



UNIVERSIDAD DE CUNDINAMARCA. BENAVIDES VARGAS. IMPLEMENTACION DE UNA BASE DE DATOS GEOGRAFICOS CON MODELO MULTIDIMENSIONAL DE COBERTURA EDUCATIVA Y CALIDAD DE VIDA SEGÚN LOS DATOS PUBLICADOS POR EL MINISTERIO DE EDUCACION Y LOS INDICADORES DEL DANE.

IMPLEMENTACION DE UNA BASE DE DATOS GEOGRAFICOS CON MODELO MULTIDIMENSIONAL DE COBERTURA EDUCATIVA Y CALIDAD DE VIDA SEGÚN LOS DATOS PUBLICADOS POR EL MINISTERIO DE EDUCACIÓN Y LOS INDICADORES DEL DANE

Yuli Alexandra Benavides Vargas
yulibenavides1128@hotmail.com

Tecnóloga en cartografía Universidad de Cundinamarca.
MSc. Saúl Becerra Ospina
sbecerrao@mail.unicundi.edu.co

Resumen: Este trabajo consiste en la implementación de una base de datos con un modelo multidimensional en el cual se integró los datos públicos del Ministerio de educación de cobertura educativa y los datos del Departamento Administrativo Nacional de Estadística de Colombia, DANE, obtenidos de la encuesta de calidad de vida llevada a cabo en el 2015.

El modelo de datos es implementado sobre PostgreSQL y se desarrollan algunos componentes en PL/PgSQL para analizar y presentar los datos y la inclusión de dicha información al Software Quantum GIS y ArcGIS Desktop para la elaboración de mapas temáticos sobre Cobertura educativa y calidad de vida en Colombia.

El modelo multidimensional facilita el análisis, consulta y procesamiento para convertir los datos en información útil y significativa para luego adquirir la comprensión necesaria para apoyar los procesos de toma de decisiones. Facilitando la consulta de datos relevantes provenientes de diferentes

fuentes permitiendo combinarlos para fortalecer el análisis.

Abstract: In this work we implemented a Data Base to integrate official data about education and life quality indicators by the National Administrative Department of Statistics and the Education Minister.

The datamodel is implemented on the DBMS PostgreSQL. We develop some components using PL/PgSQL language to analysis data and to produce thematic cartography analysis.

Palabras claves: Bases de datos, cobertura educativa, calidad de vida.

Keywords: Databases, Education coverage, Quality of life.

1. INTRODUCCIÓN

Con el presente trabajo se pretende aportar mediante la integración de datos, a la construcción de herramientas apropiadas para el



UNIVERSIDAD DE CUNDINAMARCA. BENAVIDES VARGAS. IMPLEMENTACION DE UNA BASE DE DATOS GEOGRAFICOS CON MODELO MULTIDIMENSIONAL DE COBERTURA EDUCATIVA Y CALIDAD DE VIDA SEGÚN LOS DATOS PUBLICADOS POR EL MINISTERIO DE EDUCACION Y LOS INDICADORES DEL DANE.

análisis de datos usando herramientas modernas para tal fin.

La educación en Colombia teniendo en cuenta la cobertura y calidad, implica una posición de diferentes conceptos políticos y sociales, que a largo plazo pretende equilibrar las necesidades tanto económicas sociales y medio ambientales, que operan en un entorno. La calidad educativa como política de estado plantea el desarrollo del individuo como ser social, hábil y competente para desenvolverse en un ámbito laboral eficientemente y posibilitando una posición dentro de una sociedad funcional que aporte al desarrollo. (Educacion, 2000).

La calidad de vida se mide mediante el estudio de algunos factores como lo son materiales, ambientales y de relacionamiento. Los factores materiales son los recursos que uno tiene como ingresos disponibles, posición en el mercado de trabajo, salud, nivel de educación, etc. Los factores ambientales son las características del lugar donde vives presencia y acceso a servicios, condiciones del hogar, grados de seguridad y criminalidad, transporte y movilización, habilidad para el manejo de nuevas tecnologías. Los factores de relacionamiento hacen referencia a las relaciones con la familia, amigos y redes sociales. Se toma la iniciativa a nivel internacional de realizar un estudio de varios aspectos sociales y de bienestar que van asociados al progreso de las personas. (Palomba, 2002)

Para realizar nuestro proyecto implementamos una base de datos geográficos con un modelo multidimensional para los datos del Ministerio de educación y las

encuestas de calidad de vida en Colombia realizadas por el DANE.

Se diseña un modelo de datos multidimensional para alojar los datos recolectados; Y luego implementarlos en la base de datos geográficos sobre el sistema gestor de bases de datos PostgreSQL. Por último se genera cartografía temática a partir de la base de datos implementada.

2. METODOLOGIA

En este trabajo se diseña, construye e implementa una base de datos geográficos con modelo multidimensional de cobertura educativa y calidad de vida para identificar en qué lugares hay deficiencia de educación por lo tanto el índice de calidad de vida es muy bajo.

A la información obtenida del MEN y el DANE, se implementa una serie de métodos que a continuación se explica la utilización de cada uno:

2.1 Extraer, Transformar y cargar

El proceso de ETC (Extraer, Transformar y Cargar) (Ver Figura 1 Extracción, Transformación y carga (Ramírez Farfán & Benavides Arteaga)) es la principal actividad de esta fase. Periódicamente se importan los datos desde los sistemas de la organización a la bodega de datos sistemas (ERP) y otros sistemas relacionados con el negocio. Estas tareas son la materia prima de la bodega y nos debe proporcionar información consistente,



ya que la calidad de los datos es un factor determinante en el éxito del proyecto. (Ramírez Farfán & Benavides Arteaga)



Figura 1 Extracción, Transformación y carga (Ramírez Farfán & Benavides Arteaga)

2.2 Estructuración de datos

La estructuración de datos se realizó ordenando y desglosando en grupos relevantes toda la información obtenida por el MEN y DANE. Por ejemplo, la encuesta de calidad de vida se hace en las diferentes regiones y zonas del país dividiéndolas en dos grupos los cuales son (cabecera y centro poblado y resto).

Todos los datos son descargados en archivos de Excel con formatos, es decir, que cuentan con colores, datos no atómicos, celdas combinadas, fórmulas y demás características que complican la búsqueda y análisis de datos.

Esto conduce a una de las principales y más larga labor dentro del desarrollo del presente proyecto, puesto que es necesario estructurar los datos en tablas y posibilitar la migración a un modelo de datos adecuado, teniendo en cuenta tipos de datos y conceptos.

En total se estructuraron más de quinientos mil registros encontrados en matrices y tablas publicadas por el ministerio de educación y por el DANE.

2.3 Creación de modelo multidimensional

La bodega de datos es definida por Bill Inmon como:

- ✓ Orientado a temas: Los datos en la base de datos se encuentran organizados de tal manera de que los elementos relativos al mismo evento u objeto se encuentran relacionados.
- ✓ Variante en el tiempo: Los cambios de los datos en el tiempo quedan almacenados en la bodega para que en los informes se vean reflejadas estas variaciones.
- ✓ No volátil: Los datos no son modificados ni eliminados, al almacenar el dato este se convierte de solo lectura y se mantiene para futuras consultas.
- ✓ Integrado: Contiene todos los datos de los sistemas



UNIVERSIDAD DE CUNDINAMARCA. BENAVIDES VARGAS. IMPLEMENTACION DE UNA BASE DE DATOS GEOGRAFICOS CON MODELO MULTIDIMENSIONAL DE COBERTURA EDUCATIVA Y CALIDAD DE VIDA SEGÚN LOS DATOS PUBLICADOS POR EL MINISTERIO DE EDUCACION Y LOS INDICADORES DEL DANE.

operacionales de la organización, los cuales deben ser consistentes.

Define la metodología Top-Down como un resumen del sistema sin especificar detalles, luego se refinan los detalles y cada vez con mayor detalle hasta que es lo suficientemente detallado para validar el modelo. Así es planteado por Bill Inmon ya que este autor afirma que la bodega de datos debe responder a todas las necesidades de los usuarios de la organización y no a un pequeño grupo.

Ralph Kimball define la bodega de datos como una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis, la unión de todos los datamarts de una organización. Define la metodología Botton-Up que plantea la construcción de Data-marts independientes que luego se enlazaran para conformar el sistema completo.

Ralph Kimball plantea una metodología denominada el ciclo de vida de Kimball la cual se ha convertido en un estándar en el apoyo en la toma de decisiones empresariales. (Inmon, 2002).

2.4 DANE

Las encuestas de calidad de vida surgen como respuesta a la necesidad de caracterizar la población en los

diferentes aspectos involucrados en el bienestar de los hogares. El DANE implementa la Encuesta Nacional de Calidad de Vida (ECV) incorporando en su diseño y ejecución lo postulado por la metodología para la edición de las condiciones de vida (The Living Standards Measurement Study -LSMS-). La ECV 2015 es una investigación que permite recoger información sobre diferentes aspectos y dimensiones del bienestar de los hogares, incluyó variables relacionadas con las características físicas de las viviendas (material de paredes y pisos); el acceso a servicios públicos, privados o comunales; salud; atención integral de niños y niñas menores de 5 años; tecnologías información y comunicación; tenencia y financiación de la vivienda; condiciones de vida del hogar y variables demográficas como: sexo, edad, parentesco, estado civil, entre otras. De acuerdo con el diseño estadístico de la ECV 2015, el tamaño de la muestra fue de 23.005 hogares con encuestas completas, el universo de la encuesta está conformado por la población civil no institucional, residente en el territorio nacional, excluyendo la parte rural de los departamentos creados por la constitución de 1991. (DANE, 2015)

2.5 Homologación e inserción a la BD

Para insertar los datos es necesario realizar un proceso de homologación, es decir, unificar descripciones, nombres, dominios y demás valores



encontrados en las tablas insumo. En particular se revisaron los tipos de datos y codificaciones.

En el caso de los datos geográficos, se presenta mayor resolución espacial en los datos suministrados por el ministerio de educación, mientras que los datos del DANE se encuentran a nivel regional y algunos departamentos.

2.6 Implementación métodos estadísticos

2.6.1 Análisis exploratorio de datos

Puede definirse como el conjunto de técnicas que describen y visualizan las distribuciones espaciales, identifican localizaciones atípicas o “atípicos espaciales” (“spatial outliers”), descubren esquemas de asociación espacial, agrupamientos (“clusters”) o puntos calientes (“hot spots”) y sugieren estructuras espaciales u otras formas de heterogeneidad espacial (Anselin, 1999). Por tanto, el AEDE se correspondera con los métodos de estadística descriptiva espacial. (Yrigoyen, 2003)

2.6.2 Clasificación por cuantiles

La usamos para la clasificación de datos y lograr un resumen estadístico (Myers, 1999). Dicho resumen se realiza con base en los cuantiles, el número de clases se determina usando la regla de Sturgers.

2.7 Producción de graficas estadísticas

2.7.1 Histograma

Es una gráfica de la distribución de un conjunto de datos. Es un tipo especial de gráfica de barras, en la cual una barra va pegada a la otra, es decir no hay espacio entre las barras. Cada barra representa un subconjunto de los datos. Muestra la acumulación o tendencia, la variabilidad o dispersión y la forma de la distribución. (México, 2009)

2.8 Producción de la Cartografía

Con la información recopilada se hace un manejo y procesamiento adecuado de la misma para la creación de cartografía temática acerca de la cobertura educativa y calidad de vida en Colombia.

Los datos geográficos o cartografía digital utilizada son almacenados sobre la base de datos PostgreSQL usando el soporte adicionado por PostGIS. Esto es posible puesto que el sistema gestor de bases de datos es relacional orientado a objetos. Esto facilita la generación de consultas usando lenguaje SQL.

3. Resultados

3.1 Recolección de Datos Estadísticos MEN y DANE.

A continuación se anexa una tabla de los datos que se usaron para los análisis y la producción cartográfica.



DATOS	AÑO
Calidad De Vida	2015
Docentes	(2007-2013)
Matriculados	(2000-2013)

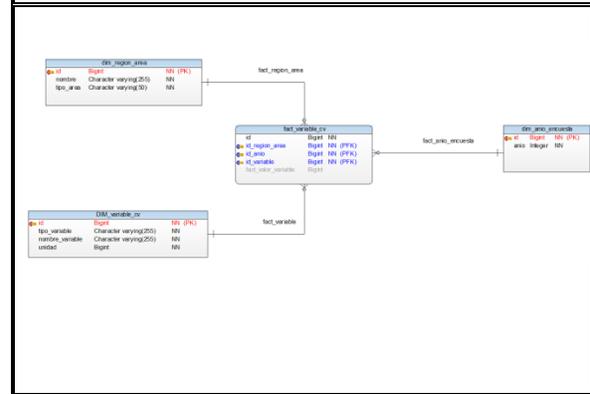
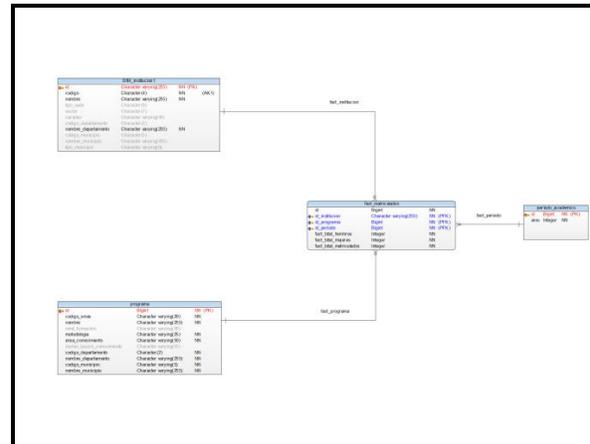
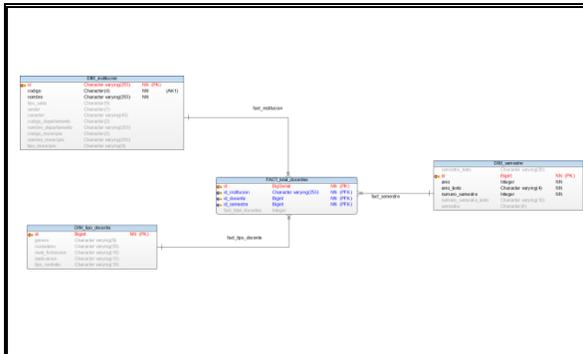
Tabla 1 Datos usados para el análisis y la producción cartográfica

3.2 Modelo De Datos Multidimensional

El modelo de datos multidimensional fue creado a partir de la información recolectada en las entidades correspondientes (MEN y DANE) aplicando herramientas y programas de Inteligencia de negocios para facilitar su análisis y mejor toma de decisiones.

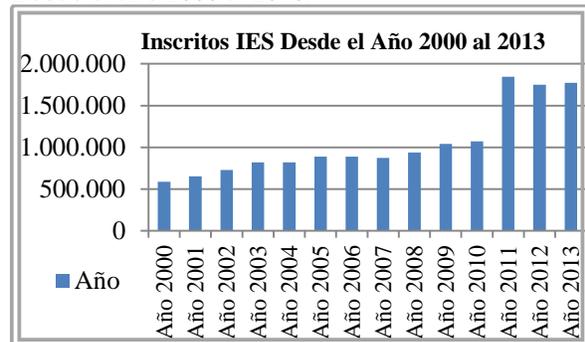
El anterior modelo físico fue implementado sobre PostgreSQL 9.4 y PostGIS 2.1. No se presentan requerimientos de sistema de hardware más que los necesarios para instalar el sistema gestor de base de datos. El sistema se puede instalar sobre Windows, GNU/Linux y Mac OS X. En el marco de este trabajo, ha sido probado sobre Windows 8 y sobre el sistema GNU/Linux en la distribución Ubuntu 16.04.

Figura 2 Modelos Bases de Datos Multidimensional.



3.3 Análisis de cobertura educativa a nivel profesional.

Figura 3 Personas Inscritas a los IES Nacional Desde el año 2000 al 2013.

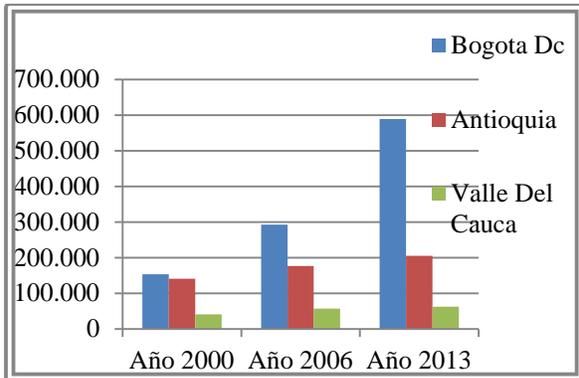


Fuente: Elaboración propia con información del MEN.

Figura 4 Crecimiento de estudiantes inscritos en los IES años (2000-2006-2013).para Bogotá D.C, Antioquia y Valle del Cauca.

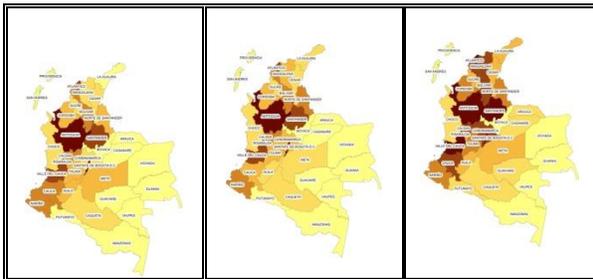


UNIVERSIDAD DE CUNDINAMARCA. BENAVIDES VARGAS. IMPLEMENTACION DE UNA BASE DE DATOS GEOGRAFICOS CON MODELO MULTIDIMENSIONAL DE COBERTURA EDUCATIVA Y CALIDAD DE VIDA SEGUN LOS DATOS PUBLICADOS POR EL MINISTERIO DE EDUCACION Y LOS INDICADORES DEL DANE.



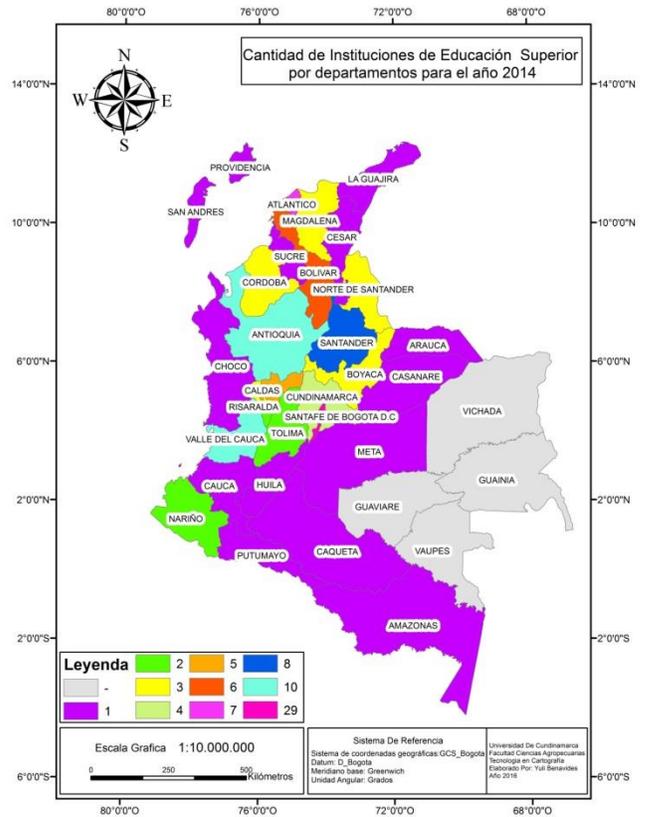
Fuente: Elaboración propia con información del MEN

Figura 5 Relación Mapas Coropleticos IES Por Departamentos Y Distrito Capital Años (2000-2006-2013).



Fuente: Elaboración propia con información del MEN.

Figura 6 Distribución por números según la cantidad de IES que hay por departamentos para el año 2014

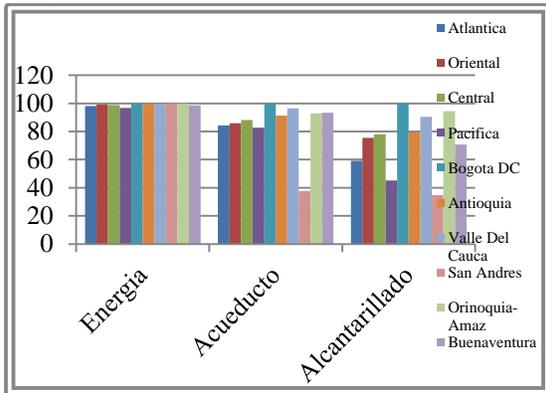


Fuente: Elaboración propia con base de datos del MEN.



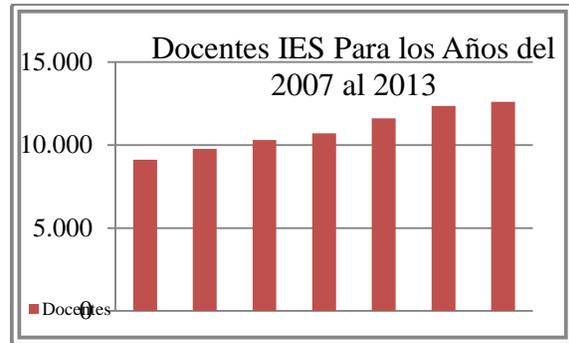
UNIVERSIDAD DE CUNDINAMARCA. BENAVIDES VARGAS. IMPLEMENTACION DE UNA BASE DE DATOS GEOGRAFICOS CON MODELO MULTIDIMENSIONAL DE COBERTURA EDUCATIVA Y CALIDAD DE VIDA SEGUN LOS DATOS PUBLICADOS POR EL MINISTERIO DE EDUCACION Y LOS INDICADORES DEL DANE.

Figura 7 Porcentaje de cobertura de los servicios energía, Acueducto, Alcantarillado según la ECV del 2015.



Fuente: Elaboración propia con información del DANE.

Figura 8 Docentes IES Para los Años del 2007 al 2013.



Fuente: Elaboración propia con información del MEN.

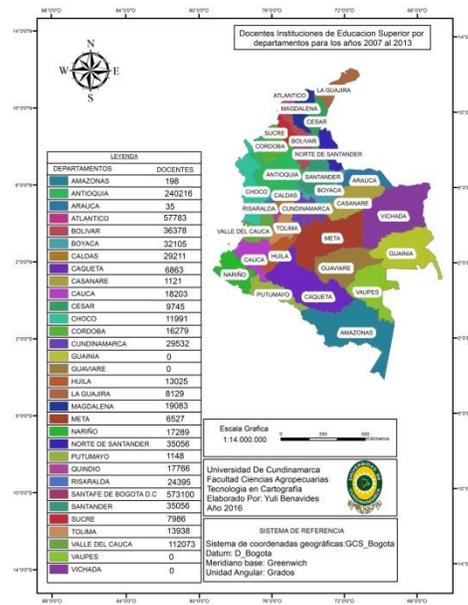


Figura 9 Figura 8 Docentes IES Para los Años del 2007 al 2013.



4. Bibliografía

- DANE. (2015). *Departamento Administrativo Nacional De Estadística* . Obtenido de <http://www.dane.gov.co/index.php/estadisticas-sociales/calidad-de-vida-ecv>
- Educacion, I. N. (2000). *Desafios de la educacion*. Buenos Aires: UNESCO.
- Inmon, W. H. (2002). *Building The Data Warehouse*. Wiley Computer Publishing.
- México, U. N. (2009). HISTOGRAMA . 1-4.
- Myers, R. H.-S. (1999). *Probabilidad Y Estadística Para Ingenieros*.
- Palomba, R. (2002). *Calidad de Vida: Conceptos y medidas*. Roma, Italia : Institute of Population Research and Social Policies.
- Ramírez Farfán, C. E., & Benavides Arteaga, L. N. *Datamart De Inventarios Para La Empresa Sodimac Colombia*. Universidad De Bogotá Jorge Tadeo Lozano, Bogota D.C.
- Yrigoyen, C. C. (2003). Análisis Exploratorio De Datos Espaciales Al Servicio Del Geomarketing. 4-5.