

Datawarehouse de una base de datos de migración

Datawarehouse of migration data base

Angel Agustín Caymayo Cuesta^{1*}, Juan Pablo Cuenca Tapia¹

¹ Universidad Católica de Cuenca

*aacaimayoc60@est.ucacue.edu.ec

DOI: https://doi.org/10.26871/killkana_tecnica.v4i1.588

Resumen

La institución en estudio necesita extraer la información de los datos almacenados en una Base de Datos de Migración, a un archivo de Excel, para conocer cómo se desarrolló la migración registrada en el año 2017 y presentar un proceso de toma de decisiones más adecuado. En la investigación que se expone se plantea ejecutar la metodología HEFESTO, con la finalidad de obtener los requerimientos, las necesidades y los pedidos que los usuarios competentes realizaron. Con los siguientes softwares: las Herramientas de Tableau (permite presentar los resultados), DbSchema (posibilita la manipulación y la muestra de los datos), Power Designer (crea el cubo con sus dimensiones y hechos), y Knime (proporciona el trabajo con la minería de datos), se puede mostrar a los directivos la información solicitada, mediante la representación en gráficos, con una ilustración eficiente y entendible. Datawarehouse es una ciencia que mediante la manipulación de los datos, se puede extraer información a partir de los datos guardados, y facilita la toma de decisiones y el análisis de los resultados. La institución presta los debidos recursos físicos, lógicos, económicos, la disponibilidad y el tiempo, los cuales quedaron establecidos en las reuniones con los interesados para la realización del presente trabajo y obtener un producto final eficaz.

Palabras clave: Base de Datos, Datawarehouse, migración, herramientas de consulta e informe.

Abstract

The institution under study needs to extract information from the data stored in a Migration Database, to transfer it to an Excel document to find out how the migration registered in the year 2017 was done and present a more suitable decision-making process. The HEFESTO methodology is proposed in this research so that the competent users' needs and requirements can be satisfied. To present the requested information to the authorities, the following software programs: Tableau Tools (for results presentation), DbSchema (for data manipulation and display), Power Designer (for cube dimensions and facts creation), and Knime (provides the work with data mining), and the representation of efficient and easy to understand illustrations and graphs, will be used. Datawarehouse is a science that, through data manipulation, allows extracting information from stored data, facilitating decision making and the analysis of the results. The institution provides the appropriate physical, logical, economic resources, availability, and time; which were established in the meetings with the interested parties to carry out this work, and thus get a useful final product.

Keywords: Database, Datawarehouse, migration, query and reporting tools.

I. INTRODUCCIÓN

En el mundo la migración es un factor de mucho interés y preocupación, en especial en los países donde más se desarrolla como: Colombia, México, Perú, Venezuela, Ecuador, países de centro América, países del continente africano, etc., por distintas razones, que hay que resolver. En el Ecuador institutos como el INEC o el Ministerio del Interior, registran los datos de los migrantes para tomar medidas y con ello reducir la cantidad y/o ayudar a la comunidad migrante.

En otros países como Colombia, existe D. A. N. E. que es un Departamento encargado de llevar las estadísticas, incluida las de migración, e informar al gobierno sobre los resultados de cada año.

Para el análisis de los datos se va a aplicar Datawarehouse, para obtener información pertinente para los Directivos de la empresa.

Aplicar las técnicas que Datawarehouse tiene, para extraer un producto final eficiente, requerido y de calidad, que ayudara a conocer el estado de un proceso y tomar decisiones al respecto.

II. MARCO TEÓRICO

A. Migración

Desplazamiento geográfico de individuos o grupos, generalmente por causas económicas o sociales. [1]

Los movimientos de entradas y salidas de ecuatorianos y extranjeros es una investigación orientada a cuantificar el

movimiento internacional ocurrido en el país, según las vías de transporte utilizadas. [2]

La información de carácter individual es recolectada por las Jefaturas de Migración que funcionan en el territorio nacional a través de la Dirección Nacional de Migración. [2]

El INEC es el encargado del procesamiento. [2]

B. ¿Qué es Datawarehouse?

Un Datawarehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. [3]

Un dashboard extrae los datos de su almacén de datos y los transforma en una serie de cuadros, gráficos y otras visualizaciones que se actualizan en tiempo real para que pueda usarlos para tomar decisiones comerciales clave. [4]

III. METODOLOGÍA

Se utiliza la metodología llamado HEFESTO, para desarrollar ordenadamente, cada fase de un Datawarehouse, con los requerimientos y necesidades de información de los usuarios finales, para la toma de decisiones.

Los pasos de HEFESTO son:

- Análisis de requerimientos
- Análisis de los OLTP
- Modelo lógico del Datawarehouse
- Integración de datos

IV. DESARROLLO

A. Análisis de requerimientos

Se hizo reuniones con los usuarios, empleados competentes y directivos.

Los directivos supieron manifestar que necesitan de datos que les permita tomar decisiones o conocer el flujo de migración. Realizar el estudio y análisis de las actividades diarias importantes, que soporta algún tipo de OLTP. Enfocarse en la cantidad de migrantes de los registros disponibles en periodos de tiempo cortos, ya que los registros son tomados del año 2017.

1. Identificar preguntas

Existen datos que son de mucha importancia, por lo que se debe extraer información a partir de ella.

Se plantean las siguientes preguntas, atendiendo a los requerimientos que los distintos tipos de usuarios supieron manifestar:

- 1) ¿Cuántas personas con residencia ecuatoriana, por mes del año 2017, han migrado?
- 2) ¿Cuántas personas, teniendo en cuenta el género, han migrado a los diferentes países de procedencia de destino, de cada mes del año 2017?

- 3) ¿Cuál es el promedio de edad de las personas, por nacionalidad y género, que han migrado en los 3 primeros trimestres del 2017?
- 4) ¿Cuántas personas que han residido en los diferentes continentes han migrado en los semestres del 2017?
- 5) ¿Cuáles son las ocupaciones que generaron más migrantes, en los cuatrimestres del 2017?
- 6) ¿Cuál es el índice de cada cantón del Ecuador, que los migrantes han necesitado hacer escalas, de cada bimestre, en el 2017?
- 7) ¿Cantidad de migrantes que han hecho escalas en los diferentes países, de cada bimestre, en el 2017?
- 8) ¿Cuáles son las jefaturas de provincias y cantones donde se registraron procesos migratorios en los cuatrimestres del 2017 y cuantos por cada cantón?

El resultado de esta etapa nos muestra un conjunto de preguntas que al resolver van a atender a los requerimientos de los usuarios. Las presuntas van a servir para la siguiente etapa ya que son el modelo de negocio, con el objetivo de extraer las perspectivas y los indicadores.

2. Identificador indicadores y perspectivas

Las preguntas obtenidas permitieron la extracción de las perspectivas y los indicadores, como se muestran a continuación:

TABLA I: Identificación de Perspectivas e indicadores

PERSPECTIVAS	INDICADORES
Migrantes	Cantidad
Tiempo	Promedio
Viajes	
Detalle	

Fuente: Autor del proyecto

Los indicadores y perspectivas son de importancia ya que dan la dirección que la metodología HEFESTO necesita, además permite graficar e ilustrar el modelo conceptual que se efectúa en la siguiente fase.

3. Modelo conceptual

Se procede a graficar el modelo conceptual, como se muestra a continuación:

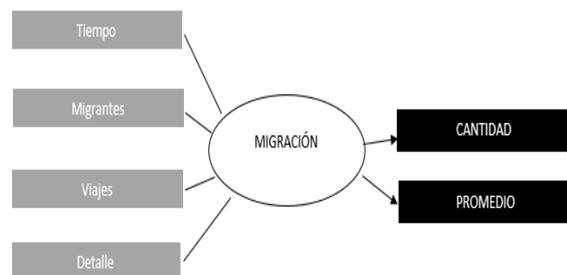


Fig. 1: Modelo Conceptual de la empresa

Fuente: Autor del proyecto

En la parte izquierda se muestra las perspectivas (Tiempo Migrantes, Viajes y Detalle), en el centro la relación (MIGRACIÓN) y en la parte derecha se muestra los indicadores (CANTIDAD y PROMEDIO)

El Modelo conceptual es el resultado final de la etapa A, que se refiere a el análisis de requerimientos.

B. Análisis de los OLTP

En la anterior fase se obtuvieron los indicadores:

- CANTIDAD de migrantes
- PROMEDIO de edad

1. Conformar indicadores

Con los indicadores establecidos, se procede a conformarlos, demostrando como se va a extraer o calcular, y que operadores son los que se van a usar.

TABLA II: Indicadores, función y operación

INDICADOR	HECHOS	FUNCIÓN DE SUMARIZACIÓN	OPERACIÓN
CANTIDAD de migrantes	cod_migran	COUNT	COUNT(cod_migran)
PROMEDIO de edad	edad	AVG	AVG(edad)

Fuente: Autor del proyecto

C. Establecer correspondencias

En el OLTP de la institución analizada, tiene el proceso de migración digitalizada y los datos almacenados en una Base de datos, además, está estructurada la B. D. como se muestra en el siguiente Diagrama de Entidad-Relación (Figura 3):

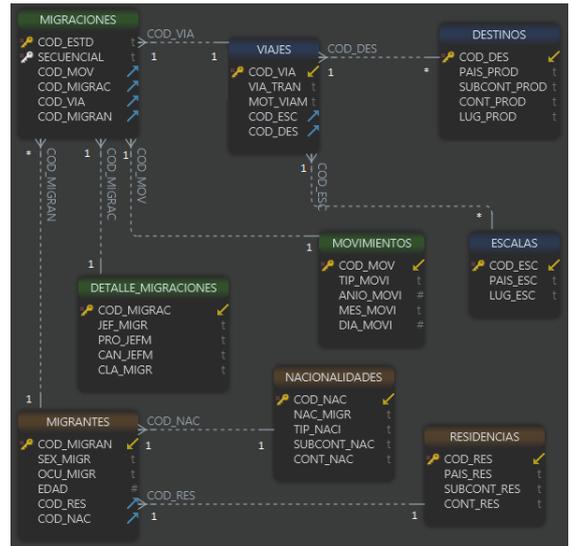


Fig. 3: Modelo Entidad-Relación con la cardinalidad, extraída de DbSchema

Fuente: Autor del proyecto

En la fase anterior se pudo saber de dónde se va a extraer los datos para los indicadores, ahora vamos a proceder a identificar la relación entre los componentes de los modelos conceptual y el modelo de entidad-relación. La correspondencia entre los modelos se puede apreciar en la siguiente figura 2

Al ya tener la correspondencia entre modelos que se extrajo como resultado de esta fase, servirá para poder tener referencia de dónde se va a tomar los datos.

1. Nivel de granularidad

Se buscan los campos que van a ser necesarios para la construcción del Modelo Conceptual ampliado, estos campos tienen que estar ligados a las preguntas antes extraídas y al modelo de negocio de la empresa.

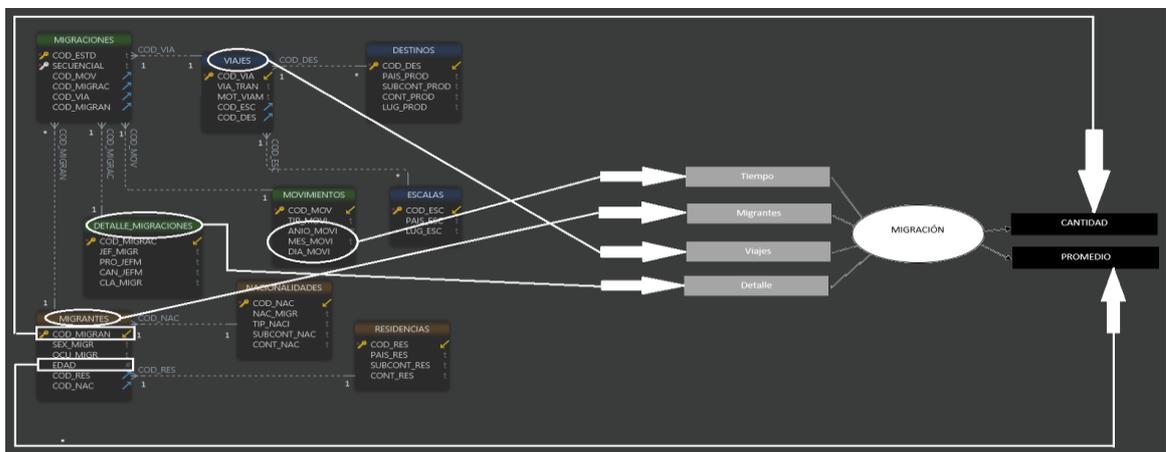


Fig. 2: Correspondencia del Modelo Entidad-Relación y el Modelo Conceptual

Fuente: Autor del proyecto

Los campos son muy necesarios ya que a través de estos se analizarán y filtrarán los indicadores. Es necesario conocer que significa cada campo y con qué tipo de datos se los puede llenar.

Es importante definir el tiempo, la granularidad que se utilizara para poder agrupar los datos o incluso para sumarizar.

Para entender lo que significa cada campo mostrado en el Modelo Entidad-Relación se procedió a reunirse con el Administrador de la Base de Datos, quien gentilmente, ayudo a comprender las características y el sentido de cada campo, su función, además su relevancia. Así también, entrego el diccionario de datos de la Base de Datos.

Al analizar el nivel de granularidad, tomando en cuenta las preguntas y también en coordinación con los usuarios y directivos, determinó los campos necesarios que se tendrán en cuenta para poder filtrar los datos.

TABLA III: Nivel de granularidad, obtenida del Diccionario de datos de Excel

No.	Nombre del Campo (Encabezado Columna)	Descripción del Campo		NIVEL DE GRANULARIDAD
1	cod_estd	Código Estadística		
2	secuencial	Secuencial		
3	cod_mov	Código de movimiento		
4	tip_movi	Tipo de movimiento		
5	anio_movi	Año de movimiento		
6	mes_movi	Mes de movimiento		
7	día_movi	Día de movimiento		
8	cod_nac	Código de Nacionalidad		
9	nac_migr	Nacionalidad		
10	tip_naci	Tipo de Nacionalidad		
11	subcont_nac	Subcontinente de Nacionalidad		
12	cont_nac	Continente de Nacionalidad		
13	cod_via	Código del viaje		
14	via_tran	Medio de transporte		
15	mot_viam	Motivo de viaje		
16	cod_des	Código de Procedencia Destino		
17	pais_pro	País de Procedencia Destino		
18	subcont_pro	Subcontinente de Procedencia destino		
19	cont_pro	Continente de Procedencia Destino		
20	lug_pro	Lugar de Procedencia Destino		
21	cod_res	Código de Residencia		
22	pais_res	País de Residencia		
23	subcont_res	Subcontinente de Residencia		
24	cont_res	Continente de Residencia		
25	cod_migrac	Código de Migración		
26	jef_migr	Jefatura de Migración		
27	pro_jefm	Provincia Jefatura de Migración		
28	can_jefm	Cantón Jefatura de Migración		
29	cla_migr	Clase de Migración		
30	cod_migran	Código de Migrante		
31	sex_migr	Sexo		
32	ocu_migr	Ocupación		
33	edad	Edad		
34	cod_esc	Código de Escala		
35	pais_esc	País de Escala		
36	lug_esc	Lugar de Escala		
37	cod_tiem	Código del tiempo		
38	mes_let	Mes en letras		
39	mes_num	Mes en números		
40	bimestre	Bimestre del año		
41	trimestre	Trimestre del año		
42	cuatrimestre	Cuatrimestre del año		
43	semestre	Semestre del año		
44	anio	Año		

Fuente: Autor del proyecto

En la última columna se determina los campos que se van a tomar en cuenta.

Como se aprecia en la tabla III: de color verde, están los campos que pertenecen a la perspectiva MIGRANTES; de color morado opaco, los campos que van a pertenecer a la perspectiva de VIAJES; de color celeste, los campos que perteneces al TIEMPO; y de color café opaco los campos que pertenecen a la perspectiva de DETALLE.

Teniendo los campos de cada perspectiva se puede incluir en el Modelo Conceptual ampliado que se efectúa en la siguiente fase.

2. Modelo conceptual ampliado

Al ya definir las columnas que se van a utilizar en cada perspectiva y las operaciones de cada indicador podemos ya agregarlos en el Modelo Conceptual, como se puede apreciar en la figura 4:

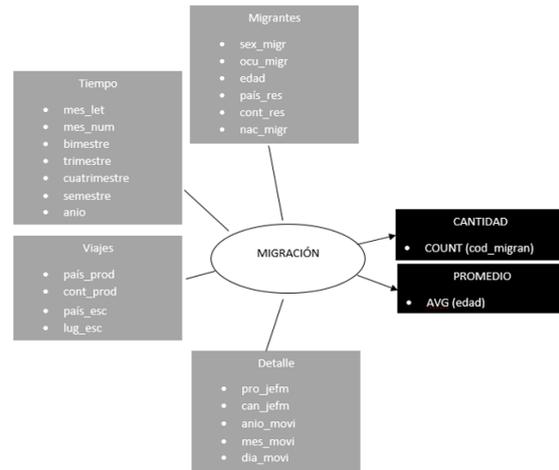


Fig. 4: Modelo Conceptual ampliado

Fuente: Autor del proyecto

Al contar con un modelo conceptual ampliado con los campos que se van a utilizar por perspectiva y las operaciones de cada indicador, que se tomaron en consideración en la relación, permitirá realizar las etapas posteriores.

D. Modelo lógico del DW

Con el Modelo Conceptual ampliado se puede pasar a efectuar el Modelo lógico de la estructura del Datawarehouse.

1. Tipo de Modelo Lógico del DW

En esta fase se escoge escoger el esquema más idóneo para el negocio. En el caso de la institución en estudio, se determinó que el esquema en estrella es el eficiente ya que tiene los mejores tiempos de respuesta, el diseño es modificable, facilita el análisis y facilita el uso de herramientas de consulta.

Para efectuar el esquema de estrella necesitamos desnormalizar el modelo Entidad-Relación.

2. Tablas de dimensiones

Las tabas de dimensiones son las que van a permitir filtrar la información (figura 5).

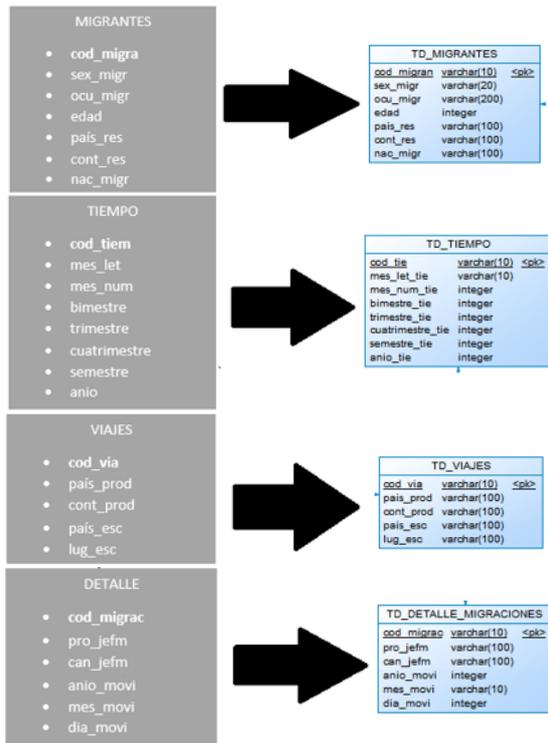


Fig. 5: Perspectivas y las respectivas Tablas de Dimensiones
Fuente: Autor del proyecto

Cada perspectiva constituye una tabla de dimensión. En la parte izquierda de la figura se encuentran las perspectivas y en la parte derecha las tablas de dimensiones con los respectivos campos que se obtuvieron en las perspectivas. Además, se incluye en cada uno la clave primaria.

3. Tablas de hechos

La tabla de hechos contendrá los indicadores de estudio (figura 6)

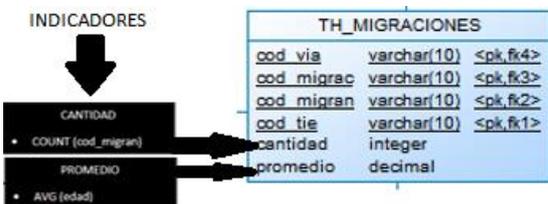


Fig. 6: Tabla de hechos, obtenida de PowerDesigner
Fuente: Autor del proyecto

La tabla de Hechos está formada por los campos primarios de las tablas de dimensiones, que, pasan a ser la clave primaria compuesta de la tabla de hechos de migración. Se agregan los indicadores que en este caso son CANTIDAD y PROMEDIO.

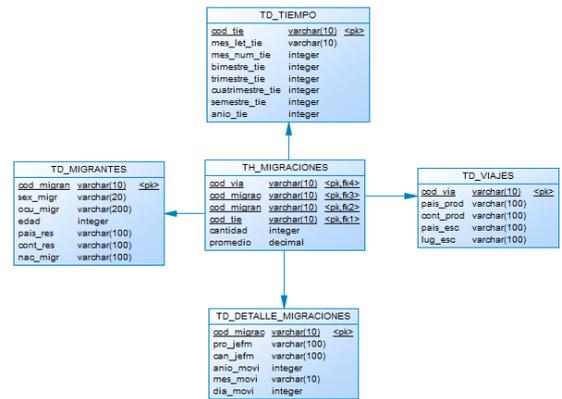


Fig. 7: Uniones de las tablas de dimensiones a la tabla de hechos, obtenida de PowerDesigner
Fuente: Autor del proyecto

Determinado: el esquema del negocio, que es en estrella; las tablas de dimensiones, que son 4; y la tabla de Hechos de migración, se desarrolla las respectivas uniones, que como se observa en la figura, es de las tablas de dimensiones a la tabla de hechos y con esto se obtiene el diseño del Modelo lógico del Datwarehouse.

E. Integración de los datos
1. Carga Inicial

Con la tabla de hecho y las dimisiones con sus respectivas uniones, se procede a realizar el proceso ETL como se muestra en la siguiente figura:

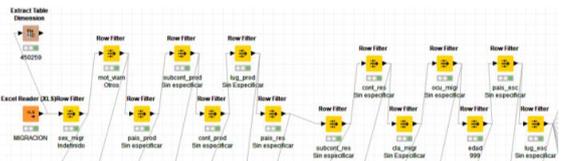


Fig. 8: Proceso ETL sobre la Base de Datos de Migración, obtenida de KNIME
Fuente: Autor del proyecto

No existe datos nulos o en blanco en la Base de Datos, pero existen campos llenados con “Indefinidos”, “Sin Especificar”, “Otros”, “999”, que fueron extraídos del proceso de DWH, ya que no permiten tener un resultado de calidad.

Se carga de las tablas de dimensiones con sus respectivos atributos (primero la tabla TD_MIGRANTES, segundo TD_VIAJES, tercero TD_DETALLE_MIGRACIONES y finalmente TD_TIEMPO con la granularidad según análisis) y la tabla de hechos de TH_MIGRACIONES. Este proceso se cumple, gracias a la ayuda de la herramienta KNIME.

2. Actualización

Finalizado el proceso ETL, se procede a la creación de los siguientes indicadores:

- “Cantidad de Migrantes” = COUNT(TD_MIGRANTES.CANTIDAD)
- “Promedio de Edad” = AVG(TD_MIGRANTES.PROMEDIO)

Se define la jerarquía de las fechas:

- SEMESTRE_TIE, CUATRIMESTRE_TIE, TRIMESTRE_TIE o BIMESTRE_TIE y
- MES_LET_TIE o MES_NUM_TIE

Para cada pregunta obtenida en la fase 1 se va a efectuar un cubo multidimensional, el cual va a responder y satisfacer el requerimiento solicitado, como se demuestra a continuación:

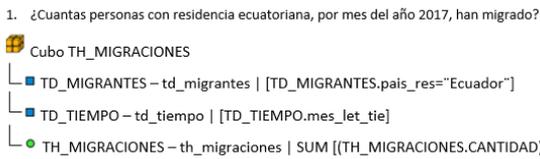


Fig. 9: Cubo multidimensional de la pregunta 1
Fuente: Autor del proyecto

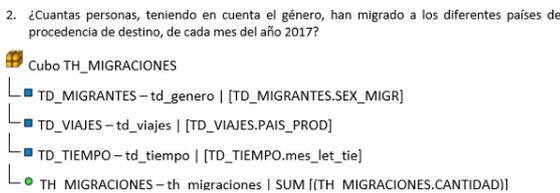


Fig. 10: Cubo multidimensional de la pregunta 2
Fuente: Autor del proyecto

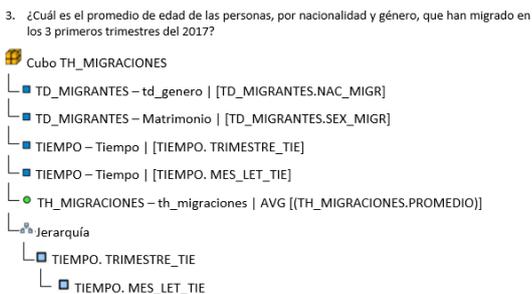


Fig. 11: Cubo multidimensional de la pregunta 3
Fuente: Autor del proyecto

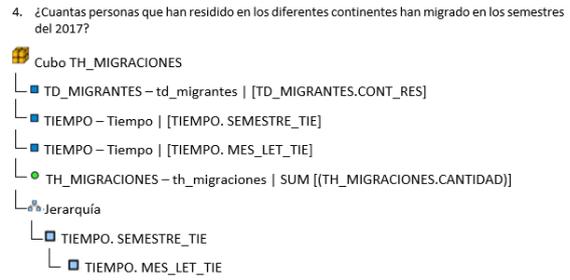


Fig. 12: Cubo multidimensional de la pregunta 4
Fuente: Autor del proyecto

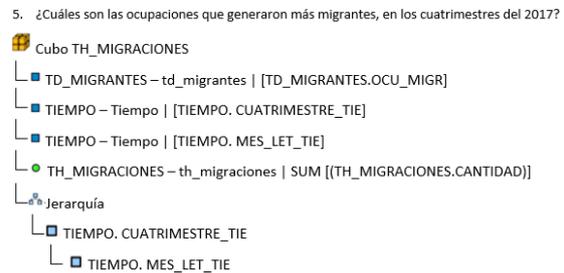


Fig. 13: Cubo multidimensional de la pregunta 5
Fuente: Autor del proyecto

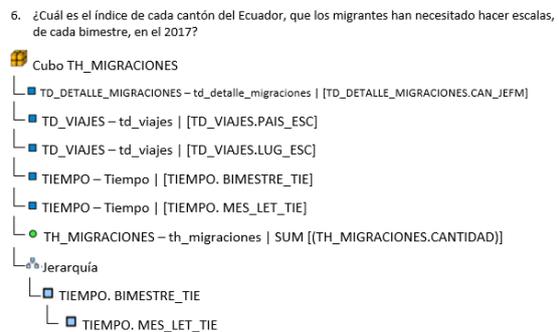


Fig. 14: Cubo multidimensional de la pregunta 6
Fuente: Autor del proyecto

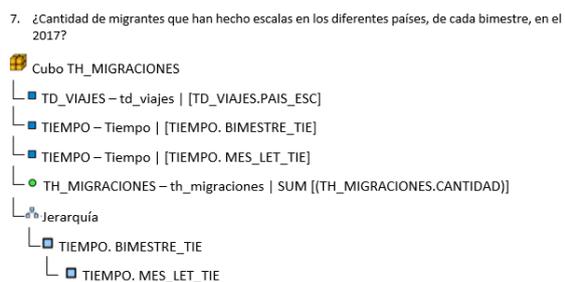


Fig. 15: Cubo multidimensional de la pregunta 7
Fuente: Autor del proyecto

8. ¿Cuáles son las jefaturas de provincias y cantones donde se registraron procesos migratorios en los cuatrimestres del 2017 y cuantos por cada cantón?

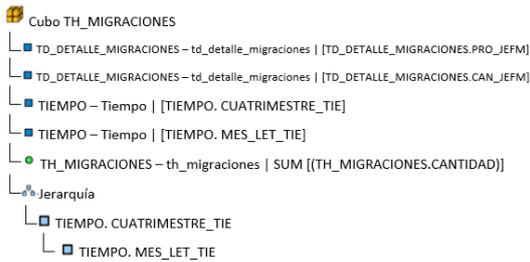


Fig. 16: Cubo multidimensional de la pregunta 8

Fuente: Autor del proyecto

Se tiene el producto final de esta fase que son los cubos multidimensionales de cada pregunta.

V. RESULTADOS

Con la ayuda de la herramienta de software TABLEAU que permite crear dashboards, se desarrollan los pertinentes, que permiten resolver las preguntas y requerimientos.

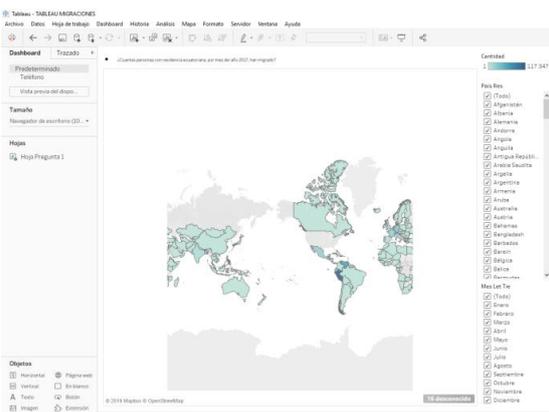


Fig. 17: Creación del Dashboard de la Pregunta 1, obtenida de TABLEAU

Fuente: Autor del proyecto



Fig. 18: Creación del Dashboard de la Pregunta 2, obtenida de TABLEAU

Fuente: Autor del proyecto



Fig. 19: Creación del Dashboard de la Pregunta 3, obtenida de TABLEAU

Fuente: Autor del proyecto

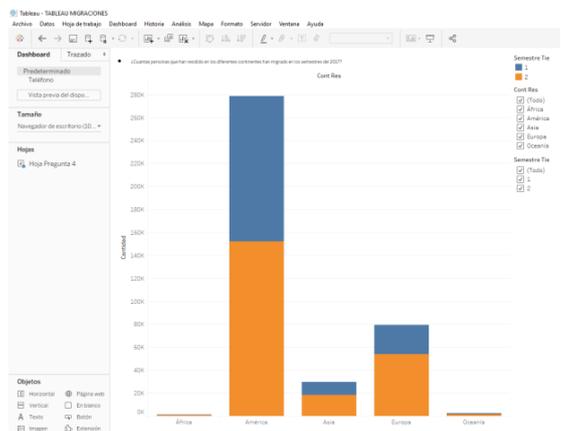


Fig. 20: Creación del Dashboard de la Pregunta 4, obtenida de TABLEAU

Fuente: Autor del proyecto

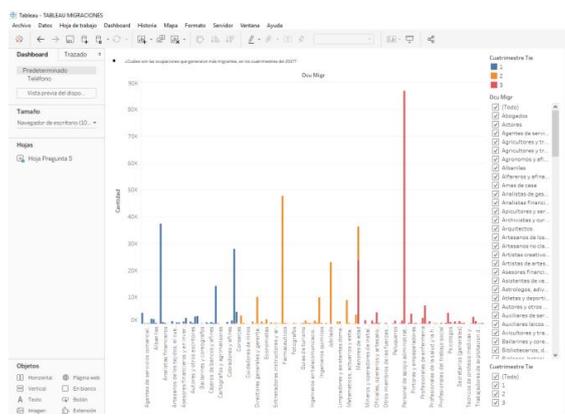


Fig. 21: Creación del Dashboard de la Pregunta 5, obtenida de TABLEAU

Fuente: Autor del proyecto

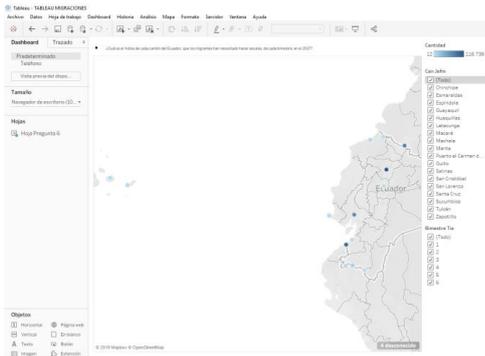


Fig. 22: Creación del Dashboard de la Pregunta 6, obtenida de TABLEAU

Fuente: Autor del proyecto

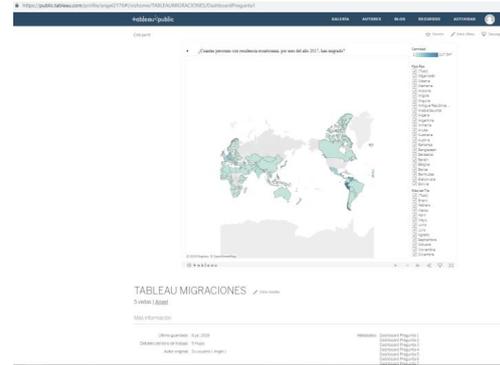


Fig. 25: Dashboards cargados en el Internet, obtenida de TABLEAU public

Fuente: Autor del proyecto

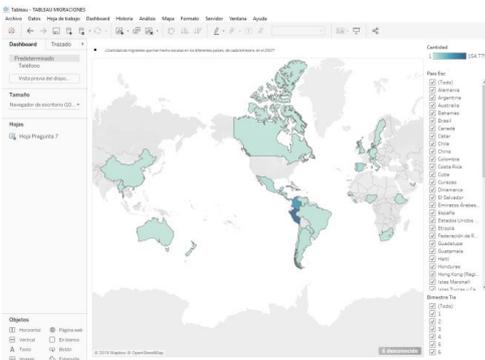


Fig. 23: Creación del Dashboard de la Pregunta 7, obtenida de TABLEAU

Fuente: Autor del proyecto

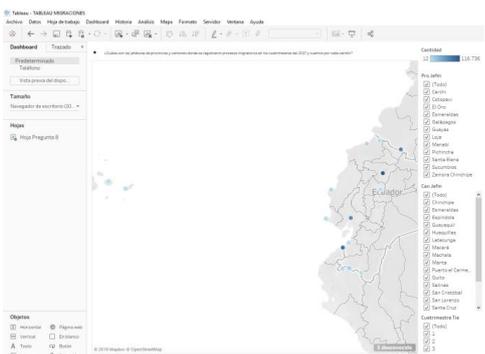


Fig. 24: Creación del Dashboard de la Pregunta 8, obtenida de TABLEAU

Fuente: Autor del proyecto

Una de las ventajas de Tableau es que carga los dashboards creados, en el internet, se lo puede visualizar en el siguiente [Tableau Dashboards](#).

Con los gráficos elaborados se muestra y entrega a los directivos correspondientes, que son quienes van a utilizar las gráficas para la toma de decisiones.

Se realizó una reunión con los directivos para presentar el producto terminado, mostrando estos una satisfacción por el trabajo cumplido.

VI. CONCLUSIONES

Datawarehouse es una ciencia que permite tratar los datos y obtener información a partir de esos datos. Las herramientas que permiten llevar a cabo el tratamiento de los datos facilitaron el trabajo, con la ayuda de la teoría se extrajo el material requerido por la institución, con lo que se pudo determinar que Datawarehouse es un medio idóneo para poder extraer información para la toma de decisiones para las distintas instituciones.

La institución en curso es analizando los datos extraídos y próximamente darán a conocer las decisiones que han tomado frente a los resultados que se les ha proporcionado.

REFERENCIAS

- [1] Real Academia de la Lengua Española, *Diccionario de la lengua española*. Real Academia Española, 2000.
- [2] INEC, “[ecuadorencifras.gob.ec](#).”
- [3] Sinnexus, “[sinnexus.com](#).”
- [4] Klipfolio, “[klipfolio.com](#).”

Recibido: 25 de septiembre de 2019

Aceptado: 5 de noviembre de 2019

