

**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**

**UNIVERSIDAD DE CUNDINAMARCA  
FACULTAD DE CIENCIAS AGROPECUARIAS  
TECNOLOGÍA EN CARTOGRAFÍA  
FUSAGASUGÁ  
2016**

**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**  
**Cód. 190212204**

**Proyecto para optar por el título de Tecnólogo en Cartografía**

**Director**  
**MSc. Saúl Becerra Ospina**

**UNIVERSIDAD DE CUNDINAMARCA**  
**FACULTAD DE CIENCIAS AGROPECUARIAS**  
**TECNOLOGÍA EN CARTOGRAFÍA**  
**FUSAGASUGÁ**  
**2016**

**Nota de aceptación**

---

---

---

---

---

**Saúl Becerra Ospina**  
**Director de proyecto**

---

**Diego Méndez**  
**Jurado**

---

**Sócrates Cardona**  
**Jurado**

## AGRADECIMIENTOS

A Dios por haberme acompañado y guiado a lo largo de la carrera, por brindarme una vida llena de aprendizajes, experiencias, por fortalecer mi corazón e iluminar mi mente y por haberme puesto en el camino a aquellas personas que han sido soporte y compañía durante todo el periodo de estudio.

A mis padres por ser el pilar fundamental en todo lo que soy, por su educación tanto académica como de la vida y por ser incondicional apoyo durante el tiempo para la culminación de esta etapa que es de vital importancia en mi vida.

A los docentes que hicieron parte de mi proceso de aprendizaje, por su dedicación, conocimientos, apoyo y gran sentido de motivación; al director de tesis Saúl Becerra Ospina por su persistencia y acompañamiento durante el transcurso de este ciclo.

## TABLA DE CONTENIDO

TABLA DE CONTENIDO .....	3
TABLA DE FIGURAS.....	6
GLOSARIO.....	7
ABREVIATURAS .....	8
RESUMEN.....	9
INTRODUCCION.....	11
1. OBJETIVOS .....	13
1.1 Objetivo General .....	13
1.2 Objetivos Específicos .....	13
2. MARCO REFERENCIAL.....	14
2.1 Modelo multidimensional .....	14
2.1.1 Usos de la Inteligencia de negocios. ....	16
2.1.2 Beneficios de la Inteligencia de Negocios.....	16
2.2 Bodegas de Datos Estructuradas .....	17
2.2.1 Componentes de una Bodega de datos .....	17
2.2.2 Modelo Relacional .....	17
2.2.3 DataMart.....	18
2.3 Bases de Datos Geográficos .....	18
2.3.1 Datos espaciales .....	18

2.3.2	Datos no espaciales .....	18
2.3.3	Capas Geográficas.....	18
2.3.4	Entidad .....	18
2.3.5	Representación Geométrica.....	19
2.3.6	Modelos de datos.....	19
2.4	Calidad De Vida (CV).....	19
2.5	MEN.....	25
3.	METODOLOGIA .....	26
3.1	ETC.....	26
3.1.1	Extraer.....	26
3.1.2	Transformar.....	27
3.1.3	Carga.....	27
3.2	Estructuración de datos .....	28
3.3	Creacion modelo multidimensional .....	28
3.4	Homologación e inserción a la BD .....	29
3.5	Implementación métodos estadísticos .....	30
3.5.1	Análisis exploratorio de datos .....	30
3.5.2	Clasificación por cuantiles .....	30
3.6	Producción graficas estadísticas .....	30
3.6.1	Histograma .....	30

3.7 Producción de la Cartografía .....	31
4. RESULTADOS.....	32
4.1 Recolección Datos Estadísticos MEN y DANE.....	32
4.2 Modelo De Datos Multidimensional .....	32
4.2 Análisis de cobertura educativa a nivel profesional. ....	36
5. CONCLUSIONES.....	43
6. LISTA DE REFERENCIAS.....	44
7. ANEXOS .....	46
A.1. Cartografía temática .....	46
A.2. Diccionario de datos. ....	46

## TABLA DE FIGURAS

Figura 1. Un cubo tridimensional de datos de ventas que tiene dimensiones tienda, Tiempo del producto y una cantidad medida (Zimányi, 2008) .....	14
Figura 2 Inteligencia de negocios. ....	15
Figura 3 Regiones ECV.....	23
Figura 4 Regiones ECV (Región Pacifico Incluyendo Valle).....	24
Figura 5 Extracción, Transformación y carga (Ramírez Farfán & Benavides Arteaga) .....	26
Figura 6 Modelos Bases de Datos Multidimensional.....	33
Figura 7 Personas Inscritas a los IES Nacional Desde el año 2000 al 2013.....	36
Figura 8 Crecimiento de estudiantes inscritos en los IES años (2000-2006-2013).para Bogotá D.C, Antioquia y Valle del Cauca. ....	37
Figura 9 Relación Mapas Coropleticos IES Por Departamentos Y Distrito Capital Años (2000-2006-2013) .....	38
Figura 10 Distribución por números según la cantidad de IES que hay por departamentos para el año 2014.....	39
Figura 11 Porcentaje de cobertura de los servicios energía, Acueducto, Alcantarillado según la ECV del 2015.....	40
Figura 12 Docentes IES Para los Años del 2007 al 2013.....	41
Figura 13 Docentes IES por departamentos para los años del 2007 al 2013.....	42

## GLOSARIO

**Inteligencia de negocios:** proceso de analizar los datos acumulados existentes en una organización o empresa para extraer y generar un conocimiento de ellos.

**Data Mart:** es simplemente un subconjunto de datos de un Data Warehouse para un área específica

**PostgreSQL:** Es el sistema gestor de bases de datos con licencia GPL más robusto.

**Base de datos:** Una base de datos es una colección de relación lógica con los datos que apoya las actividades de una organización.

**OLAP:** OnLine Analytical Processing. Sistemas de procesamiento analítico.

**Modelo Relacional:** El modelo e-r se considera un modelo conceptual ya que permite a un nivel alto el ver con claridad la información utilizada en algún problema o negocio.

**Base de datos geográficos:** Una Base de Datos Geográfica (BDG) es un conjunto de datos geográficos organizados de tal manera que permiten la realización de análisis y la gestión del territorio dentro de aplicaciones de Sistemas de Información Geográfica (SIG). Además, una BDG se utiliza de soporte para la implantación de servicios geográficos relacionados con las Infraestructuras de Datos Espaciales (IDE), y su contenido es la base fundamental en los procesos de producción cartográficos. (España)

**Base de datos:** Colección compartida de los datos relacionados lógicamente, y una descripción de estos datos, diseñado para satisfacer las necesidades de información de una organización y para apoyar sus actividades.

## **ABREVIATURAS**

- MEN: Ministerio de educación nacional.
- DANE: Departamento administrativo nacional de estadística.
- IGAC: Instituto geográfico Agustín Codazzi.
- NBI: Necesidades básicas insatisfechas.
- IES: Instituciones de educación superior.
- SIG: Sistemas de información geográfica.
- ECV: Encuesta Calidad de vida.
- OLAP: Procesamiento analítico en línea.
- OMS: Organización mundial de la salud.
- BI: Businnes intelligence.
- BD: Bases de datos.
- BDG: Base de datos geográficos.
- QGIS: Quantum GIS.
- ETC: Extraer transformar y cargar.
- DPTO: Departamento.

## **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 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.

## **ABSTRAC**

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.

## INTRODUCCION

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.

La calidad de vida es un estado que todo ser humano pretende alcanzar para satisfacer necesidades básicas pero este es proporcional a su grado de escolaridad o educación. Según la organización mundial de la salud (OMS) la calidad de vida es una posición que tenemos los individuos con respecto a una consecución de metas que se asumen como procesos que se articulan a través de factores externos e internos. En sus comentarios SCHALOCK 1996 manifiesta en su ensayo que cada individuo labra su propio destino según las necesidades planteadas estableciendo su criterio personal para alcanzar un nivel de calidad de vida de acuerdo a su contexto social demográfico etnográfico donde se manejan relaciones personales, inclusión social, desarrollo personal, bienestar físico autodeterminación bienestar material o emocional y derechos y deberes. (Santos, 2009).

Según los estudios realizados por Rosella Palomba para muchos autores la definición de bienestar o calidad de vida es muy pocas veces dada primero es necesario definir algunos dominios específicos de la calidad de vida.

Para manejar el concepto de calidad de vida se deben estudiar variables multidimensionales como lo son las condiciones de vida y el bienestar del individuo, todo esto debe de llevar a la satisfacción a muchas necesidades.

Para medir la calidad de vida se hace 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.

Según la conferencia de la unión europea y otros regímenes internacionales tenían como objetivo delimitar algunas problemáticas sociales tomando varios aspectos como la salud, trabajo, educación entre otros y así poder sacar unas estadísticas para el apoyo de mejor tomas de decisiones. En el 2009 realizan un informe para la medición del progreso social y económico llamada Stiglits dentro de ella lleva unas recomendaciones dirigidas a las oficinas con el fin de la medición del progreso.

El comité del sistema estático europeo (CSSE) lleva a cabo las recomendaciones del informe “Medición del progreso, Bienestar y el desarrollo sostenible” y crea un Sponsorship Group (SG)

En el 2013 Eurostat publica el primer conjunto de indicadores de calidad de vida basado en el informe publicado en el 2011 por el CCSE.

El consejo superior de estadística de España inicia en julio del 2011 trabajando sobre el informe con el objetivo de analizar y ver cómo se puede trasladar estas iniciativas al sistema estático español. El informe actualizado por el consejo finaliza en el 2012 incluyendo 60 recomendaciones.

Las características multidimensionales como cobertura y calidad son ejes esenciales que permite posesionar a una sociedad en un contexto integral social, que perfila la competitividad de un país donde se debe guardar relación estrecha entre calidad de vida, cobertura educativa y de calidad. Estos parámetros planteados por el estado en diferentes documentos apuntan hacia procesos de participación, investigación y acción social que son insumo preponderante para el bienestar que debe tener el ser humano dentro la sociedad.

En toda sociedad una óptima condición de salud económica y educativa permite identificarla como un fortalecimiento anexo a valores sociales como la solidaridad cooperación respeto tolerancia y otros que permiten un desarrollo de una comunidad activa y progresiva, además de características que no son fáciles de alcanzar cuando todos sus miembros no se identifican con la solución de problemas característicos de la sociedad moderna.

Con el presente trabajo se pretende aportar mediante la integración de datos, a la construcción de herramientas apropiadas para el análisis de datos usando herramientas modernas para tal fin.

En este documento se encuentran los objetivos, marco teórico, metodología así como los recursos necesarios en el desarrollo del proyecto.

## **1. OBJETIVOS**

### **1.1 Objetivo General**

- ✓ Implementar 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.

### **1.2 Objetivos Específicos**

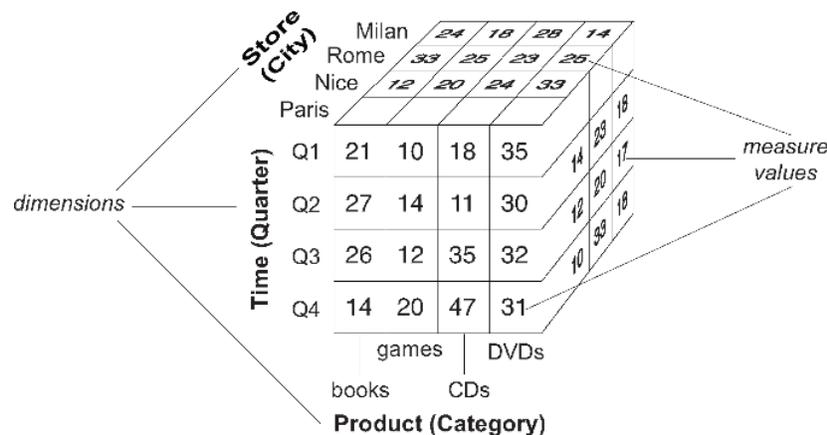
- ✓ Recolectar datos estadísticos proporcionados por el Ministerio de Educación y el DANE.
- ✓ Diseñar un modelo de datos multidimensional para alojar los datos recolectados.
- ✓ Implementar la base de datos geográficos sobre el sistema gestor de bases de datos PostgreSQL.
- ✓ Generar cartografía temática a partir de la base de datos implementada.

## 2. MARCO REFERENCIAL

### 2.1 Modelo multidimensional

Los almacenes de datos y sistemas OLAP se basan en un modelo multidimensional. Este modelo permite una mejor comprensión de los datos para fines de análisis y proporciona un mejor rendimiento para consultas analíticas complejas. Los datos de vistas de modelos multidimensionales en un espacio  $n$  - dimensional, generalmente llamados un cubo de datos o un hipercubo.

Un cubo de datos está definido por las dimensiones y hechos. Las dimensiones son diferentes perspectivas que se utilizan para analizar los datos. Por ejemplo, el cubo de datos en (Figura 1) se utiliza para analizar las cifras de ventas y tiene tres dimensiones: las tiendas, Tiempo, y del producto. (Zimányi, 2008).



**Figura 1.** Un cubo tridimensional de datos de ventas que tiene dimensiones tienda, Tiempo del producto y una cantidad medida (Zimányi, 2008)

### 1.2 OLAP

El asistente de creación de informes OLAP permite crear un informe en el que los datos OLAP se muestran como un objeto de cuadrícula. Aunque es similar al asistente para la creación de informes de tablas de referencias en muchos aspectos, el asistente de creación de informes OLAP es algo diferente debido a los requisitos de trabajar con orígenes de datos OLAP. Primero puede especificar la ubicación de los datos OLAP y luego elegir las dimensiones que desea incluir en la cuadrícula. A continuación, puede filtrar los datos del informe y elegir el estilo del objeto de cuadrícula, que también puede personalizar. Finalmente, puede definir etiquetas para la cuadrícula e insertar un gráfico, si lo desea. (SAP, 2011)

### 1.3 Inteligencia de negocios

La inteligencia de negocios es un proceso interactivo para explorar y analizar información estructurada sobre la bodega de datos (Datawarehouse), Para descubrir tendencias o patrones, a partir de los cuales deriva ideas y extrae conclusiones.

El proceso de Business intelligence incluye la comunicación de los descubrimientos y efectuación de los cambios. Las áreas incluyen clientes, proveedores, productos, Servicios y competidores. (Cano, 2007)

La Inteligencia de negocios, permite a las organizaciones entender mejor, analizar e incluso predecir lo que está ocurriendo en su empresa. BI ayuda a las empresas a convertir sus datos en información útil y significativa para luego distribuir esta información dentro de las áreas que la necesiten, cuando la soliciten para que puedan tomar decisiones oportunas y mejor informadas. (Ver Figura 2 Inteligencia de negocios.). Igualmente permite a las organizaciones combinar datos de una amplia variedad de fuentes, para lograr tener una visión de 360 grados, y saber cuáles son las falencias de la misma. (Arroyo, 2015)

Esto es especialmente importante para las empresas de tamaño medio, ya que al no tener los vastos recursos de industrias gigantes, suelen ser capaces de implementar con mayor rapidez las decisiones empresariales. Operaciones y análisis son dos caras de la empresa y BI permite ser un socio valioso en ambos campos. (Arroyo, 2015)

La inteligencia de negocios o Business Intelligence (en inglés) es la aplicación de un conjunto de conceptos, métodos, estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una empresa u organización. La inteligencia de negocios brinda una ventaja competitiva a las organizaciones, ya que el conocimiento generado tiene como objetivo principal reducir amenazas y aprovechar oportunidades en miras de cumplir y fortalecer los objetivos organizacionales de la compañía. (Arroyo, 2015)



**Figura 2 Inteligencia de negocios.**

### **2.1.1 Usos de la Inteligencia de negocios.**

- ✓ BI puede ser utilizado por una organización así:
- ✓ Para determinar el nivel de inventario de un producto o una pieza.
- ✓ Identificar sus productos más vendidos, y revisar si esto se cumple todos los puntos de venta.
- ✓ Identificar los clientes que están reduciendo sus compras, de manera que se puedan ofrecer incentivos especiales para retenerlos.
- ✓ Implementar cuadros de mando para que los ejecutivos y supervisores puedan reconocer rápidamente excepciones operativas, o cuando los gastos probablemente superen el presupuesto.
- ✓ Establecer y monitorear las métricas de rendimiento y tomar acciones correctivas si se encuentran en peligro de no ser cumplido.
- ✓ Comparar las ventas del año actual contra las ventas establecidas en el año anterior para la misma época y así prever lo que es probable que ocurra para todo el año.
- ✓ Integrar datos de diversas fuentes como lo son la hoja de cálculo y los datos históricos para fines de análisis, y así dar consistencia a la única versión de la información para la organización. (Ramírez Farfán & Benavides Arteaga)

### **2.1.2 Beneficios de la Inteligencia de Negocios.**

Una parte importante del trabajo de cualquier gerente es tomar decisiones y generar la eficacia de la organización. Debido a esto, BI fue referido inicialmente como un sistema de soporte de decisión.

BI permite a los usuarios de negocios analizar y comprender mejor los planes y los resultados de su organización. Esto da una idea de lo que está funcionando correctamente, mientras que la identificación de áreas con problemas potenciales a tiempo para las acciones correctivas que se deben tomar. Se puede utilizar para reconocer las oportunidades, así como problemas, y alertar a su organización a los problemas potenciales

cuando las ventas caen un 20% por debajo del pronóstico o inventario cae por debajo de un valor umbral.

Mientras que muchos gerentes y supervisores se enorgullecen de su intuición, BI proporciona herramientas para ayudar a verificar sus áreas de negocio, sino incluso descubrir otros nuevos. Permite a los usuarios explorar los resultados a un nivel alto y luego profundizar para analizar los detalles subyacentes. La inteligencia de negocios es una de las claves principales para la toma de decisiones. (Ramírez Farfán & Benavides Arteaga)

## **2.2 Bodegas de Datos Estructuradas**

Una bodega de datos contiene los datos que son útiles para la organización, almacenados en un repositorio de datos que han sido transformados en información útil para el usuario, entregando la información correcta, en el momento preciso y en el formato adecuado. La bodega de datos debe dar respuesta a las necesidades de los usuarios expertos, que utilizan Sistemas de soporte a la decisión (DSS), sistemas de información ejecutiva (EIS) o herramientas de consulta e informes, permitiéndoles a los usuarios hacer consultas sin afectar la operación del sistema. (Ramírez Farfán & Benavides Arteaga)

### **2.2.1 Componentes de una Bodega de datos**

Una Bodega de datos está compuesta por varios Datamarts los cuales son conjuntos de datos que ayudan a áreas específicas del negocio, para ayudar al proceso de toma de decisiones. Estos procesos son orientados hacia la consulta de herramientas OLAP (Procesamiento Analítico en Línea) que proporcionan una visión multidimensional de la información. Los sistemas OLAP tienen como objetivo agilizar la consulta de grandes volúmenes de datos, para lo cual utilizan estructuras multidimensionales o cubos OLAP, son sistemas que aumentan considerablemente el rendimiento de las consultas SQL tipo SELECT.

Los Cubos OLAP están compuestos de hechos, datos numéricos o medidas contenidos en una tabla central y dimensiones que son las perspectivas desde las cuales se va a analizar la información. (Ramírez Farfán & Benavides Arteaga)

### **2.2.2 Modelo Relacional**

El modelo relacional propone una estructura de datos simple, una relación (o tabla) compuesto de uno o varios atributos (o columnas). (Zimányi, 2008)

### **2.2.3 DataMart**

Versión de Datawarehouse con fines específicos. (Bartolini, 2009)

## **2.3 Bases de Datos Geográficos**

“Una base de datos geográfica es una colección de datos organizados de tal manera que sirvan efectivamente para una o varias aplicaciones SIG. Esta base de datos comprende la asociación entre sus dos principales componentes: Datos espaciales atributos o datos no espaciales”. (ESRI, 1998)

### **2.3.1 Datos espaciales**

Los datos o atributos espaciales son las características geográficas de los objetos descritos (Ubicación, Dimensión, Forma), Es decir, Los puntos que conforman el perímetro de una población están almacenados en cierto tipo de archivos que interpretan las aplicaciones geográficas que se encuentran en el mercado. (ESRI, 1998)

### **2.3.2 Datos no espaciales**

Los datos no espaciales o atributos son las características cuantitativas asociadas al objeto que se desea describir, generalmente se almacenan en tablas y se administran por algún manejador de bases de datos. También son llamados datos descriptivos. (ESRI, 1998)

### **2.3.3 Capas Geográficas**

Las capas son las características geográficas del evento o área que se desea modelar, organizadas en temas que se desean facilitar la información. Por ejemplo, Un mapa puede ser organizado en varios temas o capas, tales como división política, hidrología, caminos, contornos o puntos de control. Dichas capas pueden ser almacenadas en archivos separados, Pues sus atributos son diferentes. En el caso que utilicen bases de datos se almacenan en tablas separadas. (ESRI, 1998)

### **2.3.4 Entidad**

En general una entidad es una cosa (objeto, persona, evento, concepto) Distinguible de lo que le rodea acerca de la cual se requiere información para propósitos de la BDG, Una entidad es la representación digital del componente descriptivo de un dato geográfico. Se le asocia un nombre con el fin de distinguirla de otras entidades (ejemplos: carretera, presa, línea de transmisión, eje estructural, área agrícola). (Korth, 2002)

### **2.3.5 Representación Geométrica**

Constituye la representación digital del componente espacial de un rasgo geográfico. La BDG sustenta tres tipos diferentes y básicos de representación geométrica: Punto Línea y área. Cada entidad puede estar asociada con distintos tipos de representación geométrica. Por ejemplo, una localidad puede estar representada ya sea como punto, o como área, dependiendo de sus dimensiones y de la escala. (INEGI, 1997).

### **2.3.6 Modelos de datos**

Los modelos de datos son un conjunto de herramientas conceptuales para describir datos, sus relaciones, su significado y sus restricciones de consistencia los dos tipos de modelados importantes son los modelados orientados a registros también llamados relacionales y los modelos orientados a objetos. (Ver Figura 6 Modelos Bases de Datos Multidimensional.) (Korth, 2002)

## **2.4 Calidad De Vida (CV)**

La educación como proceso de mejoramiento de la calidad de vida de los individuos y de la comunidad; La integridad de todos los seres humanos está articulada también a valores sociales y de relaciones con otros individuos es importante resaltar que todas las actividades de salud como de educación identifican un medio que pueden ser favorables o desfavorables para un individuo para que se desarrolle en un ambiente de productividad comunitaria social y científica en un determinado medio. (Vincezi & Tudesco, 2009)

Algunas apreciaciones de distintos autores me permiten analizar y concluir que la perspectiva de cada ser humano se relaciona proporcionalmente con la educación, con el rol de las personas y el apoyo de los entornos donde participa este como individuo. La calidad de vida se dice que son las condiciones por las que tiene que pasar un individuo en el transcurso de su existencia.

Para saber la calidad de vida y el progreso de un individuo necesitamos estudiar muchas variables que hacen dependencia de cada una de las actividades que realiza el mismo.

En cuanto a la calidad de vida desde el que hacer pedagógico lo menciona VERDUGO 2001 es la prioridad que se imparte en la escuela es la calidad de vida y un objetivo primordial para conseguir la meta de la satisfacción y toda acción educativa que correlaciona los conocimientos y valores se enmarcan dentro de un sistema educativo que consta de diseños curriculares, programas, textos individuos y gobierno.

El DANE mide la calidad de vida de los Colombianos a través de una encuesta que se ha venido realizando desde el 2010 hasta la actualidad donde inicialmente hacia el estudio de algunas dimensiones. La encuesta del 2015 la más actual si posee el estudio de diferentes dimensiones o variables como lo son servicios básicos como lo son (Ver Figura 11

Porcentaje de cobertura de los servicios energía, Acueducto, Alcantarillado según la ECV del 2015) entre otros (Ver

Tabla 1. Variables incluidas en la encuesta del DANE de la Calidad de vida.) Por lo tanto es más exacto medir la calidad de vida de los individuos en cada región o departamento.

Este trabajo usa las mediciones cuantitativas adelantadas por el DANE de las condiciones de vida de los colombianos incluyendo variables relacionadas con la vivienda (material de paredes, pisos y servicios públicos, entre otros). Los hogares que involucra variables como: tenencia de bienes y percepción del jefe o del cónyuge sobre las condiciones de vida en el hogar y las personas para los que se incluyen variables de: educación, salud, cuidado de los niños, fuerza de trabajo, gastos e ingresos, etc. (DANE, 2015).

**Tabla 1. Variables incluidas en la encuesta del DANE de la Calidad de vida.**

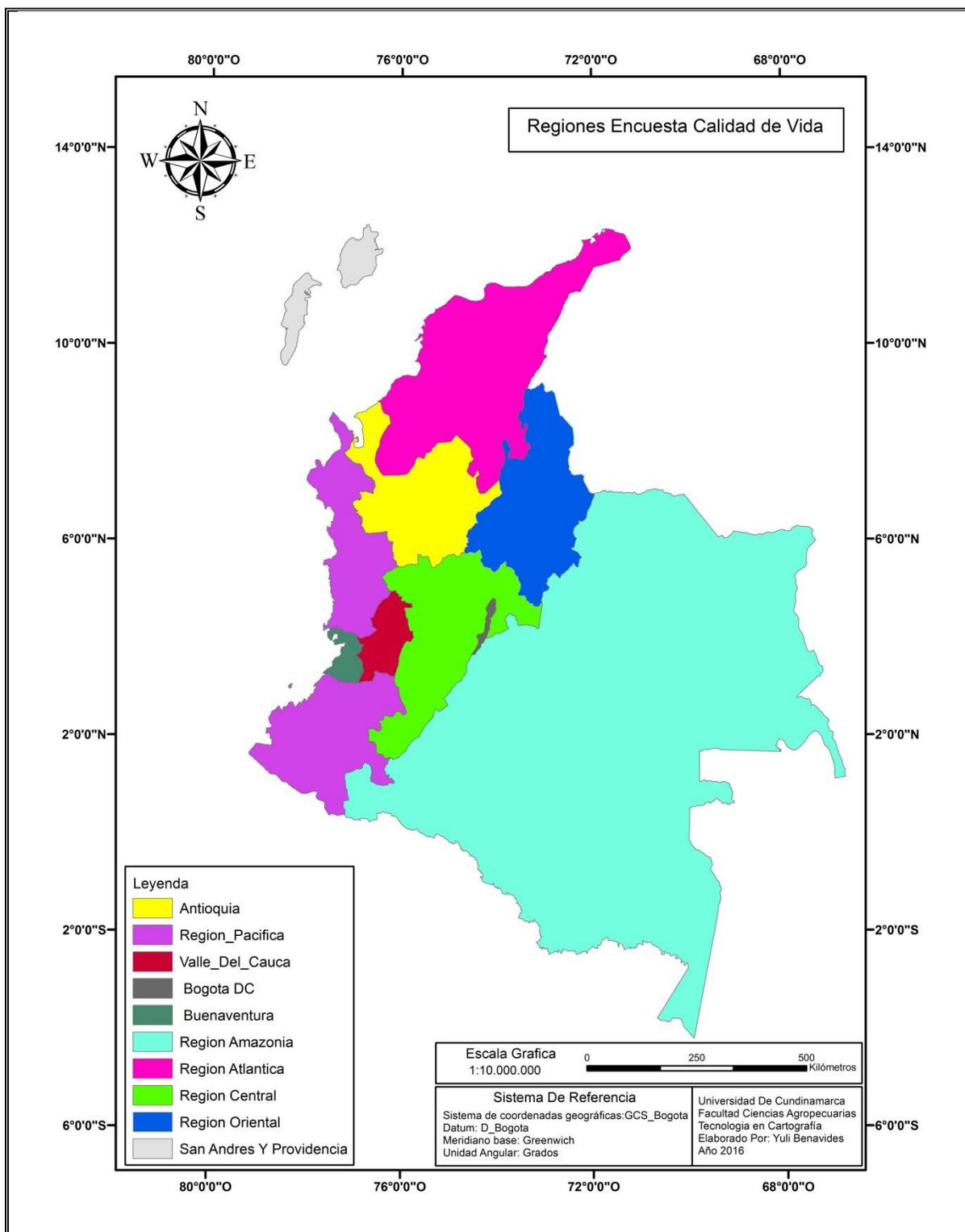
<b>VARIABLES ENCUESTA CV</b>
VIVIENDAS, HOGARES Y PERSONAS, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES POR ACCESO A SERVICIOS PÚBLICOS, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
TOTAL DE HOGARES Y DISTRIBUCION PORCENTUAL POR ACCESO A SERVICIOS PÚBLICOS, SEGÚN REGIONES DEL PAIS Y ÁREA (CAB - CP Y RURAL)
HOGARES QUE CLASIFICAN LAS BASURAS POR TIPO DE MATERIAL QUE CLASIFICAN, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
HOGARES POR TIPO DE PRÁCTICAS QUE REALIZAN PARA REDUCIR EL CONSUMO DE AGUA Y ENERGÍA DE ELÉCTRICA, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
HOGARES CON SERVICIO DE TELÉFONO FIJO Y TELÉFONO CELULAR, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
NÚMERO DE PERSONAS DE 5 AÑOS Y MÁS CON TELÉFONO CELULAR, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES QUE PREPARAN ALIMENTOS DE ACUERDO CON EL COMBUSTIBLE (ENERGÍA) QUE UTILIZAN PARA COCINAR, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
HOGARES QUE PREPARAN ALIMENTOS Y DISTRIBUCIÓN PORCENTUAL DE ACUERDO CON EL COMBUSTIBLE (ENERGÍA) QUE UTILIZAN PARA COCINAR, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
HOGARES POR TENENCIA DE LA VIVIENDA, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
HOGARES POR JEFATURA (MASCULINA O FEMENINA) SIN CÓNYUGE Y CON HIJOS MENORES DE 18 AÑOS, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
POBLACIÓN AFILIADA AL SISTEMA GENERAL DE SEGURIDAD SOCIAL EN SALUD (S.G.S.S.S.), POR REGÍMENES, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
NIÑOS Y NIÑAS MENORES DE 5 AÑOS POR SITIO O PERSONA CON QUIEN PERMANECEN LA MAYOR PARTE DEL TIEMPO ENTRE SEMANA, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
NIÑOS Y NIÑAS MENORES DE 5 AÑOS POR TIPO DE ESTABLECIMIENTO AL QUE ASISTEN, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
PROMEDIO DE AÑOS DE EDUCACIÓN DE LAS PERSONAS DE 5 AÑOS Y MÁS, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
PROMEDIO DE AÑOS DE EDUCACIÓN DE LAS PERSONAS DE 5 AÑOS Y MÁS, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
HOGARES POR OPINIÓN DEL JEFE O DEL CÓNYUGE SOBRE LOS INGRESOS DE SU HOGAR, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES POR OPINIÓN DEL JEFE O DEL CÓNYUGE RESPECTO A SI SE CONSIDERA POBRE, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES POR LOS BIENES Y SERVICIOS QUE POSEEN, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
DISTRIBUCIÓN (%) DE HOGARES SEGÚN PRINCIPAL MEDIDA TOMADA ANTES DE CONSUMIR EL AGUA PARA BEBER, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES EN LOS QUE ALGUNA PERSONA SABE QUE ES EL SERVICIO DE TELEVISIÓN DIGITAL TERRESTRE -TDT, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES SEGÚN OPINIÓN DEL JEFE O DEL CÓNYUGE RESPECTO A CÓMO SE SIENTE EN EL BARRIO, PUEBLO O VEREDA DONDE VIVE, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES POR PRÁCTICAS PARA REDUCIR EL CONSUMO DE AGUA Y ENERGÍA ELÉCTRICA, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
NÚMERO DE PERSONAS DE 5 A 24 AÑOS POR GRUPOS DE EDAD Y ASISTENCIA ESCOLAR, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
HOGARES CON SERVICIO DE INTERNET, POR TIPO DE CONEXIÓN, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
NÚMERO DE PERSONAS DE 5 AÑOS O MÁS QUE USAN INTERNET, POR FRECUENCIA DE USO, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
DISTRIBUCIÓN DE LA POBLACIÓN POR SEXO Y GRUPOS DE EDAD SEGÚN TIPO DE HOGAR DONDE HABITAN, REGIONES DEL PAÍS Y ÁREA (CAB Y CP Y RURAL DISPERSO)
TASA DE ACTIVIDAD DOMÉSTICA EN LA POBLACIÓN FEMENINA DE DOCE AÑOS Y MÁS SEGÚN EL NÚMERO DE MENORES DE 5 AÑOS PRESENTES EN EL HOGAR, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
ASISTENTES DE 5 AÑOS Y MÁS EN LOS NIVELES PREESCOLAR, BÁSICA Y MEDIA POR TIEMPO QUE GASTAN PARA IR AL ESTABLECIMIENTO EDUCATIVO, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
ASISTENTES EN LOS NIVELES PREESCOLAR, BÁSICA Y MEDIA POR SECTOR DEL ESTABLECIMIENTO EDUCATIVO DONDE ESTUDIAN Y JORNADA, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
HOGARES POR NÚMERO DE PERSONAS, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
ASISTENTES A ESTABLECIMIENTOS EDUCATIVOS OFICIALES EN LOS NIVELES PREESCOLAR A MEDIA QUE MANIFESTARON HABER TENIDO DURANTE ESTE AÑO ESCOLAR GASTOS POR PENSIÓN O MATRÍCULA, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
NÚMERO DE PERSONAS QUE POR PREVENCIÓN CONSULTAN POR LO MENOS UNA VEZ AL AÑO AL MÉDICO O AL ODONTÓLOGO, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
MENORES DE 5 AÑOS POR ACTIVIDADES REALIZADAS CON LA PERSONA QUE PERMANECEN LA MAYOR PARTE DEL TIEMPO EN EL HOGAR, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
USO DEL COMPUTADOR, INTERNET, CELULAR Y RADIO SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)

<b>VARIABLES ENCUESTA CV</b>
USO DE DISPOSITIVOS PARA ACCEDER A INTERNET, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES POR LOS BIENES Y SERVICIOS TIC QUE POSEEN, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
ASISTENTES DE 5 AÑOS Y MÁS EN LOS NIVELES PREESCOLAR, BÁSICA Y MEDIA POR PRINCIPAL MEDIO DE TRANSPORTE PARA IR AL ESTABLECIMIENTO EDUCATIVO, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB-CP Y RURAL DISPERSO)
NÚMERO DE PERSONAS POR OPINIÓN SOBRE LA CALIDAD DEL SERVICIO DE LA ENTIDAD DE LA SEGURIDAD SOCIAL EN SALUD EN LA QUE ESTÁ AFILIADO, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)
HOGARES POR NÚMERO DE PERSONAS, SEGÚN REGIONES DEL PAÍS Y ÁREA (CAB - CP Y RURAL DISPERSO)

En la

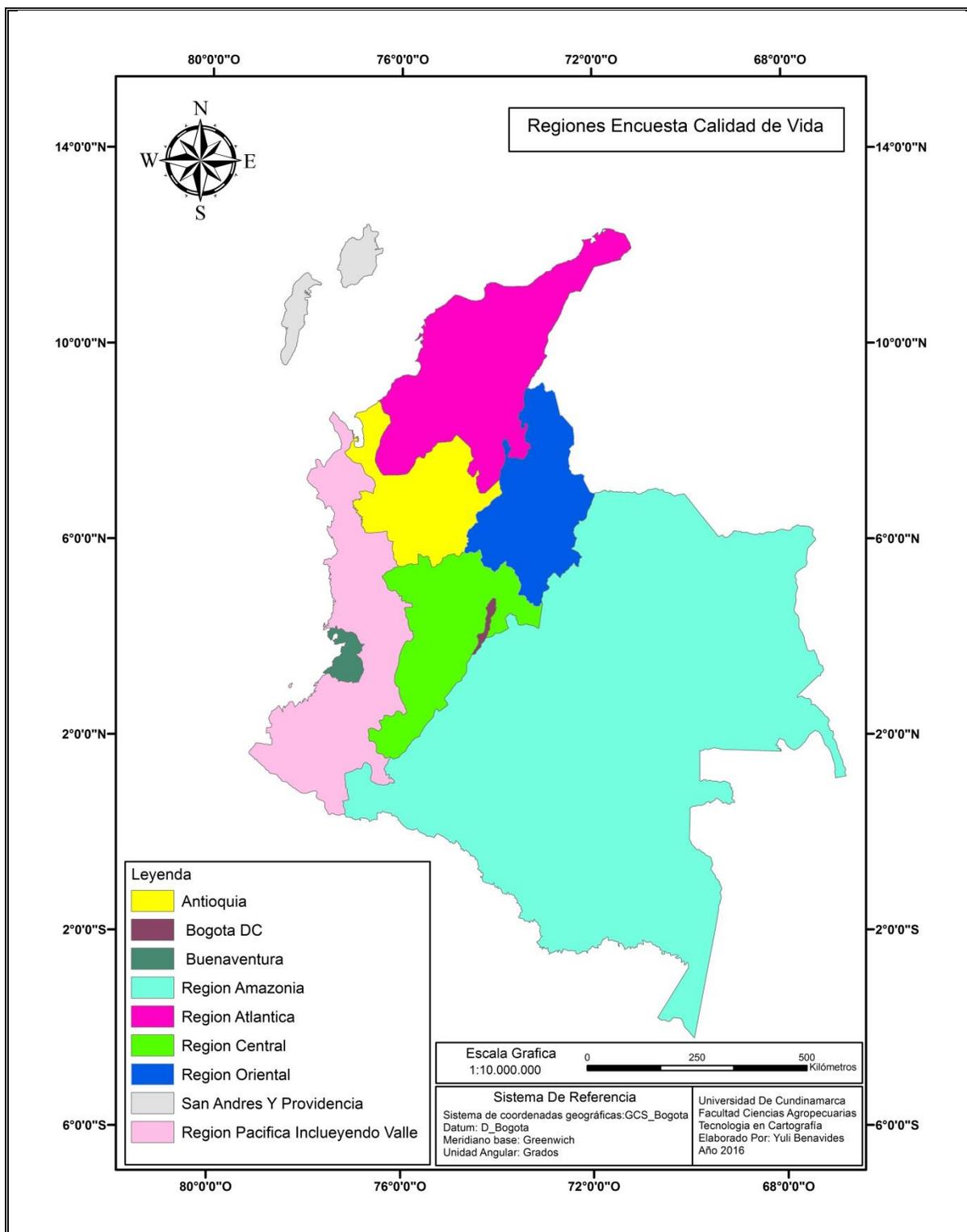
Tabla 1 se presentan las variables que actualmente mide el DANE en la encuesta de calidad de vida. Dichas variables están medidas discriminando la cabecera municipal y los centros poblados y rural disperso, de forma absoluta y relativa. En la Figura 3 Regiones .

Figura 3 Regiones ECV.



Fuente: Elaboración propia con base de datos INS.

**Figura 4 Regiones ECV (Región Pacifico Incluyendo Valle)**



Fuente: Elaboración propia con base de datos INS.

## **2.5 MEN**

El Ministerio de Educación Nacional tiene como fin lograr una educación de calidad, que forme mejores seres humanos, ciudadanos con valores éticos, competentes, respetuosos de lo público, que ejercen los derechos humanos, cumplen con sus deberes y conviven en paz. Una educación que genere oportunidades legítimas de progreso y prosperidad para ellos y para el país. Mediante una educación competitiva, pertinente, que contribuya a cerrar brechas de inequidad y en la que participa toda la sociedad. (Nacional).

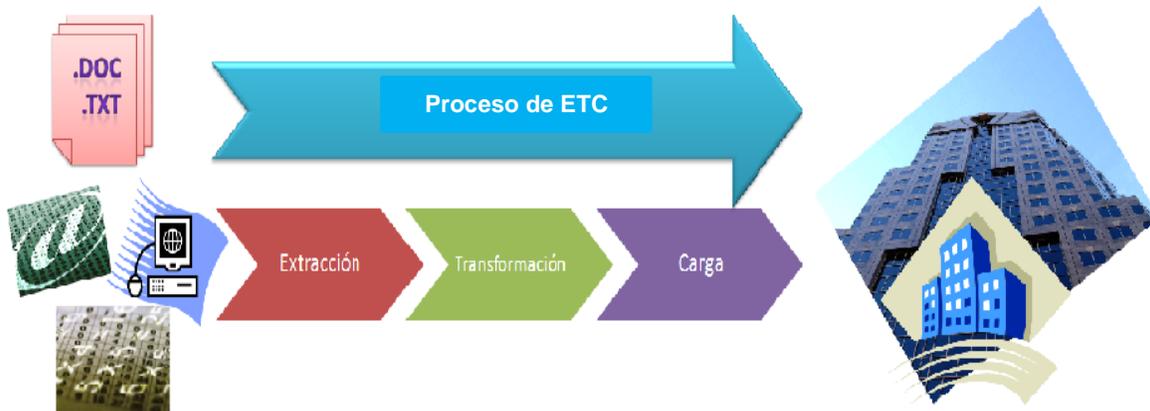
En el capítulo de resultados se presenta el análisis exploratorio y su correspondiente representación en mapas Coropléticos ver (Figura 9 Relación Mapas Coropléticos IES Por Departamentos Y Distrito Capital Años (2000-2006-2013)).

### 3. 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:

#### 3.1 Extraer, Transformar y cargar

El proceso de ETC (Extraer, Transformar y Cargar) (Ver Figura 5 Extracción, Transformación y carga) 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 5 Extracción, Transformación y carga (Ramírez Farfán & Benavides Arteaga)*

##### 3.1.1 Extraer.

La primera parte del proceso del ETL es extraer y consiste en obtener los datos de los sistemas origen. La mayoría de los datos provienen de diferentes sistemas origen generalmente. Cada sistema puede tener formatos distintos y pertenecer a organizaciones diferentes, los formatos se encuentran en bases de datos relacionales, archivos planos, bases de datos no relacionales y estructuras diferentes. Analizar los datos extraídos implica recibirlos en el formato esperado de no ser así los datos son rechazados. El proceso de extracción convierte los datos a un formato preliminar para la transformación.

Los sistemas de extracción deben causar un impacto mínimo en los sistemas origen, si la extracción tiene un alto impacto en los sistemas origen estos se pueden perder rendimiento e incluso colapsar, por esta razón las operaciones de extracción son programadas en horarios y días donde el impacto sobre el sistema origen sea el mínimo. (Ramírez Farfán & Benavides Arteaga)

### **3.1.2 Transformar.**

En la etapa de transformación es donde se aplican las reglas de negocio, donde los datos son transformados o convertidos en datos que serán cargados en la bodega de datos. Algunos datos no necesitan transformaciones o las transformaciones pueden ser mínimas, pero en otros casos se pueden aplicar otras transformaciones descritas a continuación:

- ✓ Dependiendo de las reglas del negocio podemos o no cargar las columnas nulas o seleccionar solo algunas columnas para cargar.
- ✓ Traducción de códigos, dependiendo de los datos almacenados en el origen por ejemplo en el origen hombre se encuentra con M y mujer con F y definimos que en la bodega se deben guardar como 1 y 2.
- ✓ Codificar valores libres, por ejemplo Sr en 1.
- ✓ Obtener valores calculados por ejemplo la fecha de carga tomarla del sistema o valores como totales (total vendido = Cantidad \* precio de venta).
- ✓ Unir datos entre dos orígenes para completar la información.
- ✓ Calcular valores de múltiples datos.
- ✓ Dividir una columna en varias, por ejemplo el nombre si está completo dividirlo en Nombre, Segundo Nombre, Primer Apellido y Segundo Apellido.
- ✓ Entrega de los datos validados.
- ✓ Políticas de manejo de excepciones por ejemplo rechazar los registros con error o dejarlos nulos dependiendo de las políticas de negocio.

### **3.1.3 Carga.**

Es la fase en la cual tomamos los datos transformados para ser cargado en el sistema destino Datamarts. En la mayoría de bases de datos se sobrescribe la información, pero en las bodegas de datos se debe mantener la historia de los registros de tal manera que se pueda hacer una auditoria de los datos y se vea su historia a lo largo del tiempo. El proceso de carga se puede dividir en:

- ✓ Acumulación Simple: Consiste en realizar un resumen de las transacciones dentro de un periodo de tiempo y transportar los datos en una sola transacción hacia la bodega de datos.
- ✓ Rolling: Se aplica en el caso en los que se pueda mantener varios niveles de granularidad, lo cual implica almacenar la información a diferentes niveles, que pueden corresponder a la unidad de tiempo o diferentes niveles jerárquicos.

### **3.2 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.

### **3.3 Creación 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 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)

## **3.4 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.

### **3.5 Implementación métodos estadísticos**

#### **3.5.1 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, Análisis Exploratorio De Datos Espaciales Al Servicio Del Geomarketing, 2003)

#### **3.5.2 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 Sturges.

### **3.6 Producción graficas estadísticas**

#### **3.6.1 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)

### **3.7 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 es almacenada 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.

## 4. RESULTADOS

### 4.1 Recolección 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)

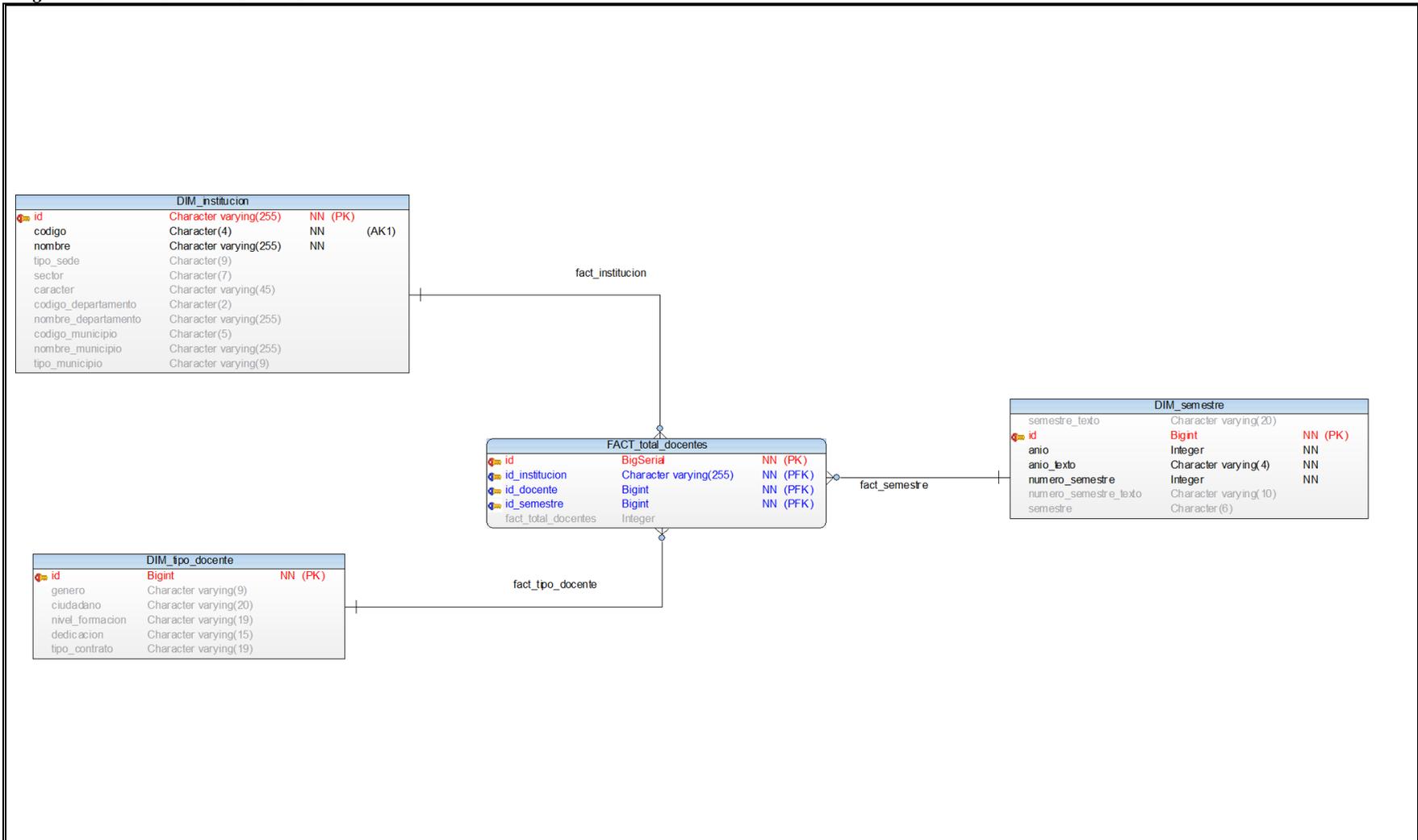
*Tabla 2 Datos usados para el análisis y la producción cartográfica*

### 4.2 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 6 Modelos Bases de Datos Multidimensional.



DIM_institucion1		
id	Character varying(255)	NN (PK)
codigo	Character (4)	NN (AK1)
nombre	Character varying(255)	NN
tipo_sede	Character (9)	
sector	Character (7)	
caracter	Character varying(45)	
codigo_departamento	Character (2)	
nombre_departamento	Character varying(255)	NN
codigo_municipio	Character (5)	
nombre_municipio	Character varying(255)	
tipo_municipio	Character varying(9)	

programa		
id	Bigint	NN (PK)
codigo_snies	Character varying(20)	NN
nombre	Character varying(255)	NN
nivel_formacion	Character varying(20)	
metodologia	Character varying(25)	NN
area_conocimiento	Character varying(50)	NN
nucleo_basico_conocimiento	Character varying(55)	
codigo_departamento	Character (2)	NN
nombre_departamento	Character varying(255)	NN
codigo_municipio	Character varying(5)	NN
nombre_municipio	Character varying(255)	NN

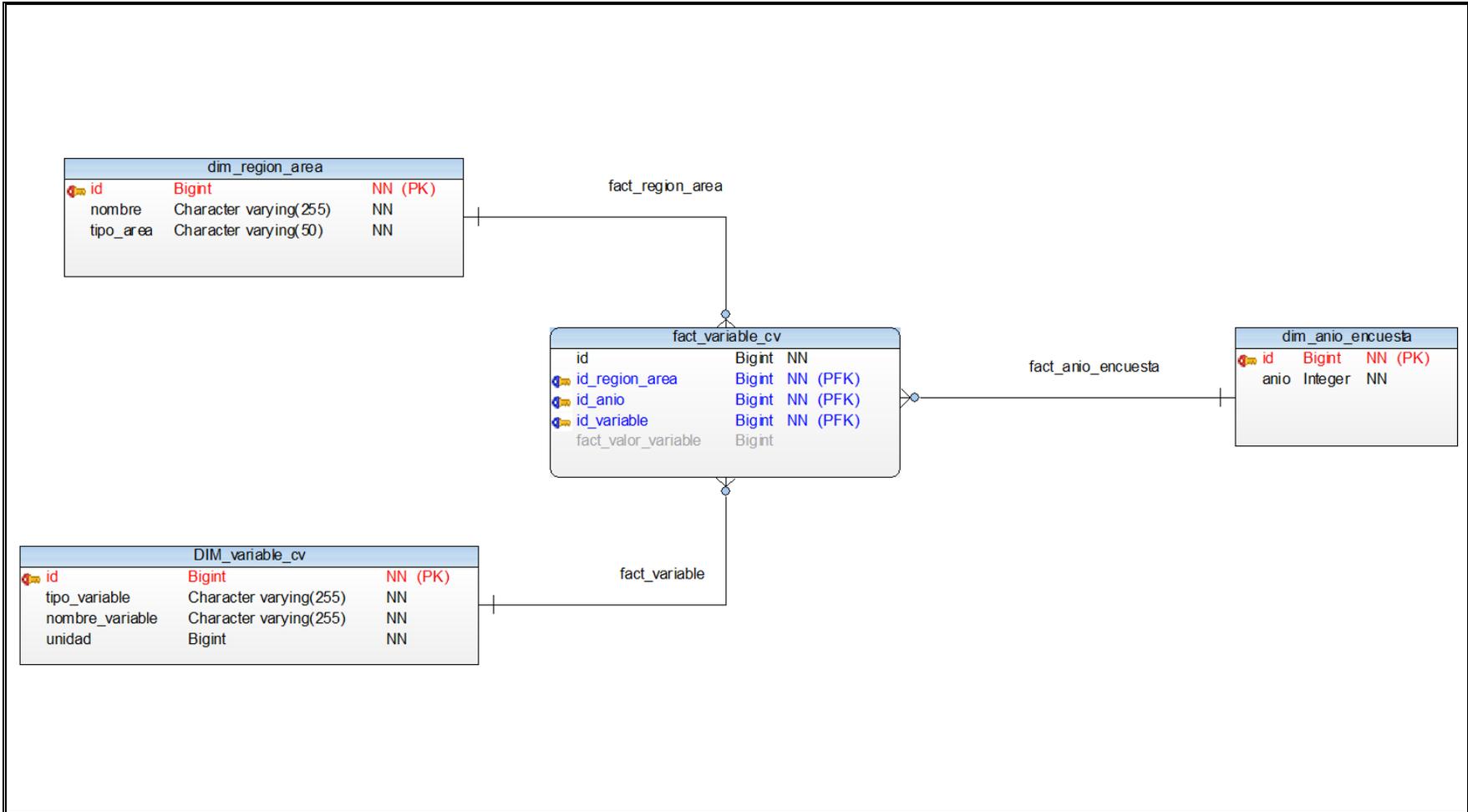
fact_matriculados		
id	Bigint	NN
id_institucion	Character varying(255)	NN (PFK)
id_programa	Bigint	NN (PFK)
id_periodo	Bigint	NN (PFK)
fact_tdtal_hombres	Integer	NN
fact_tdtal_mujeres	Integer	NN
fact_tdtal_matriculados	Integer	NN

periodo_academico		
id	Bigint	NN (PK)
anio	Integer	NN

fact\_institucion

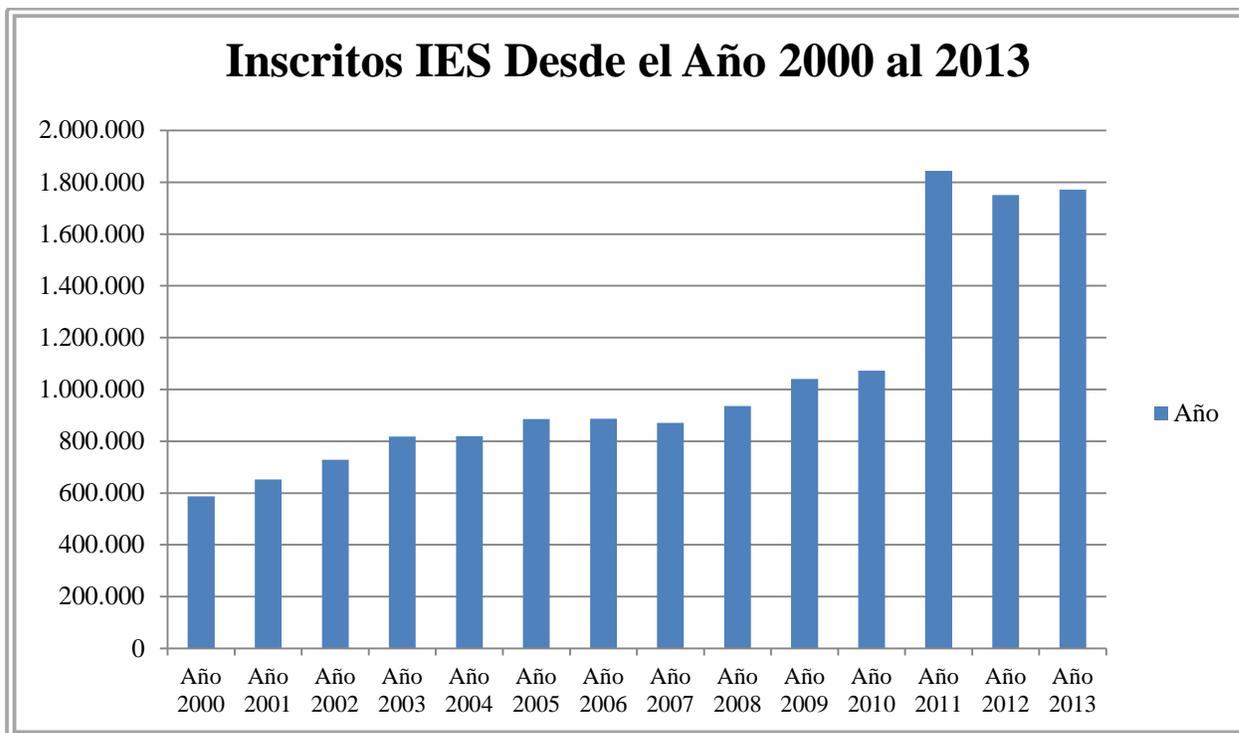
fact\_periodo

fact\_programa



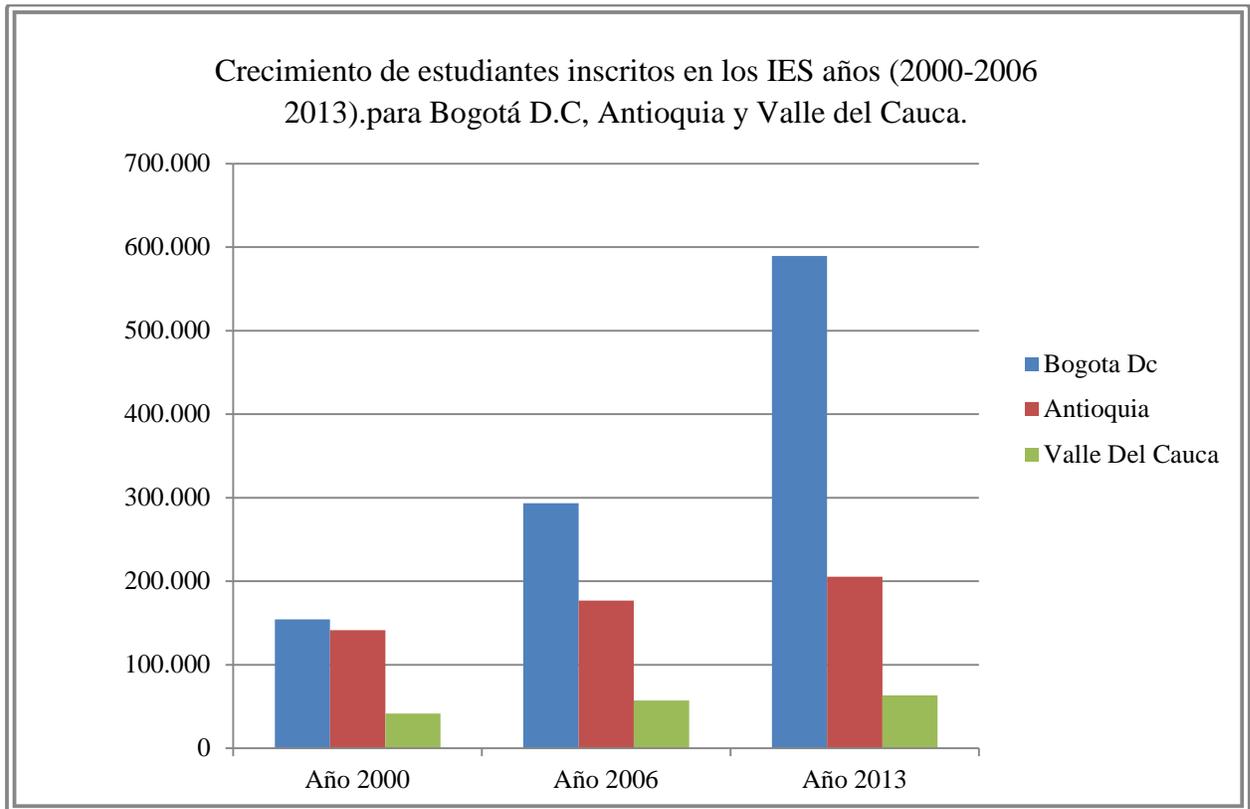
## 4.2 Análisis de cobertura educativa a nivel profesional.

*Figura 7 Personas Inscritas a los IES Nacional Desde el año 2000 al 2013.*



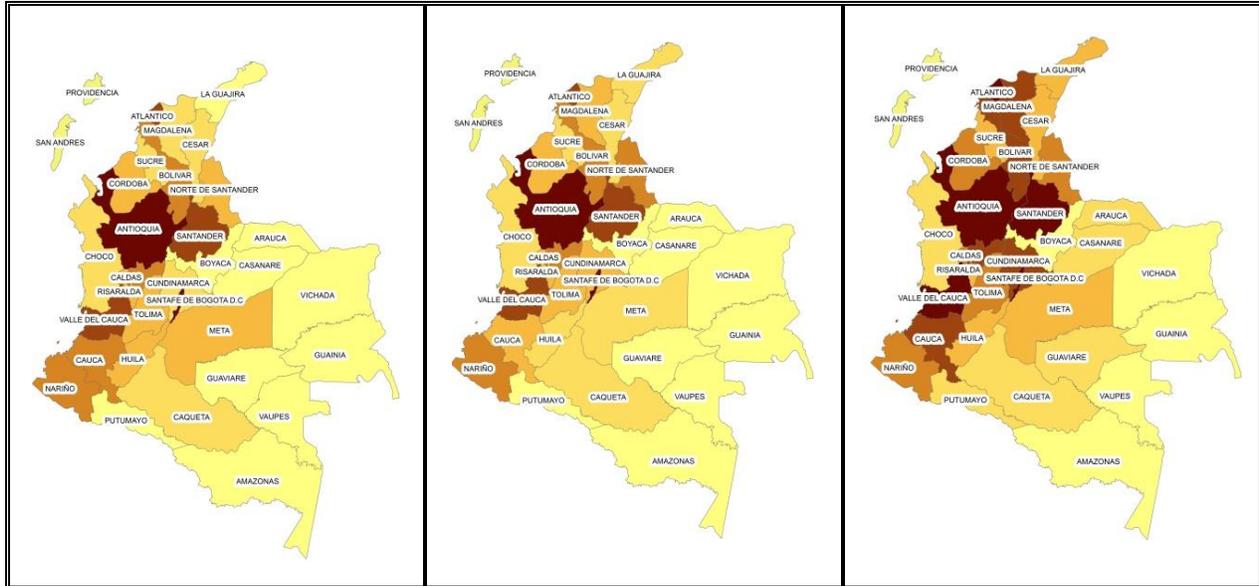
Fuente: Elaboración propia con información del MEN.

**Figura 8 Crecimiento de estudiantes inscritos en los IES años (2000-2006-2013).para Bogotá D.C, Antioquia y Valle del Cauca.**



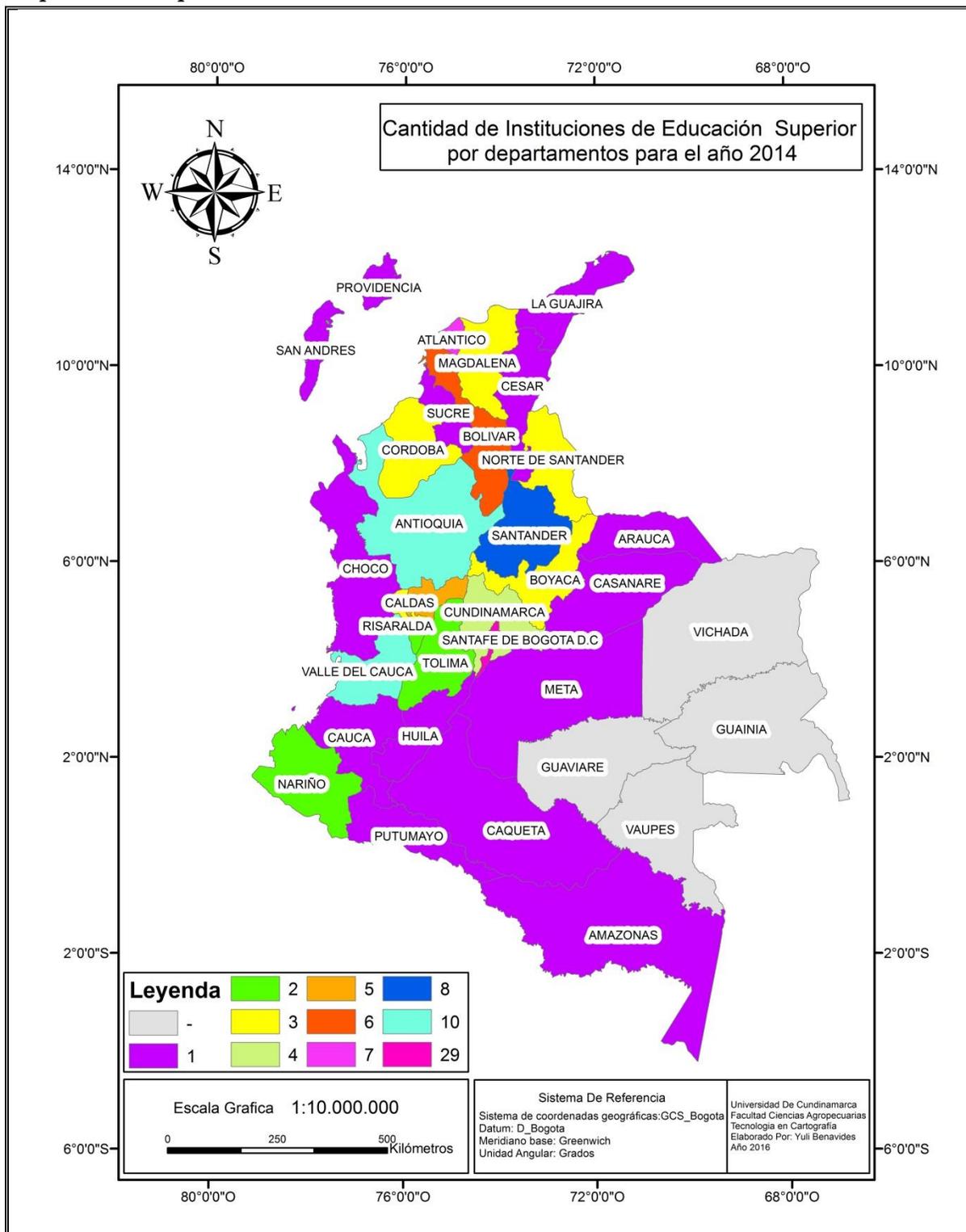
Fuente: Elaboración propia con información del MEN.

**Figura 9 Relación Mapas Coropleticos IES Por Departamentos Y Distrito Capital Años (2000-2006-2013)**



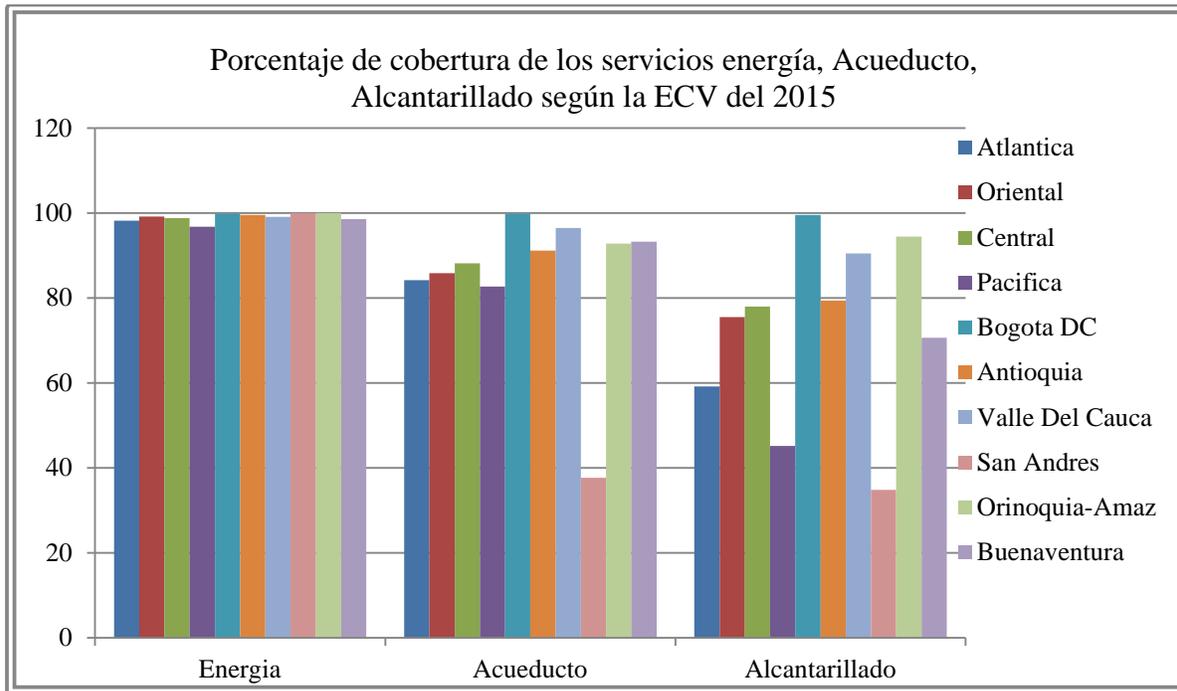
Fuente: Elaboración propia con información del MEN.

**Figura 10 Distribución por números según la cantidad de IES que hay por departamentos para el año 2014**



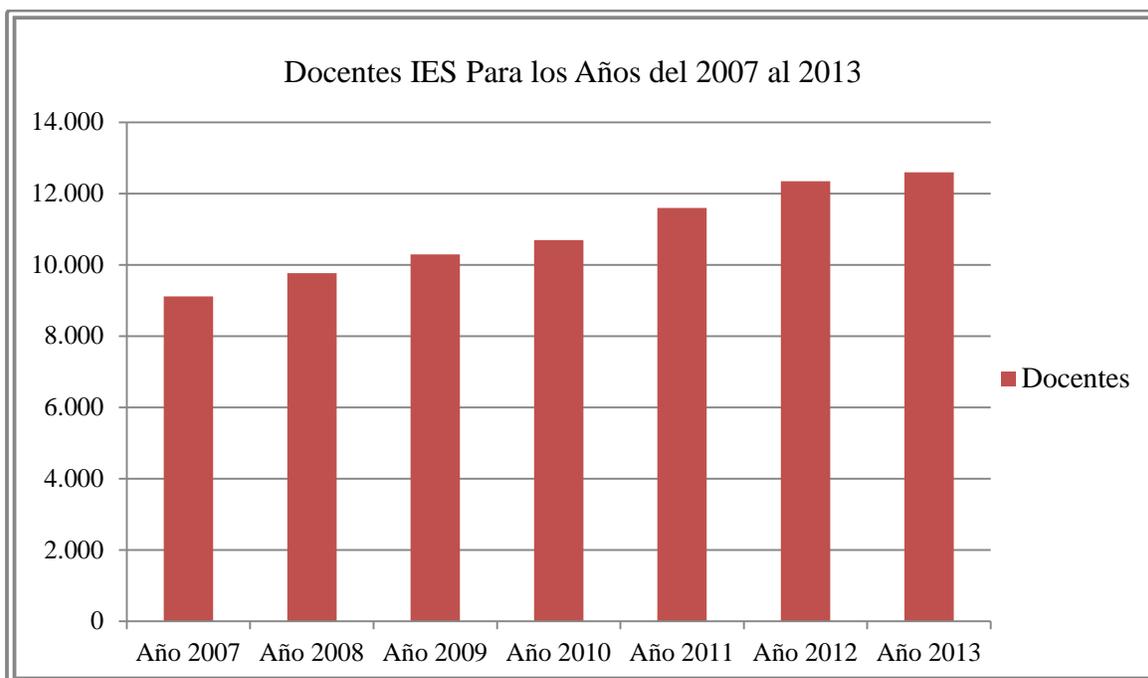
Fuente: Elaboración propia con base de datos del MEN.

**Figura 11 Porcentaje de cobertura de los servicios energía, Acueducto, Alcantarillado según la ECV del 2015**



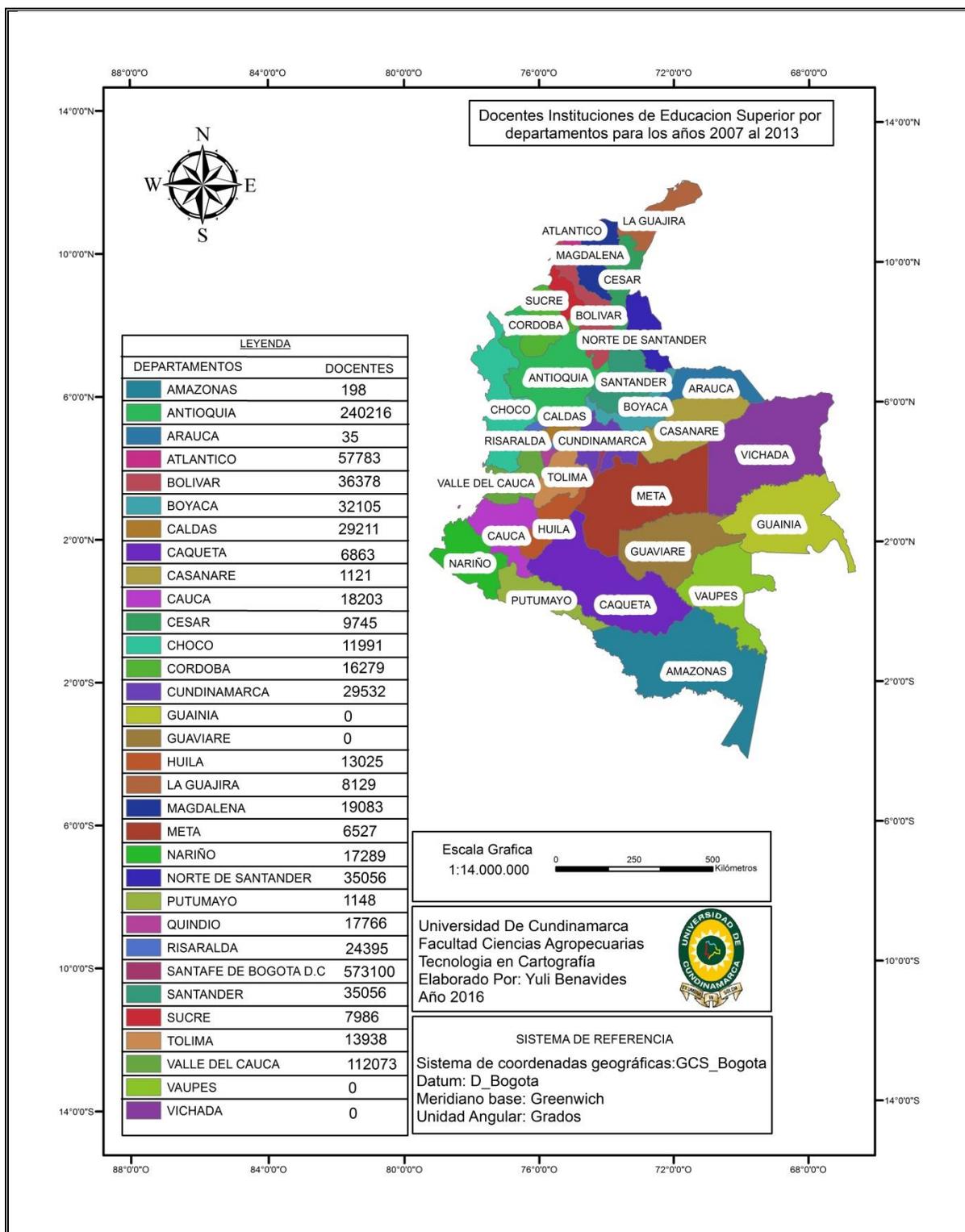
Fuente: Elaboración propia con información del DANE.

**Figura 12 Docentes IES Para los Años del 2007 al 2013**



Fuente: Elaboración propia con información del MEN.

**Figura 13 Docentes IES por departamentos para los años del 2007 al 2013**



Fuente: Elaboración propia con base de datos del MEN.

## 5. CONCLUSIONES

- ✓ Se ha desarrollado un modelo de datos multidimensional para los datos públicos del Ministerio de Educación y los indicadores de calidad de vida del DANE. Se aplicaron las técnicas de diseño de bases de datos en sistemas orientados a la consulta y de esta manera se realizó un eficiente análisis de datos exploratorio el cual se refleja en diferentes mapas temáticos que evidencian la funcionalidad del modelo de datos realizado.
- ✓ El producto de este trabajo constituye un insumo útil para futuros estudios. Se ha logrado una base de datos con más de quinientos mil registros en unos sesenta megabits. Adicionalmente, la base de datos soporta la inclusión de nuevos datos lo cual facilita la ampliación de los análisis y el uso de técnicas más robustas de análisis estadístico.
- ✓ La base de datos fue implementada sobre un sistema gestor de datos escalable a ambientes de producción e integrable con otras tecnologías de BI e independiente de plataforma, es decir que funciona sobre Windows, Solaris, GNU/Linux y Mac OS X.
- ✓ El modelo de datos también es multiescala, es decir que se puede aumentar la resolución tanto espacial como temporal.

## 6. LISTA DE REFERENCIAS

- Antioquia, U. d. (2008). *Cubos gráficos OLAP*. Medellin, Antioquia: Universidad de Antioquia. Facultad de Ciencias Sociales y Humanas.
- Arriazu, R. (2009). *Mejor Educacion Asegura Mayor Calidad De Vida*.
- Arroyo, T. d. (2015). *Inteligencia de negocios, una introduccion*.
- Bartolini, G. (2009). *Data Warehousing With PostgreSQL*. Paris, Francia.
- Cano, J. L. (2007). *Business Intelligence: Competir Con Informacion*.
- DANE. (2015). *Departamento Administrativo Nacional De Estadistica* . Obtenido de <http://www.dane.gov.co/index.php/estadisticas-sociales/calidad-de-vida-ecv>
- Duschatzky, A. B.-S. (1999). *Los Condicionantse De La Calidad Educativa*. Buenos Aires , Arg: Novedades Educativas .
- España, I. G. (s.f.). *Bases De Datos Geograficos*.
- ESRI. (1998). *ESRI Shapefile Technical Description*. ESRI White Paper.
- Estadística, I. N. (2015). *Indicadores Calidad De Vida*. Obtenido de [www.ine.es/dyrt3/metadatos/es/RespuestaDatos.htm?oe=30471](http://www.ine.es/dyrt3/metadatos/es/RespuestaDatos.htm?oe=30471)
- INEGI. (1997). *Características Metodológicas de la Encuesta Nacional de la Dinámica Demográfica (ENADID) 1997*.
- Inmon, W. H. (2002). *Building The Data Warehouse*. Wiley Computer Publishing.
- (s.f.). *Introducción a la Cartografía Temática*.
- Juan Carlos Trujillo, J. N. (s.f.). *Diseño Y Explotacion De Almacenes De Datos*. Editorial Club Universitario.
- Korth, H. (2002). *Fundamentos De Bases De Datos*. Madrid: Concepción Fernández Madrid.
- Londoño, D. H. (2007). *SIG y teoría de percolación aplicados al estudio de fenómenos de propagación en epidemiología*. 1-3.
- Maria Sueli Almeida, M. I. (1999). *Getting Started with Data Warehouse*. (IBM) International Business Machines Corporation.
- México, U. N. (2009). *HISTOGRAMA* . 1-4.
- Myers, R. H.-S. (1999). *Probabilidad Y Estadística Para Ingenieros*.
- Nacional, M. D. (s.f.). *MINEDUCACION*. Obtenido de <http://www.mineducacion.gov.co/1759/w3-propertyvalue-55301.html>

- Polembo, R. (2012). *Calidad de Vida: Conceptos y medidas*. Santiago, Chile: CELADE.
- Rafanelli, M. (2003). *Multidimensional Databases: Problems a Solutions*. Italia : Idea Group Publishing.
- 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.
- SAP (Sistemas, A. y. (2011). SAP Crystal Reportes del usuario 2011. En SAP.
- Tiempo, E. (30 de Octubre de 2014). 14 De 100 Colombianos Viven En Tugurios.
- Vincezi, A. D. (2009). La Educacion Como Proceso De Mejoramiento De La Caludad De Vida De Los Individuos Y De La Comunidad. *Revista Iberoamericana De Educacion*.
- Yrigoyen, C. C. (2003). Análisis Exploratorio De Datos Espaciales Al Servicio Del Geomarketing. 4-5.
- Zimányi, E. M.-E. (2008). *Advanced Data Warehouse Design*.

## 7. ANEXOS

### A.1. Cartografía temática

Se anexan los siguientes mapas:

- ✓ Mapa Regiones ECV
- ✓ Mapa Coropleticos de los estudiantes inscritos para los años 2000 al 2013 a escala 1:10.000.000.
- ✓ Mapa de distribución por números según las IES que hay por departamentos para el año 2014

### A.2. Diccionario de datos.

Entidades.

Etiqueta	Nombre	Comments
FACT_total_docentes	FACT_total_docentes	
DIM_institucion	DIM_institucion	Sede en la que se ofertan programas de educación superior, está localizada en un municipio, aunque en oportunidades pertenece a una institución de orden nacional o departamental.
DIM_tipo_docente	DIM_tipo_docente	En esta dimensión se alojan los posibles tipo de docentes que se tienen, en cuanto a genero, nivel educativo y nacionalidad, entre otros aspectos.
DIM_semestre	DIM_semestre	Dimension que incluye el tiempo. La resolución es semestral.
DIM_institucion1	DIM_institucion1	Sede en la que se ofertan programas de educación superior, está localizada en un municipio, aunque en oportunidades pertenece a una institución de orden nacional o departamental.
Programa	programa	Programa académico.
periodo_academico	periodo_academico	Año para el cual se conoce el total de matriculados.
fact_matriculados	fact_matriculados	Total de matriculados en el año por género.
dim_region_area	dim_region_area	Regiones o áreas geográficas donde se especifica la variable de encuesta de calidad de vida del DANE.
dim_anio_encuesta	dim_anio_encuesta	Dimensión temporal para las encuestas de calidad de vida.
DIM_variable_cv	DIM_variable_cv	Total viviendas reportadas por región o área.
fact_variable_cv	fact_variable_cv	Tabla de hechos de las variables de calidad de vida de la encuesta del DANE.

**Columnas:**

Table	Etiqueta	Nombre	Tipo de dato	Primary Key	Not Null	Descripción
FACT_total_docentes	Id	id	BigSerial	VERDADERO	VERDADERO	
FACT_total_docentes	id_institucion	id_institucion	Character varying(255)	VERDADERO	VERDADERO	Llave foranea de las instituciones o sedes.
FACT_total_docentes	id_docente	id_docente	Bigint	VERDADERO	VERDADERO	Llave foranea de los tipos de docentes.
FACT_total_docentes	id_semestre	id_semestre	Bigint	VERDADERO	VERDADERO	Llave foranea del semestre.
FACT_total_docentes	fact_total_docentes	fact_total_docentes	Integer	FALSO	FALSO	
DIM_institucion	id	id	Character varying(255)	VERDADERO	VERDADERO	
DIM_institucion	codigo	codigo	Character(4)	FALSO	VERDADERO	Código de la institución ante el ministerio de educación.
DIM_institucion	nombre	nombre	Character varying(255)	FALSO	VERDADERO	
DIM_institucion	tipo_sede	tipo_sede	Character(9)	FALSO	FALSO	Tipo de sede, puede ser PRINCIPAL o SECCIONAL.
DIM_institucion	sector	sector	Character(7)	FALSO	FALSO	Campo para especificar si la institución es oficial o es privada.
DIM_institucion	caracter	caracter	Character varying(45)	FALSO	FALSO	El carácter de la institución educativa es para saber si es universidad, institución universitaria o escuela tecnológica, institución tecnológica profesional o Institución técnica profesional,
DIM_institucion	codigo_departamento	codigo_departamento	Character(2)	FALSO	FALSO	Código DANE del departamento. Esta compuesto de dos caracteres. Es numérico.

Table	Etiqueta	Nombre	Tipo de dato	Primary Key	Not Null	Descripción
DIM_institucion	nombre_departamento	nombre_departamento	Character varying(255)	FALSO	FALSO	Nombre de la unidad orgánica territorial de orden departamental en la que se encuentra ubicada la institución o la sede a la que se hace referencia.
DIM_institucion	codigo_municipio	codigo_municipio	Character(5)	FALSO	FALSO	Código DANE del municipio. Corresponde a los dos caracteres del código departamental más tres del código del municipio.
DIM_institucion	nombre_municipio	nombre_municipio	Character varying(255)	FALSO	FALSO	Nombre de la unidad orgánica territorial de orden municipal en la que se encuentra ubicada la institución o la sede a la que se hace referencia.
DIM_institucion	tipo_municipio	tipo_municipio	Character varying(9)	FALSO	FALSO	Especifica si el municipio es capital. Tiene dos posibles valores: MUNICIPIO y CAPITAL.
DIM_tipo_docente	id	id	Bigint	VERDADERO	VERDADERO	Identificador único de cada tipo de docente.
DIM_tipo_docente	genero	genero	Character varying(9)	FALSO	FALSO	Genero de los docentes, puede ser FEMENINO o MASCULINO.
DIM_tipo_docente	ciudadano	ciudadano	Character varying(20)	FALSO	FALSO	Tipo de ciudadano, NACIONAL o EXTRANJERO.

Table	Etiqueta	Nombre	Tipo de dato	Primary Key	Not Null	Descripción
DIM_tipo_docente	nivel_formacion	nivel_formacion	Character varying(19)	FALSO	FALSO	El nivel de formación de los docentes, por ejemplo, técnico o doctorado.
DIM_tipo_docente	dedicacion	dedicacion	Character varying(15)	FALSO	FALSO	Tiempo de dedicación de los docentes.
DIM_tipo_docente	tipo_contrato	tipo_contrato	Character varying(19)	FALSO	FALSO	Tipo de contratación del docente.
DIM_semestre	semestre_texto	semestre_texto	Character varying(20)	FALSO	FALSO	Semestre expresado en texto. Por ejemplo: "PRIMER SEMESTRE DE 2010"
DIM_semestre	id	id	Bigint	VERDADERO	VERDADERO	Llave primaria del semestre.
DIM_semestre	anio	anio	Integer	FALSO	VERDADERO	Año para el cual se está especificando el semestre.
DIM_semestre	anio_texto	anio_texto	Character varying(4)	FALSO	VERDADERO	el año pero en formato de texto.
DIM_semestre	numero_semestre	numero_semestre	Integer	FALSO	VERDADERO	Número que identifica el semestre, 1 o 2.
DIM_semestre	numero_semestre_texto	numero_semestre_texto	Character varying(10)	FALSO	FALSO	El indicador del semestre del año en formato de texto. Puede ser PRIMERO y SEGUNDO.
DIM_semestre	semestre	semestre	Character(6)	FALSO	FALSO	Semestre especificado con año y número de semestre. Por ejemplo "2010-2"
DIM_institucion1	id	id	Character varying(255)	VERDADERO	VERDADERO	
DIM_institucion1	codigo	codigo	Character(4)	FALSO	VERDADERO	Código de la institución ante el ministerio de educación.
DIM_institucion1	nombre	nombre	Character varying(255)	FALSO	VERDADERO	
DIM_institucion1	tipo_sede	tipo_sede	Character(9)	FALSO	FALSO	Tipo de sede, puede ser PRINCIPAL o SECCIONAL.

Table	Etiqueta	Nombre	Tipo de dato	Primary Key	Not Null	Descripción
DIM_institucion1	sector	sector	Character(7)	FALSO	FALSO	Campo para especificar si la institución es oficial o es privada.
DIM_institucion1	caracter	caracter	Character varying(45)	FALSO	FALSO	El carácter de la institución educativa es para saber si es universidad, institución universitaria o escuela tecnológica, institución tecnológica profesional o Institución técnica profesional,
DIM_institucion1	codigo_departamento	codigo_departamento	Character(2)	FALSO	FALSO	Código DANE del departamento. Esta compuesto de dos caracteres. Es numérico.
DIM_institucion1	nombre_departamento	nombre_departamento	Character varying(255)	FALSO	VERDADERO	Nombre de la unidad orgánica territorial de orden departamental en la que se encuentra ubicada la institución o la sede a la que se hace referencia.
DIM_institucion1	codigo_municipio	codigo_municipio	Character(5)	FALSO	FALSO	Código DANE del municipio. Corresponde a los dos caracteres del código departamental más tres del código del municipio.
DIM_institucion1	nombre_municipio	nombre_municipio	Character varying(255)	FALSO	FALSO	Nombre de la unidad orgánica territorial de orden municipal en la que se encuentra ubicada la institución o la sede a la que se hace referencia.

Table	Etiqueta	Nombre	Tipo de dato	Primary Key	Not Null	Descripción
DIM_institucion1	tipo_municipio	tipo_municipio	Character varying(9)	FALSO	FALSO	Especifica si el municipio es capital. Tiene dos posibles valores: MUNICIPIO y CAPITAL.
programa	id	id	Bigint	VERDADERO	VERDADERO	Identificador único del programa.
programa	codigo_snies	codigo_snies	Character varying(20)	FALSO	VERDADERO	Código único en el Sistema Nacional de Información de la Educación Principal.
programa	nombre	nombre	Character varying(255)	FALSO	VERDADERO	Nombre del programa académico para el Ministerio de Educación.
programa	nivel_formacion	nivel_formacion	Character varying(20)	FALSO	FALSO	Nivel de formación del programa.
programa	metodologia	metodologia	Character varying(25)	FALSO	VERDADERO	Metodología del programa, puede ser: presencial, a distancia (virtual), a distancia (tradicional) y sin metodología definida.
programa	area_conocimiento	area_conocimiento	Character varying(50)	FALSO	VERDADERO	Área del conocimiento según el Ministerio de Educación.
programa	nucleo_basico_conocimiento	nucleo_basico_conocimiento	Character varying(55)	FALSO	FALSO	Núcleos básicos del conocimiento según el SNIES.
programa	codigo_departamento	codigo_departamento	Character(2)	FALSO	VERDADERO	Código DANE del departamento donde se oferta el programa.
programa	nombre_departamento	nombre_departamento	Character varying(255)	FALSO	VERDADERO	Nombre del departamento donde se oferta el programa.
programa	codigo_municipio	codigo_municipio	Character varying(5)	FALSO	VERDADERO	Código DANE del municipio donde se oferta el programa.

Table	Etiqueta	Nombre	Tipo de dato	Primary Key	Not Null	Descripción
programa	nombre_municipio	nombre_municipio	Character varying(255)	FALSO	VERDADERO	Nombre del municipio donde se oferta el programa.
periodo_academico	id	id	Bigint	VERDADERO	VERDADERO	Identificador único del periodo académico. Actualmente esta con una resolución temporal anual.
periodo_academico	anio	anio	Integer	FALSO	VERDADERO	El año en números, cuatro dígitos, por ejemplo 2016.
fact_matriculados	id	id	Bigint	FALSO	VERDADERO	Identificador único.
fact_matriculados	id_institucion	id_institucion	Character varying(255)	VERDADERO	VERDADERO	Llave foranea de la institución o sede donde se encuentran matriculados los estudiantes.
fact_matriculados	id_programa	id_programa	Bigint	VERDADERO	VERDADERO	Llave foranea con el programa académico.
fact_matriculados	id_periodo	id_periodo	Bigint	VERDADERO	VERDADERO	Llave foranea con el periodo académico.
fact_matriculados	fact_total_hombres	fact_total_hombres	Integer	FALSO	VERDADERO	Total de hombres matriculados.
fact_matriculados	fact_total_mujeres	fact_total_mujeres	Integer	FALSO	VERDADERO	Total de mujeres matriculadas al programa.
fact_matriculados	fact_total_matriculados	fact_total_matriculados	Integer	FALSO	VERDADERO	Total de matriculados al programa.
dim_region_area	id	id	Bigint	VERDADERO	VERDADERO	Identificador de la dimensión.
dim_region_area	nombre	nombre	Character varying(255)	FALSO	VERDADERO	Nombre de la región o área geográfica.
dim_region_area	tipo_area	tipo_area	Character varying(50)	FALSO	VERDADERO	Precisa si es a nivel de cabecera o centro poblado y rural disperso.
dim_anio_encuesta	id	id	Bigint	VERDADERO	VERDADERO	Identificador único y llave primaria.

Table	Etiqueta	Nombre	Tipo de dato	Primary Key	Not Null	Descripción
dim_anio_encuesta	anio	anio	Integer	FALSO	VERDADERO	Año en cuatro dígitos para el cual esta disponible la encuesta.
DIM_variable_cv	id	id	Bigint	VERDADERO	VERDADERO	Identificador único.
DIM_variable_cv	tipo_variable	tipo_variable	Character varying(255)	FALSO	VERDADERO	Identifica el tipo de variable de la encuesta de calidad de vida.
DIM_variable_cv	nombre_variable	nombre_variable	Character varying(255)	FALSO	VERDADERO	Nombre de la variable de calidad de vida.
DIM_variable_cv	unidad	unidad	Bigint	FALSO	VERDADERO	Unidad en la que se mide la variable, por ejemplo, porcentajes.
fact_variable_cv	id	id	Bigint	FALSO	VERDADERO	Identificador de hechos.
fact_variable_cv	id_region_area	id_region_area	Bigint	VERDADERO	VERDADERO	Relación con la dimensión región o área.
fact_variable_cv	id_anio	id_anio	Bigint	VERDADERO	VERDADERO	Relación con la dimensión año.
fact_variable_cv	id_variable	id_variable	Bigint	VERDADERO	VERDADERO	Relación con la variable de calidad de vida según la encuesta del DANE.
fact_variable_cv	fact_valor_variable	fact_valor_variable	Bigint	FALSO	FALSO	Medida de cada variable de calidad de vida.