadores resultantes de los interrogativos y sus respectivas perspectivas de análisis, mediante las cuales se construirá el modelo conceptual de datos del Data Warehouse.

Identificar preguntas

El análisis de los requerimientos es el punto de partida de esta metodología, proporcionando una guía para la investigación, encaminando está a un desarrollo que refleje claramente lo que se espera de los datos, en relación a sus funciones y cualidades.

El objetivo principal de esta fase, es la de obtener e identificar las necesidades de información clave de alto nivel, que es esencial para llevar a cabo las metas y estrategias de la investigación, y que facilitará una eficaz y eficiente toma de decisiones.

Con la información que se posee sobre tendencia política en redes sociales, se plantean las siguientes preguntas:

- ¿Cuál es la tendencia Política en Redes Sociales?
- ¿Cuáles son los partidos políticos más representativos?
- ¿Cómo establecer la cantidad de personas por ideología política?
- ¿Cuál es la cantidad de personas afiliadas a partidos políticos?
- ¿Cuál es la actividad en redes sociales de las personas en temas políticos?
- ¿Cuáles son las redes sociales más usadas con fines Políticos?
- ¿Qué personaje político tiene mayor proyección en futuras elecciones presidenciales?
- ¿Qué tiempo emplean las personas en redes sociales para fines políticos?

Identificar los indicadores y perspectivas

Una vez que se han establecido las preguntas de la investigación, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

Para ello, se debe tener en cuenta que los indicadores, para que sean realmente efectivos son en general valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas, etc.

En cambio, las perspectivas se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros, etc. Cabe destacar, que el Tiempo es muy comúnmente una perspectiva.

Los indicadores son:

- ✓ Cantidad total de personas participantes en el análisis.
- ✓ Cantidad de tiempo empleado en actividades políticas en redes sociales.

Las perspectivas de análisis son:

- ✓ Personas
- ✓ Partidos políticos
- ✓ Redes Sociales
- ✓ Actividad red social
- ✓ Proyección futuras elecciones
- ✓ Tendencia Política
- ✓ Ideología política
- ✓ Afiliación partido político
- ✓ Tiempo

Modelo Conceptual

En esta etapa, se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas en el paso anterior.

El Modelo Conceptual es una descripción de alto nivel de la estructura de la base de datos, en la cual la información es representada a través de objetos, relaciones y atributos.

A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, además al poseer un alto nivel de definición de los datos, permite que pueda ser presentado ante los usuarios y poder explicado con facilidad.

En la figura 7, se puede ver la representación gráfica del modelo conceptual que es la siguiente:

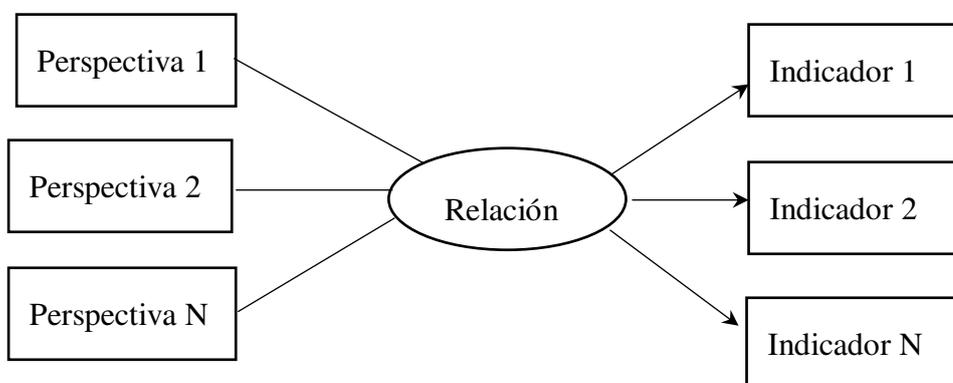


Figura 7. Modelo conceptual

A la izquierda se colocan las perspectivas seleccionadas, que serán unidas a un óvalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación, constituye el proceso o área de estudio elegida. De dicha relación y entrelazadas con flechas, se desprenden los indicadores, estos se ubican a la derecha del esquema.

Como puede apreciarse en la figura anterior, el modelo conceptual permite de un solo vistazo y sin poseer demasiados conocimientos previos, comprender cuáles serán los resultados que se obtendrán, cuáles serán las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos, como se puede ver en la figura 8.

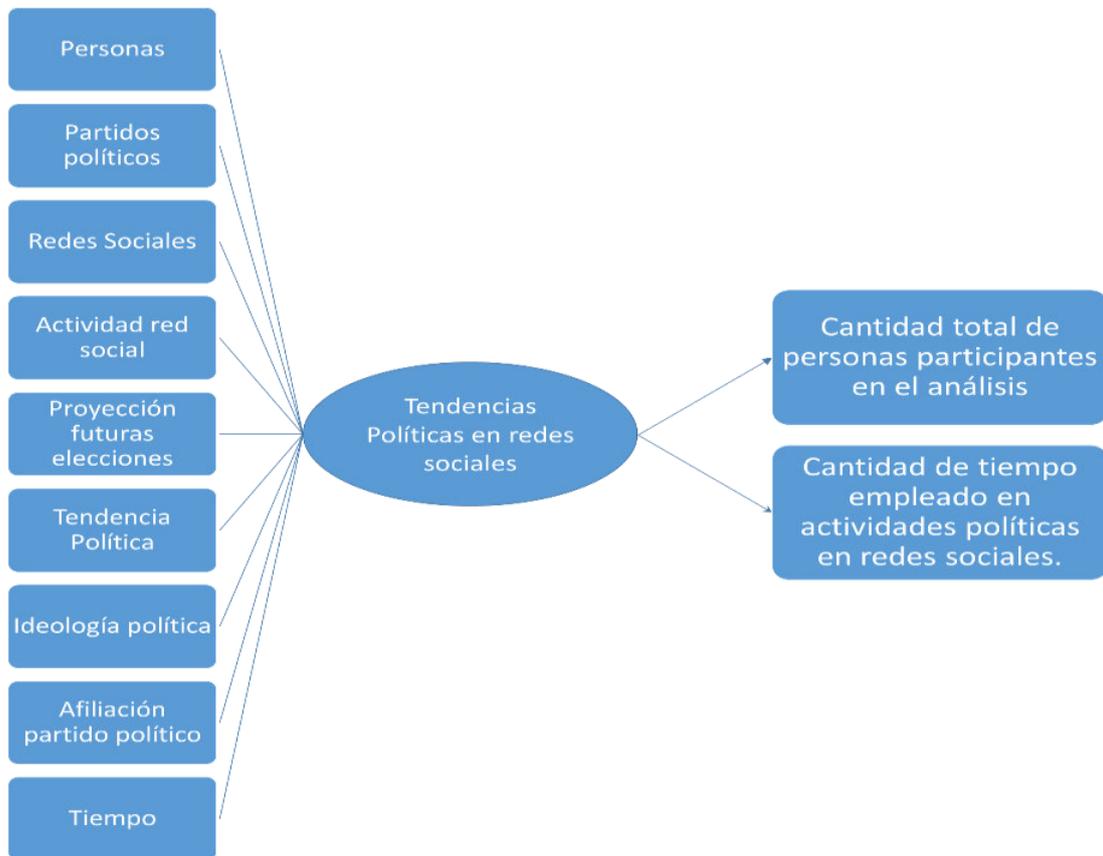


Figura 8. Aplicación del modelo conceptual

3.6.2.3 Análisis de data sources

Después, se analizarán los Data Sources en pos de determinar cómo se construirán los Indicadores, señalando el mapeo correspondiente y seleccionando los campos de estudio de cada perspectiva.

a) Conformar indicadores

En este paso se explica cómo se calculan los indicadores, definiendo los siguientes conceptos para cada uno de ellos.

Los indicadores se calcularán de la siguiente manera:

“Cantidad de Personas participantes en el análisis”

- **Hechos:** Cantidad de Personas participantes en el análisis.
- **Función de sumarización:** SUM.
- **Aclaración:** el indicador “Cantidad de Personas participantes en el análisis” representa la sumatoria de las diferentes personas que participaron en el análisis de tendencias políticas en redes sociales.

“Cantidad de tiempo empleado en actividades políticas en redes sociales”

- **Hechos:** Cantidad de tiempo empleado en actividades políticas en redes sociales.
- **Función de sumarización:** SUM.
- **Aclaración:** el indicador “Cantidad de tiempo empleado en actividades políticas en redes sociales” representa la sumatoria del tiempo que las personas participantes en el análisis emplean en redes sociales.

b) Establecer Correspondencias

En este paso se debe examinar los Data Sources e identificar sus características propias, y asegurar que los Data Sources disponibles contengan los datos requeridos.

Luego, se debe establecer cómo serán obtenidos los elementos que se han definido en el Modelo Conceptual, estableciendo de esta manera una correspondencia directa entre elementos del Modelo Conceptual y Data Sources.

c) Nivel de granularidad

Una vez que se han establecido el Mapeo con los Data Sources, se deben seleccionar los campos que contendrá cada Perspectiva, ya que a través de estos se analizarán los Indicadores.

Con respecto a la **Perspectiva Persona**, los datos disponibles son los siguientes:

- ID_persona: es la clave primaria de la tabla Persona, y representa unívocamente a una persona en particular.
- sexo: representa el sexo de una persona.
- edad: representa la edad de una persona.
- etnia: representa la edad de una persona.
- provincia_zona_3: representa la provincia de procedencia de una persona ubicada en la zona 3 según la división política de la república del Ecuador.
- area: área define si el lugar de residencia esta designado como zona urbana o rural.
- estatus_economico: representa el estado financiero de una persona.
- ideologia_politica: representa la ideología política de una persona.
- afiliacion_politica: establece si una persona pertenece o no a un partido político.
- ingreso_hogar: representa la cantidad de ingresos monetarios que posee una persona.
- nivel_educacion: representa la formación académica de una persona.

En la **Perspectiva Partidos políticos**, los datos que se pueden utilizar son los siguientes:

- ID_partido_politico: es la clave primaria de la tabla Partidos Políticos, y representa unívocamente a un Partido Político en particular.
- partido_politico: indica el nombre de un partido político.
- presidente: representa el nombre del presidente de un partido político.
- eslogan: representa una frase representativa de un partido político.
- ideologia: representa la tendencia y pensamiento crítico de un partido político.
- posicion: representa la posición política al que pertenece un partido político.
- sede: representa una o varias instalaciones que posee un partido político.
- fundacion: representa el año de creación o fundación de un partido político.
- lista: representa a un partido político de forma numérica.
- siglas: representa una palabra que se forma con las letras y representa un partido político.
- asambleista: representa el número de asambleístas obtenidos por un partido político.
- prefecto: representa el número de prefecto obtenidos por un partido político.
- alcalde: representa el número de alcaldes obtenidos por un partido político.

En la **Perspectiva Red Social**, los datos que se pueden utilizar son los siguientes:

- ID_red_social: es la clave primaria de la tabla Red Social, y representa unívocamente a una Red Social en particular.
- red_social: representa a una red social en la que las personas expresan tendencias políticas.

En la **Perspectiva Tiempo**, los datos que se pueden utilizar son los siguientes:

- ID_tiempo_red: es la clave primaria de la tabla Tiempo, y representa unívocamente el tiempo que una persona utiliza una o varias Redes Sociales en temas políticos.
- hora_dia: representa el tiempo en horas que una persona utiliza una o varias Redes Sociales en temas políticos.

En la **Perspectiva Actividad Red Social**, los datos que se pueden utilizar son los siguientes:

- ID_actividad_red: es la clave primaria de la tabla Actividad Red Social, y representa unívocamente a una Actividad en Redes Sociales relacionados con temas políticos.
- grado: representa la actividad de una persona en redes sociales con fines políticos. Por ejemplo: Alto, Medio, Bajo, Ninguno.
- grado_actividad: representa la descripción del grado de actividad que una persona mantiene en redes sociales con fines políticos.
- actividad_red_1: describe actividades que se puede realizar en redes sociales con fines políticos.
- actividad_red_2: describe actividades que se puede realizar en redes sociales con fines políticos.
- actividad_red_3: describe actividades que se puede realizar en redes sociales con fines políticos.
- actividad_red_4: describe actividades que se puede realizar en redes sociales con fines políticos.
- actividad_red_5: describe actividades que se puede realizar en redes sociales con fines políticos.

En la **Perspectiva Voto 2021**, los datos que se pueden utilizar son los siguientes:

- ID_partido_politico: es la clave primaria de la tabla Voto 2021, y representa unívocamente a un candidato que representara a un Partido Político en particular en las próximas elecciones presidenciales 2021.
- candidato: representa el nombre de un candidato que participara en las próximas elecciones presidenciales 2021.
- partido_politico: indica el nombre de un partido político.

d) Modelo conceptual Ampliado

En este paso, y con el fin de graficar los resultados obtenidos en los pasos anteriores, se ampliará el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo, (Ver figura 9):

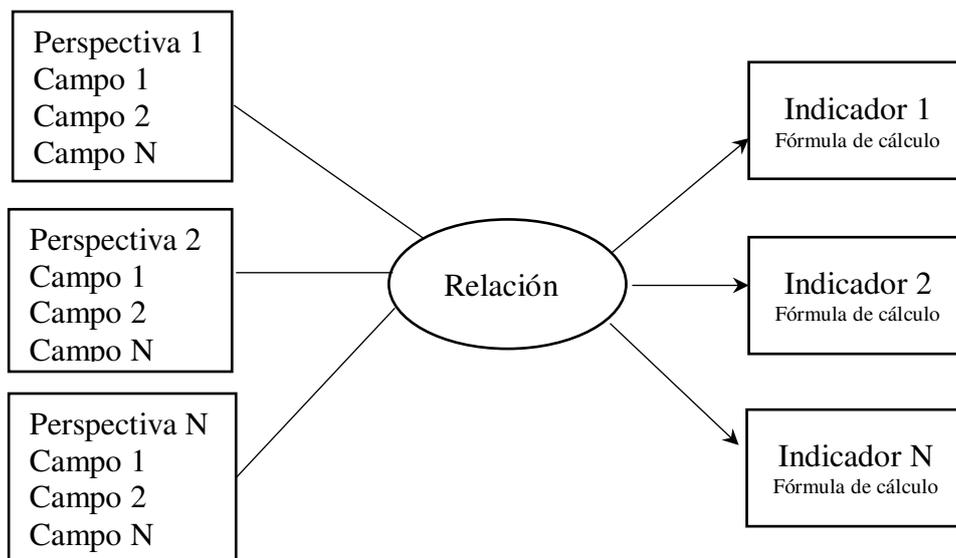


Figura 9. Modelo conceptual Ampliado

Caso práctico:

Teniendo esto en cuenta, se completa el diseño del diagrama conceptual como se puede visualizar en la figura 10:

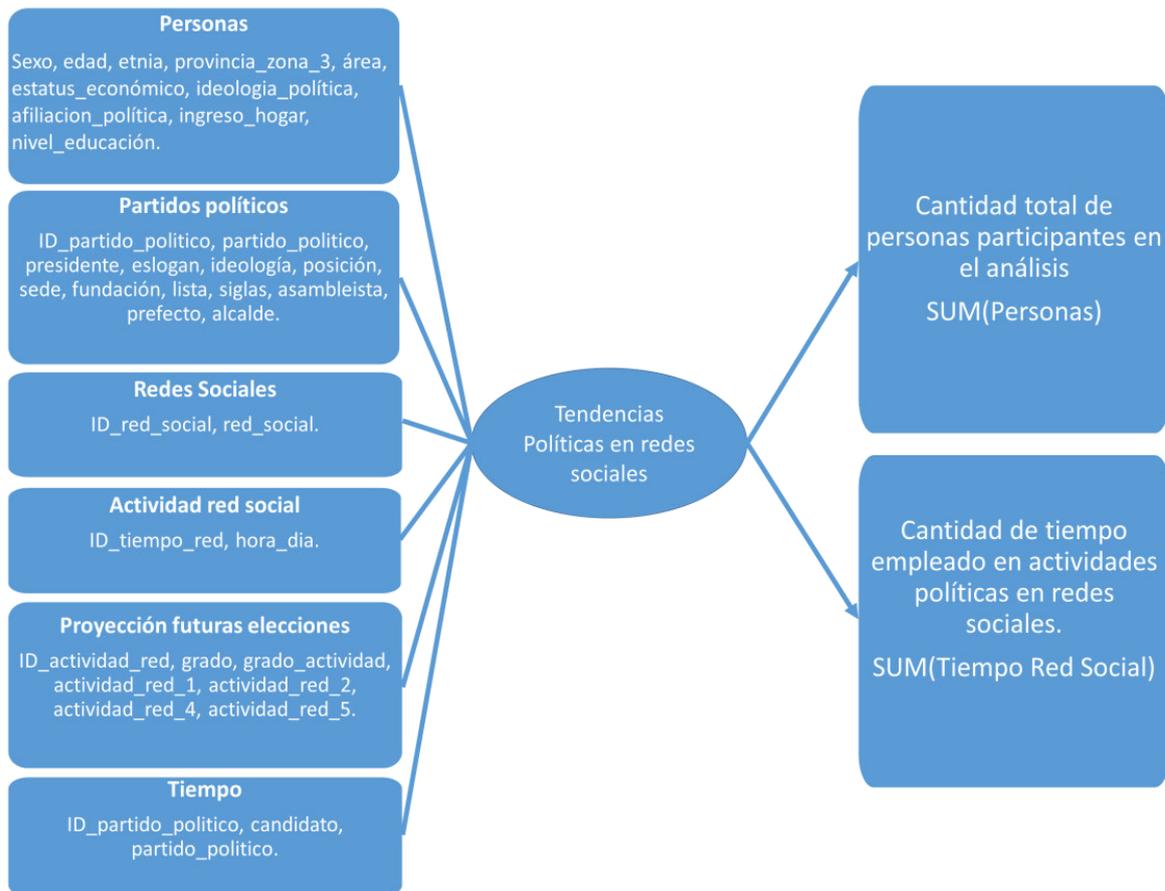


Figura 10. Diseño de diagrama conceptual

3.6.2.4 Modelo lógico del Data Warehouse

A continuación, se desarrollará el Modelo Lógico de la estructura del DW, teniendo como base el Modelo Conceptual que ya ha sido creado.

Un Modelo Lógico es la representación de una estructura de datos, que puede procesarse y almacenarse en algún SGBD.

Inicialmente, se definirá el tipo de Modelo Lógico que se utilizará y luego se diseñarán las tablas de Dimensiones y de Hechos con sus respectivas relaciones.

a) Tipo de Modelo lógico del Data Warehouse

Se debe seleccionar el tipo de Esquema que mejor se adapte a los requerimientos y necesidades de la investigación, Además es muy importante definir objetivamente si se

empleará un Esquema en Estrella, Copo de Nieve o Constelación, ya que esta decisión afectará considerablemente la elaboración del Modelo Lógico.

Se ha seleccionado el Esquema en Estrella ya que cumple con los requerimientos planteados y es simple de implementar y comprender.

b) Tablas de Dimensiones

En este paso se diseñan las tablas de Dimensiones que formarán parte del Data Warehouse.

Para los tres tipos de esquemas, cada perspectiva definida en el Modelo Conceptual se constituirá en una Tabla de Dimensión. Para ello, a partir de cada perspectiva y sus campos deben realizarse el siguiente proceso (Ver figura 11):

- Se elegirá un nombre que identifique la tabla de Dimensión.
- Se añadirá un campo que represente su clave principal.
- Se redefinirán los nombres de los campos si es que no son lo suficientemente intuitivos.

Dim_Persona	
# <u>ID_persona</u>	<u>Integer</u>
o sexo	Variable characters (50)
o edad	Integer
o etnia	Variable characters (50)
o provincia_zona_3	Variable characters (50)
o area	Variable characters (50)
o estatus_economico	Variable characters (50)
o ideologia_politica	Variable characters (50)
o afiliacion_politica	Variable characters (50)
o ingreso_hogar	Variable characters (50)
o nivel_educacion	Variable characters (50)

a) Dimensión para Personas

Dim_Partido_Politico	
# <u>ID_partido_politico</u>	<u>Integer</u>
o partido_politico	Variable characters (50)
o presidente	Variable characters (50)
o eslogan	Variable characters (50)
o ideologia	Variable characters (80)
o posicion	Variable characters (50)
o sede	Variable characters (50)
o fundacion	Integer
o lista	Integer
o siglas	Variable characters (50)
o asambleista	Integer
o prefecto	Integer
o alcalde	Integer

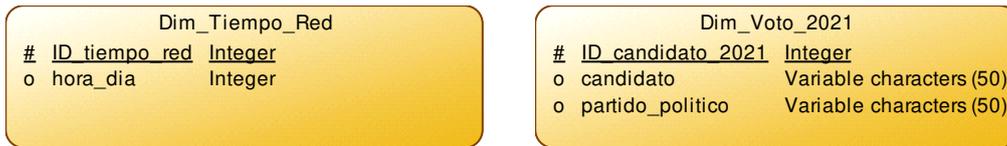
b) Dimensión para Partido Político

Dim_Actividad_Red	
# <u>ID_actividad_red</u>	<u>Integer</u>
o grado	Variable characters (50)
o grado_actividad	Variable characters (80)
o actividad_red_1	Variable characters (80)
o actividad_red_2	Variable characters (80)
o actividad_red_3	Variable characters (80)
o actividad_red_4	Variable characters (80)
o actividad_red_5	Variable characters (80)

c) Dimensión Actividad Red

Dim_Red_Social	
# <u>ID_red_social</u>	<u>Integer</u>
o red_social	Variable characters (50)

d) Dimensión Red Social



e) Dimensión Tiempo Red

f) Dimensión Voto 2021

Figura 11. Tablas de Dimensiones

c) Tablas de Hechos

Para los Esquemas en Estrella, se realizará lo siguiente (Ver figura 12):

- Se deberá asignar un nombre a la tabla de Hechos que represente la información que contiene, área de investigación, negocio enfocado, etc.
- Se definirá su clave primaria, que se compone de la combinación de las claves primarias de cada tabla de Dimensión relacionada.
- Se crearán tantos campos de Hechos como Indicadores se hayan definido en el modelo conceptual y se les asignará un nombre.



Figura 12. Tabla de hechos factor Tendencia Política

d) Uniones

Para los esquemas tipo estrella (Ver figura 13), se realizarán las uniones correspondientes entre sus tablas de Dimensiones y sus tablas de Hechos.

Se realizarán las uniones necesarias según corresponda:

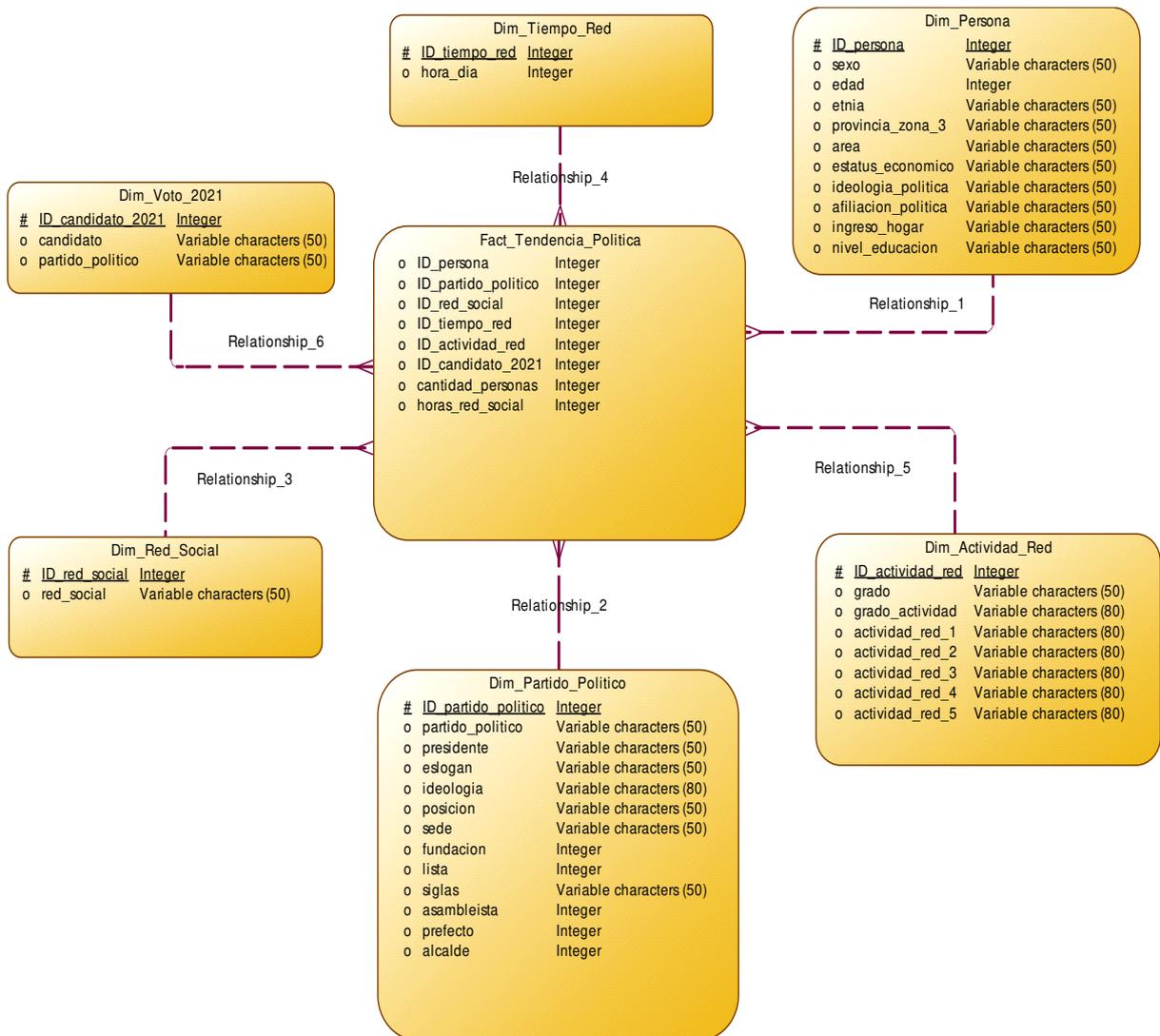


Figura 13. Unión Tabla de dimensión con tabla de hechos

3.6.2.5 Integración de datos

Por último, utilizando técnicas de limpieza, calidad de datos y procesos ETL, se realizó la carga inicial de información al Data Warehouse.

En este paso se debe realizar la Carga Inicial del DW, poblando el modelo construido en pasos anteriores. Para lo cual se debe llevar adelante una serie de tareas básicas, tales como asegurar la limpieza y calidad de los datos, procesos ETL, etc.

En muchos casos, las tareas antes mencionadas tienen una lógica compleja. Afortunadamente, en la actualidad existen muchas herramientas de software que se pueden emplear y que facilitan en gran parte el trabajo.

A continuación, se puede visualizar en la figura 14, la creación de la Base de Datos a partir del Diagrama Físico del modelado.

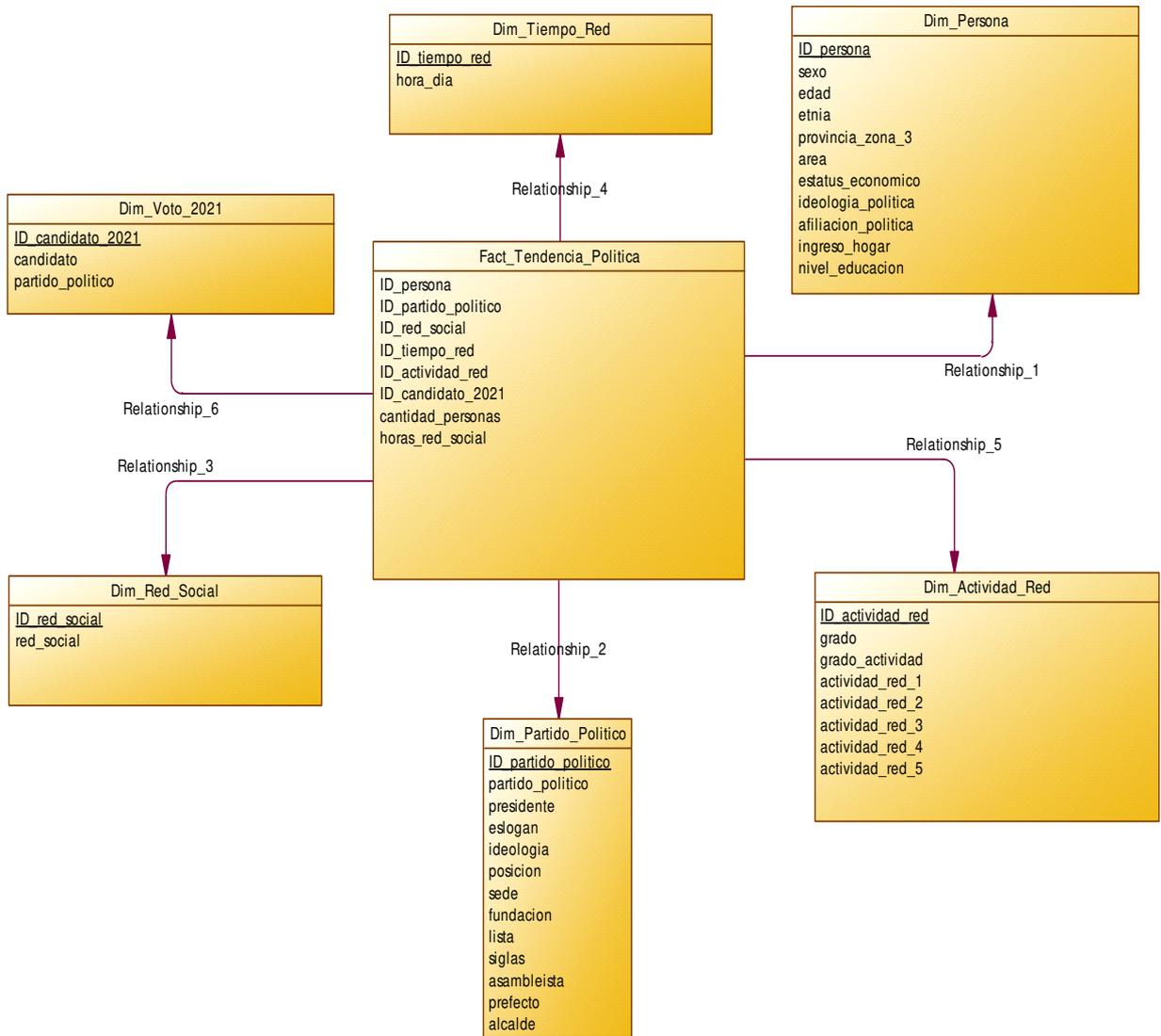


Figura 14. Diagrama físico del modelado.

Una vez conformado el diagrama físico, se procede a exportarlo a una base de datos. Para esta investigación se establece como BD principal PostgreSQL, misma que se puede ver en la figura 15.

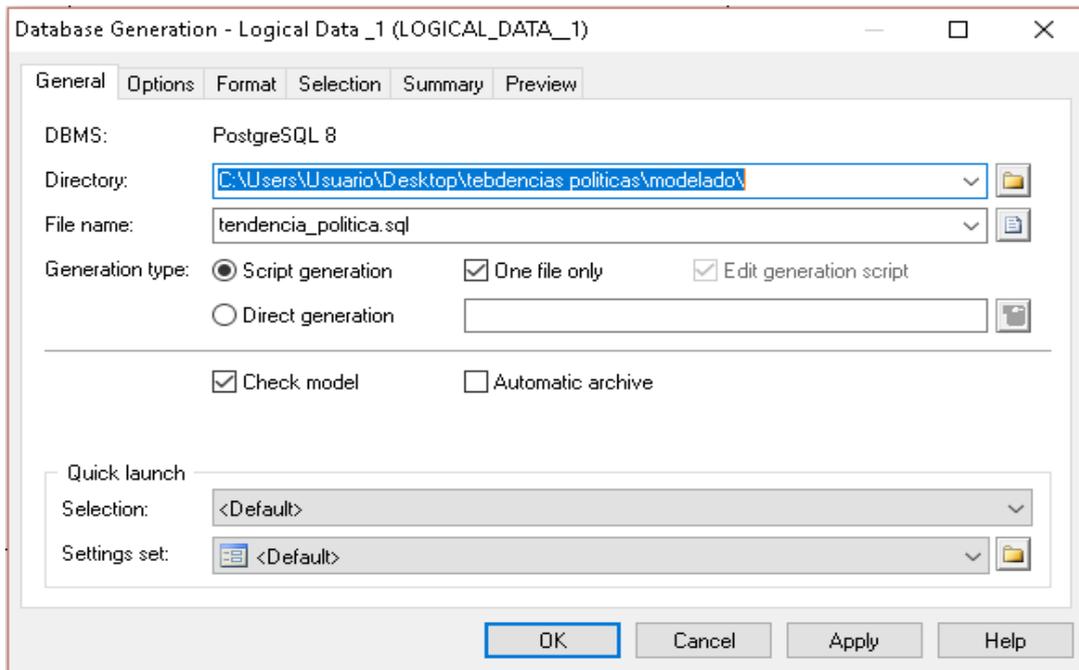


Figura 15. Exportación del Diagrama Físico

En la figura 16 y 17, se puede visualizar previamente el código y el Script generado, así como también las tablas correspondientes al diseño a implementar.

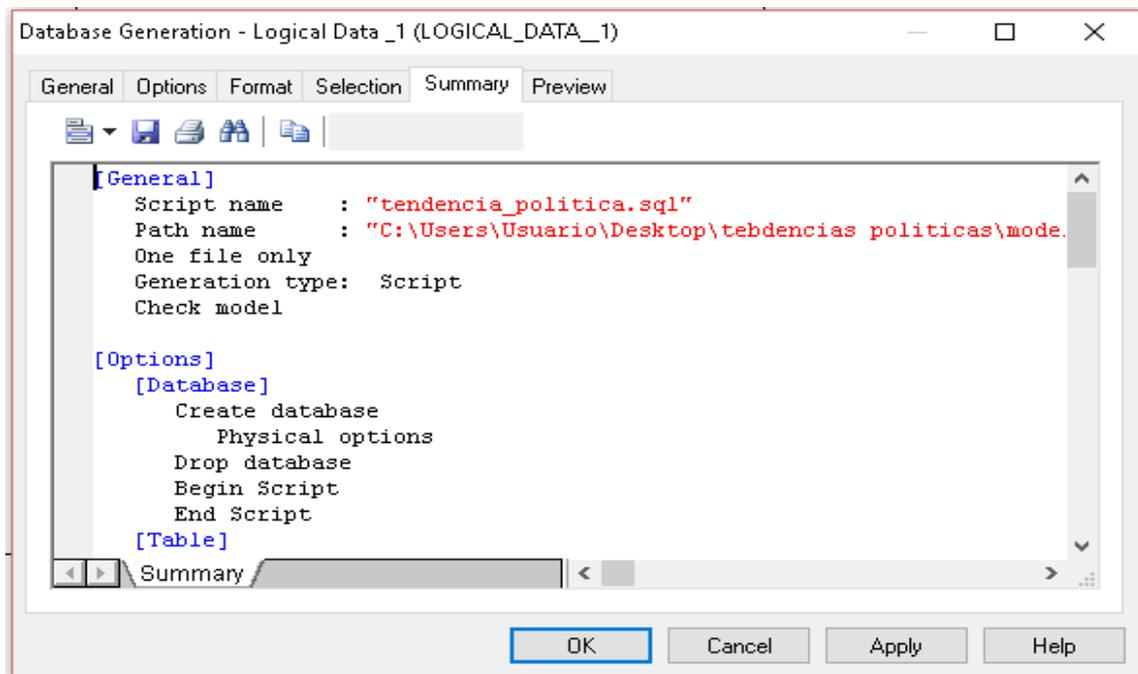


Figura 16. Código script a generar

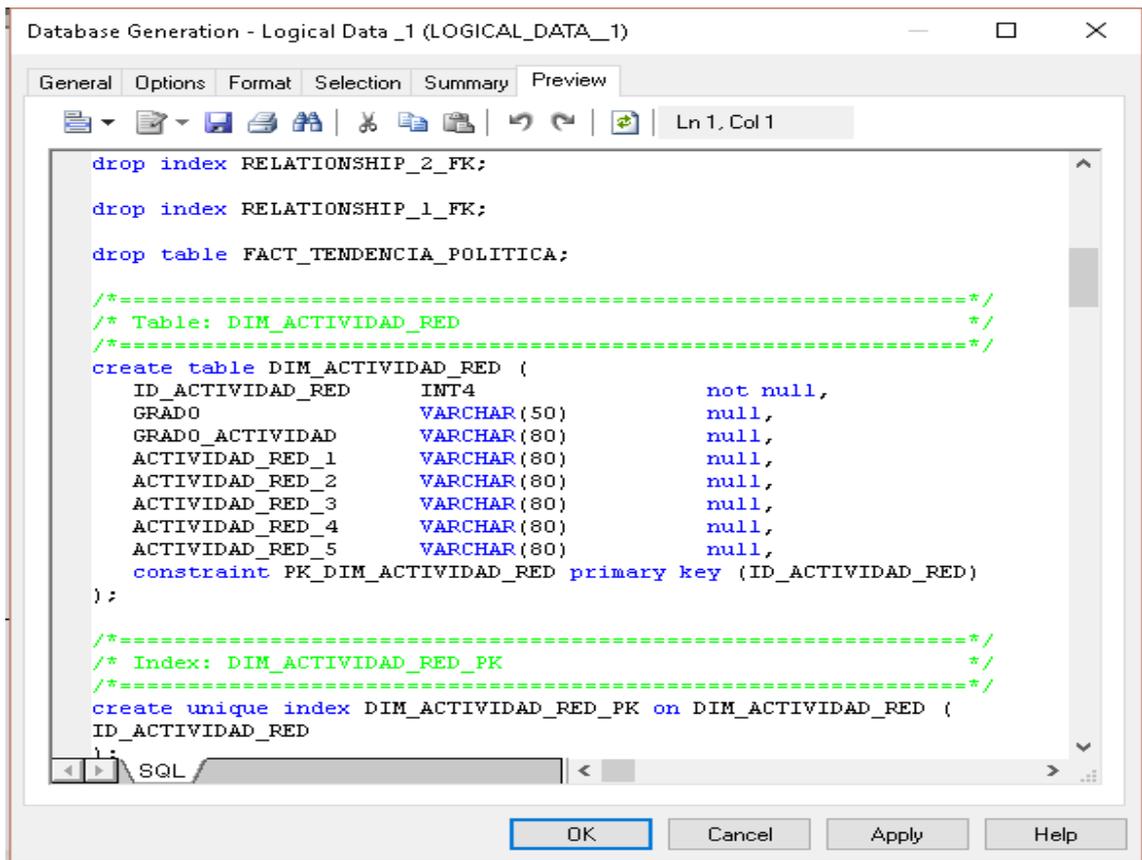


Figura 17. Tabla script generada

Una vez generado el Script, se procede a la creación de la base de datos en PostgreSQL, como se puede ver en la figura 18.

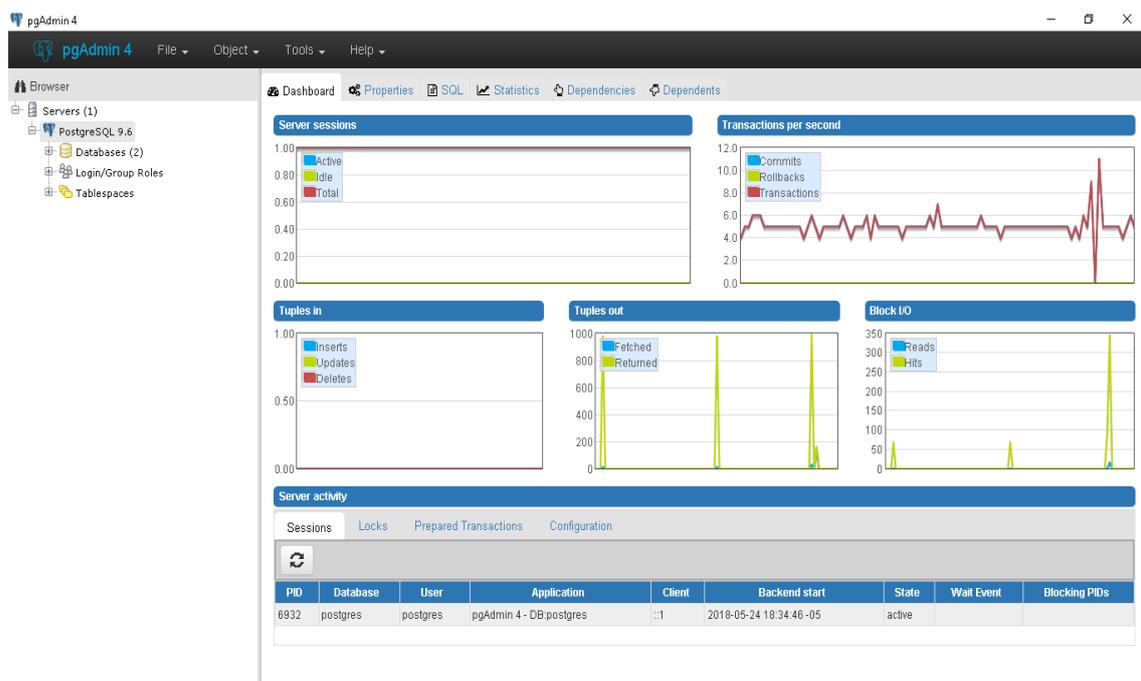


Figura 18. Creación de la base de datos

En la figura 19, se muestra el procedimiento para crear una nueva base de datos en PostgreSQL, la cual se le otorgó el nombre de “DWH_TENDENCIA” a la base de datos a la cual se carga en el diseño desarrollado.

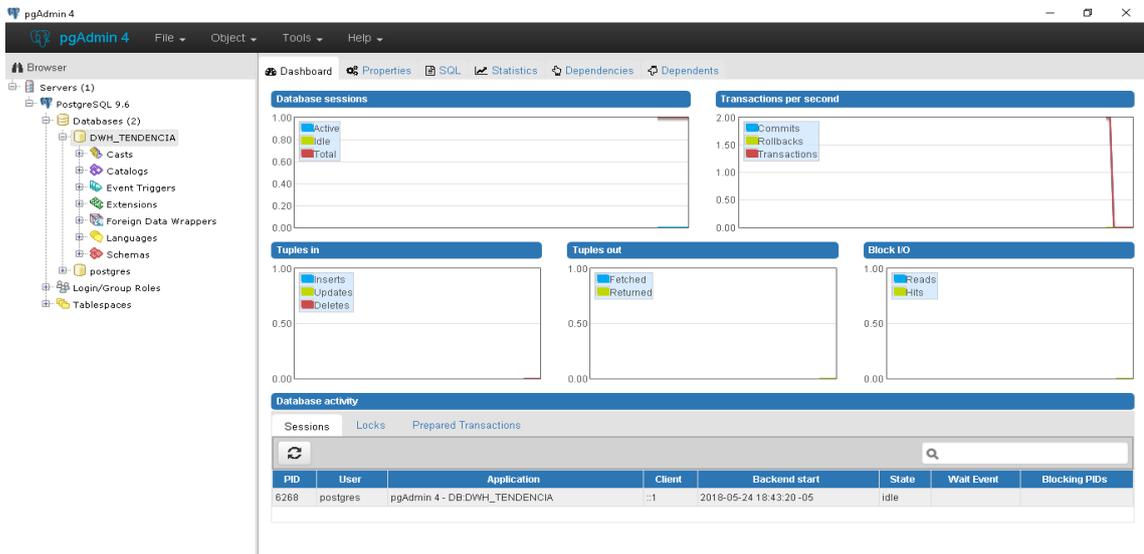


Figura 19. Base de datos creada

Se procede a cargar el Script que contiene el modelado y una vez abierto el Script es necesario ejecutarlo. Después de la carga se debe apreciar el mensaje que indica carga satisfactoria, por ello en la figura 20, se puede apreciar la creación de las diferentes dimensiones y la tabla de hechos.

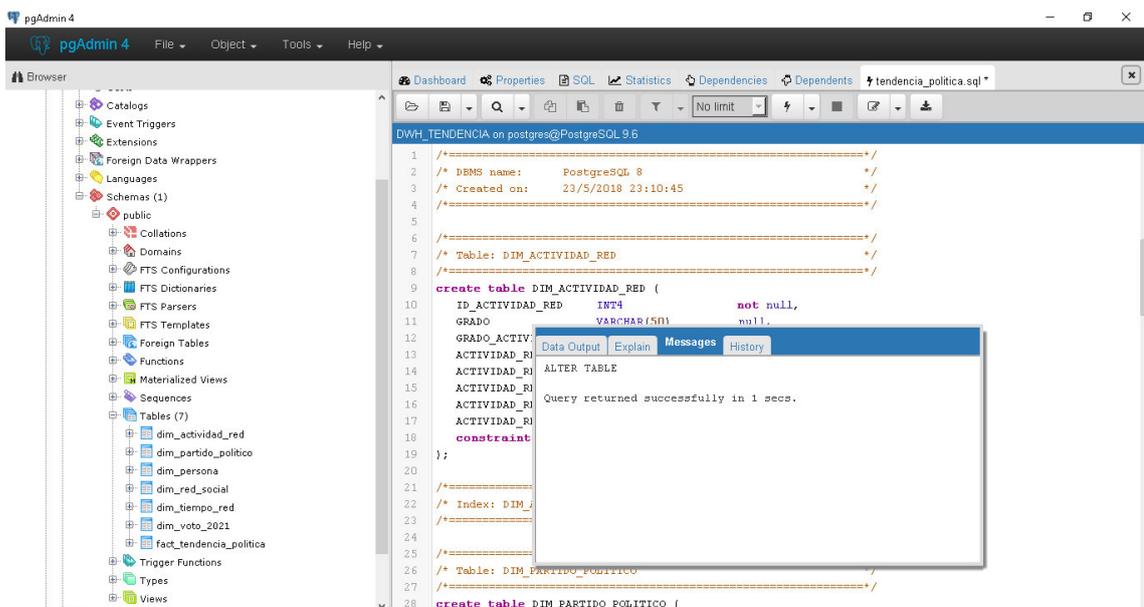


Figura 20. Carga Satisfactoria

En la figura 21, se puede apreciar que se han cargado los diferentes parámetros y relaciones que contienen las tablas de dimensiones y de hechos. Sin embargo, es necesario cargar la información a la base de datos correspondiente a tendencias políticas en redes sociales.

En este paso se debe realizar la Carga Inicial al DataWarehouse, poblando el modelo de datos que se ha construido anteriormente. Para lo cual se debe llevar adelante una serie de tareas básicas, tales como limpieza de datos, calidad de datos, procesos ETL, etc.

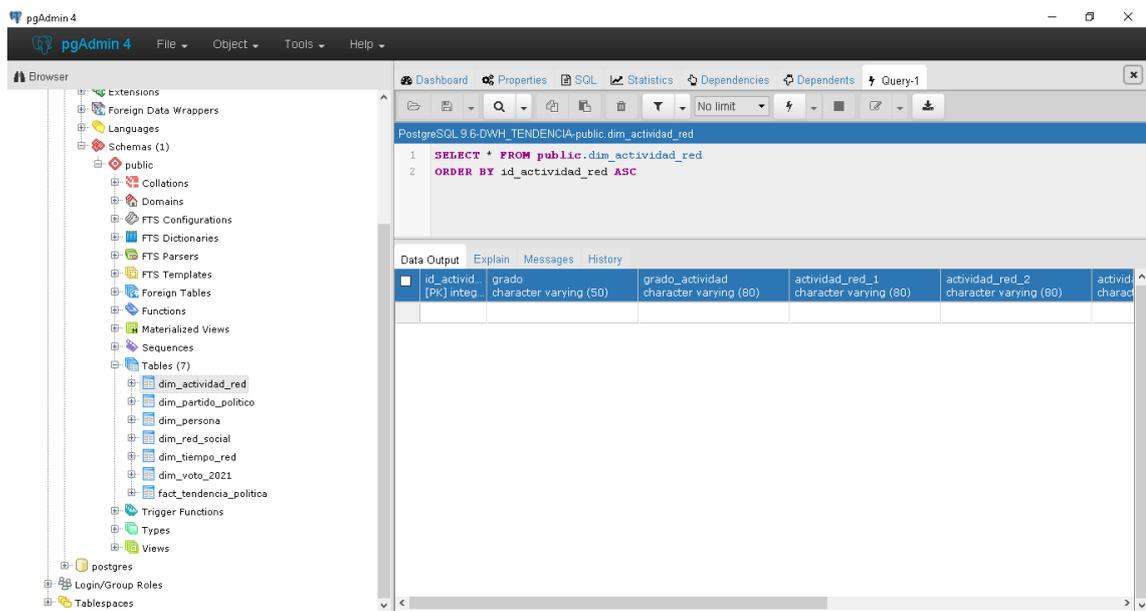


Figura 21. Carga inicial al Data Warehouse

3.6.2.6 Creación del proceso Data mining – ETL (Extraer, transformar y cargar)

Extraer, transformar y cargar (ETL) es un conducto de datos utilizado para recopilar datos de diversas fuentes, transformar los datos de acuerdo con las reglas comerciales y cargarlos en un almacén de datos de destino. El trabajo de transformación en ETL se lleva a cabo en un motor especializado, y a menudo implica el uso de tablas intermedias para mantener temporalmente los datos a medida que se transforman y, finalmente, se cargan en su destino (Ver figura 22).

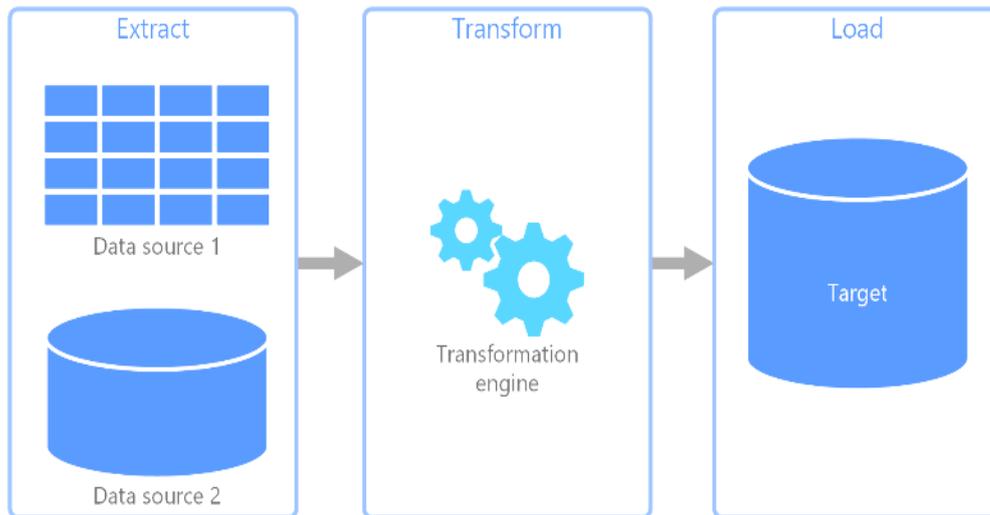


Figura 22. Gráfica del proceso ETL

a) Extraer

Mediante la realización de encuestas virtuales se obtuvo una gran cantidad de información relacionada con tendencias políticas en redes sociales en el Ecuador, logrando obtener el consolidado de tendencias políticas en redes sociales en una matriz de Excel, la cual servirá para proveer toda la información a extraer.

Una vez que se ha adquirido una gran cantidad de información se requiere realizar una interpretación y limpieza de la información, para esto se utilizaron funciones de Excel que permite conocer la longitud máxima de una cadena de carácter, tipo de dato y cantidad total de registros, logrando así analizar la calidad de datos obtenida.

Posteriormente se debe encontrar patrones, valores faltantes, conjuntos de caracteres y otras características de los datos analizados, con el fin de generar una estructura adecuada para la extracción de la información.

A continuación, en la figura 23, se puede apreciar la estructuración de la información a extraer para cada una de las dimensiones y tabla de hechos según el modelado realizado:

Dim_Persona

	A	B	C	D	E	F	G	H	I	J
1	ID_persona	sexo	edad	etnia	provincia_zona_3	area	estatus_economico	ideologia_politica	afiliacion_politica	ingreso_hogar
2	1	Masculino	16	Mestizo/a	Cotopaxi	urbana	medio-bajo	ninguna	si	medio-bajo
3	2	Femenino	65	Mestizo/a	Chimborazo	rural	medio-bajo	Izquierda	no	medio-bajo
4	3	Femenino	50	Blanco/a	Tungurahua	rural	alto	ninguna	no	alto
5	4	Femenino	53	Indigena	Pastaza	urbana	medio	Centro-derecha	si	medio
6	5	Femenino	18	Mestizo/a	Tungurahua	rural	medio-bajo	Centro-Izquierda	si	medio-bajo
7	6	Femenino	18	Otro/a	Tungurahua	urbana	bajo	ninguna	no	bajo
8	7	Masculino	58	Blanco/a	Cotopaxi	rural	medio-alto	Izquierda	no	medio-alto
9	8	Femenino	19	Mestizo/a	Pastaza	urbana	medio-bajo	Centro-derecha	si	medio-bajo
10	9	Femenino	33	Mestizo/a	Chimborazo	rural	alto	Centro-derecha	no	alto
11	10	Masculino	21	Blanco/a	Cotopaxi	urbana	alto	Centro-derecha	no	alto
12	11	Femenino	45	Mestizo/a	Tungurahua	rural	alto	Derecha	si	alto
13	12	Masculino	30	Indigena	Pastaza	rural	medio	Centro-Izquierda	no	medio
14	13	Masculino	34	Indigena	Tungurahua	rural	alto	Centro	no	alto
15	14	Masculino	46	Mestizo/a	Pastaza	rural	medio-bajo	Centro-Izquierda	si	medio-bajo
16	15	Femenino	56	Indigena	Chimborazo	rural	alto	Centro-derecha	no	alto
17	16	Femenino	25	Indigena	Tungurahua	urbana	medio	Centro	no	medio

Figura 23. Estructuración de la información a extraer

b) Transformaciones

Spoon es el diseñador gráfico de transformaciones y trabajos del sistema de ETLs de Pentaho Data Integration (PDI), también conocido como Kettle (acrónimo recursivo: "Kettle Extraction, Transformation, Transportation, and Load Environment "). Está diseñado para ayudar en los procesos ETLs, que incluyen la Extracción, Transformación, Transporte y Carga de datos.

Spoon es una Interfaz Gráfica de Usuario (GUI), que permite diseñar transformaciones y trabajos que se pueden ejecutar con las herramientas de Kettle (Pan y Kitchen).

Para ingresar algún tipo de información se utiliza las herramientas contenidas en la carpeta "Input", como se puede visualizar en la figura 24.

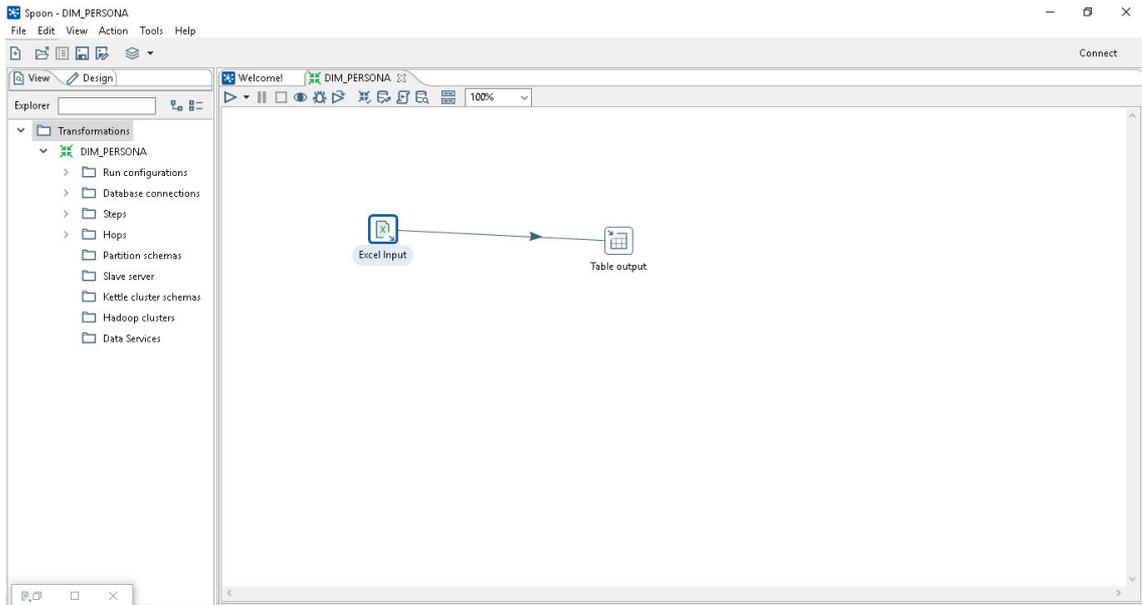


Figura 24. Transformaciones y trabajos del sistema de ETL

Debido al modelado y la estructura de la información, se utilizar un “Excel Input” para obtener la información, además es necesario señalar la ubicación del archivo y agregarlo, (Ver figura 25).

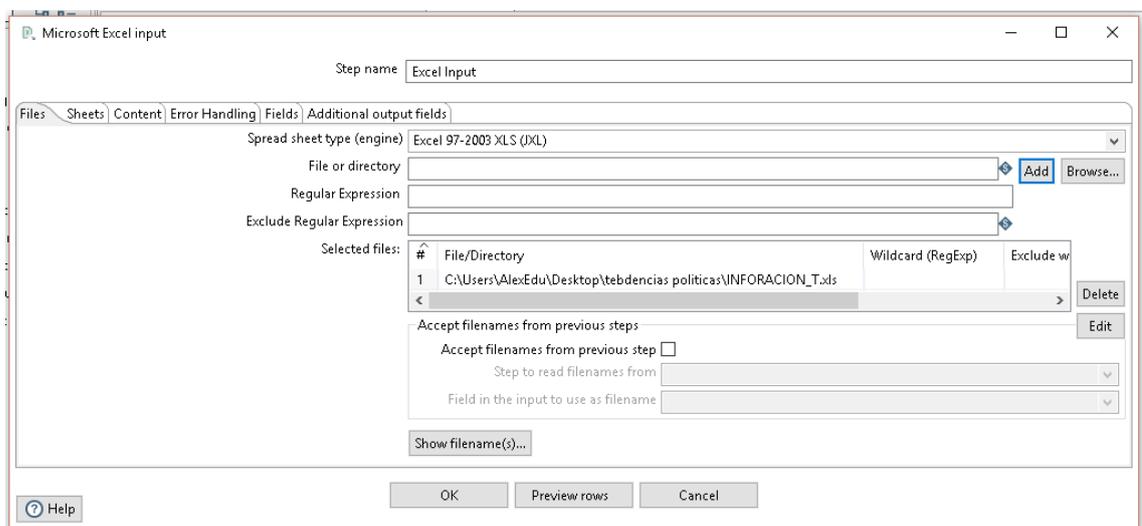


Figura 25. Apertura de Microsoft Excel Input

Una vez agregado el libro de Excel, es necesario seleccionar el libro que contiene la información que se requiera, para esto se presiona “Gets sheetname(s)”, como se puede ver en la figura 26.

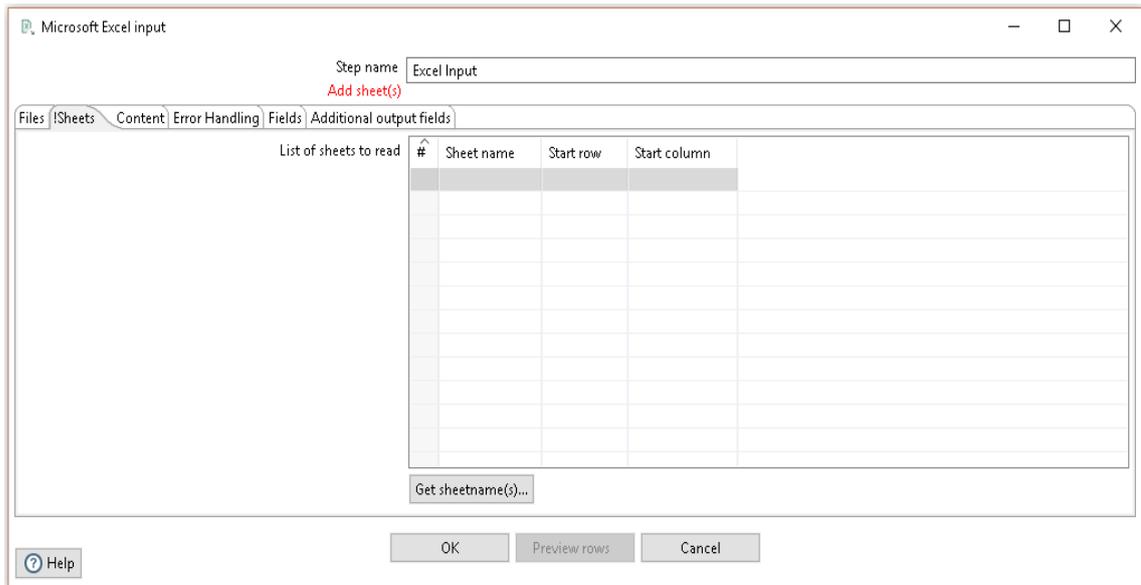


Figura 26. Carga de libros de información

En la figura 27, se puede identificar los diferentes libros contenidos, seleccionamos el más adecuado a la transformación que se va a implementar.

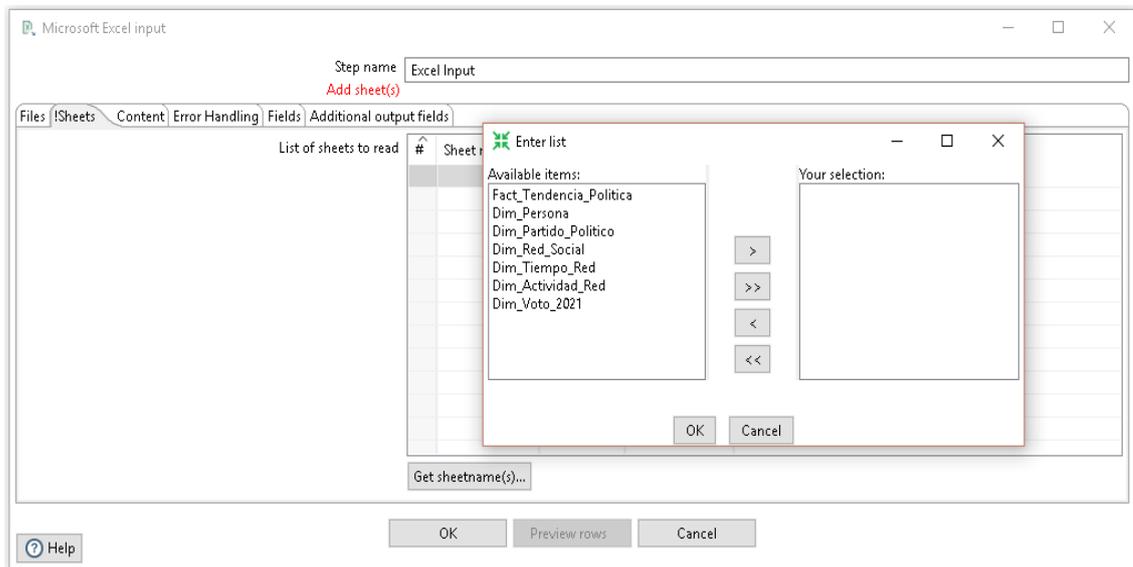


Figura 27. Selección de la información

La figura 28, presenta un ejemplo que está dirigido a la creación de la transformación del Dim_Persona.

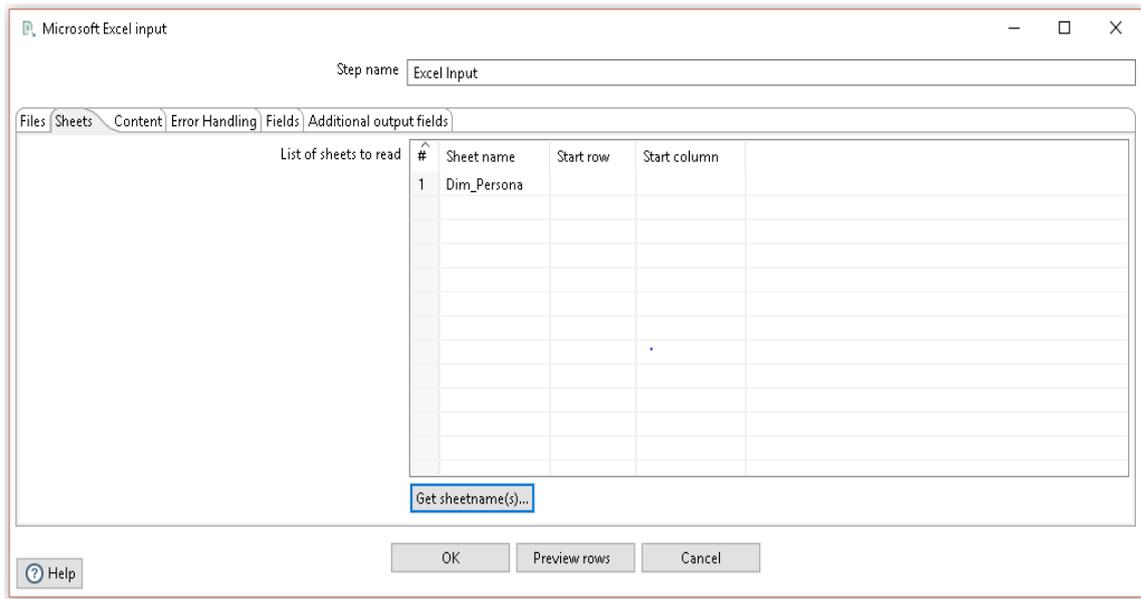


Figura 28. Selección - Dim Persona

Para poder visualizar la información y extraer la misma se presiona “Get fields from header row...”, como se ve en la figura 29.

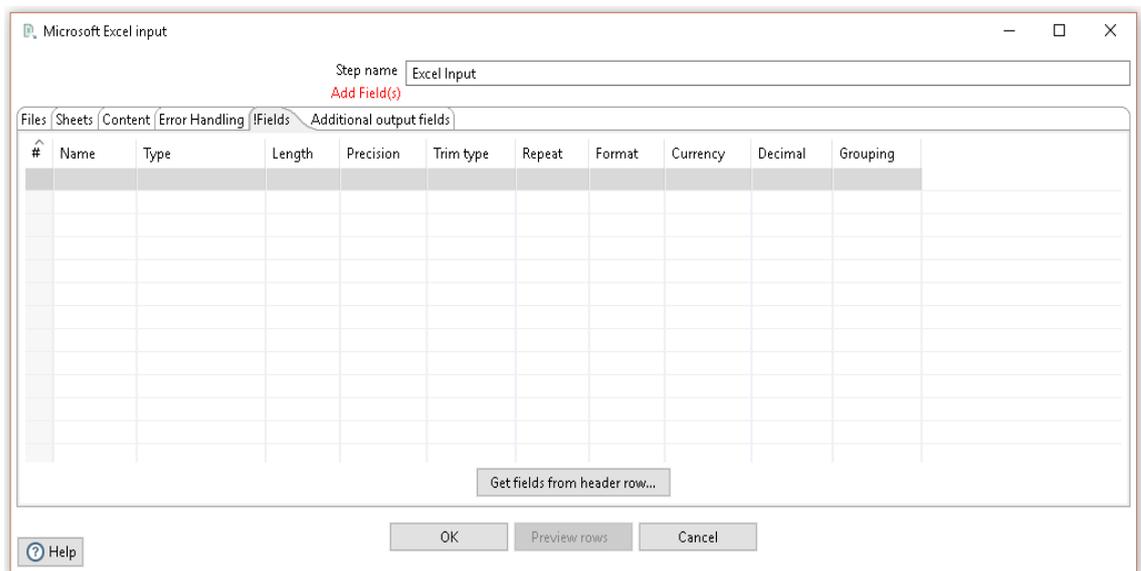


Figura 29. Extracción de la información Dim persona

En la figura 30, se puede apreciar toda la información contenida en el libro Dim_Persona, misma que se visualiza a continuación.

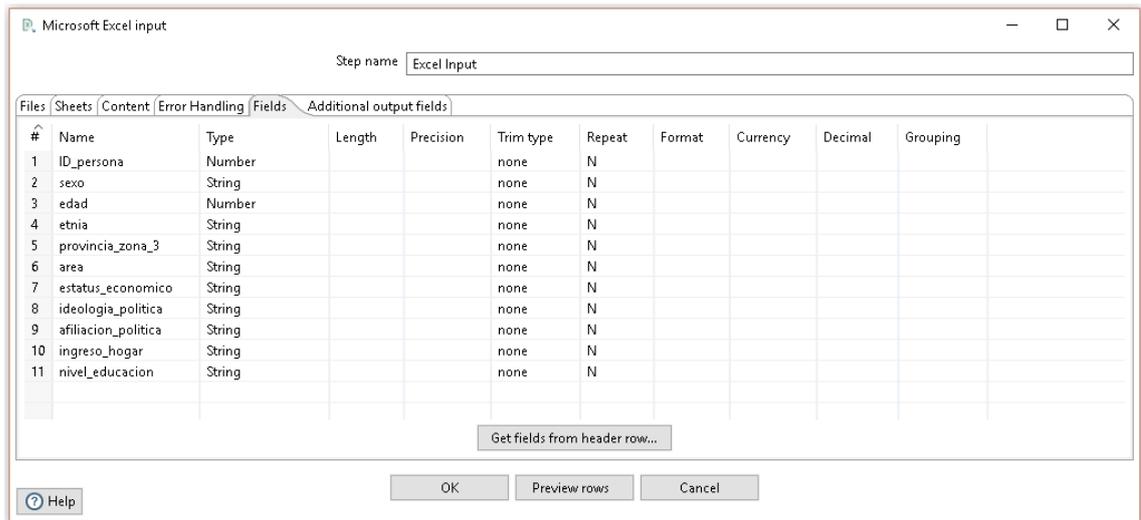


Figura 30. Información Dim persona

Es necesario identificar el tipo de dato de cada parámetro a extraer y de ser necesario colocar el más adecuado como se puede ver en la figura 31.

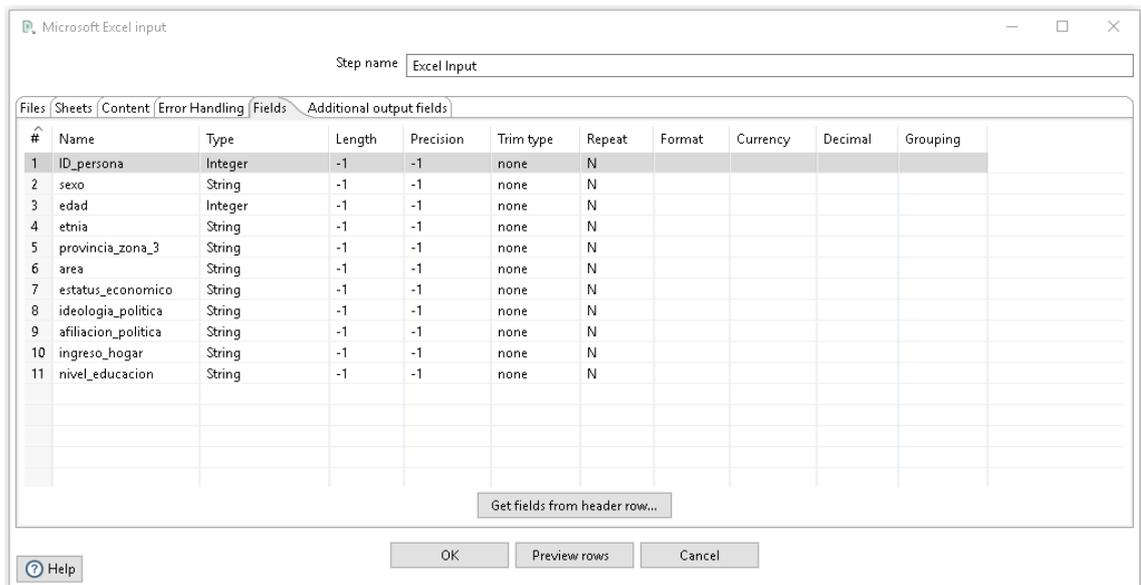


Figura 31. Identificación de datos

Una vez configurados todos los parámetros de ingreso de información, es necesario proporcionar un almacén a los datos obtenidos, para esto se utiliza “table output”.

Esta función toma la información obtenida y la almacena en la base de datos creada anteriormente, (Ver figura 32).

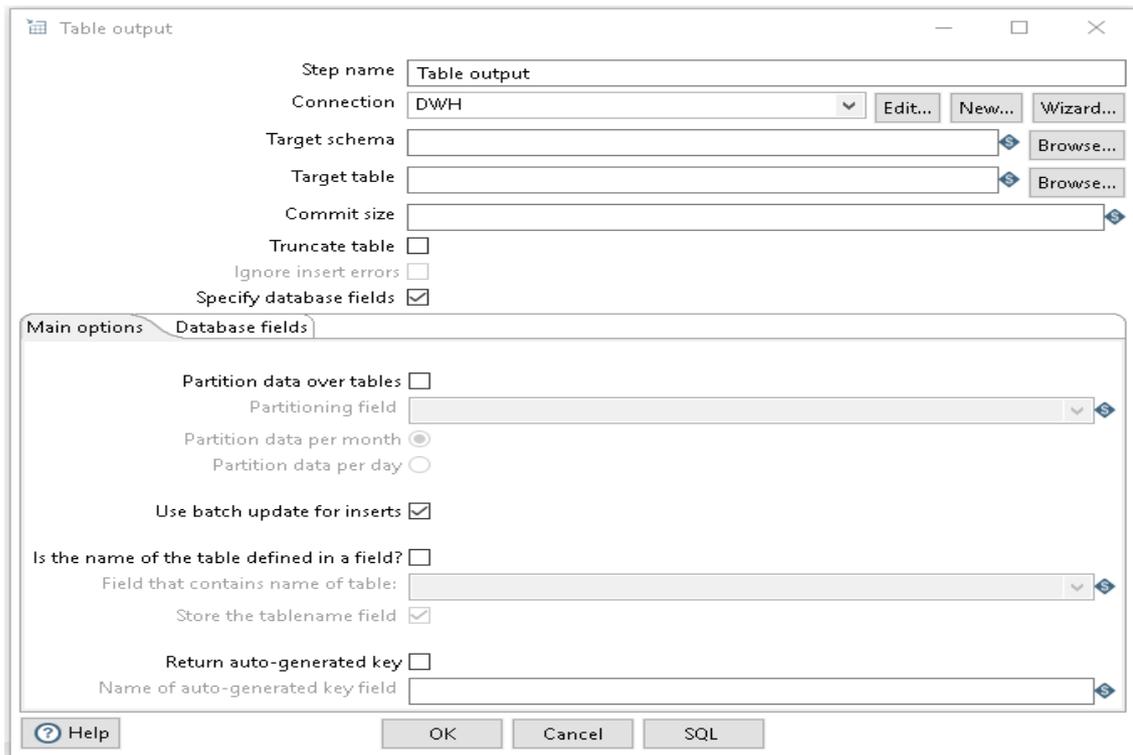


Figura 32. Table output

Es importante configurar los parámetros necesarios para establecer la conexión entre “table output” y la base de Datos denominada DWH_TENDENCIA, (Ver figura 33).

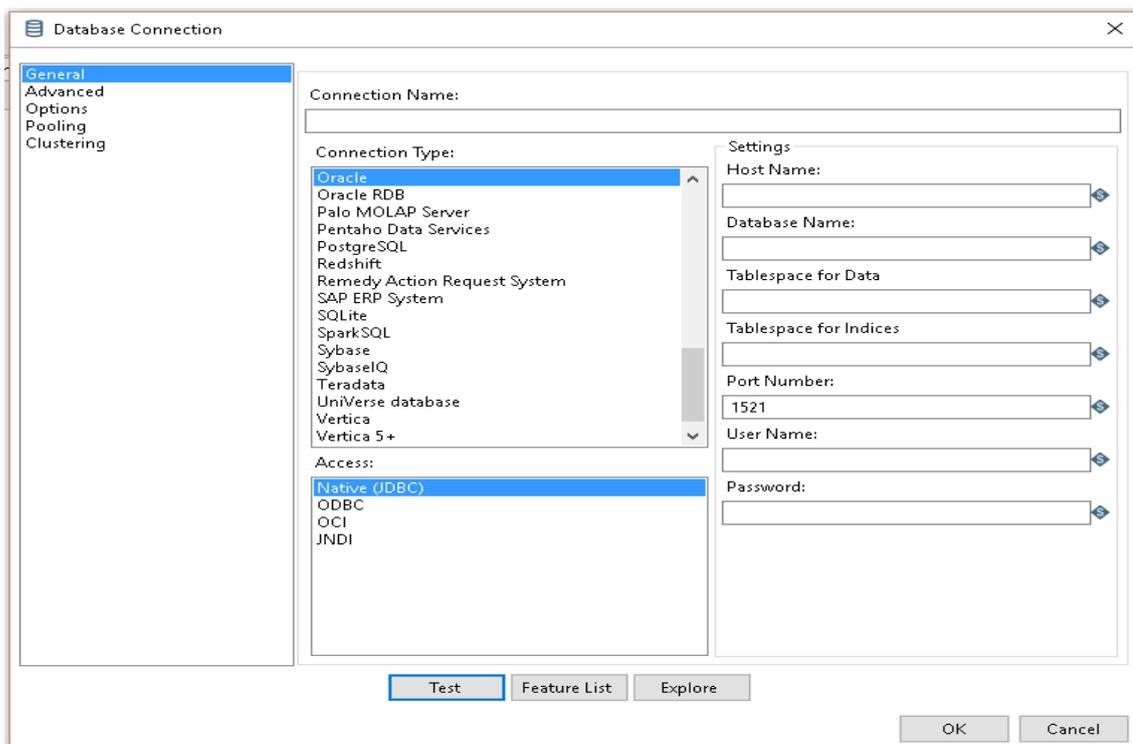


Figura 33. Configuración de parámetros

En la figura 34, se muestran todos los parámetros requeridos para establecer una conexión a la base de datos DWH_TENDENCIA, la cual está desarrollada en PostgreSQL.

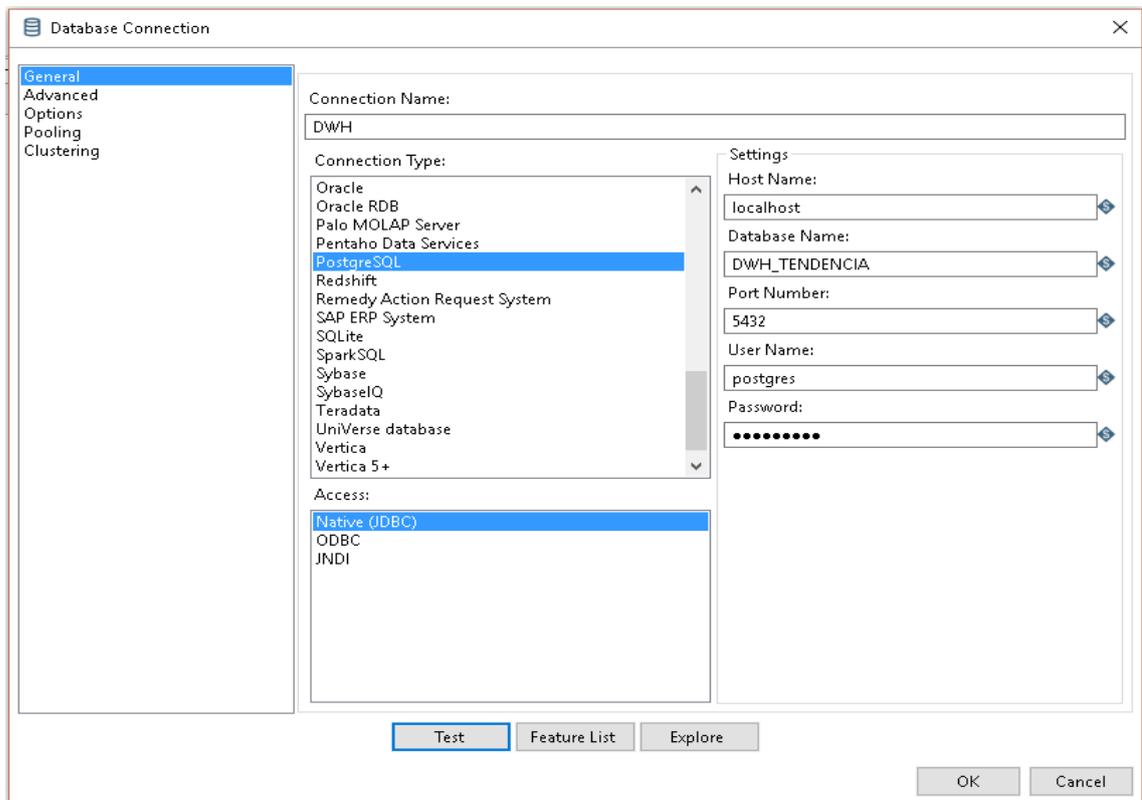


Figura 34. Parámetros requeridos para establecer una conexión

De realizar correctamente la configuración se obtiene un mensaje satisfactorio a conexión, (Ver figura 35).

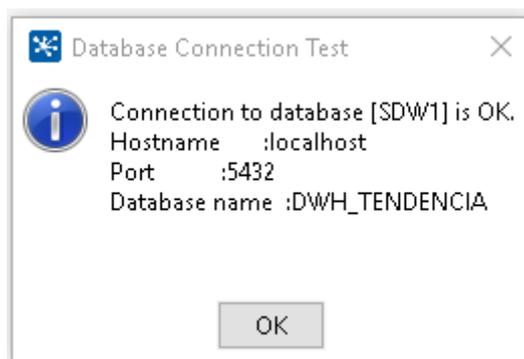


Figura 35. Mensaje de configuración correcta

Posteriormente se direcciona la ubicación de la tabla en donde se desea cargar la información, para este ejemplo se apunta a la tabla dim_persona de la base de datos DWH_TENDENCIA, (Ver figura 36).

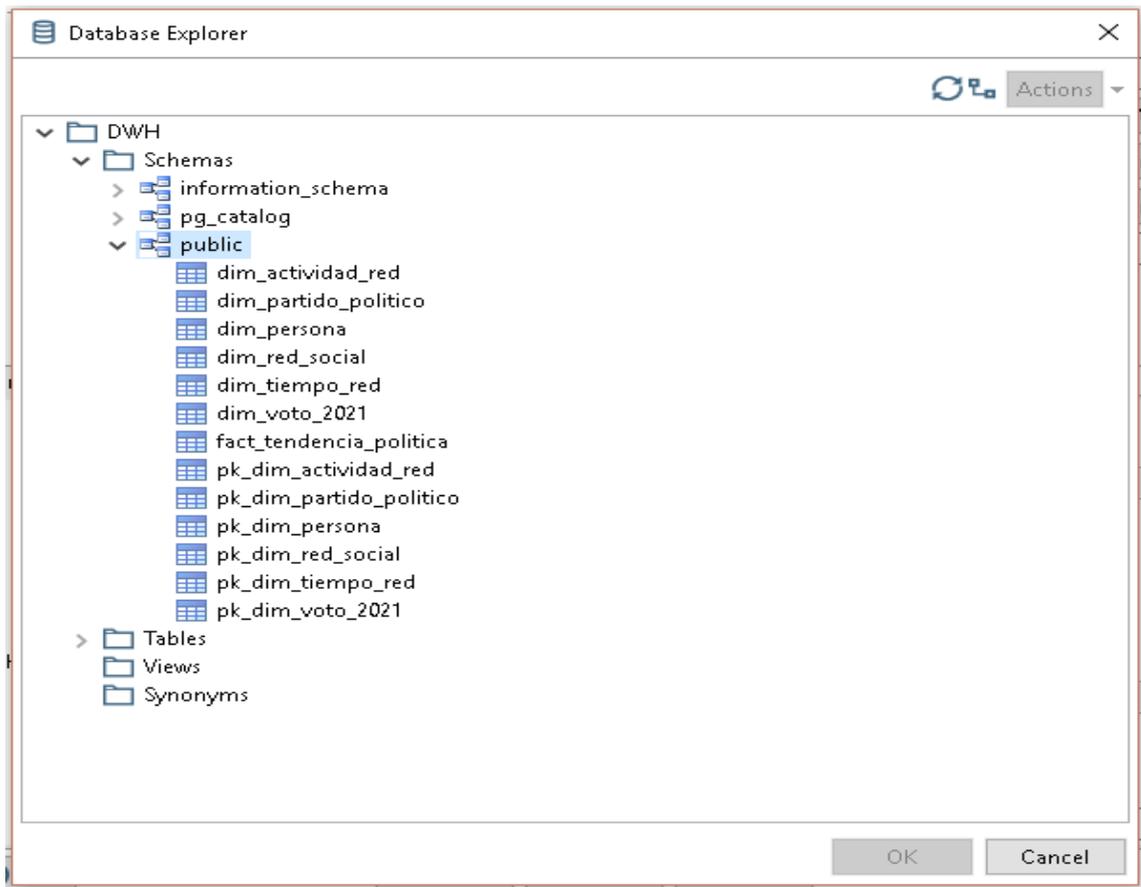


Figura 36. Direccionamiento de la tabla para carga de información

En la figura 37, se puede ver que hay que presionar “Database fields”, en esta sección es necesario realizar las correspondencias exactas entre los parámetros a ser extraídos mediante “Excel Input” y los parámetros contenidos en la base de datos.

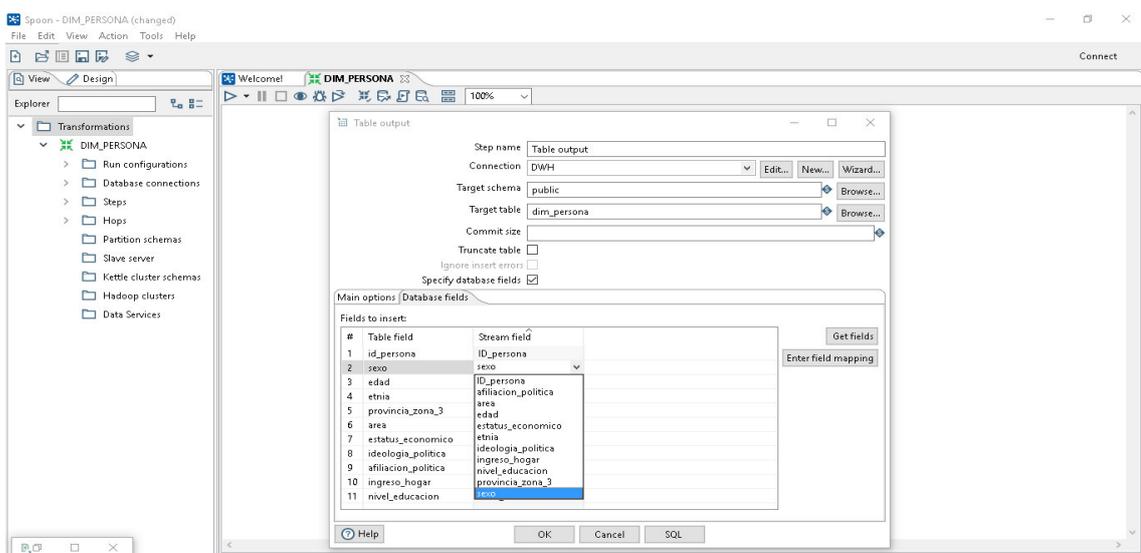


Figura 37. Database fields

c) Carga

Primero se cargarán los datos de las dimensiones y luego los de las tablas de hechos, teniendo en cuenta siempre la correcta correspondencia entre cada elemento.

Una vez creadas las transformaciones simplemente es necesario ejecutarlas, el presente informe carga la dimensión dim_persona a la base de datos, (Ver figura 38).

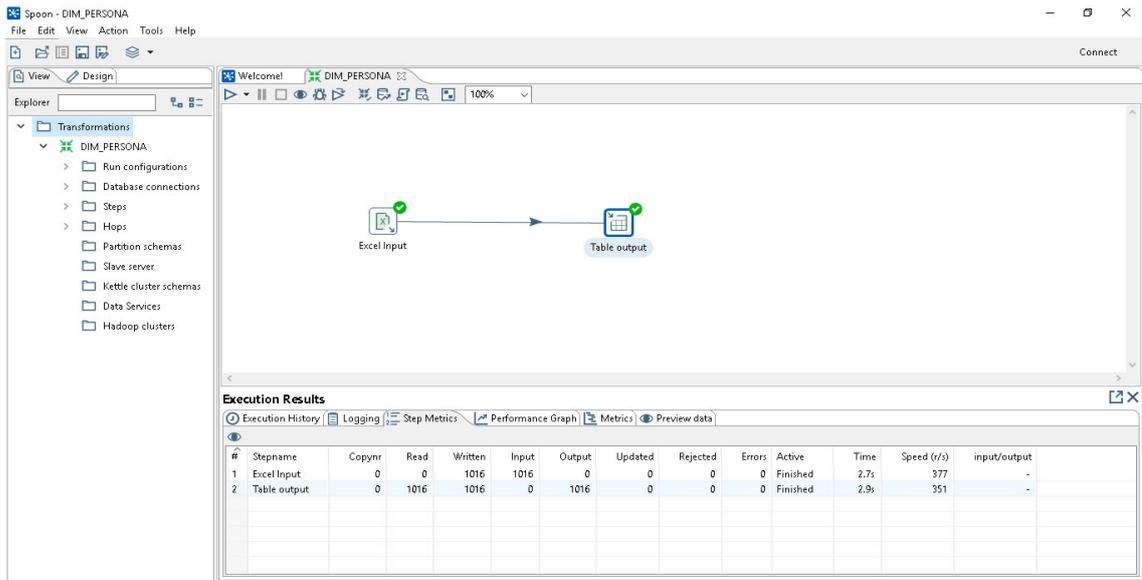


Figura 38. Carga dimensión dim_persona a la base de datos.

En la figura 39 y 40, se abre la base de datos y se revisa las tablas, lo cual permitirá apreciar que los datos han sido cargados de acuerdo al modelado realizado.

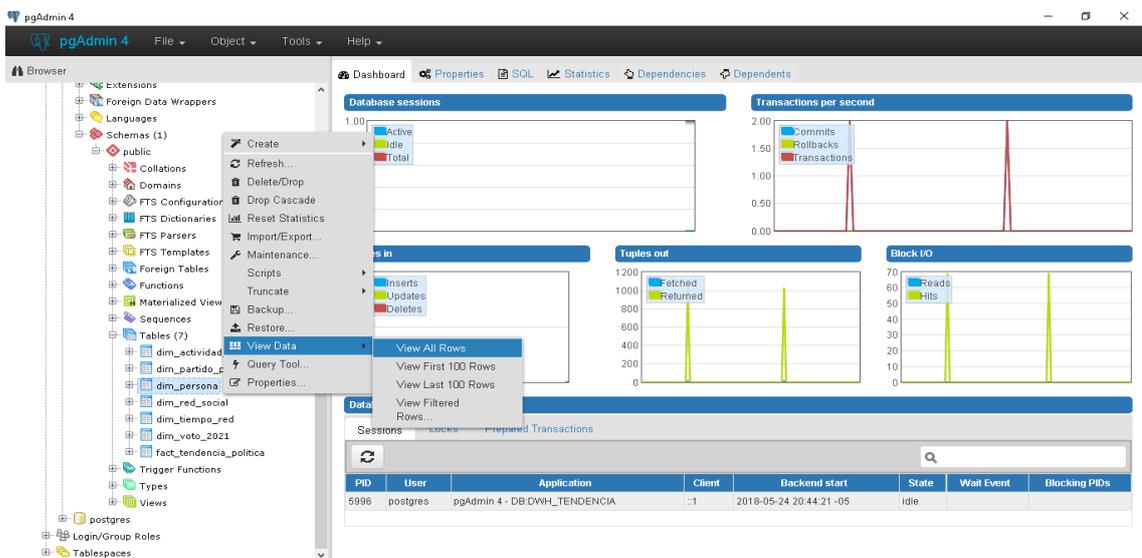


Figura 39. Apertura de la base de datos

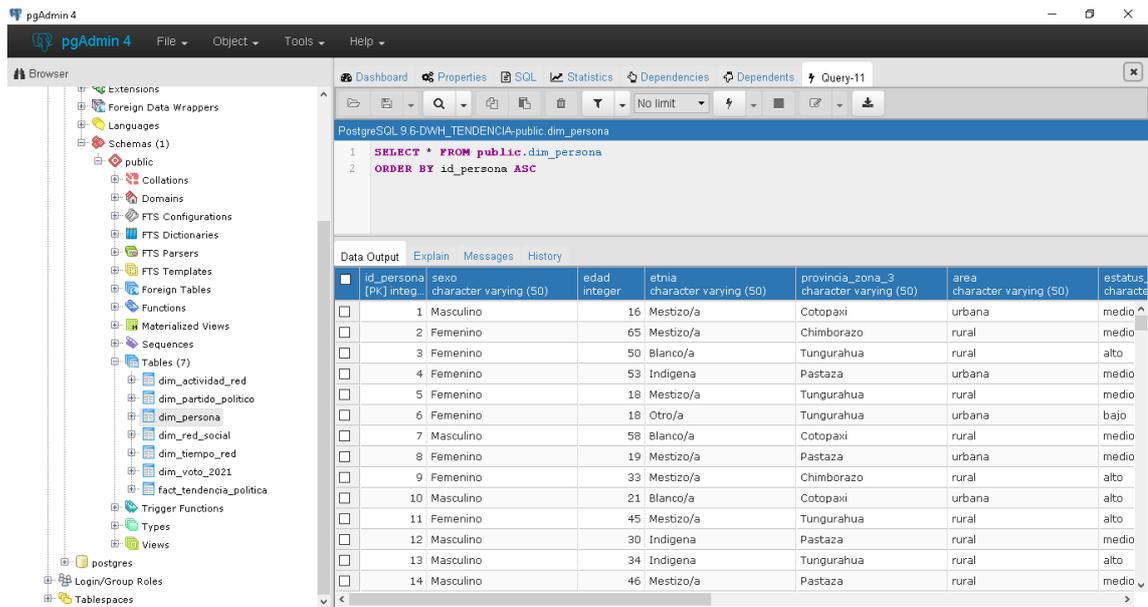


Figura 40. Visualización de la base de datos

A continuación, se muestra la carga de las diferentes dimensiones a la base de datos denominada DWH_TENDENCIA, (Ver figura 41 y 42).

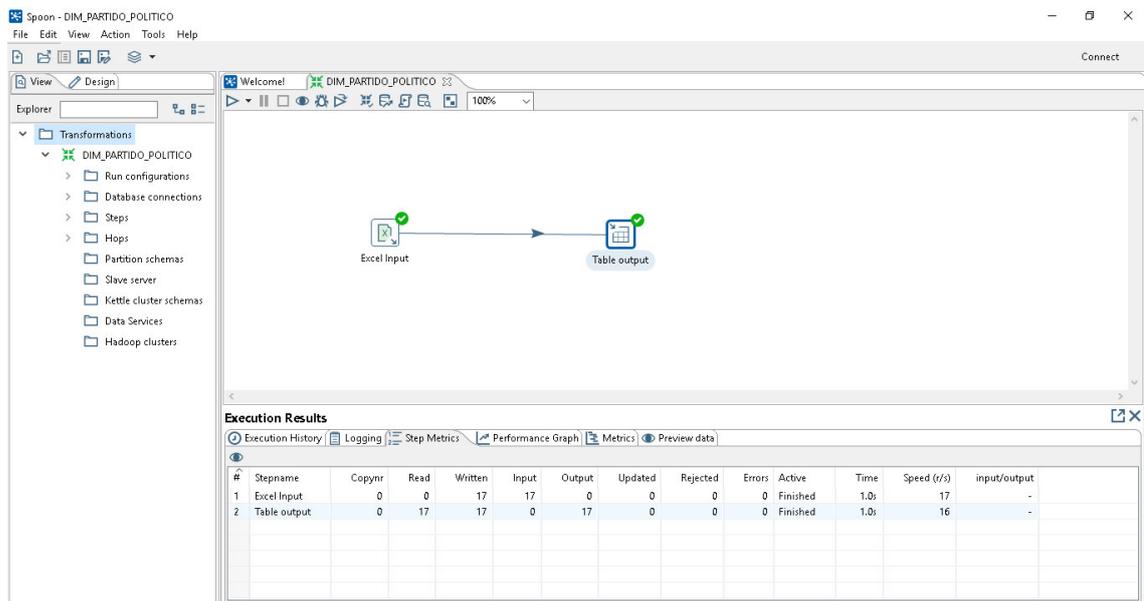


Figura 41. Carga de dimensiones

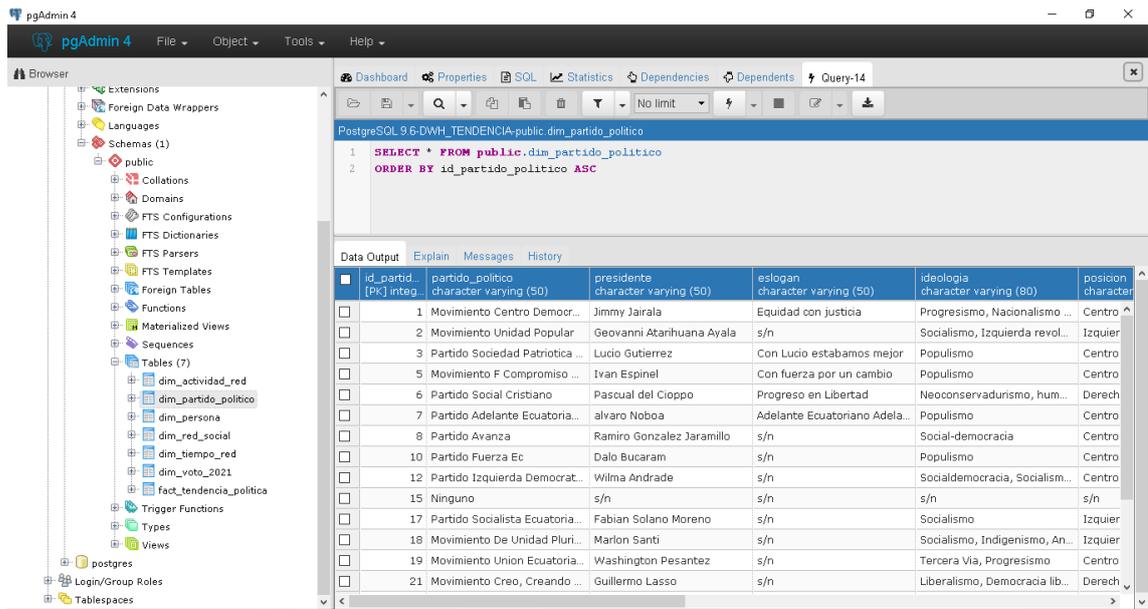


Figura 42. DWH_Tendencia

Cabe recalcar que se debe cargar la transformación que contiene la información de la tabla de hechos después de todas las transformaciones correspondientes a las dimensiones, con el fin de evitar errores de correspondencia, (Ver figura 43 y 44).

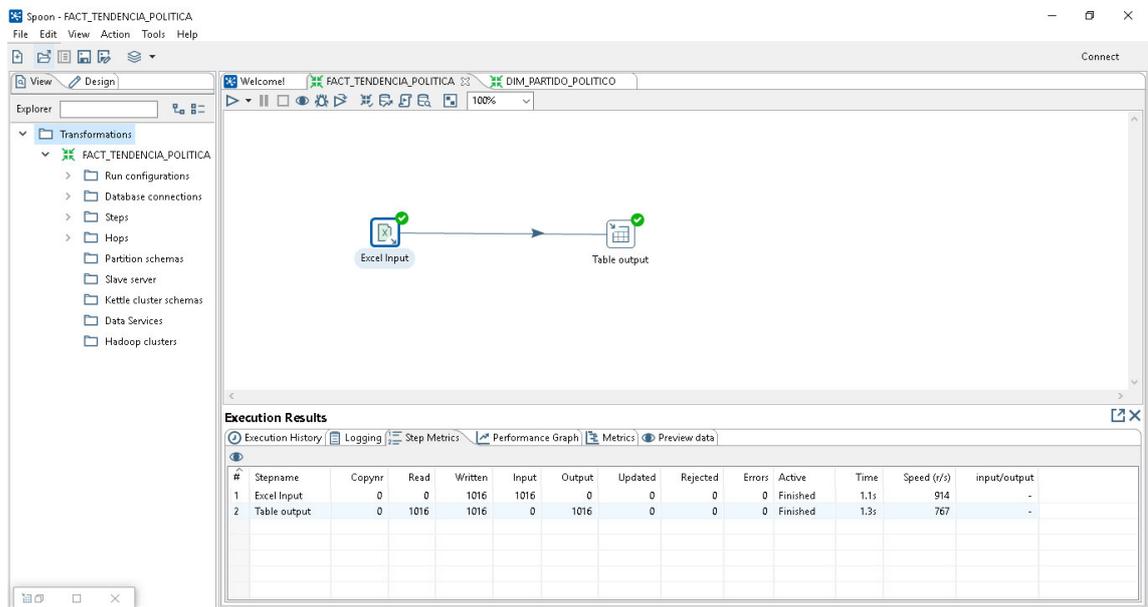


Figura 43. DWH_Tendencia 1-5

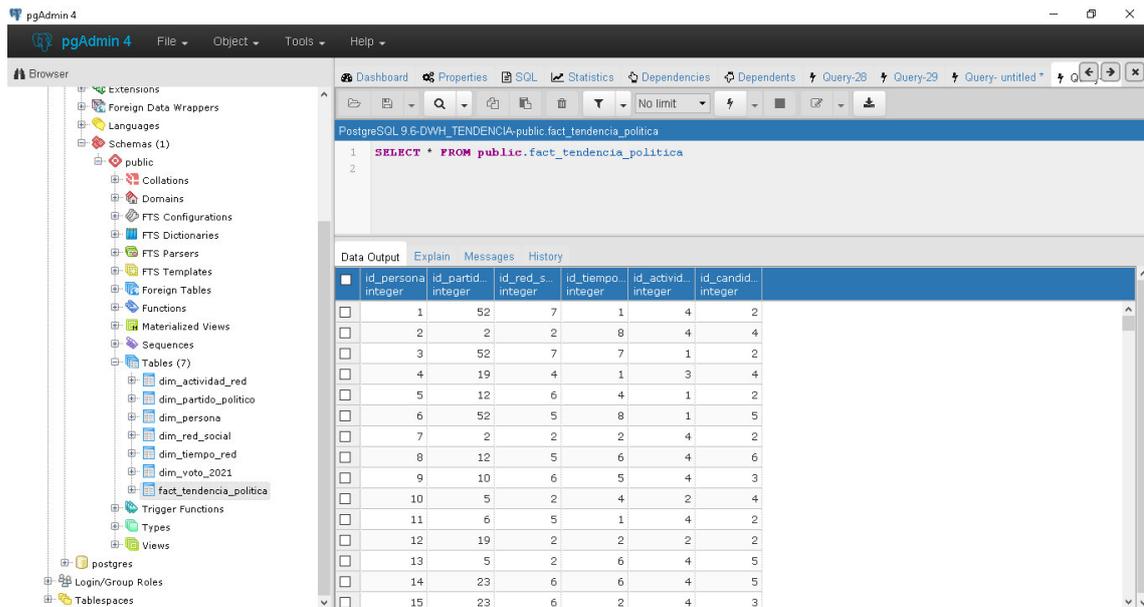


Figura 44. DWH_Tendencia 1-6

3.6.2.7 Creación de Cubos de información

Para la implementación y visualización de los cubos del Data Warehouse se utilizó la herramienta SpagoBI, la cual es una multiplataforma integrada para la Inteligencia de negocios (Business Intelligence) desarrollada enteramente de acuerdo con la filosofía del software libre y de código abierto (FOSS), esta multiplataforma cubre y satisface todos los requisitos de BI (Business Intelligence), tanto en términos de análisis y de gestión de datos, administración y seguridad.

En el mundo analítico ofrece soluciones para la presentación de informes, análisis multidimensional (OLAP), minería de datos (Data Mining), tableros de mando (Dashboard) y consultas ad-hoc. Añade módulos originales para la gestión de procesos de colaboración a través de análisis dossiers y el análisis de geo-referencia.

A continuación, en la figura 45, se visualiza la interfaz gráfica de SpagoBiMeta, la cual está lista para crear un nuevo proyecto.

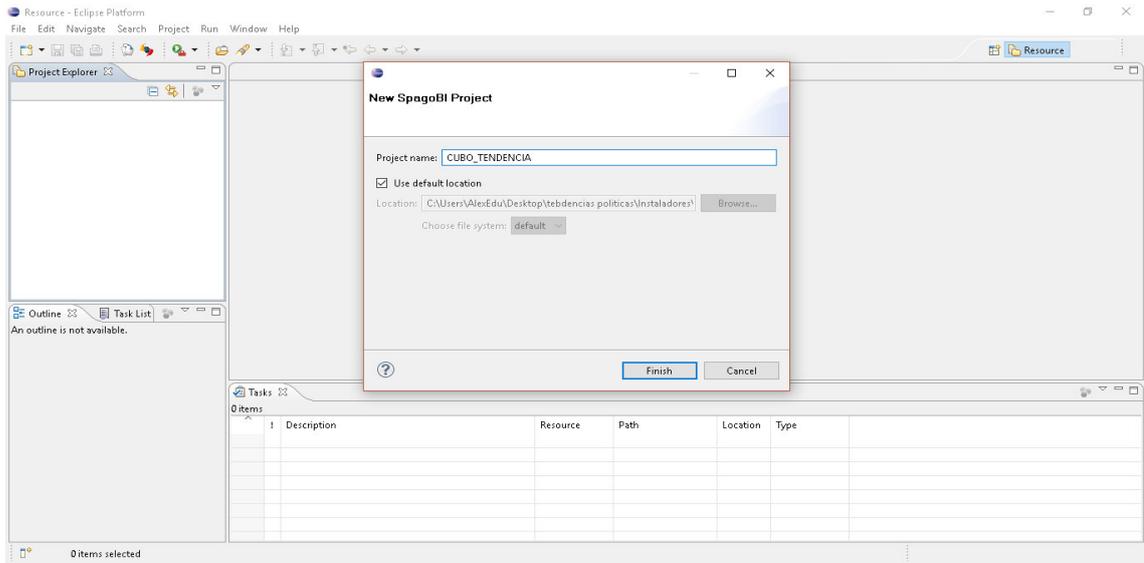


Figura 45. Interfaz gráfica de SpagoBiMeta

Es necesario agregar un servidor el cual establece la conexión con SpagoBI Server, como se visualiza en la figura 46.

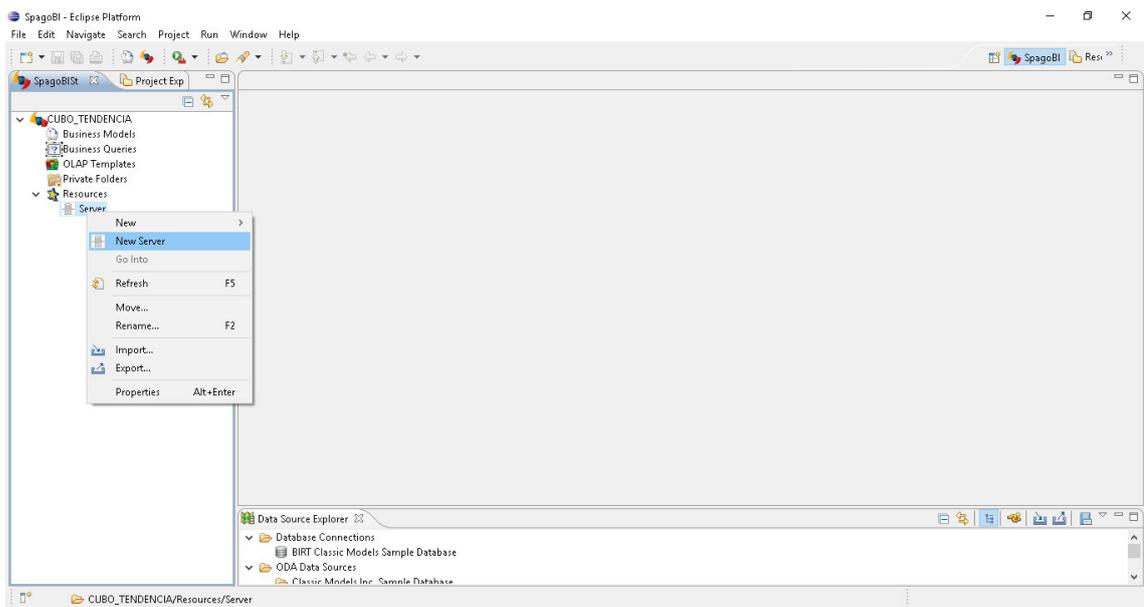


Figura 46. Conexión con SpagoBI Server.

Es importante recalcar que, es necesario agregar los parámetros para establecer una nueva conexión (Ver figura 47).

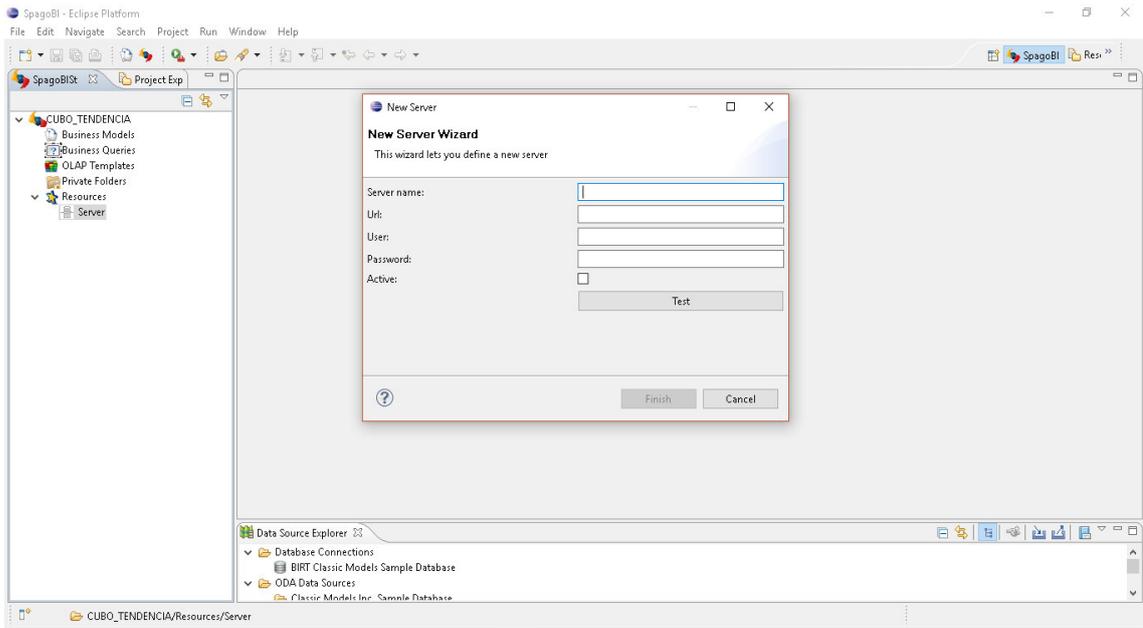


Figura 47. Agregar los parámetros necesarios para establecer una nueva conexión.

Si todos los parámetros de configuración son correctos el test de conexión será satisfactorio y se visualizará la conexión al servidor creada, (Ver figura 48).

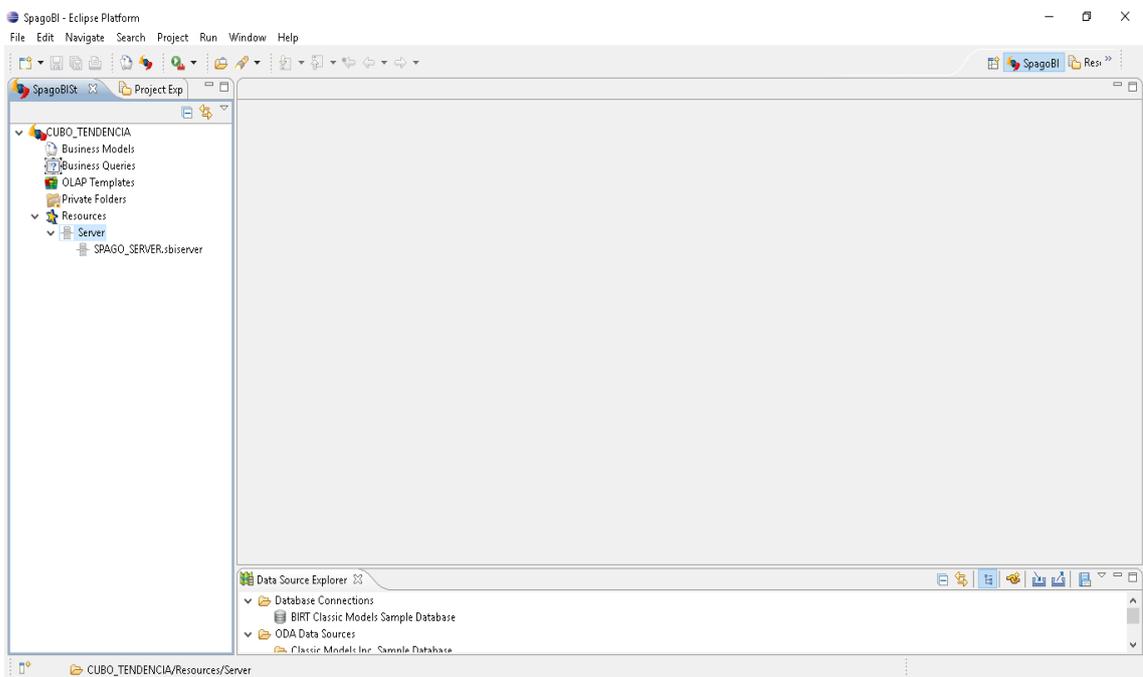


Figura 48. Visualización de la conexión al servidor creada.

A continuación, en la figura 49 se visualiza como que se debe generar una conexión con la base de datos creada denominada DWH_TENDENCIA.

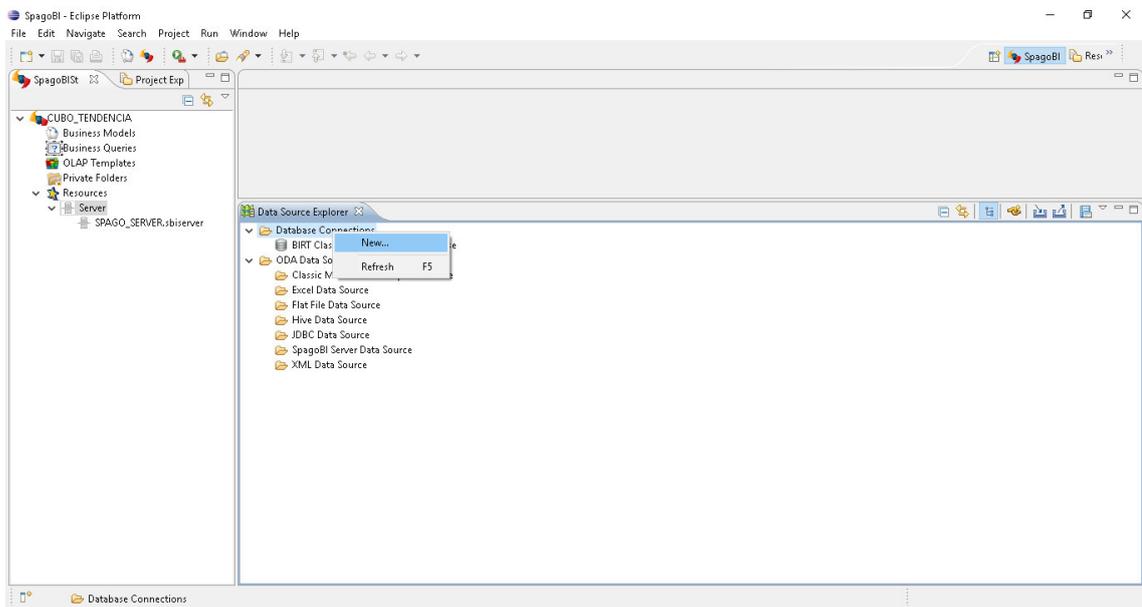


Figura 49. Conexión con la base de datos

Es importante describir que, ya que el Data Warehouse está implementado en PostgreSQL, es necesario configurar todos los parámetros para esta base de datos en especial, (Ver figura 50).

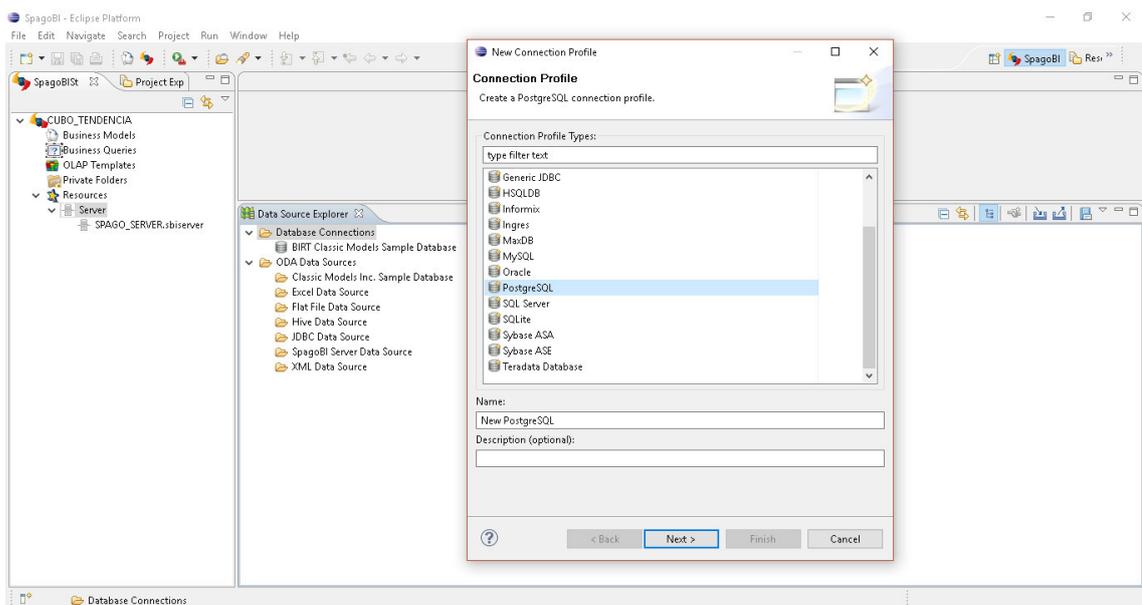


Figura 50. Configuración de parámetros

Es importante instalar los drivers para establecer conexiones y se requiere descargarse los instaladores de la página oficial de PostgreSQL, una vez descargado se direcciona la ubicación de este para poder ejecutarlo, (Ver figura 51).

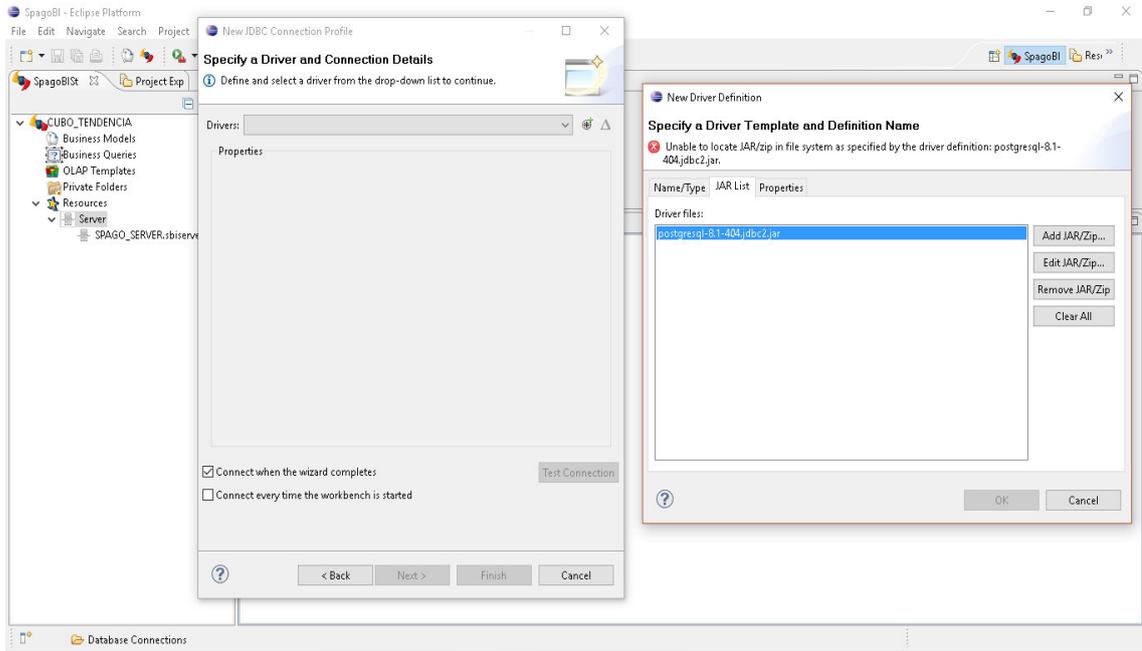


Figura 51. Descarga los drivers de la página oficial de PostgreSQL

Una vez instalados los drivers, se debe configurar los parámetros que son necesarios para establecer la conexión con la base de datos DWH_TENDENCIA, como se puede ver en la figura 52.

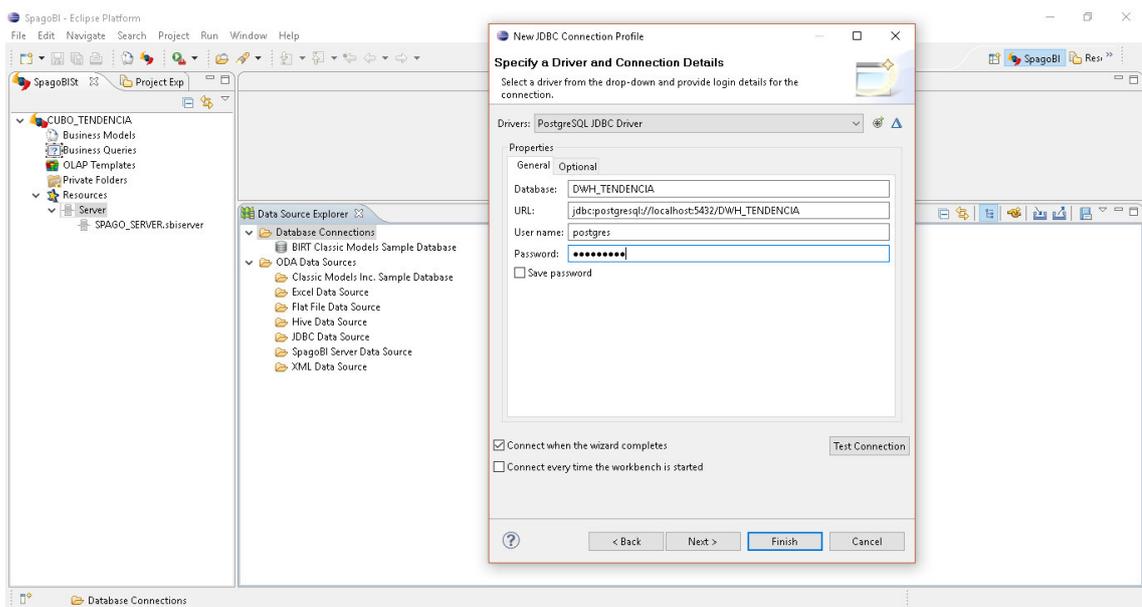


Figura 52. Configuración de parámetros necesarios para establecer conexión con la BD

Con el fin de comprobar el estado de la conexión, se deberá realizar una prueba presionando “Test Connection” y una vez que se ha creado de forma exitosa la conexión a DWH_TENDENCIA se tiene que crear un nuevo modelo, (Ver figura 53).

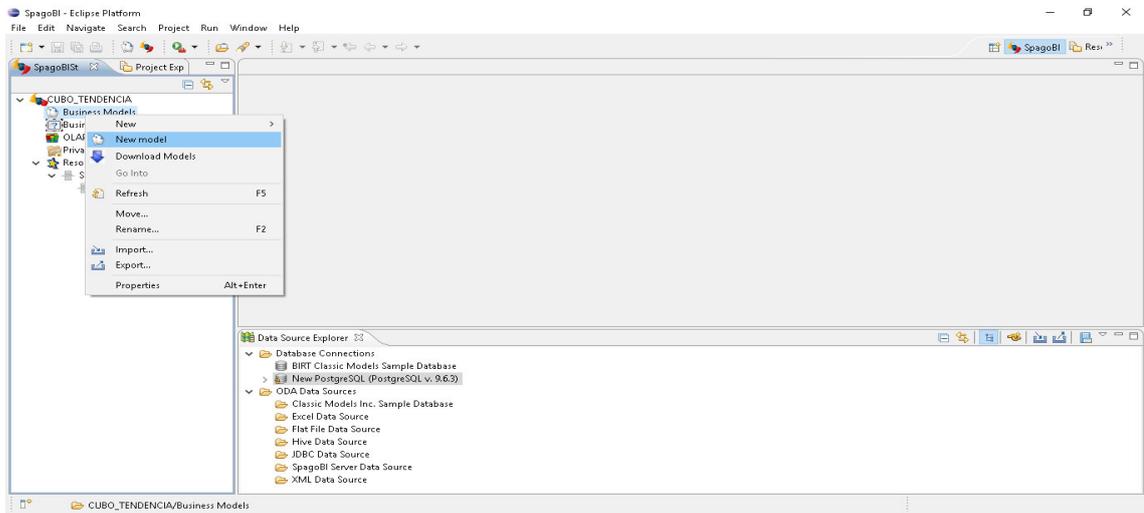


Figura 53. Creación de un nuevo modelo

Una vez realizado las pruebas, se debe proporcionar la ubicación al nuevo modelo creado en “Bussiness Models” y agregar un nombre al modelo, también se deberá seleccionar “New PostgreSQL” que corresponde a la conexión con el DataWareHouse y se debe extraer las dimensiones y la tabla de hechos que conformaran el modelo físico del modelo a desarrollarse.

El siguiente paso es extraer las dimensiones y la tabla de hechos que conformaran la forma del modelo de negocios a desarrollarse, como se puede ver en la figura 54.

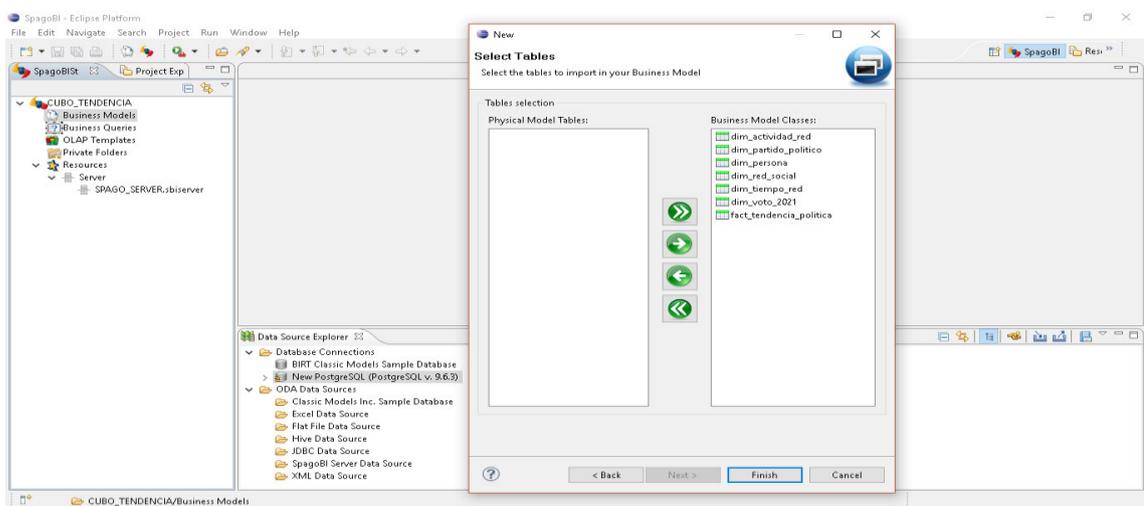


Figura 54. Extracción de las dimensiones y la tabla

Se aprecia las tablas de dimensiones y hechos lista para ser configuradas y lo primero que se requiere es definir a las tablas como dimensiones y a la tabla de hechos como Cubo, (Ver figura 55).

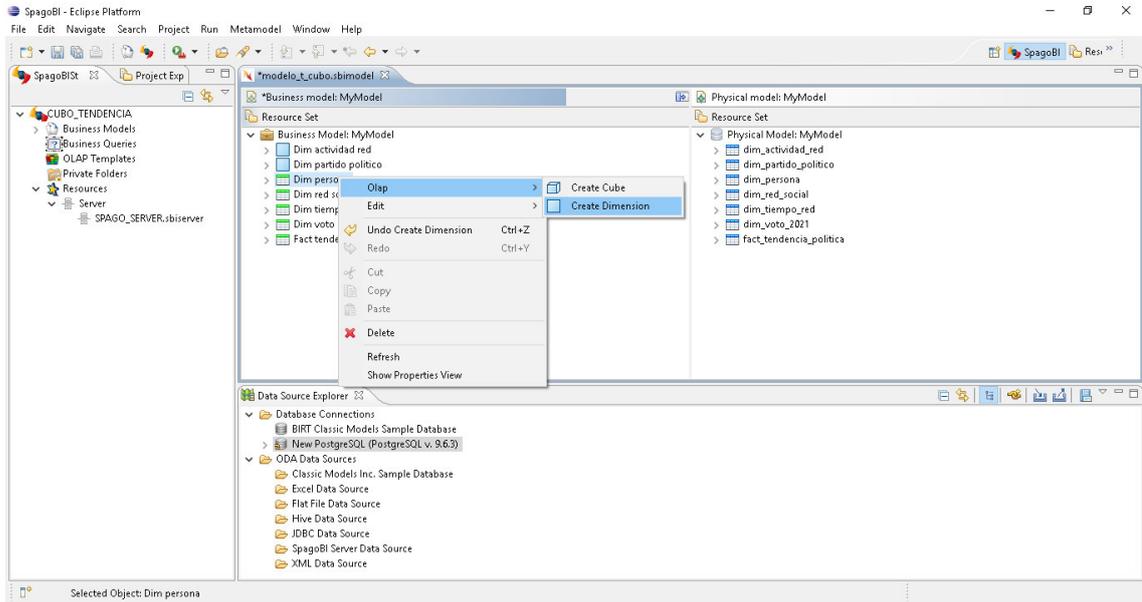


Figura 55. Definir a las tablas como dimensiones y a la tabla de hechos como Cubo

Ahora en la figura 56, se informa que es necesario editar las jerarquías de cada una de las dimensiones, esta configuración establece que parámetros serán los más importantes durante la visualización del cubo y se presionamos “Add Hierarchy” para comenzar con la edición.

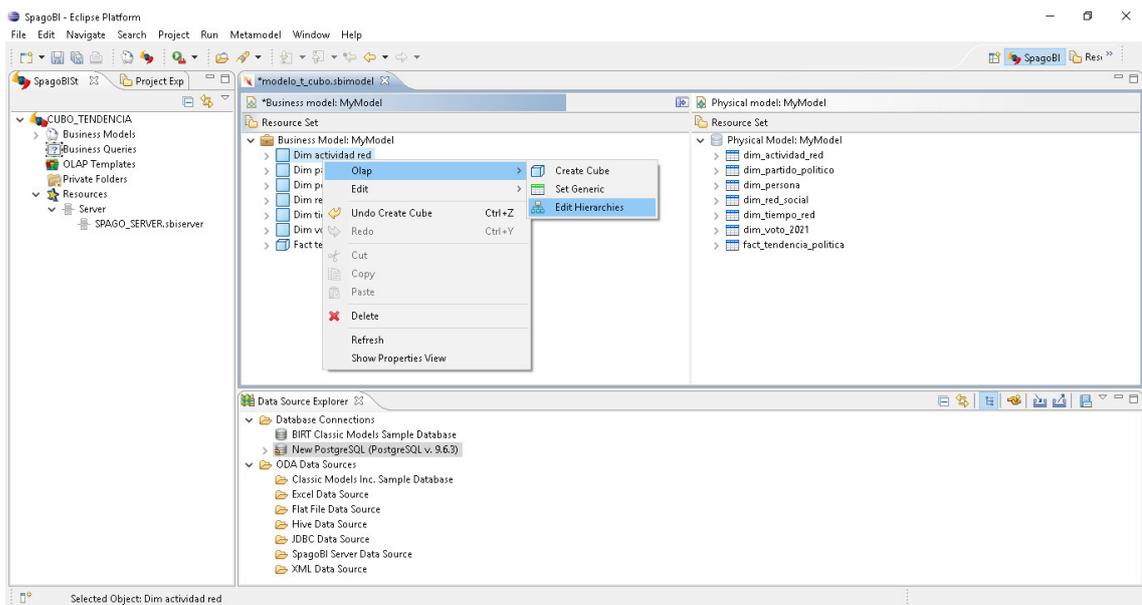


Figura 56. Editar las jerarquías de cada una de las dimensiones

A continuación en la figura 57, se expone como ejemplo la jerarquía establecida para Dim_actividad_red y se muestran las jerarquias para las demás dimensiones.

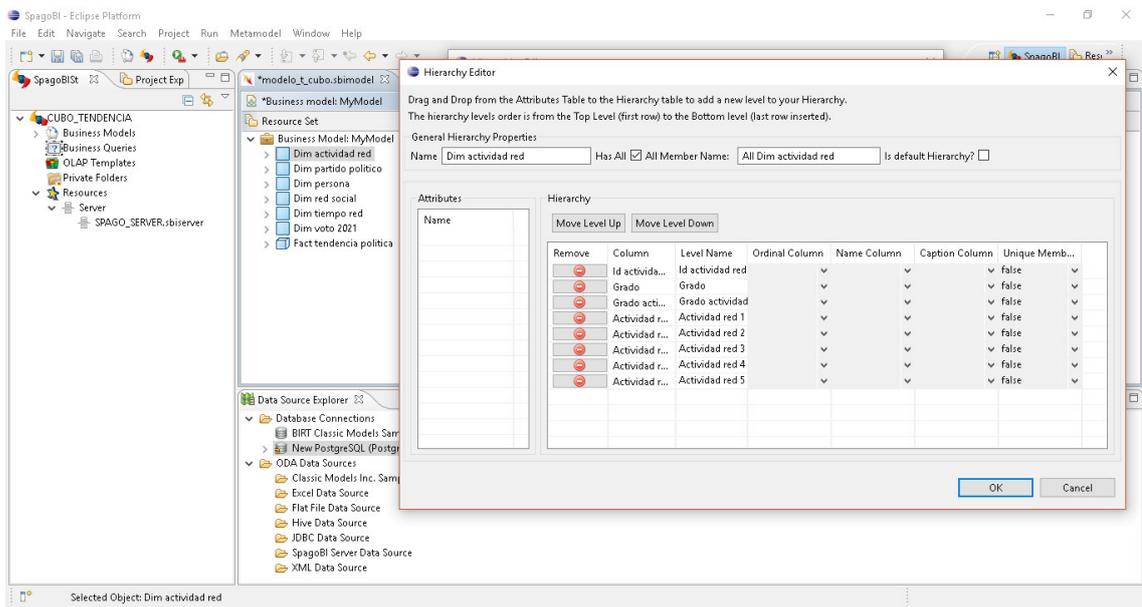


Figura 57. Jerarquía establecida para Dim_actividad_red

En este apartado se agregan las relaciones existentes entre las dimensiones y la tabla de hechos, dicho paso se realizará con cada una de las dimensiones y con la tabla de hechos, cada indicador debe ser colocado con su correspondiente y generar la relación. Se deberá repetir el paso por cada uno de los indicadores (Ver figura 58).

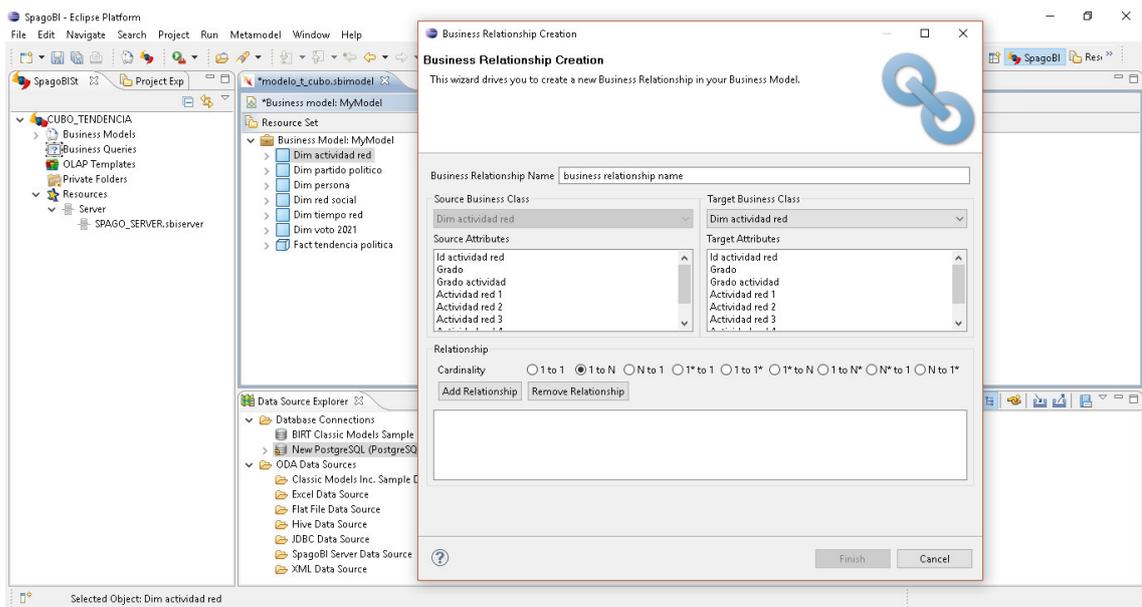


Figura 58. Generación de la relación

A continuación, se aprecia las relaciones para las demás dimensiones que son: Dim partido político, Dim persona, Dim red social, Dim tiempo red, Dim voto 2021, (Ver figura 59).

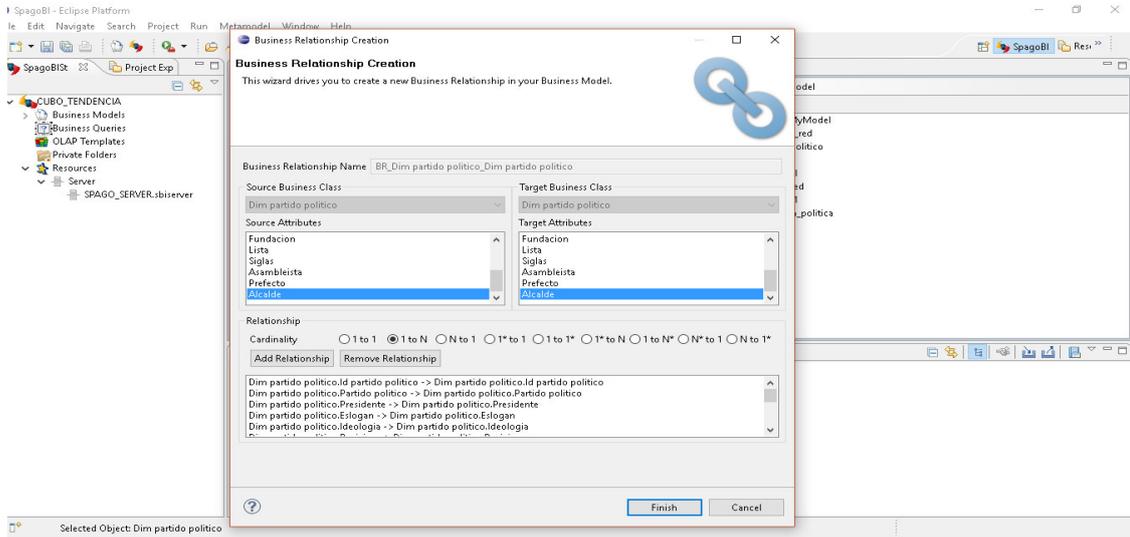


Figura 59. Dim partido político

La relación entre la tabla de hechos debe relacionarse con el ID correspondiente a cada una de las dimensiones, por lo cual en la figura 60, se muestra las diferentes relaciones entre: Relación Fact tendencia política – Id persona, Relación Fact tendencia política – Id partido político, Relación Fact tendencia política – Id red social, Relación Fact tendencia política – Id tiempo red, Relación Fact tendencia política – Id actividad red, Relación Fact tendencia política – Id candidato 2021.

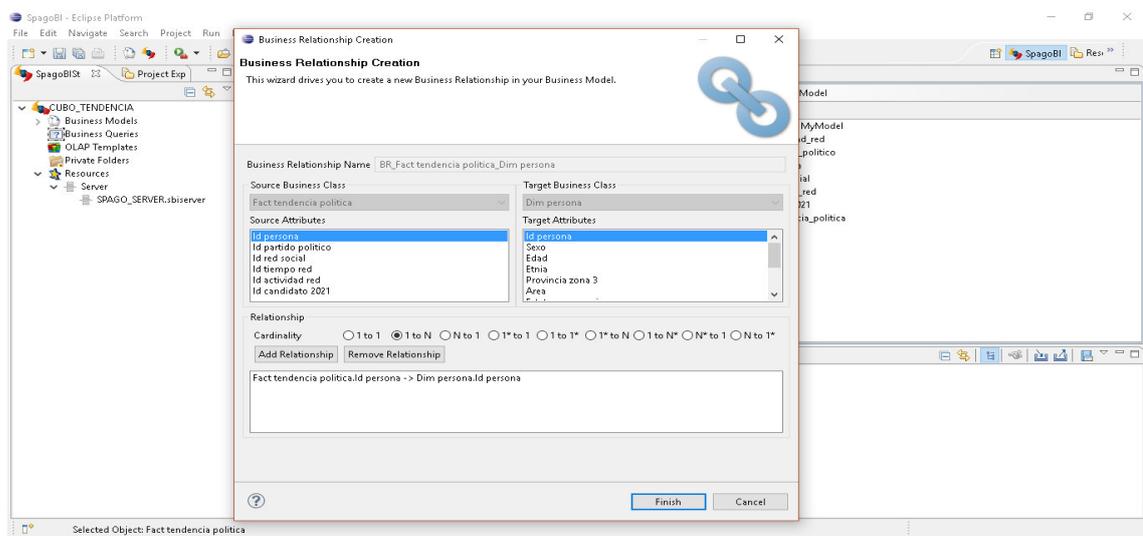


Figura 60. Relación Fact tendencia política – Id persona

Es necesario dar un clic derecho sobre “Business Model: MyModel” y coger Create – Mondriam Template, también se debe seleccionar la ubicación en “OLAP Templates” y añadir un nombre, (Ver figura 61).

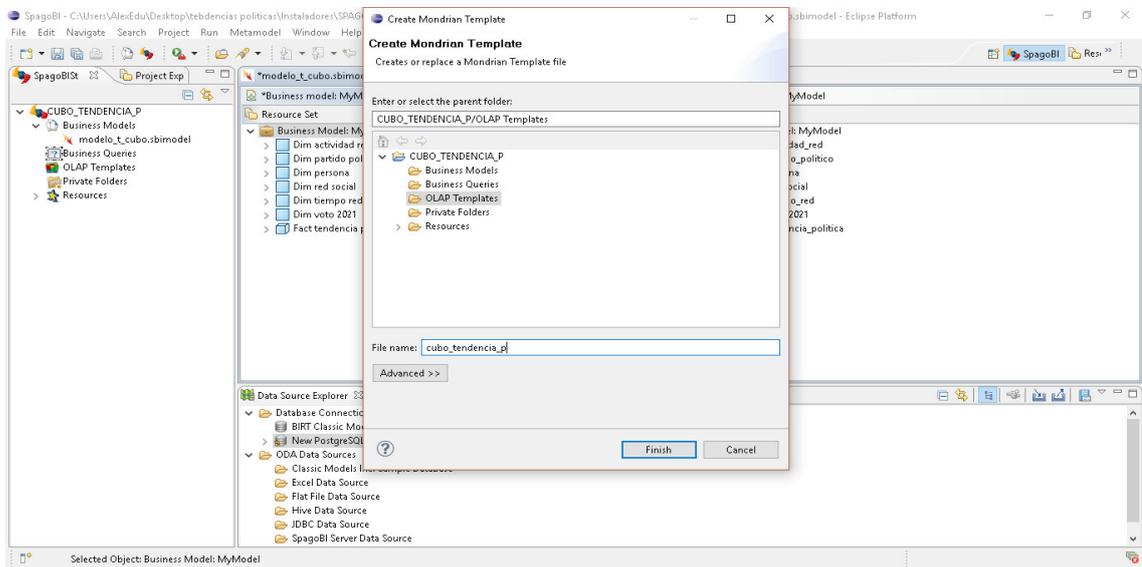


Figura 61. Selección de la ubicación del OLAP

Seguido del paso anterior, hay que ubicarse sobre el archivo creado y abrir con un editor de texto. Es necesario corregir este archivo, para esto se tiene que localizar esta línea de código: “<Hierarchy hasAll=“” y añadir “<Hierarchy hasAll=“true””, revisar con cuidado y corregir todas las apariciones de este error, guardar los cambios la cual se va a repetir varias veces, (Ver figura 62).

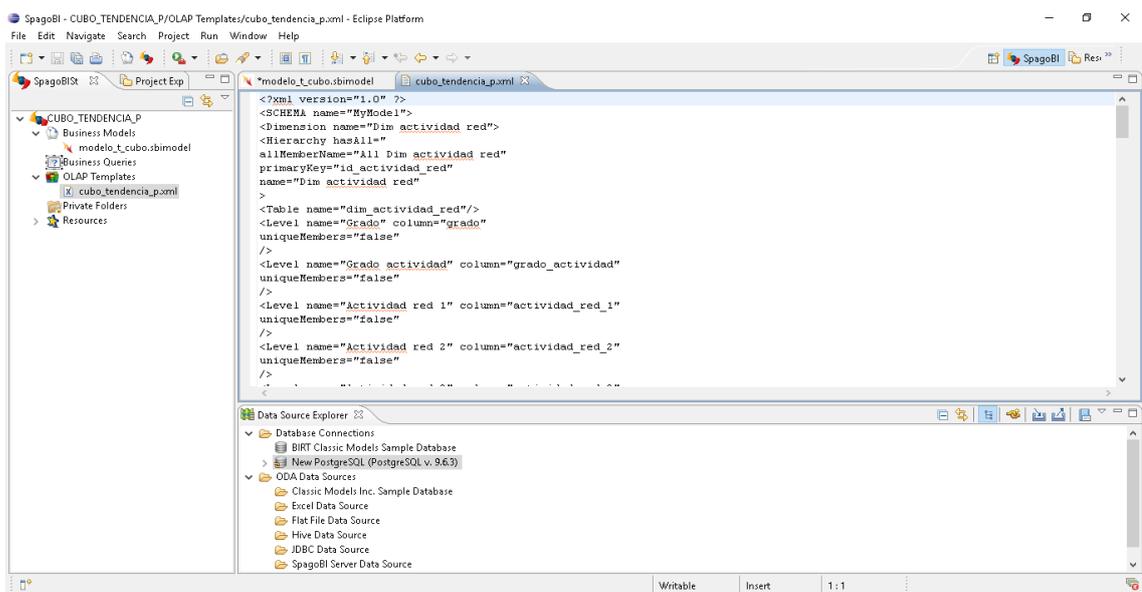


Figura 62. Corrección del archivo y errores

En la última parte, en la figura 63, se especifica que se deberá dar clic derecho sobre “cubo_tendencia.xml”, seleccionar “Deploy Olap Template” y darle un nombre al cubo desarrollado, seleccionar el Data Warehouse, en dicho caso se creará “DWH_TENDENCIA”, y presionar “Finish” para finalizar y finalmente se visualiza que el esquema olap ha sido cargado correctamente.

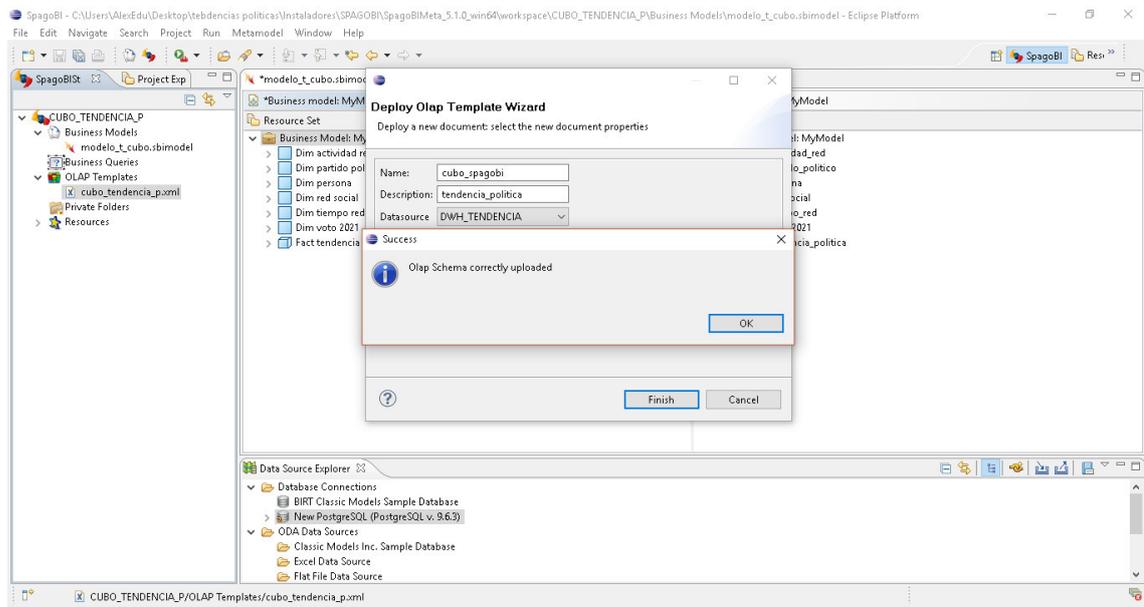


Figura 63. Esquema Olap final

3.6.2.8 Validación del procesamiento de la información

Para la visualización de los datos se utilizó la plataforma SpagoBI, misma que es una multiplataforma integrada para la resolución de procesos en Inteligencia de negocios (Business Intelligence), la cual esta desarrollada bajo código abierto (FOSS).

A continuación, en la figura 64, se visualiza la interfaz de trabajo de SpagoBi Server.



Figura 64. Interfaz de trabajo de SpagoBI Server.

En la figura 65, se visualiza el cubo desarrollado, el cual está ubicado dentro del usuario “biadmin”

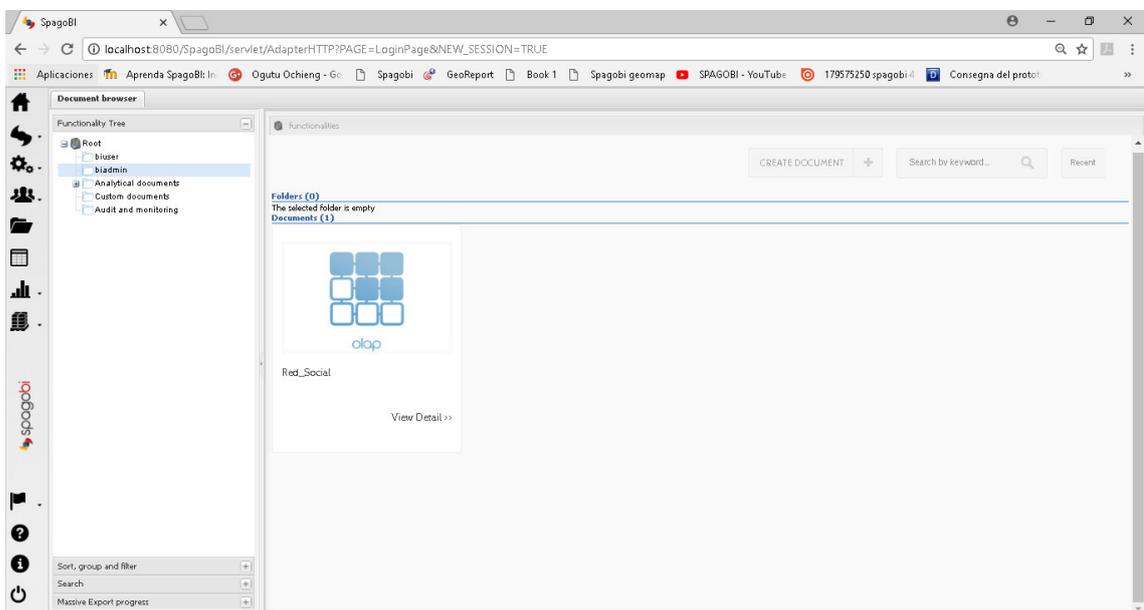


Figura 65. Visualización del cubo

Al Ingresar al cubo de información se puede visualizar las barras de herramientas que será útil para interactuar con el cubo. Estas barras permiten realizar cruces con las distintas dimensiones, desplazarse por la información, generar gráficos, entre otras funciones, (Ver figura 66).

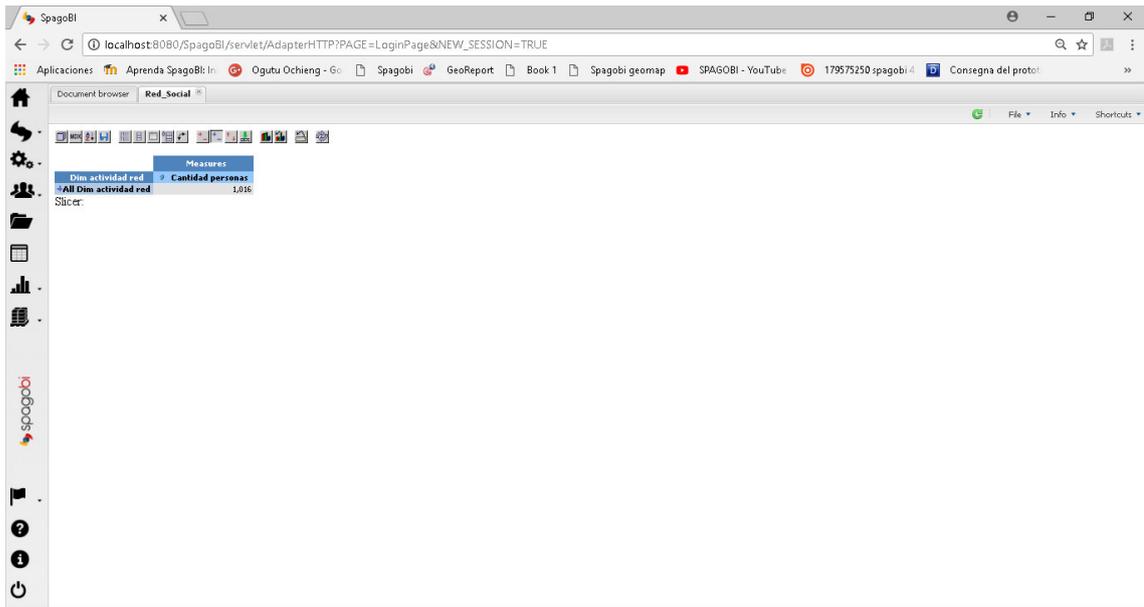


Figura 66. Ingreso al cubo y visualización de barras de herramientas

Una vez ingresado los datos, la visualización del análisis de cruzar las dimensiones Dim_actividad_red con Dim_persona, muestra información del grado de actividad que realizan las personas en las redes sociales, así como también la ideología política que las personas indicaron tener. Además, se realiza la interpretación gráfica de distintos modelos, cabe recalcar que las herramientas del cubo permiten girar las dimensiones permitiendo otra perspectiva de análisis como se visualiza en la figura 67.

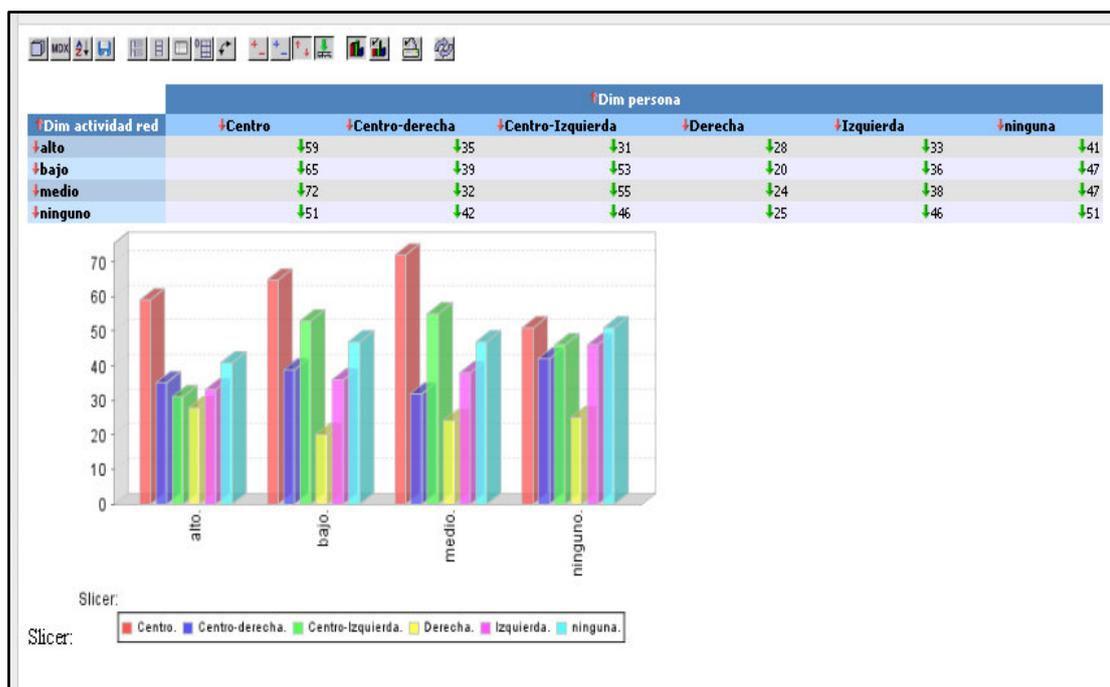


Figura 67. Interpretación gráfica agrupada de Dim personas

Es importante recalcar que se puede visualizar que el 24.3% del total de encuestados son de tendencia de centro, 14.6% son de centro derecha, 18.2% son de centro izquierda, 9.5% son de derecha, 15.1% son de izquierda y el 18.3% de los votantes tomaron como opción que no pertenecen a ninguna tendencia política.

En cuanto a la Interpretación gráfica agrupada de Dim referente a la Actividad red, se puede visualizar en la figura 68 que los porcentajes de votación referente a las actividades que los usuarios usan la red son: 22.3% de los encuestados tiene un porcentaje alto, el 25.6% bajo, el 26.4% medio y el 25.7% prefiere abstener en la respuesta de la encuesta realizada.



Figura 68. Interpretación gráfica agrupada de Dim. Actividad red

En la figura 69, se puede describir que las tendencias son las siguientes:

- Centro: El 5,8% estipulan que son de alta prioridad, el 6,4% especifican baja, el 7,1% denominan como prioridad media y el 5,0% no contesta.
- Centro derecha: El 3,4% especifica una alta importancia, el 3,8% como proceso bajo, el 3,1% dice que es medianamente de centro y el 4,1% no sabe o no conoce.
- Centro Izquierda: En este punto, el 3,0% denomina como prioridad alta, el 5,4% media, 5,2% baja y el 4,5% no responde.

- Derecha: El 2,8% especifica que son de derecha con dicho porcentaje, el 2,0% especifica que su tendencia es baja y el 2,4% como prioridad media.
- Izquierda: en la tendencia de izquierda, se puede ver que el 3,2% especifica como muy alta su prioridad, el 3,7% describe como media y el 3,5% especifica que es muy baja la prioridad de izquierda



Figura 69. Interpretación gráfica agrupada horizontal de Dim actividad red

En las figuras 70, 71 y 72, se muestra que la diferencia del promedio entre los votantes no es muy grande, por ello, en centro derecha e izquierda el porcentaje es de 1.3% de diferencia (Izquierda tiene el porcentaje más elevado), mientras que la tendencia de centro tiene un porcentaje 15.5% más que centro izquierda.

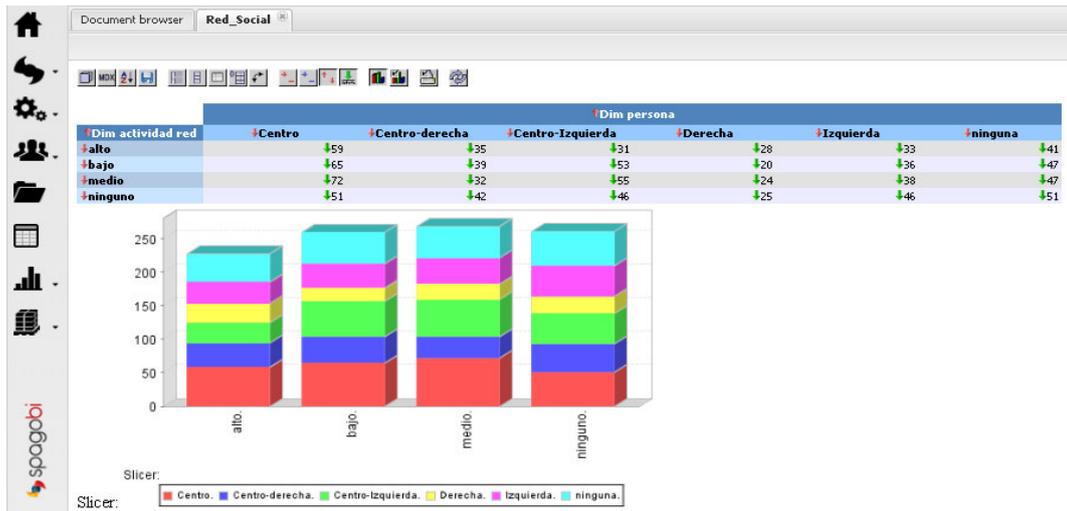


Figura 70. Interpretación gráfica apilada de Dim personas

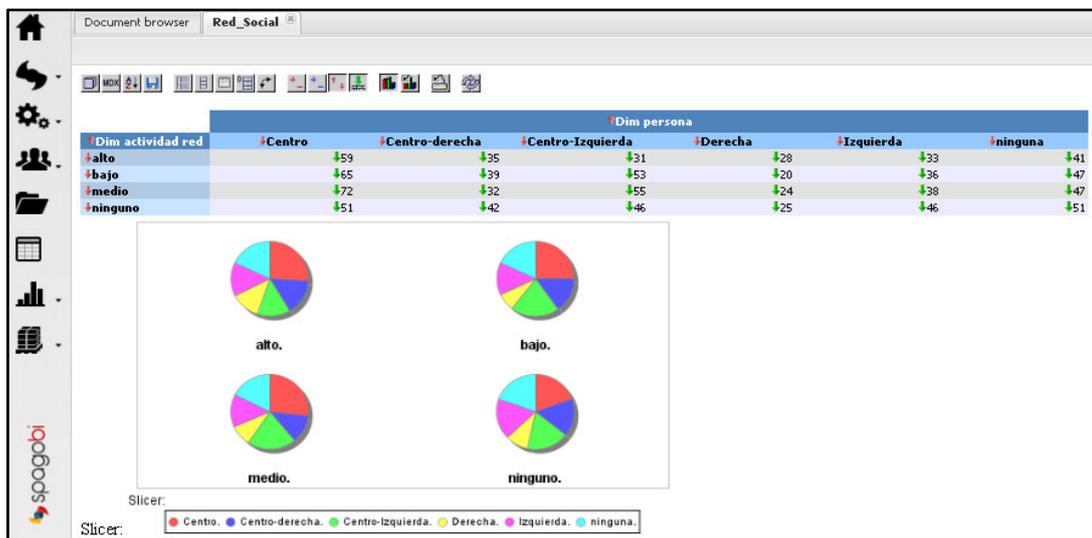


Figura 71. Tabulación de datos Dim personas



Figura 72. Tabulación de datos Dim actividad red

En las figuras 73 y 74, se ha hecho el análisis de la cantidad de personas que se definieron con ideología política centro. Dicho proceso se realizó para saber de este grupo de personas el tiempo ocupado en redes sociales con respecto a su género, con lo cual se especifica que los hombres tienen una tendencia del 53%, mientras que las mujeres un 47% del total.

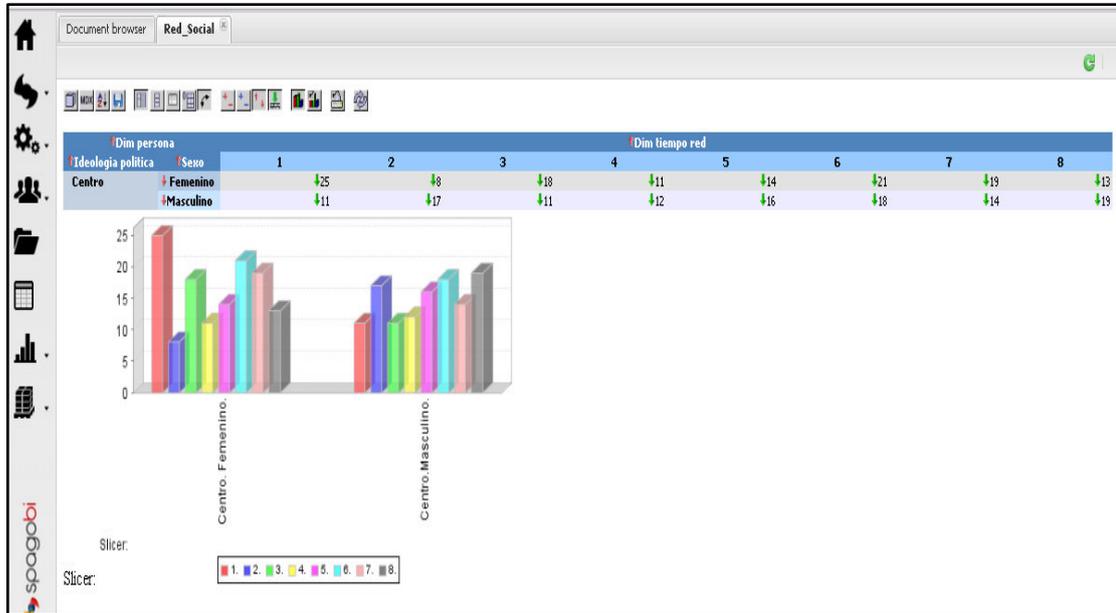


Figura 73. Gráfica del Dim tiempo red (masculino y femenino)



Figura 74. Gráfica del Dim persona (masculino y femenino)

En el siguiente análisis se realizó el cruce entre las dimensiones Dim partido político con Dim persona, esto permitió analizar la cantidad de personas que apoyan a un partido político en particular respecto a la ideología política que estas personas poseen, (Ver figuras 75 y 76).

Dim partido político	Dim persona					
	Centro	Centro-derecha	Centro-Izquierda	Derecha	Izquierda	ninguna
Movimiento Alianza País						29
Movimiento Centro Democrático	52	19	15			
Movimiento Concertación	42	13	25			
Movimiento Creo, Creando Oportunidades				55		
Movimiento De Unidad Plurinacional Pachakutik					47	
Movimiento F Compromiso Social	55	15	21			
Movimiento Sociedad Unida Mas Accion, Suma	53	11	17			
Movimiento Unidad Popular					42	
Movimiento Union Ecuatoriana		20	23			
Ninguno						186
Partido Adelante Ecuatoriano Adelante		14	22			
Partido Avanza		15	19			
Partido Fuerza Ec		11	12			
Partido Izquierda Democratica		15	17			
Partido Social Cristiano				42		
Partido Socialista Ecuatoriano					35	
Partido Sociedad Patriótica 21 De Enero	45	15	14			

Figura 75. Análisis de personas que apoyan a un partido político

Dim persona	Dim partido político																
	Movimiento Alianza País	Movimiento Centro Democrático	Movimiento Concertación	Movimiento Creo, Creando Oportunidades	Movimiento De Unidad Plurinacional Pachakutik	Movimiento F Compromiso Social	Movimiento Sociedad Unida Mas Accion, Suma	Movimiento Unidad Popular	Movimiento Union Ecuatoriana	Ninguno	Partido Adelante Ecuatoriano Adelante	Partido Avanza	Partido Fuerza Ec	Partido Izquierda Democratica	Partido Social Cristiano	Partido Socialista Ecuatoriano	Partido Sociedad Patriótica 21 De Enero
Centro	52	19	42			55	53										45
Centro-derecha	19	13	13			15	11		20		14	15	11	15			15
Centro-Izquierda	15	25				21	17		23		22	19	12	17			14
Derecha					55					42					42		
Izquierda																	35
ninguna										186							

Figura 76. Personas que apoyan a un partido político desde otra posición

CONCLUSIONES

- En el estudio se puede evidenciar que las redes sociales han llegado a la sociedad para quedarse, pues permiten mantener la comunicación entre personas de diferentes culturas, etnias, creencias, costumbres, entre otros, extendiendo sus campos y llegando a todas los segmentos del mundo, ya que miles de usuarios expresan sus sentimientos, conocimientos y afinidades a través de estas, información que puede ser aprovechada para proyecciones en diversos campoco como la política.
- La forma de agrupación de la información en la minería de datos permite observar de diferentes perspectivas a las bases de datos, por ejemplo, una agrupación de la base de datos de tendencias políticas podría exponer todos los partidos políticos que existen dentro de las regiones, mientras que otro grupo podría mostrar los partidos políticos a los que los usuarios tienen mayor afinidad.
- Con la metodología Data Warehouse que es una base de datos corporativos encargada de integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista, se pudo determinar que los usuarios si utilizan las redes sociales para compartir contenido político, siendo evidente que existe mayor afinidad por los partidos como Movimiento Centro Democrático, Partido Sociedad Patriótica, Movimiento Alianza País, entre otras.

RECOMENDACIONES

- Es importante explorar dentro del campo de la minería de datos, debido a que se constituye en una herramienta de gran interés tanto en los negocios como en su dimensión social, puesto que puede ser combinada con otras fuentes que permitan obtener una información más completa en el ámbito que se le quiera aplicar.
- La minería de datos es una herramienta que ha despertado el interés de diferentes organizaciones e instituciones debido a que permite conocer mejor las opiniones de los clientes, usuarios u otras personas sobre un tema determinado, lo que permitirá tener un alto porcentaje de acertación en la toma de decisiones y mejorar así la productividad de cualquier empresa pública o privada.
- Para que la información sea completa y se obtenga los resultados requeridos, es necesario diseñar una herramienta con datos precisos, pues estos ayudarán a una proyección detallada de la información que se busca, por lo cual es importante el uso de diferentes herramientas tecnológicas, mismas que ayudarán en la obtención y el análisis de datos a gran escala.

Bibliografía

- Benalcázar, J. (2017). *Análisis comparativo de metodologías de minería de datos y su aplicabilidad a la industria de servicios*. Quito: Universidad de las Américas. Recuperado el mayo de 2018, de <http://dspace.udla.edu.ec/bitstream/33000/7547/6/UDLA-EC-TMGSTI-2017-11.pdf>
- Contreras, L., & Rosales, K. (2016). *Análisis del comportamiento de los clientes en las redes sociales mediante técnicas de Minería de Datos*. Colombia: Universidad Popular del Cesar. Recuperado el mayo de 2018, de <http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/661/COMTEL%202016%20-%20Paper18.pdf?sequence=1&isAllowed=y>
- Corbella, J. (2018). Las noticias falsas se difunden más rápido que las ciertas. *la vanguardia*. Obtenido de <https://www.lavanguardia.com/ciencia/20180309/441373100634/noticias-falsas-se-difunden-mas-rapido-twitter.html>
- Cortés, R., Zapata, A., Menéndez, V., & Canto, P. (mayo de 2015). El estudio de los hábitos de conexión en redes sociales virtuales, por medio de la minería de datos. *Innovación Educativa*, 15(68). Recuperado el mayo de 2018, de <http://www.scielo.org.mx/pdf/ie/v15n68/v15n68a7.pdf>
- Del Canto, E., & Silva, A. (2013). Metodología cuantitativa: abordaje desde la complementariedad en Ciencias Sociales. *Revista de Ciencias Sociales*, 3(141), 25-34. Recuperado el 2018, de <http://www.redalyc.org/pdf/153/15329875002.pdf>
- Díaz, P. (2017). *Aplicación de técnicas de minería de datos para el análisis de sentimientos en las redes sociales sobre productos de la marca fideos Cayambe*. Quito: Escuela Superior Politécnica del Ejército. Recuperado el mayo de 2018, de <https://repositorio.espe.edu.ec/bitstream/21000/13528/1/T-ESPE-053887.pdf>
- Flores, J., Morán, J., & Rodríguez, J. (2015). *Las redes sociales*. Lima: Universidad de San Martín de Porres.
- Hüt Herrera, H. (2012). *Las redes sociales una nueva herramienta de difusión*. Costa Rica: Universidad de Costa Rica. Recuperado el mayo de 2018, de <http://www.redalyc.org/pdf/729/72923962008.pdf>

- Ibarra, M. (2006). *Procesamiento Analítico en Línea(OLAP)*. Universidad Nacional del Nordeste. Recuperado el mayo de 2018, de <http://exa.unne.edu.ar/informatica/SO/OLAPMonog.pdf>
- Jaramillo, S., Cardona, S., & Fernández, A. (2015). Minería de datos sobre streams de redes sociales, una herramienta al servicio de la Bibliotecología. *Información, cultura y sociedad*, 33, 63-74. Recuperado el mayo de 2018, de <http://www.redalyc.org/pdf/2630/263042678005.pdf>
- Kadushin, C. (2012). *Comprender las redes sociales*. Madrid: Centro de investigaciones sociológicas.
- Martínez, C. (2012). *Aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión en entel*. Santiago de Chile: Universidad de Chile. Recuperado el mayo de 2018, de http://repositorio.uchile.cl/bitstream/handle/2250/112065/cf-martinez_ca.pdf?sequence=1
- Martínez, R., & Rodríguez, E. (2016). *Manual de metodología de la Investigación científica*. Recuperado el junio de 2018, de http://www.sld.cu/galerias/pdf/sitios/cielam/manual_de_metodologia_deinvestigaciones._1.pdf
- Molina, J., & García, J. (2006). *Aplicaciones prácticas utilizando microsoft excel y weka*. Madrid, España: Universal Carlos Tercero de Madrid. Recuperado el mayo de 2018, de http://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf
- Morales, Ó. (2004). *Fundamentos de la investigación documental y la monografía*. Web del profesor. Recuperado el junio de 2018, de <http://webdelprofesor.ula.ve/odontologia/oscarula/publicaciones/articulo18.pdf>
- Muñoz, V. (2002). *Técnicas de investigación de campo*. México: Escuela Nacional de Biblioteconomía y Archivonomía.
- Nieto, M. (2016). *Plan de negocios para la creación de una empresa de servicios de minería de datos en Colombia*. Bogotá: UNIR. Recuperado el mayo de 2018, de <https://reunir.unir.net/bitstream/handle/123456789/4838/NIETO%20PATARROY%20MARCO%20ANTONIO.pdf>

- NN. (2015). La política y las redes sociales. *La vanguardia*. Obtenido de <https://www.lavanguardia.com/vida/20150521/54431779373/politica-redes-sociales.html>
- Otzen, T., & Manterola, C. (2017). Técnicas de Muestreo sobre una Población a Estudio. *Int. J. Morphol*, 35(1), 227-232. Recuperado el junio de 2018, de <https://scielo.conicyt.cl/pdf/ijmorphol/v35n1/art37.pdf>
- Pérez, C. (2007). *Minería de datos técnicas y herramientas*. Madrid, España: Paraninfo. S.A. Recuperado el Mayo de 2018, de https://books.google.com.ec/books?id=wz-D_8uPFCEC&printsec=frontcover&hl=es#v=onepage&q&f=false
- Pérez, M., & López, F. (2015). *El fenómeno de las redes sociales: Evolución y perfil de usuario*. Madrid: Universidad Camilo José Cela. Recuperado el mayo de 2018, de https://www.researchgate.net/publication/287209549_El_fenomeno_de_las_redes_sociales_evolucion_y_perfil_del_usuario
- Roales, N. (2014). *Detección de tendencias en twitter utilizando minería de datos adaptativa*. Universidad Autónoma de Madrid. Recuperado el mayo de 2018, de https://repositorio.uam.es/bitstream/handle/10486/662510/roales_gonzalez_natalia_tfg.pdf?sequence=1&isAllowed=y
- Rodríguez, Y., & Díaz, A. (julio-diciembre de 2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 3(3-4), 77. Recuperado el mayo de 2018, de <http://www.redalyc.org/pdf/3783/378343637009.pdf>
- Secretaría Marina. (2016). *Metodología de la investigación*. México: Universidad Naval. Recuperado el junio de 2018, de https://www.gob.mx/cms/uploads/attachment/file/133491/METODOLOGIA_DE_INVESTIGACION.pdf
- Timarán, S. R., Hernández, I., Caicedo, S. J., Hidalgo, A., & Alvarado, J. (2016). *El proceso de descubrimiento de conocimiento en base de datos*. Bogotá: Universidad Cooperativa de Colombia. Recuperado el mayo de 2018, de ediciones.ucc.edu.co/index.php/ucc/catalog/download/36/40/230-1

APLICACIÓN DE LA MINERÍA DE DATOS PARA ANÁLISIS DE TENDENCIAS POLÍTICAS EN REDES SOCIALES

Esmeralda Proaño A.¹, Ing. Fidel David Parra Balza M. Ph. D.

Universidad Tecnológica Israel, Ecuador

e-mail: esmeraldap2007@gmail.com¹

RESUMEN

Actualmente las Redes Sociales se encuentran inmersas en todos los campos del diario vivir, en áreas como ingeniería, arquitectura, salud, economía, política, etc., por lo que se han constituido en un factor de gran peso en la decisión del usuario, por tal motivo el presente trabajo se desarrolla con la finalidad de utilizar la Minería de Datos como herramienta para transformar la información recopilada de las redes sociales sobre las tendencias políticas. La investigación parte del estudio teórico de la minería de datos con la definición de conceptos necesarios para el proceso, posteriormente se diseñó la herramienta para la recopilación de información, la cual fue aplicada a los usuarios de redes sociales como Facebook, Twitter, WhatsApp e Instagram, debido a que se estudió acerca de las tendencias políticas en dichas redes. Del total de la población, mediante un muestreo no probabilístico por conveniencia, se seleccionó un total de 1016 personas para la aplicación de la encuesta, con la cual se pudo obtener la información necesaria para realizar el análisis requerido con la técnica de minería de datos. Posteriormente para el procesamiento de información se utilizó la metodología Data Warehouse que es una base de datos corporativos encargada de integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista, con lo cual se pudo determinar que los usuarios si utilizan las redes sociales para compartir contenido político, siendo evidente que existe mayor afinidad por los partidos como Movimiento Centro Democrático, Partido Sociedad Patriótica, Movimiento Alianza País, entre otras.

Palabras claves: Data Warehouse, minería de datos, política, redes sociales

ABSTRACT

Currently Social Networks have been immersed in all fields of daily life, in areas such as engineering, architecture, health, economics, politics, etc., so they have become a factor of great weight in the decision of the user, for this reason the present work is developed with the purpose of using Data Mining as a tool to transform the information collected from social networks on political trends. The investigation starts from the theoretical study of data mining with the definition of necessary concepts for the process, later the tool was designed for the collection of information, which was applied to users of social networks such as Facebook, Twitter, WhatsApp and Instagram, because we studied about the political trends in these networks. Out of the total population, through a non-probabilistic sampling for convenience, a total of 1016 people were selected for the application of the survey, with which it was possible to obtain the necessary information to perform the required analysis with the data mining technique. Later, for the processing of information, the Data Warehouse methodology was used, which is a corporate database in charge of integrating and debugging the information and then processing it, allowing the analysis of the information from different points of view, with which it was possible to determine that the users if they use social networks to share political content, being evident that there is greater affinity for parties such as Democratic Center Movement, Patriotic Society Party, Alianza País Movement, among others.

Keywords: Data Warehouse, data mining, politics, social networks

1. INTRODUCCIÓN

En los últimos años, la industria de las comunicaciones ha centrado su atención en sistemas que transportan datos a largas distancias. La industria de la informática y de las redes WAN (Wide Area Network) es un sector estable en constante desarrollo abarcando cada vez más y más actividades personales, profesionales y empresariales.

Tradicionalmente las personas disponían de su propia computadora, con la cual ejecutaban sus propias aplicaciones y administraban sus actividades e información de forma independiente conectada a una red de recursos, pronto surgió la necesidad de compartir esos recursos con el entorno social acoplándose a la tendencia de las comunicaciones globales para lograr los objetivos establecidos.

El factor tecnológico disponible solo al entorno profesional, pronto llegó al sector personal cubriendo un entorno social muy diverso con la aplicación de las TIC (Tecnologías de Información y Comunicación) que engloban un amplio rango de las Redes Sociales (RRSS) disponibles, las cuales están siendo utilizadas por personas sin diferenciar nivel social, género, ni edad.

Con estas aseveraciones se deduce que desde hace años atrás se ha evidenciado la necesidad de administrar la información que generan todo tipo de organizaciones, lo que exige llevar de manera ordenada datos que faciliten la toma de decisiones, por tal motivo se promueve la generación masiva de información que creen históricos de los movimientos que realizan las organizaciones, sin embargo actualmente

en la mayoría de los casos este control se lleva de forma no muy desarrollada, por lo que los resultados no son los que se desean, (Benalcázar, 2017).

Dentro de las organizaciones mundiales se encuentran las redes sociales las cuales indiscutiblemente juegan un papel importante en la vida social, económica y política de las personas, debido a que han permitido transformar la dinámica para la comunicación y adquisición de información, llegando a ser un factor de gran peso en la decisión del usuario despreciando la veracidad o falsedad de la información.

De acuerdo a ello, en la actualidad la técnica de minería de datos está siendo aplicada de forma muy notoria en diversas áreas, tales como el marketing, finanzas, medicina, salud pública, educación, entre otras, debido a que la aplicación de esta técnica sobre los datos provenientes de fuentes como las redes sociales pueden revelar patrones sobre individuos inmersos en el ambiente compartido y producir conocimiento que años atrás no era factible encontrar, pues la información era muy variada y compleja, (Jaramillo, Cardona, & Fernández, 2015).

Por lo expuesto anteriormente, el presente trabajo de investigación tiene como finalidad analizar las redes sociales de mayor impacto en conjunto con los diferentes algoritmos de minería de datos que puedan utilizarse para generar conocimiento a partir de la información de dichas redes.

2. DESARROLLO

Aproximación al concepto de minería de datos

Inicialmente la minería de datos se puede definir como un proceso en el cual se analiza una gran cantidad de datos con la finalidad de descubrir nuevas relaciones, tendencias y patrones significativos, (Pérez, 2007).

El análisis de datos toma el nombre de *Data Mining* (Minería de datos) gracias a la disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas cuyo objetivo es descubrir de manera automática el conocimiento contenido en una extensa base de datos.

Las técnicas que se utilizan en esta metodología tienen como meta descubrir patrones, perfiles y tendencias mediante el análisis de los datos haciendo uso de tecnologías de reconocimiento de patrones, lógica difusa, redes neuronales, algoritmos genéticos y otras técnicas especiales de análisis de datos.

EIS: Executive Information System

Es un sistema que permite a los ejecutivos tomar decisiones en tiempo real, el cual no solo consta del software y de hardware sino también de toda la infraestructura que permite tener la información completa al alcance de su mano representada de manera gráfica al estilo de hojas de cálculo (ventas semanales, inventarios, balances, etc), (Nieto, 2016).

OLAP (On-Line Analysis Planning)

Se denominan sistemas OLAP (Procesamiento analítico en línea) a aquellos que deben analizar datos desde diferentes perspectivas y soportar

requerimientos complejos de análisis en un volumen ingente de datos.

Este sistema se caracteriza por ser una técnica de estudio multidimensional de datos en el cual el usuario puede navegar por los mismos de modo asistido, (Molina & García, 2006)

En este sistema la información es vista como cubos los cuales están formados por categorías descriptivas (dimensiones) y valores cuantitativos (medidas) (Ver figura 1 y 2).

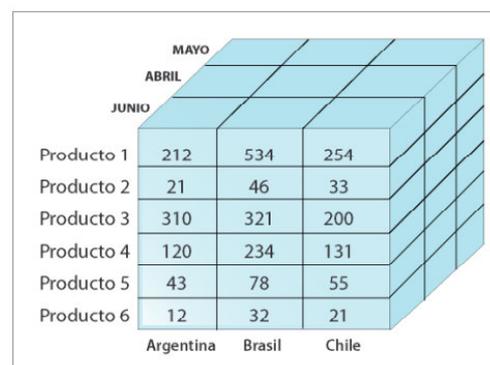


Figura 1. Vista de los datos en forma de cubo
Fuente: (Ibarra, 2006)

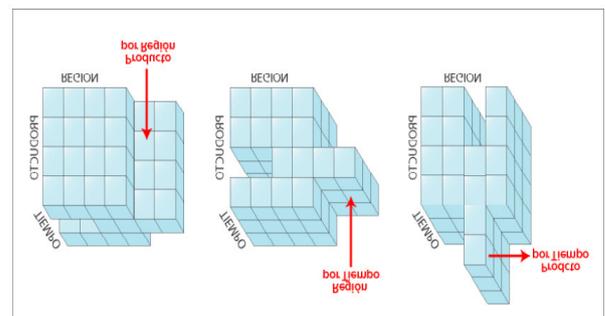


Figura 2. Vista de los datos en forma de cubo
Fuente: (Ibarra, 2006)

Las ventajas de presentar de esta manera la información es que facilita a los usuarios diseñar consultas complejas, ordenar datos en un reporte, filtrar o disgregar datos en subconjuntos significativos y cambiar de datos resumidos a datos detallados.

Diseño de una Data Warehouse

Se han propuesto diferentes metodologías para el diseño y construcción de DW las cuales tienen en común las siguientes etapas:

1. **Analizar los requerimientos del usuario final:** es necesario que las herramientas y aplicaciones entreguen información útil de acuerdo a la decisión estratégica solicitada por el cliente.
2. **Seleccionar las fuentes de información:** luego de haber definido el requerimiento final del cliente se buscan las fuentes de información que permitan cumplir con el objetivo planteado.
3. **Elaborar un modelo lógico de datos:** los más utilizados son el modelo cubo y el modelo estrella.
4. **Generar un prototipo para el cliente:** la meta es indicar al usuario una idea preliminar de DW a desarrollar para que este apruebe y asegure que el modelo se ajusta a sus requerimientos.
5. **Seleccionar el sistema administrador de base de datos (SABD):** este sistema se encarga de la velocidad de respuesta en las consultas por lo cual se convierte en una de las variables más importantes dentro del desarrollo del DW.
6. **Diseño físico de la BDD:** es la implementación física del modelo lógico de datos en base a la estructura que entregue el SABD (relaciones entre el modelo de datos y las tablas).
7. **Almacenamiento:** guardar la información a través de un trabajo de extracción, conversión y carga de datos.

8. **Mejora continua del desempeño obtenido:** si es necesario realizar modificaciones a la estructura interna de datos para mejorar el tiempo de respuesta ante las consultas, (Martínez, 2012).

Proceso de extracción del conocimiento a partir de datos (KDD)

El *Knowledge Discovery in Databases* es un proceso automático en el que se combinan análisis y descubrimiento. La tarea consiste en extraer patrones en forma de reglas o funciones para que el cliente los revise.

Las etapas básicas de este proceso son, reprocesar los datos, hacer minería de datos y presentar resultados.

Etapas del proceso KDD

- **Selección:** De acuerdo a los requerimientos del usuario y metas del proceso se crea un grupo de datos objetivo, escogiendo todo el conjunto de datos o una muestra de este, sobre el cual se efectuará el proceso de análisis.
- **Procesamiento / Clasificación:** En esta etapa se disgrega los datos de menor de calidad aplicando la técnica de remoción de datos ruidosos, datos nulos, datos duplicados. En esta fase en de suma importancia el trabajo a la par entre el usuario y el analista.
- **Conversión/reducción:** Se utilizan técnicas de reducción de dimensiones para disminuir el número de variables simplificando una tabla o base de datos vertical u horizontalmente. Las más utilizadas son agregaciones, compresión de datos, segmentación, histogramas, muestreo, etc.
- **Data mining (Minería de datos):** El objetivo de esta fase es descubrir patrones inimaginables e importantes

aplicando las técnicas anteriormente expuestas.

- **Análisis de resultados /evaluación:** Interpretación de patrones descubiertos y generalmente se repite el ciclo en busca de nuevos hallazgos.

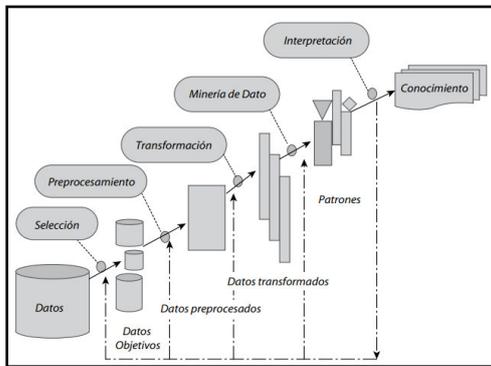


Figura 3. Etapas del proceso KDD

Fuente: (Timarán, Hernández, Caicedo, Hidalgo, & Alvarado, 2016)

Las redes sociales

El término *networking* se ha convertido en una palabra común, todas las personas van a una fiesta para gestionar su red de contactos, para un gran porcentaje de población la *World Wide Web* tiene como principal objetivo establecer contactos, (Kadushin, 2012).

Las redes sociales han sido parte de nuestras vidas desde la época de los recolectores y cazadores, los individuos se generaban vínculos entre sí a través de sus relaciones unas con otras y su dependencia entre ellos. Las relaciones familiares y el parentesco son redes sociales, el barrio las ciudades son redes sociales. Fuera del parentesco familiar las personas de la sociedad moderna dependen unas de otras para el servicio de taxis, la venta del periódico el dentista, etc. Es decir, el parentesco en comunidades se forma a través de las necesidades y en torno a las redes sociales geográficamente dispersas es por ellos que se dice que llevamos milenios siendo *networkers*.

El *networking* es el uso de una de red de contactos para alcanzar diferentes objetivos personales, por lo tanto una red es simplemente un grupo de relaciones entre objetos que pueden ser organizaciones, personas, naciones en las cuales se comparte ideas de amistad, dinero, poder e incluso la enfermedad, (Flores, Morán, & Rodríguez, 2015).

Hace solo algunos años (2003) se ha venido manejando el concepto de las redes sociales, pero como se describe en los párrafos anteriores esto no es del todo nuevo, sin embargo, lo que es relativamente nuevo es la manera de referirse a ellas describiéndolas analizándolas y relacionándolas con instancias sociales más formales como organizaciones y gobiernos.

La red social viéndolo desde el lado informático se denomina como un servicio que brinda la oportunidad a los individuos de construir un perfil semipúblico o público dentro de un sistema delimitado, socializar con otros usuarios con los que comparten una conexión, ver y recorrer su historial de conexiones y consultar las de los demás usuarios del sistema, (Flores, Morán, & Rodríguez, 2015).

3. RESULTADOS

En la presente investigación se tomó como población a todos los usuarios de redes sociales como Facebook, Twitter, WhatsApp e Instagram, debido a que se estudió acerca de las tendencias políticas en dichas redes. Del total de la población, mediante un muestreo no probabilístico por conveniencia, que de acuerdo a Otzen & Manterola (2017) permite seleccionar aquellos casos accesibles que acepten ser

incluidos en el proceso o se encuentren disponibles en el tiempo o periodo de investigación, se seleccionó un total de 1016 personas para la aplicación de la encuesta, con la cual se pudo obtener la información necesaria para realizar el análisis requerido con la técnica de minería de datos.

Para el procesamiento de la información se utilizó la metodología DataWarehouse, traducido como almacén de datos, misma que se trata de una base de datos corporativa que se caracteriza por integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista. La creación de un DataWarehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

Para el desarrollo del análisis y construcción del Data Warehouse se utilizó la Metodología Hefesto.

Es importante describir que también se utilizaron técnicas de limpieza y calidad de datos, procesos ETL y se realizó la carga inicial de información al Data Warehouse.

En este paso se debe realizar la Carga Inicial del DW, poblando el modelo construido, para lo cual se debe llevar adelante una serie de tareas básicas, tales como asegurar la limpieza y calidad de los datos, procesos ETL, etc.

En muchos casos, las tareas antes mencionadas tienen una lógica compleja, afortunadamente, en la actualidad existen muchas herramientas de software que se pueden emplear y que facilitan en gran parte el trabajo.

Una vez realizado la carga de información, se genera la Base de Datos a partir del Diagrama Físico del modelado, (Ver figura 4).

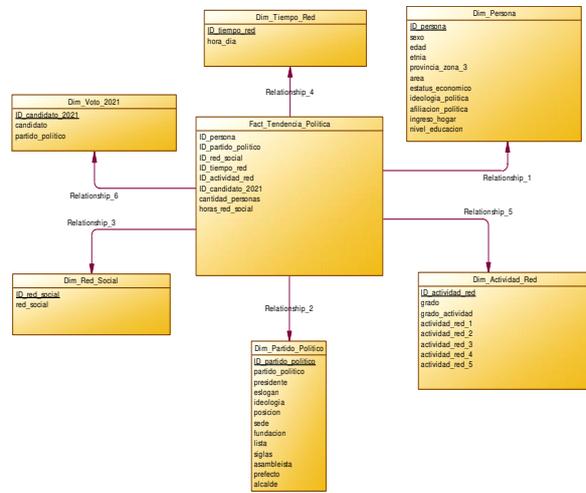


Figura 4. Diagrama físico del modelado.

Después de haber generado el modelo físico y haber realizado la carga de información, se muestra el grado de actividad de las personas en redes sociales, respecto a la ideología política que las personas indicaron tener. Además, se realiza la interpretación gráfica de distintos modelos, cabe recalcar que las herramientas del cubo permiten girar las dimensiones permitiendo otra perspectiva de análisis.

Una vez realizado el proceso Data Warehouse junto con la ayuda del software Spagobi, se puede visualizar información que representa la tendencia política de los usuarios encuestados, (Ver figura 5):



Figura 5. Tendencias políticas

En este caso se ha hecho el análisis de la cantidad de personas que se definieron con ideología política centro, para saber de este grupo de personas el tiempo ocupado en redes sociales con respecto a su género, (Ver figura 6).

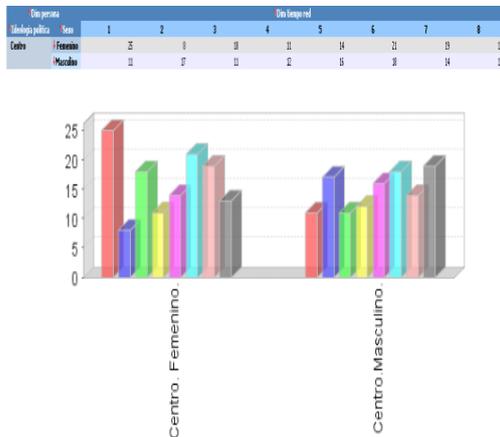


Figura 6. Ideología relacionada con el género

4. CONCLUSIONES

- En el estudio se puede evidenciar que las redes sociales han llegado a la sociedad para quedarse, pues permiten mantener la comunicación entre personas de diferentes culturas, etnias, creencias, costumbres, entre otros, extendiendo sus campos y llegando a todas los segmentos del mundo, ya que miles de usuarios expresan sus sentimientos, conocimientos y afinidades a través de esta información que puede ser aprovechada para proyecciones en diversos campo como la política.
- La forma de agrupación de la información en la minería de datos permite observar diferentes perspectivas a las bases de datos, por ejemplo, una agrupación de la base de datos de tendencias

políticas podría exponer todos los partidos políticos que existen dentro de las regiones, mientras que otro grupo podría mostrar los partidos políticos a los que los usuarios tienen mayor afinidad.

- Con la metodología Data Warehouse que es una base de datos corporativos encargada de integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista, se pudo determinar que los usuarios si utilizan las redes sociales para compartir contenido político, siendo evidente que existe mayor afinidad por los partidos como Movimiento Centro Democrático, Partido Sociedad Patriótica, Movimiento Alianza País, entre otras.

5. BIBLIOGRAFÍA

- Benalcázar, J. (2017). *Análisis comparativo de metodologías de minería de datos y su aplicabilidad a la industria de servicios*. Quito: Universidad de las Américas.
- Flores, J., Morán, J., & Rodríguez, J. (2015). *Las redes sociales*. Lima: Universidad de San Martín de Porres.
- Ibarra, M. (2006). *Procesamiento Analítico en Línea(OLAP)*. Universidad Nacional del Nordeste.
- Jaramillo, S., Cardona, S., & Fernández, A. (2015). Minería de datos sobre streams de redes sociales, una herramienta al servicio de la

Bibliotecología. *Información, cultura y sociedad*, 33, 63-74.

Kadushin, C. (2012). *Comprender las redes sociales*. Madrid: Centro de investigaciones sociológicas.

Martínez, C. (2012). *Aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión en entel*. Santiago de Chile: Universidad de Chile.

Molina, J., & García, J. (2006). *Aplicaciones prácticas utilizando microsoft excel y weka*. Madrid, España: Universal Carlos Tercero de Madrid.

Nieto, M. (2016). *Plan de negocios para la creación de una empresa de servicios de minería de datos en Colombia*. Bogotá: UNIR.

Pérez, C. (2007). *Minería de datos técnicas y herramientas*. Madrid, España: Paraninfo. S.A.

Timarán, S. R., Hernández, I., Caicedo, S. J., Hidalgo, A., & Alvarado, J. (2016). *El proceso de descubrimiento de conocimiento en base de datos*. Bogotá: Universidad Cooperativa de Colombia.