UNIVERSIDAD DE GUAYAQUIL

FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

ANÁLISIS DE LA IMPORTANCIA DEL BIG DATA – STORAGE SOBRE PROBLEMÁTICAS REALES DE GRAN VOLUMEN DE ALMACENAMIENTO

PROYECTO DE TITULACIÓN

Previa a la obtención del Título de:

INGENIERO EN SISTEMAS COMPUTACIONALES

AUTOR: FÁTIMA VIZUETA QUINTO

TUTOR: Ing. MAIKEL LEYVA VAZQUEZ PhD.

GUAYAQUIL – ECUADOR

2016

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS

TITULO : "Análisis de la importancia del big data – storage sobre problemáticas reales de gran volumen de almacenamiento"	
AUTOR: Fátima Cynthia Vizueta Quinto	**REVISORES:** Ing. Alfonso Aníbal Guijarro Rodríguez, Ing. Paola del Rocio Quinche Villon
INSTITUCIÓN: Universidad de Guayaquil	**FACULTAD:** Ciencias Matemáticas y Físicas
CARRERA: Ingeniería en Sistemas Computacionales	
FECHA DE PUBLICACION: Noviembre , 2016	**Nº DE PÁGS:** 163
AREAS TEMATICAS: Tecnología, Base de datos, Big Data	
PALABRAS CLAVE: instalación, Big Data, Neo4j,almacenamiento	
RESUMEN: El tema "Análisis de la importancia del Big Data Storage sobre problemáticas reales de gran volumen de almacenamiento" Big data es uno de los conceptos de moda en el mundo informático; se lo define como el conjunto de herramientas informáticas destinadas a la manipulación, gestión y análisis de grandes cantidades de datos estructurados, semiestructurados o sin estructurar y su inmediato análisis. Su importancia radica que en que éste impacta tanto en la industria, como en el negocio e incluso en nuestra sociedad a la hora de manejar grandes cantidades de información en donde esta tecnología actúa almacenando y procesando todos estos datos para generar información útil y a tiempo real.	
Nº DE REGISTRO (en base de datos):	**Nº DE CLASIFICACION:**
DIRECCION URL (tesis en la web):	

ADJUNTO PDF:	**SI** x	**NO**
CONTACTO CON AUTOR: Fátima Vizueta Quinto	Teléfono: 0992203003	E-mail: fatima.vizuetaq@ug.edu.ec
CONTACTO EN LA INSTITUCIÓN: Universidad de Guayaquil	**Nombre:** Ab. Juan Chávez	
	Teléfono: 2307729	
	E-mail: juanchaveza@ug.edu.ec	

APROBACIÓN DEL TUTOR

En mi calidad de Tutor del trabajo de investigación, **"ANÁLISIS DE LA IMPORTANCIA DEL BIG DATA – STORAGE SOBRE PROBLEMÁTICAS REALES DE GRAN VOLUMEN DE ALMACENAMIENTO"** elaborado por la Srta. FÁTIMA CINTHYA VIZUETA QUINTO egresada de la Carrera de Ingeniería en Sistemas Computacionales, Facultad de Ciencias Matemáticas y Físicas de la Universidad de Guayaquil, previo a la obtención del Título de Ingeniero en Sistemas Computacionales, me permito declarar que luego de haber orientado, estudiado y revisado, la Apruebo en todas sus partes.

Atentamente

……

Ing. MAIKEL YELANDI LEYVA VAZQUEZ PhD.

TUTOR

DEDICATORIA

Dedico este trabajo de grado a Dios por ser quien ha estado a mi lado en todo momento, por guiarme y acompañarme a lo largo de mi carrera.

A toda mi familia en especial a mis padres, el Sr. Ulbio Vizueta y la Sra. Francisca Quinto quienes me enseñaron desde pequeño a luchar por mis objetivos, porque sin su esfuerzo no hubiera sido posible alcanzar este logro en mi vida. Gracias por todos los consejos que me brindaron y por inculcarme buenos valores que me ayudaron a desarrollarme como persona. Mi triunfo es de ustedes. ¡Los Amo!
También una persona que no está físicamente pero si espiritualmente, Cristina Vizueta te Amo

AGRADECIMIENTO

Agradezco a mis padres y mis hermanos, por los ánimos y fuerzas que me dieron para que se cumpliese mi objetivo

Un especial agradecimiento al Ing. Maikel Leyva Vázquez por su excelente acompañamiento y orientación en la realización de este Proyecto de Titulación, quien estuvo siempre aportando conocimiento.

TRIBUNAL DE PROYECTO DE TITULACIÓN

Ing. Eduardo Santos Baquerizo, M. Sc.
DECANO DE LA FACULTAD
CIENCIAS MATEMÁTICAS Y
FÍSICAS

Ing. Roberto Crespo Mendoza, Mgs.
DIRECTOR DE LA CARRERA
INGENIERÍA EN SISTEMAS
COMPUTACIONALES

Lcdo. Jorge Alvarado Chang
PROFESOR REVISOR DEL
ÁREA - TRIBUNAL

Ing. Jorge Zambrano Santana,M.Sc.
PROFESOR REVISOR DEL
ÁREA - TRIBUNAL

Ing. Maikel Yelandi Leyva Vazquez PhD.
PROFESOR TUTOR DEL PROYECTO
DE TITULACIÓN

Ab. Juan Chávez Atocha Esp.
SECRETARIO

DECLARACIÓN EXPRESA

"La responsabilidad del contenido de este Proyecto de Titulación, me corresponden exclusivamente; y el patrimonio intelectual de la misma a la UNIVERSIDAD DE GUAYAQUIL"

_____

Srta. Fátima Cinthya Vizueta Quinto

.

UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

ANÁLISIS DE LA IMPORTANCIA DEL BIG DATA – STORAGE SOBRE PROBLEMÁTICAS REALES DE GRAN VOLUMEN DE ALMACENAMIENTO

Proyecto de Titulación que se presenta como requisito para optar por el título de INGENIERO EN SISTEMAS COMPUTACIONALES

Autora: Fátima Cinthya Vizueta Quinto

C.I.0922037635

Tutor: Ing. Maikel Yelandi Leyva Vazquez PhD.

Guayaquil, Noviembre de 2016

CERTIFICADO DE ACEPTACIÓN DEL TUTOR

En mi calidad de Tutor del Proyecto de Titulación, nombrado por el Consejo Directivo dela Facultad de Ciencias Matemáticas y Físicas de la Universidad de Guayaquil.

CERTIFICO:

Que he analizado el Proyecto de Titulación presentado por la estudiante FATIMA CINTHYA VIZUETA QUINTO, como requisito previo para optar por el título de Ingeniero en Sistemas Computacionales cuyo problema es: **"ANÁLISIS DE LA IMPORTANCIA DEL BIG DATA – STORAGE SOBRE PROBLEMÁTICAS REALES DE GRAN VOLUMEN DE ALMACENAMIENTO"**

Considero aprobado el trabajo en su totalidad.

Presentado por:

Fátima Cinthya Vizueta Quinto Cedula N° 0922037635

Tutor: Ing. Maikel Yelandi Leyva Vazquez PhD.

Guayaquil, Noviembre de 2016

UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

Autorización para Publicación de Proyecto de Titulación en Formato Digital

1. Identificación del Proyecto de Titulación

Nombre Alumno: Fátima Cinthya Vizueta Quinto	
Dirección: Coop. Reyna del Quinche 2	
Teléfono:0992203003	**E-mail:** fatima.vizuetaq@ug.edu.ec

Facultad: Ciencias Matemáticas y Físicas
Carrera: Ingeniería en Sistemas Computacionales
Proyecto de Titulación al que opta: Ingeniero en Sistemas Computacionales
Profesor guía: Tutor: Ing. Maikel Yelandi Leyva Vazquez PhD.

Título del Proyecto de Titulación: ANÁLISIS DE LA IMPORTANCIA DEL BIG DATA STORAGE SOBRE PROBLEMÁTICAS REALES DE GRAN VOLUMEN DE ALMACENAMIENTO

Tema del Proyecto de Titulación: Big Data, Neo4j,almacenamiento, NoSql, Graph Databases, Instalación

2. Autorización de Publicación de Versión Electrónica del de Proyecto de Titulación

A través de este medio autorizo a la Biblioteca de la Universidad de Guayaquil y a la Facultad de Ciencias Matemáticas y Físicas a publicar la versión electrónica de este Proyecto de Titulación.

Publicación electrónica:

Inmediata	**x**	Después de 1 año	

Firma Alumno:

3. Forma de envío:

El texto del presente Proyecto de Titulación debe ser enviado en formato Word, como archivo .Doc. O.RTF y .Puf para PC. Las imágenes que la acompañen pueden ser: .gif, .jpg o .TIFF.

DVDROM **x** CDROM

ÍNDICE GENERAL

ABREVIATURAS

Ing.	Ingeniero
CC.MM.FF	Facultad de Ciencias Matemáticas y Físicas
UG	Universidad de Guayaquil
INEC	Instituto Nacional de Estadística y Censos
s.f.	Sin Fecha
HW	Hardware
SW	Software
HTML	Lenguaje de Marca de salida de Hyper Texto
Http	Protocolo de transferencia de Hyper Texto
CISC	Carrera de Ingeniería en Sistemas Computacionales
Mtra.	Maestra
Msc.	Master
URL	Localizador de Fuente Uniforme
Www	World Wide Web (red mundial)
API	APPLICATION Programming Interface
NOSQL	No Structured Query Languaje
RDBMS	Relational Database Management System
SQL	Structured Query Languaje
VD	Variable Dependiente
VI	Variable Independiente

ÍNDICE DE CUADROS

ÍNDICE DE GRÁFICOS

ÍNDICE DE ILUSTRACIONES

ÍNDICE DE ANEXOS

UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

ANÁLISIS DE LA IMPORTANCIA DEL BIG DATA – STORAGE SOBRE PROBLEMÁTICAS REALES DE GRAN VOLUMEN DE ALMACENAMIENTO

Autora: Fátima Cinthya Vizueta Quinto
Tutor:Ing. Maikel Yelandi Leyva Vazquez PhD.

RESUMEN

El desarrollo del presente trabajo de fin de grado se efectuó durante el transcurso de cinco meses, es una investigación documental con modalidad bibliográfica, el objetivo principal que se planteó en el presente Proyecto es realizar una guía práctica del ambiente Big Data además este documento pretende que el lector obtenga una noción de Big Data y los conceptos tecnológicos que lo rodean. El instrumento de investigación utilizado fue un cuestionario comprendido al cual antes de su aplicación se le realizo una validación por parte de un experto, mientras que la técnica fue la encuesta, los representantes de la población la conformaron 57 personas que estudian en quinto semestre. Los cuestionarios fueron procesados y analizados para luego evaluar el resultado obtenido. Big data es uno de los conceptos de moda en el mundo informático; se lo define como el conjunto de herramientas informáticas destinadas a la manipulación, gestión y análisis de grandes cantidades de datos estructurados, semiestructurados o sin estructurar y su inmediato análisis. Su importancia radica que en que éste impacta tanto en la industria, como en el negocio e incluso en nuestra sociedad a la hora de manejar grandes cantidades de información en donde esta tecnología actúa almacenando y procesando todos estos datos para generar información útil y a tiempo real. La tecnología que se planteara es Neo4j que pertenece a la base de datos NoSQL orientado a grafos que es una herramienta de código libre en la cual se representa los datos semi-estructurado con mucha facilidad.

UNIVERSIDAD DE GUAYAQUIL
FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

ANÁLISIS DE LA IMPORTANCIA DEL BIG DATA – STORAGE SOBRE PROBLEMÁTICAS REALES DE GRAN VOLUMEN DE ALMACENAMIENTO

Autora: Fátima Cinthya Vizueta Quinto
Tutor:Ing. Maikel Yelandi Leyva Vazquez PhD.

ABSTRACT

The development of the present work of end of degree is made during the course of five months, is a documentary research bibliographic mode, the main objective raised in this project is a practical guide Big Data environment also this document assumes that the reader get a sense of Big Data and technological concepts surrounding it. The research instrument used was an understanding which before application questionnaire underwent a validation by an expert, while the technique was the survey, representatives of the population formed 57 people studying in the fifth semester. The questionnaires were processed and analyzed to then evaluate the result. Big Data is one of the fashion concepts in the computer world; it is defined as the set of tools for the handling, management and analysis of large amounts of structured, semi-structured or unstructured data and immediate analysis. Its importance lies in that it impacts both in industry and in the business and even in our society when handling large amounts of information where this technology works by storing and processing all this data to generate useful information and real time. The technology is Neo4j raised that belongs to the NoSQL database oriented graph which is an open source tool in which semi-structured data very easily shown.

INTRODUCCIÓN

Hoy en día, las empresas u organizaciones se enfrentan a grandes cantidades de datos, que son complicados en cuanto a su manipulación y no se adaptan del todo a los modelos de bases de datos relacionales tradicionales. Las empresas que tienen ingentes cantidades de datos no estructurados y semi-estructurados se ven obligadas a utilizar nuevas tecnologías como Hadoop, MapReduce y Base de datos NoSQL. Debido a que las soluciones tradicionales como SQL, Oracle, etc. ya no son suficientes para manipular este tipo de información (Vera, 2016). Estamos por lo tanto ante un fenómeno que está siendo cada vez más implantado en empresas y organizaciones ; ya que Big Data es el sector de las tecnologías de la información y la comunicación (TIC) que se preocupa de como almacenar y tratar grandes cantidades de información o conjuntos de datos.

Pero muchas personas se preguntaran que es Big Data a lo cual se lo define como la gestión y análisis de enorme volúmenes de datos que no pueden ser tratados por las herramientas informáticas tradicionales. De esta manera la presente investigación desarrollara una guía con la implementación de un ambiente Big Data.

A continuación el detalle de los capítulos:

El Capítulo I contiene la presentación del problema propuesto como tema de tesis, la situación actual, sus causas y consecuencias, delimitación, se plantea la formulación del problema, evaluación, los objetivos generales y específicos del Proyecto, su alcance y por último la justificación e importancia de la investigación.

En el capítulo II es el más importante porque se encuentra el contenido de la investigación, se detalla el marco teórico, el aspecto legal en la que se fundamenta la investigación, se establece las variables y definiciones de términos.

En el capítulo III encontramos la aplicación de la metodología utilizada en el desarrollo de la investigación: la modalidad y el tipo; la población y la muestra, la definición operacional de las variables, técnicas e instrumentos empleados para la recolección de datos, procedimiento de la investigación y las técnicas para el procesamiento y análisis de datos.

Finalmente en el capítulo IV encontraremos los resultados, conclusiones y recomendaciones del Proyecto a las que el autor ha llegado como resultado de la experiencia adquirida a través del desarrollo de la investigación.

CAPÍTULO I

EL PROBLEMA

PLANTEAMIENTO DEL PROBLEMA

Ubicación del Problema en un Contexto

Hoy en día, una empresa u organización dispone de una gran cantidad de información no estructurada que sólo puede ser aprovechada si se selecciona y se procesa adecuadamente. Esta información que algunas empresas no sabe manejar, pese a entender que podemos sacar de ella conclusiones valiosas, como por ejemplo descubrir el perfil, las necesidades y el sentir de sus clientes respecto a los productos o servicios vendidos se puede encontrar en Big Data una solución para aprovechar todos esos datos.

La situación actual del problema ocurre cuando la variedad, la velocidad y el volumen de datos superan las capacidades de procesamiento o almacenamiento de una organización, adicionalmente es posible enriquecer esta definición con un par de dimensiones adicionales: la primera se refiere a la incapacidad de la empresa de extraer valor de sus datos, y la última hace referencia a la complejidad de los análisis que son hechos con los datos.

Hasta hace unos años en las empresas lo más importante era la transacción de información pero, actualmente, esto está cambiando. Cada vez se da más importancia a la interacción constante con el mundo que nos rodea. Las empresas tienen una única opción: aprovecharse de toda la información que se genera. En consecuencia, las empresas deben aprender a optimizar y conseguir la mayor información posible de sus clientes, creando, por ejemplo, nuevos canales de comunicación para poder interactuar con ellos.

El Big Data ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos estructurados, no estructurados y semi estructurados (mensajes en redes sociales, señales de móvil, archivos de audio, sensores, imágenes digitales, datos de formularios, emails, datos de encuestas, logs, etc,) que tomaría demasiado tiempo y sería muy costoso cargarlos a una base de datos relacional para su análisis (Barranco, 2012)

La gran cantidad de datos que se están utilizando al día de hoy, son complicados en cuanto a su manipulación y no se adaptan del todo a los modelos de bases de datos relacionales tradicionales. Las empresas que deseen analizar todos estos datos se ven obligadas a utilizar nuevas tecnologías como Hadoop, Cassandra, HBase,Neo4j etc. debido a que las soluciones tradicionales como SQL, Oracle, etc. ya no son suficientes para manipular este tipo de información (Vera, 2016). Estamos por lo tanto ante un fenómeno que está siendo cada vez más implantado en empresas y organizaciones ; ya que Big Data es el sector de las tecnologías de la información y la comunicación (TIC) que se preocupa de como almacenar y tratar grandes cantidades de información o conjuntos de datos.

Situación Conflicto Nudos Críticos

El problema surge para dar respuesta a las dificultades en almacenamiento, gestión, análisis y procesamiento de esta cantidad de datos que durante la última década se han generado de forma exponencial, y de la urgencia de procesar esta información en un tiempo cada vez menor.

Otra cuestión es que toda esta cantidad de datos que existen ahora mismo son de muy diversa naturaleza: ya no contamos con datos estructurados que no cambian en el tiempo, sino que ahora mismo tenemos miles de fuentes de datos y cada una es totalmente distinta de otra, para ello es fundamental que sea transformada de tal manera que aporte valor (Sanchez J. , 2015). Las bases de

datos relacionales no son muy amigas del desorden, para analizar tanta información surge el término Big Data.

El problema es que las bases de datos tradicionales no pueden manejar el tamaño ni la complejidad de los diversos formatos de datos, ni pueden hacerlo a la velocidad que se necesita, por lo que han surgido diseños novedosos en bases de datos llamadas NoSQL o No Solo SQL que permiten resolver los problemas de manejar abrumadoras cantidades de información de una manera rápida y eficaz. Las grandes compañías de Internet (Facebook, Twitter, Google…) utilizan NoSQL como soporte fundamental al almacenamiento de datos (SAT, 2014).

Pero Big Data no es sólo almacenar los datos y procesarlos, hay que analizarlos y convertirlos en conocimiento, y aquí es donde se manifiesta la verdadera utilidad del Big Data. Gracias a este software las empresas pueden no solo almacenar datos si no acceder a ellos de manera lógica y estructurada para trabajar con ellos, usándose habitualmente software integrado que almacena y estructura datos de manera centralizada aportando al ser humano un avance tecnológico muy importante.

Causas y Consecuencias del Problema

CUADRO N. 1

CAUSAS Y CONSECUENCIAS

CAUSAS	CONSECUENCIAS
El incremento exponencial de información y la obligación de procesar dicha información prácticamente en tiempo real.	• Recopilación de información para su posterior comercialización contemplando incluso el robo de datos.
El volumen de información de las empresas excede la capacidad de almacenamiento.	• Decisiones erróneas en una organización. • Filtrado (no todos los datos son información).
Acumulación masiva de datos no estructurados.	• Informes no válidos. • Proyectos fallidos • Fallos de procesos
Las empresas no tienen acceso a la información necesaria para tomar decisiones clave.	• Información desactualizada. • Información poco confiable. • Publicación accidental
Sistemas de almacenaje de datos son hackeados.	• Pérdida de la confianza de sus clientes que corren a refugiar sus datos a empresas que les ofrezcan una mayor confianza y seguridad (Perdida de clientes). • Acceso desautorizado a la privacidad y la seguridad de la información. • Daño reputacional irremediable entre sus clientes.(Insatisfacción)

Falta de atención en el tratamiento de los datos.	• Poca exactitud en las estrategias ya que no cuenta con información certera.

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Delimitación Del Problema

La tesis a desarrollar se limitara al estudio de la aplicación de las tecnologías del Big Data a efectuarse en los siguientes aspectos:

Campo : Investigativo

Área : Tecnología

Aspecto : Definición de una guía práctica para la aplicación de las Tecnologías del Big Data.

Tema : Análisis de la importancia del Big Data – Storage sobre Problemáticas reales de gran volumen de almacenamiento

Formulación Del Problema

¿Cómo contribuir a la implementación de las tecnologías de almacenamiento relacionada al Big Data?

Evaluación Del Problema

Los aspectos generales de evaluación son:

Delimitado: Las tecnologías *Big Data* surgen para dar solución al problema que tienen tanto empresas como administraciones públicas cuando la información que manejan excede la capacidad de almacenamiento, procesado y análisis de la organización.

Evidente: En esta investigación se presentan ventajas e inconvenientes que se obtendrían con la aplicación de la tecnología Big Data. Esta tecnología se presenta como una herramienta muy útil, porque nos brinda la oportunidad de generar, capturar y almacenar grandes cantidades de información de los usuarios

Concreto: El problema se lo describe en una forma en la que se trata de explicar cuáles son las razones que me motivan a hacer una investigación sobre la aplicación de la Tecnología del Big Data enfocada a resolver el problema de las grandes cantidades de datos y dicho problema se lo explica de una manera precisa y directa.

Original: Este tema es poco conocido en Ecuador, y es muy importante su conocimiento para poder manejar grandes cantidades de datos ya que en cualquier momento podría ser necesaria su utilización.

Factible: La solución a este problema, es realizable según el alcance definido del Proyecto la cual se encuentra establecida en el cronograma en un corto tiempo que se le ha proporcionado. Los tiempos de las tareas están relacionados con los objetivos planteados.

Identifica los productos esperados: El resultado de este Proyecto busca demostrar el beneficio que tiene el Big Data en tratar de forma global la información de los clientes, de su relación con los servicios de atención, la información de negocio, de uso de los servicios, etc. Y tomar decisiones de forma rápida y certera en base al análisis de dichos datos.

Objetivos

Objetivo General

- Desarrollar una guía práctica para la implementación de un ambiente Big Data, utilizando para ello la herramienta Neo4j.

Objetivos Específicos

- Detallar las características actuales de Big Data.
- Determinar las ventajas e inconvenientes del Big Data.
- Establecer los modelos de almacenamiento de datos en Big Data.
- Identificar las herramientas de software y hardware necesarios para la implementación del ambiente Big Data.

Alcances Del Problema

El presente Proyecto tiene como alcance realizar un análisis de la importancia y desarrollo que tiene en Big Data como herramienta potencial para solucionar problemáticas que requieran la gestión de grandes volúmenes de almacenamiento. De este modo, el actual documento podrá servir a futuros investigadores académicos a profundizar sobre en Big Data en problemáticas reales.

El Proyecto culmina con el diseño y montaje del ambiente de Big Data, teniendo en cuenta que la herramienta será analizada y definida como parte del Proyecto. El desarrollo del Proyecto incluye:

- Un marco teórico conceptual sobre los modelos de almacenamiento de datos en Big Data.

- Guía paso a paso para la instalación y aplicación de la tecnología Neo4j.

Justificación e Importancia

En la actualidad, las empresas disponen de una gran cantidad de información no estructurada que sólo puede ser aprovechada si se selecciona y se procesa adecuadamente. Esta información que algunas empresas no sabe manejar, pese a entender que podemos sacar de ella conclusiones valiosas, como por ejemplo descubrir el perfil, las necesidades y el sentir de sus clientes respecto a los productos o servicios vendidos se puede encontrar en Big Data una solución para aprovechar todos esos datos.

La justificación para el desarrollo de este Proyecto se basa por la popularidad que está teniendo esta tecnología actualmente y para responder preguntas que quedan en el espacio como: ¿Dónde se almacena la información que se genera en el mundo? Big Data abarca todos los ámbitos, al referirse al procesamiento y análisis de ingentes cantidades de datos de diferentes tipos, por lo cual las empresas tienen mucho interés en esta tecnología para obtener información útil en la toma de decisiones donde además brinden velocidad al realizar consultas y esté disponible en cualquier momento.

De acuerdo a alguna campaña de publicidad en una empresa, esta tecnología permite analizar el comportamiento de nuestros cliente antes de sus acciones, el nivel de compromiso que una campaña generé, así mismo cuales asuntos tienen mayor interacción con su público, y también nos ayuda en la construcción de una estrategia de marketing enfocada en resultados.

El aporte teórico de esta investigación será en contribuir el enriquecimiento de teorías, beneficios, características sobre la tecnología Big Data, la misma que se presenta como una herramienta útil, porque nos brinda la oportunidad de generar, capturar y almacenar grandes cantidades de información. Otra importancia de este Proyecto es en impulsar de esta manera el campo de investigación dentro de la comunidad universitaria.

Esta investigación proporcionara nuevas oportunidades para las empresas que presenten situaciones similares de ingentes cantidades de información, donde tienen la opción de mejorar su situación, mediante el uso de este estudio como marco referencial.

Metodología Del Proyecto

Hipótesis a Contestarse

¿Qué ha propiciado la aparición de Big Data?

¿Qué se entiende por Big Data?

¿Cuál es el objetivo de Big Data?

¿Cuáles son sus principales beneficios?

¿Existe conciencia en las compañías sobre la relevancia de contar con Big Data?

Variables

Variable Independiente

Estudio de la tecnología Big Data

Variable Dependiente

Presentación de una guía con el prototipo para demostrar la investigación realizada.

Diseño Metodológico

El presente Proyecto: "**Análisis de la importancia del Big Data Storage sobre problemáticas reales de gran volumen de almacenamiento**" tiene como modalidad investigación de campo un 30% e investigación bibliográfica un 70%.

Será de tipo bibliográfica, al llevarse a cabo por medio de fuentes bibliográficas tales como libros, revistas, videos, consultas en internet. El porcentaje de campo corresponde a las encuestas y pruebas del tema propuesto.

Universo y Muestra

Universo

El universo está compuesto por 72 estudiantes que corresponden a dos cursos de quinto semestre de la carrera de Ingeniería en Sistemas Computacionales.

Muestra

Para saber el número de la muestra se debe aplicar la fórmula de la misma la cual nos da una muestra de 57 personas a las que se debe encuestar.

CAPÍTULO II

MARCO TEÓRICO

ANTECEDENTES DEL ESTUDIO

El tema investigativo que se estudia no existe hasta la fecha actual en los registros de tesis que reposan en la biblioteca de la universidad de Guayaquil, Carrera de Ingeniería en Sistemas Computacionales

Actualmente vivimos en un mundo conectados constantemente a una gran cantidad de información, ya sea a través del ordenador, del teléfono móvil, la Tablet, etc. Tanto en la industria, como en el negocio e incluso en nuestra sociedad se maneja grandes cantidades de información en donde es necesario pensar en la tecnología Big Data la cual actúa almacenando y procesando todos estos datos para generar información útil y a tiempo real (González & Carrillo, 2016).

Según lo expuesto anteriormente, se puede observar que la implementación de la tecnología de Big Data adquiere mayor importancia en las organizaciones, dado la necesidad de procesar y analizar grandes cantidades información para la toma de decisiones.

Como antecedente de este estudio existe información en internet sobre Big Data, pero no he encontrado mucha información sobre la tecnología a aplicar que es Neo4j. Años atrás en muchos casos de ingentes cantidades de datos se solían procesar con Hadoop, pero se escogió para el actual Proyecto la base de datos orientada a grafos llamada Neo4j que pertenece a las herramientas NoSql, una de sus ventajas es que se pueden efectuar las consultas directamente lo que hace esencialmente interesante su integración con aplicaciones web (Sánchez, 2014)

Mediante el presente trabajo investigativo se pretende contribuir con información confiable y actualizada, con el fin de promover el conocimiento de quienes lo consulten.

FUNDAMENTACIÓN TEÓRICA

Últimamente se escucha el término Big Data y de las cantidades colosales de información que se almacenan en sus bases de datos, pero ¿Dónde se almacenan los datos? ¿Qué tipo de bases de datos son las que brindan tan alto rendimiento? ¿Van a reemplazar éstas a las tradicionales bases de datos relacionales? Estas nuevas bases de datos son las que utilizan los grandes gigantes Facebook, Yahoo, Amazon y Google para almacenar sus datos (Santacruz, 2013).

En el presente capítulo, se han recopilado puntos básicos para el desarrollo del Proyecto Análisis de la importancia del Big Data Storage sobre problemáticas reales de gran volumen de almacenamiento, en el que voy a responder a estas preguntas y a dar una breve explicación sobre la definición, características, uso de estas nuevas bases de datos NoSQL.

BIG DATA

¿Qué es Big Data?

Big Data es un término de origen inglés cuya traducción corresponde a "Datos masivos", es uno de los conceptos de moda en el mundo informático. Aunque existen muchas definiciones en Internet, el Big Data se puede definir como el conjunto de herramientas informáticas destinadas a la manipulación, gestión y análisis de grandes cantidades de datos estructurados, semiestructurados o sin estructurar y su inmediato análisis para encontrar información oculta, esquemas recurrentes, nuevas correlaciones, etc. El conjunto de datos es tan grande y complejo que no pueden ser gestionados por las herramientas informáticas tradicionales (Instituto DataKey, 2015).

Es una realidad que la cantidad de información digital que se genera diariamente en nuestro planeta crece exponencialmente, gigantes y potentes bases de datos en donde se almacenan por ejemplo siempre que hagamos una búsqueda, enviemos un email, usemos un teléfono móvil, actualicemos una red social, usemos una tarjeta de crédito, vayamos al gimnasio, activemos el GPS, hagamos la compra en el supermercado dejamos detrás de nosotros una montaña de datos, huellas digitales y registros que ofrecen una información muy valiosa y cuyo estudio es el interés de todo analista (Perez V. , 2011).

Esta tecnología Big Data, permite convertir aquellos enormes datos en información útil para facilitar la correcta toma de decisiones en las empresas obteniendo ventajas competitivas con el fin de generar beneficios. Algunas de estas decisiones influyen a la hora de elaborar estrategias de marketing inteligentes, retener y fidelizar a los clientes o mejorar la gestión e innovación en la creación de productos y servicios (Fernandez S. , 2016).

Importancia del Big Data

En la actualidad vivimos conectados constantemente a una gran cantidad de información, ya sea a través del ordenador, del teléfono móvil, la Tablet, etc. En el 2012, la cantidad de información almacenada superó los 2.8 zettabytes, y en el 2020 se espera que la cantidad de información sea mucho más grande (Natividad, 2016).

La importancia de Big Data radica que en que éste impacta tanto en la industria, como en el negocio e incluso en nuestra sociedad a la hora de manejar grandes cantidades de información en donde esta tecnología actúa almacenando y procesando todos estos datos para generar información útil y a tiempo real (PowerData, 2013).

Es precisamente en ese tipo de datos donde las empresas han detectado que hay más valor. Gracias al Big Data las empresas pueden conocer de mejor forma

a sus clientes, consumidores y competidores, desarrollando así estrategias de negocio más efectivas y adaptadas a la realidad de sus usuarios. Transformar estos datos en información no solo facilita la llegada a los consumidores, también es una gran herramienta para tomar decisiones en tiempo real y reducir el riesgo de tomar decisiones erradas. Este análisis predictivos nos permiten modelar e influir en el futuro, al evitar que ocurran ciertos hechos y de esta forma cambiar el curso de las acciones. También nos permiten prever las preferencias de las personas y dar recomendaciones basándonos en ellas (Rivero, 2016).

Ventajas de la tecnología Big Data

A estas alturas nadie debe ignorar las ventajas que supone para cualquier empresa almacenar, clasificar, analizar y compartir el cúmulo masivo de información que dispone de sus clientes. Favorablemente existe el Big Data, que permiten predecir decisiones y así vender más y mejor y de la forma más óptima y personalizada.

Partamos de la base de que todas las ventajas del Big Data tienen que ver con su capacidad para recopilar y analizar datos. Estos datos nos ayudarán a conocer las fortalezas y debilidades dentro de nuestra empresa; las necesidades que tiene, dónde puede mejorar o dónde destaca más. Conociendo esto, los beneficios son claros:

Mejora en la toma de decisiones

La información como base para la toma de decisiones es fundamental y cuando podemos manejar de forma adecuada toda la información que nos proporciona el Big Data, podremos tomar decisiones inteligentes y rápidas que favorecerán a nuestro negocio (Kodak, 2016).

Retención y fidelidad de clientes

Captar nuevos clientes es más complicado que fidelizar a los que ya tenemos, por lo que es primordial utilizar los datos a nuestro alcance para dar a nuestros clientes lo que desean y para ello es primordial el Big Data. Por lo tanto, conocer a través de los datos el nivel de satisfacción de nuestros clientes, sus necesidades, etc. es la base para tener clientes fidelizados (Kodak, 2016).

Mayor eficiencia

El almacenamiento de datos en formato electrónico permite a los empleados públicos acceder a toda la información desde una ubicación centralizada, lo que a su vez reduce los errores. También, garantiza el acceso de los funcionarios a la información más actualizada (Kodak, 2016).

Costos más bajos

Los Proyectos de servicios públicos digitales facilitan la obtención de los datos más importantes. Esto permite agilizar la toma de decisiones con costos más bajos. Mejora en la accesibilidad de la información dentro de la empresa.
Las empresas que digitalicen los datos y utilicen herramientas para la búsqueda de información de forma eficaz, desarrollarán procesos productivos más rápidos y eficaces, y el trabajo en la empresa ganará en agilidad (Kodak, 2016).

Menos fraudes

Los algoritmos pueden usar la detección de patrones para detectar transacciones sospechosas en tiempo real. Cualquier acto criminal puede identificarse y monitorearse con mayor detenimiento (Kodak, 2016).

Menos Riesgos

Cuanto mayor sea el número de datos que seamos capaces de analizar en nuestra organización, podremos hacer mejores predicciones del comportamiento de nuestros clientes, de la evolución del sector, a evaluar los riesgos con anticipación así como también predecir la solución a posibles problemas o incluso hasta las acciones de nuestra competencia (Kodak, 2016).

Inconvenientes

Sin embargo, así como el Big Data tiene beneficios, también tiene algunos inconvenientes o aspectos negativos que se mencionan a continuación:

- La falta de profesionales expertos
- Rechazo por parte del personal
- Problemas de privacidad
- Insatisfacción de clientes
- Seguridad de Datos

A parte de estos, hay que considerar un gran inconveniente que es tan sencillo como saber sí: ¿Tiene mi compañía esta necesidad? ¿Contamos con los recursos para afrontar un Proyecto de Big Data? ¿Cuánto costará? Si el empleo de Big Data es beneficioso para nuestra organización, los inconvenientes no deberían importar, puesto que las ventajas que se obtienen serán mucho mayores (Smith, s.f.).

Retos del Big Data

Las personas cuya responsabilidad es la de tomar decisiones se frustran por la cantidad de tiempo que se tarda en obtener las respuestas a las preguntas que les plantea el negocio y cuya solución está entre los millones y millones de bytes. La visualización de estos datos se está convirtiendo en un módulo cada vez más importante de la analítica, en la era de grandes volúmenes de datos así como acceder a la información en una forma que se pueda fácilmente, compartir.

El mayor reto, como más adelante se verá, para Big Data es disponer de personas preparadas y formadas para ejecutar estos Proyectos y convertir una gran cantidad de datos en decisiones, estrategias y mejores experiencias para los consumidores. Por ello, el hecho de tener la información adecuada es la clave para un buen futuro.

A continuación, se describen algunos de los principales desafíos:

Satisfacer la necesidad de velocidad

Al día de hoy, las compañías no sólo tienen que buscar y analizar la información relevante que necesitan, además tienen que encontrarla rápidamente. Las herramientas de visualización ayuda a las organizaciones realizar análisis y tomar decisiones con mayor rapidez, pero el reto está en la velocidad que se necesita para acceder al nivel de detalle de los datos relevantes escondidos volúmenes de datos enormes (Demassieux, 2015).

Conocimiento de los datos

Se necesita mucho conocimiento para obtener los datos en la forma adecuada y que pueda ser utilizada para la visualización, como parte del análisis de datos. Asegúrese de que las personas que analizan los datos tienen un profundo conocimiento acerca de la procedencia de los datos, conocimiento acerca del tipo de público estará consumiendo los datos y la forma en el que éste va a interpretar la información (Demassieux, 2015).

Abordar la calidad de datos

El valor de estos datos estará en peligro sino tiene en cuenta su exactitud. Inclusive en el caso de que pueda encontrar y analizar datos de manera rápida puestos en el contexto adecuado para la audiencia que se consume la información, el valor de los datos para la toma de decisiones fines, se pondrá en peligro si los datos no llegan en tiempo y forma, además de ser lo más exactos posibles (Demassieux, 2015).

Trabajar con datos cada vez más complejos

Anteriormente, el Big Data tuvo que ver con datos simple como tablas de cifras, gráficos, etc; los datos procesados actualmente son más complejos y variados. Por consiguiente, es necesario remodelar y reinventar las herramientas y arquitecturas del Big Data para capturar, almacenar y analizar esta diversidad de datos. Por ello, el Big Data asume este desafío de recopilar datos multi-estructurados y poder analizarlos para optimizar sus procesos (Demassieux, 2015).

Mostrar resultados significativos

El trazado de puntos en un gráfico para el análisis, se vuelve difícil cuando se trata de cantidades extremadamente grandes de datos o bien una gran una variedad de categorías de información (Demassieux, 2015).

Seguridad

Todos los especialistas en la materia apuntan a la necesidad de contar con herramientas de seguridad avanzada para evitar las fugas de información, asegurarse que solo la gente adecuada accede a determinada información y en general, contar con una política adecuada de seguridad en lo que respecta al Big Data (Trazada, 2015). Si desde el primer día en nuestros Proyectos de Big Data no incluimos las normativas y componentes de seguridad estamos en un riesgo innecesario que afectara a nuestros activos o a los de nuestros clientes como pueden ser pérdida de datos, pérdida de credibilidad de la empresa. (Rayo, 2016)

¿De dónde Provienen los Datos?

En los últimos años, debido al avance de las Tecnologías de la Información, estamos siendo testigos de una verdadera explosión en la cantidad de datos disponibles, listos para ser analizados y así convertirse en información útil para la inteligencia de negocio. Gestionados y analizados de manera adecuada estos datos son oro para las empresas porque desvelan patrones de comportamiento y tendencias que ayudan a identificar nuevas oportunidades de mercado, además de mejorar productos o servicios y su impacto, con el consiguiente incremento en ventas (Ariely, 2016).

Catalogamos la fuente de los datos según las siguientes categorías:

Web y Social Media: Diariamente los seres humanos generamos una gran cantidad de datos cuando navegamos por internet como contenido web e información que es obtenida de las redes sociales como Facebook, Twitter, LinkedIn, etc, blogs.

Machine-to-Machine (M2M)*:* M2M significa a las tecnologías que permiten conectarse a otros dispositivos. estas utilizan dispositivos como sensores o medidores que capturan algún evento en particular (velocidad, temperatura, presión, variables meteorológicas, variables químicas como la salinidad, etc.) los

cuales transmiten a través de redes alámbricas, inalámbricas o híbridas a otras aplicaciones que traducen estos eventos en información significativa.

Transacciones de datos*:* Incluye registros de facturación, en telecomunicaciones registros detallados de las llamadas, etc. Estos datos transaccionales están disponibles en formatos tanto semiestructurados como no estructurados.

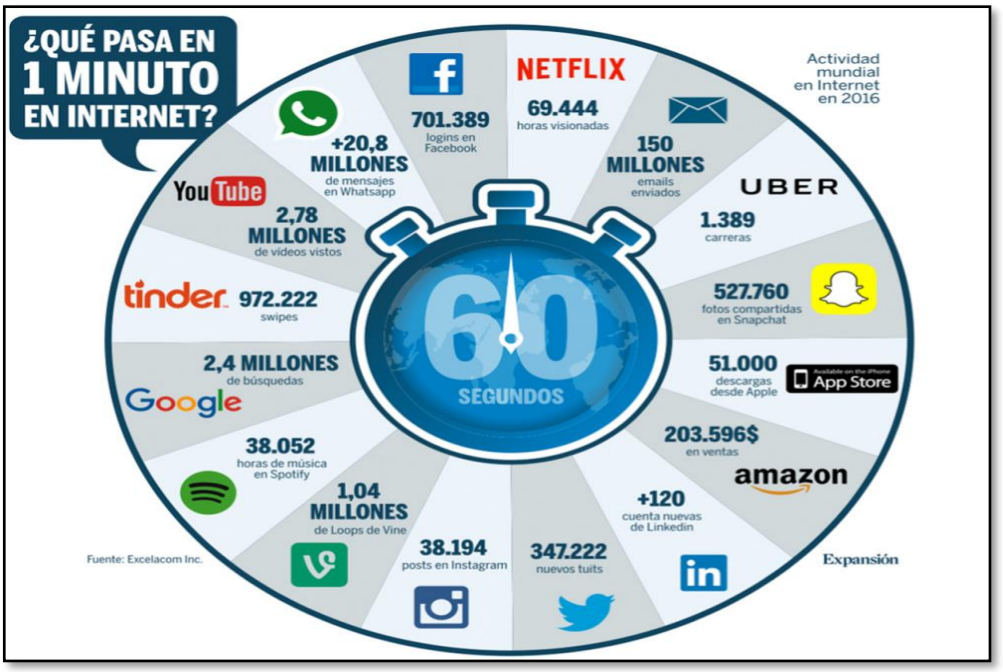
Biométrica: son Información biométrica en la que se incluye huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc. En el área de seguridad e inteligencia, los datos biométricos han sido información importante para las agencias de investigación.

Generados por las personas*:* Las personas generamos diversas cantidades de datos como publicar un estado en Facebook, tuitear contenidos, notas de voz, estudios médicos o enviar mensajes por WhatsApp etc (Barranco, 2012).

Podemos comprobar en esta página Web bit.ly/internet-30-segs lo que sucede en Internet cada 30 segundos (SAT, 2014). La cantidad de información que se genera actualmente un día cualquiera: desde todas las búsquedas de información que se realizan en Google, a todas las fotos, vídeos, mensajes que se suben a las redes sociales, blogs, webs, etc. Además cada año la compañía Excelacom, con sede en Reston, Virginia, presenta un listado que ayuda a poner en perspectiva todo lo que ocurre en sólo un minuto en internet. En la siguiente imagen podemos visualizar lo que pasa en un minuto en internet (Forbes Staff, 2016).

GRAFICO N.1

1 MINUTO EN INTERNET



Elaboración: Forbes
Fuente: http://www.forbes.com.mx/lo-pasa-minuto-internet/#gs.neFt_d0

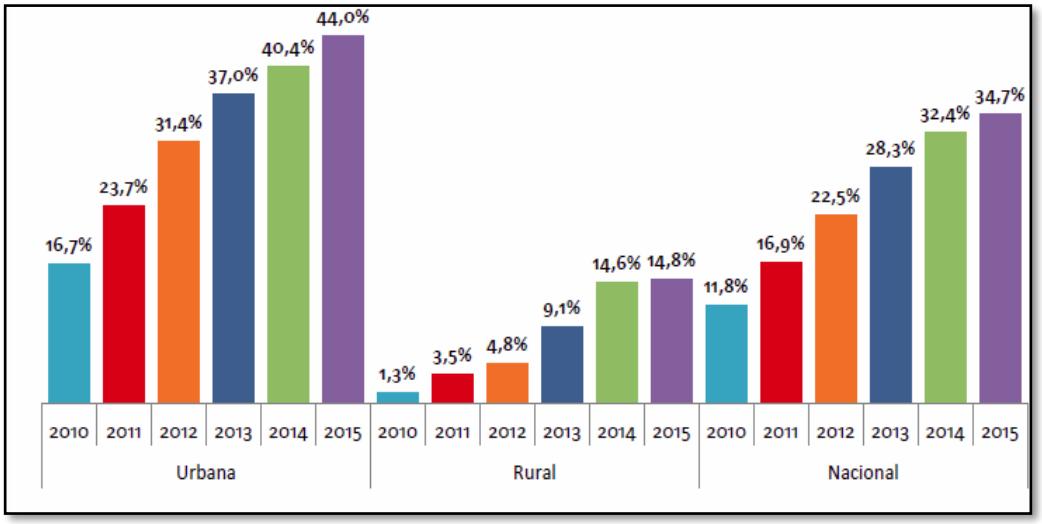
Todos nos hemos preguntado alguna vez ¿Cuantas cuentas de Facebook se crean cada segundo? En ese mismo segundo, ¿Cuantos tweets son creados? ¿Cuantos vídeos son subidos a YouTube? ¿Cuantos emails son enviados? ¿Cuantos mensajes en whatsapp son enviados? En resumen queremos saber ¿Qué está ocurriendo en **Internet** en tiempo real? La figura nos indica un promedio de la cantidad de interacciones hechas a través del internet en un intervalo de 60 segundos.

Estas cifras representan una oportunidad gigantesca para los proveedores de comunicación y los medios de comunicación, quienes, con la infraestructura adecuada y campañas innovadoras, pueden reducir costos y aumentar significativamente sus ingresos. Cómo es normal, todo sigue subiendo y cada día estas cifras aumentan, dejando una red inmensa y una población cada vez más concienciada con las nuevas tecnologías.

Centralizándonos en Ecuador visualizar que el Instituto de Estadísticas y Censos (INEC) acaba de publicar su último informe según el cual la cifra 2015 que maneja textualmente dice: “El 34,7% de los hogares a nivel nacional tienen acceso a internet, 22,9 puntos más que hace seis años. En el área urbana el crecimiento es de 27,3 puntos, mientras que en la rural de 13,5 puntos” (INEC, 2016,p.7).

GRAFICO N.2

ACCESO AL INTERNET SEGÚN ÁREA

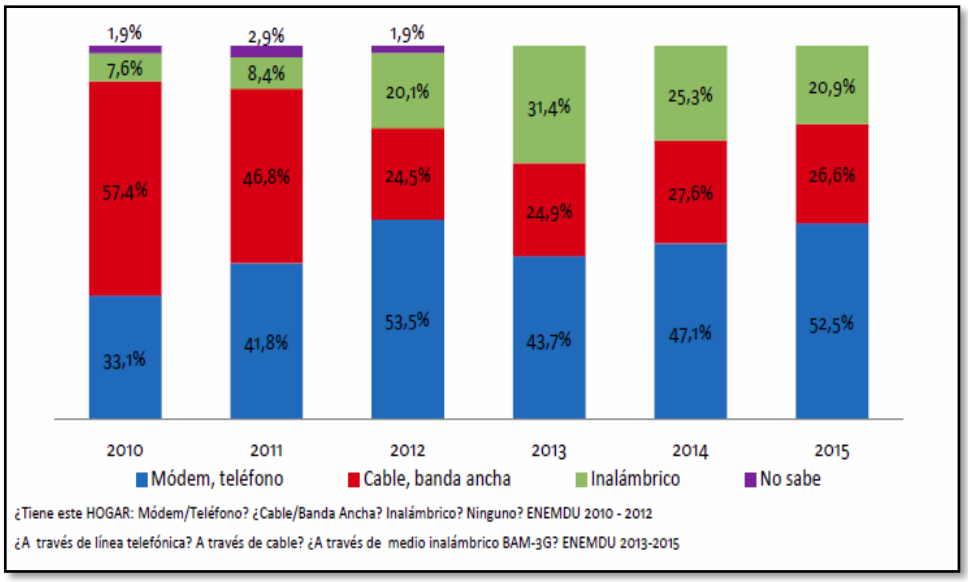


Elaboración: INEC
Fuente: http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/2015/Presentacion_TIC_2015.pdf

Donde ademas tambien se visualiza la forma de acceso a Internet:
(INEC, 2016) afirma: “El 38,8% de los hogares tiene acceso a Internet, de ellos el 20,9% accede a través de algún medio inalámbrico, modem o teléfono, 19,4 puntos más que en 2010” (p.8).

GRAFICO N.3

HOGARES QUE TIENEN ACCESO A INTERNET A NIVEL NACIONAL



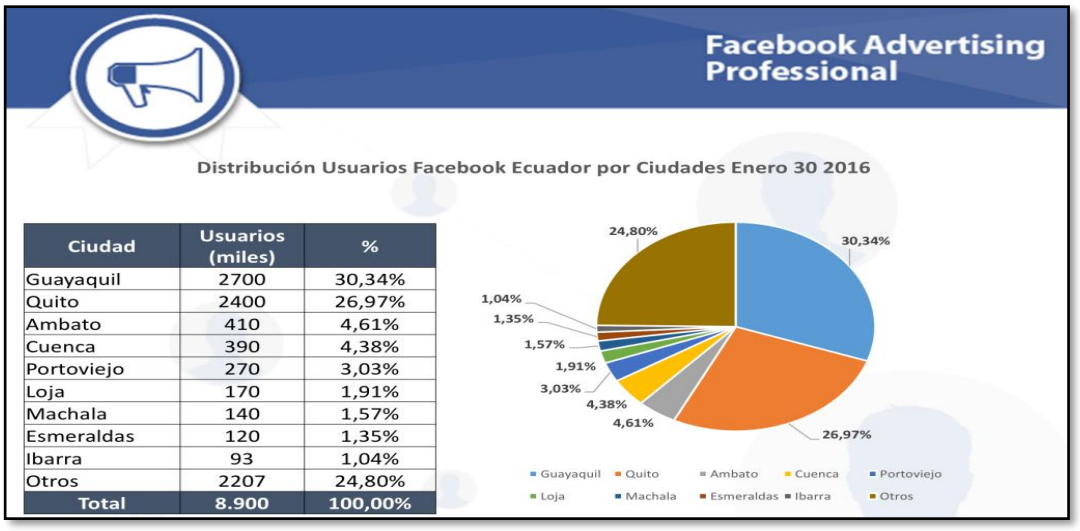
Elaboración: INEC

Fuente: http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/2015/Presentacion_TIC_2015.pdf

Mientras tanto, lo innegable es que al menos 9 de cada 10 ecuatorianos que tienen acceso a internet en Ecuador tienen una cuenta en Facebook. Con 2.700.000 de usuarios, Guayaquil es la ciudad de Ecuador con más cantidad de audiencia en Facebook, seguido de Quito con 2.400.000 usuarios y, en tercer lugar, con 410.000 de usuarios está Ambato (Alcazar, 2016).

GRAFICO N.4

USO DE FACEBOOK EN ECUADOR POR CIUDADES



Ciudad	Usuarios (miles)	%
Guayaquil	2700	30,34%
Quito	2400	26,97%
Ambato	410	4,61%
Cuenca	390	4,38%
Portoviejo	270	3,03%
Loja	170	1,91%
Machala	140	1,57%
Esmeraldas	120	1,35%
Ibarra	93	1,04%
Otros	2207	24,80%
Total	8.900	100,00%

Elaboración: Juan Pablo del Alcazar Ponce
Fuente: http://blog.formaciongerencial.com/2016/02/01/estadisticasfacebookecuador/

Características del Big Data

Las características necesarias para decir que nos referimos a Big Data, son las famosas “7V’s”, las cuales son: Volumen, Velocidad, Variedad, Variabilidad, Veracidad, Visualización y Valor.

Volumen

Es la característica principal con la que todo el mundo asocia el *Big Data*; se refiere a la cantidad de datos que se genera a través de redes sociales, correo electrónico, aplicaciones móviles, navegación cotidiana por Internet, incluyendo las transacciones de negocios y la información de los sensores o los datos de máquina a máquina.

Velocidad

Se refiere a los datos en movimiento; en la rapidez a la que se generan los datos, almacenados, analizados y procesados. El énfasis se ha puesto recientemente en apoyar el análisis de grandes volúmenes de datos en tiempo real. Por poner un ejemplo sencillo, cuando hacemos un “me gusta” en una red social estamos creando nuevos datos, cuando empleamos el GPS, cuando compramos un billete de avión, cuando pedimos hora al médico. Producimos datos de forma constante. Este dinamismo permite disponer de información en tiempo real y provoca el reto de hacer análisis bastante detallados y complejos que a menudo se integran en los procesos de trabajo y sistemas.

Variedad

Un aspecto importante de grandes volúmenes de datos es la variedad de fuentes de datos y tipos de datos que pueden ser estructurados, no estructurados y semiestructurados.

Día a día, se crean cantidades de datos, de múltiples fuentes y diversos formatos, desde estructurados y relativamente fáciles de gestionar, hasta información no estructurada en forma de documentos, vídeos, correos electrónicos provenientes de telefonía móvil, redes sociales, entradas de un blog, etc. Organizar estos datos de una manera significativa no es una tarea sencilla, especialmente cuando los datos en sí cambian rápidamente.

Variabilidad

La variabilidad es diferente de la variedad. Este se refiere a los datos cuyo significado está cambiando constantemente. Donde las diferentes consultas requieren diferentes interpretaciones.

Veracidad

No todos los datos son confiables, incluso hay muchas mentiras generadas todos los días. Es necesario analizar los datos y determinar cuál es confiable y cuál es incorrecta. Es inservible si los datos analizados son erróneos o incompletos, por lo que la información debe ser verificada para poder apoyarse en ella en la toma de decisiones.

Visualización

El uso de tablas y gráficos para visualizar grandes cantidades de datos complejos es mucho más eficaz para transmitir significado de hojas de cálculo e informes repletos de números y fórmulas. Actualmente la visualización es fundamental. Una vez procesados los datos, hace falta presentarlos de manera visible y con el mayor aporte de información posible. Este es uno de los grandes retos de Big Data.

Valor

Por ultimo tenemos la séptima V, el Valor. El valor es el propósito real; después de abordar el volumen, la velocidad, la variedad, la variabilidad, veracidad, y la visualización que tienen una gran cantidad de tiempo, esfuerzo y recursos tenemos que estar seguros de que la compañía está recibiendo el valor de los datos.

El valor se refiere a los beneficios que los enormes datos pueden ofrecer a un negocio basado en la buena gran colección de datos, gestión y análisis. Por ejemplo, un teléfono móvil puede generar gran cantidad de datos, pero si una empresa no está utilizando los datos, que no tiene valor para ellos. Los datos tienen que tener valor para un negocio (Ortoll, 2014).

Tipos de Datos en el Paradigma Big Data

Big Data no es más que el tratamiento y análisis de ingentes repositorios de datos, tan grandes que resulta imposible trabajarlos con herramientas de bases de datos convencionales.

Los datos que se procesan en la actualidad se clasifican en tres tipos:

Datos estructurados

Estos datos tienen bien definidos su longitud y su formato, como las fechas, los números o las cadenas de caracteres, las mismas que se almacenan en tablas, como las bases de datos. Un ejemplo son las bases de datos relacionales y las hojas de cálculo (THEBIGGESTDATA, 2016).

Entre ellos tenemos:

- ✓ Fechas
- ✓ Números
- ✓ Cadenas de caracteres
- ✓ Datos financieros
- ✓ Datos de entrada: información que el usuario ingrese al sistema
- ✓ Datos de los sitios web visitados
- ✓ Datos de los juegos realizados

Datos no estructurados

Su definición también es sencilla, son opuestos a los estructurados,
Datos en el formato tal y como fueron recolectados, carecen de un formato específico y no se pueden almacenar dentro de una tabla, ya que su contenido es diverso y no se ajusta a este tipo de clasificación (THEBIGGESTDATA, 2016).

Entre ellos tenemos

- ✓ Fotografía y vídeo: por ejemplo cámaras de vigilancia.
- ✓ Información interna de la compañía: documentos, presentaciones, correos electrónicos, etc.
- ✓ Redes sociales: Twitter, Facebook, LinkedIn, Flickr, Instagram, etc.
- ✓ Dispositivos móviles: Mensajes que enviamos con nuestros teléfonos móviles.
- ✓ Sitios web: Vídeos de YouTube, blog, etc.

Datos semiestructurados

Datos que no se limitan a campos determinados, pero que contiene marcadores para separar los diferentes elementos. Es una información poco regular como para ser gestionada de una forma estándar. Estos datos poseen sus propios metadatos semiestructurados que describen los objetos y las relaciones entre ellos, y pueden acabar siendo aceptados por convención. Un ejemplo es el HTML, el XML o el JSON (THEBIGGESTDATA, 2016).

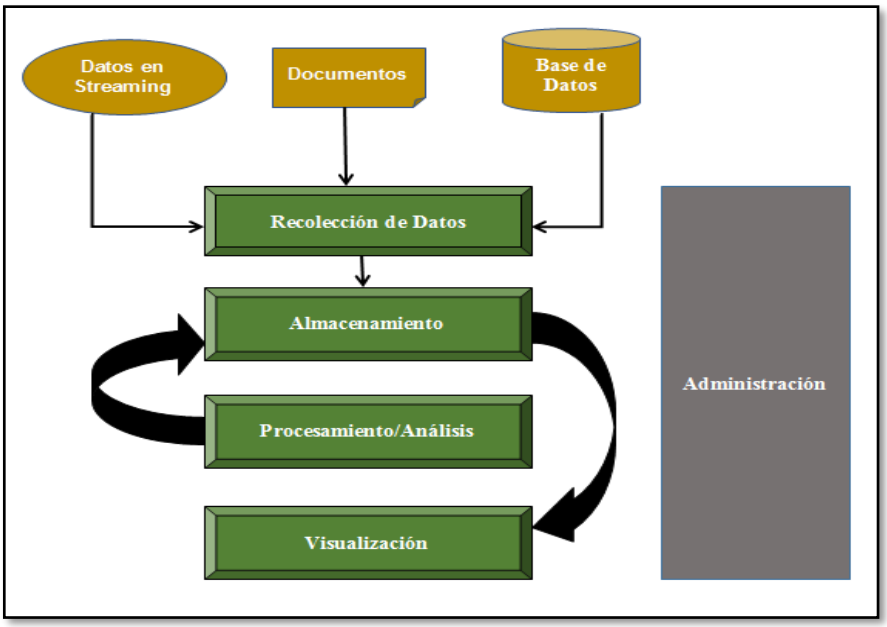
Arquitectura Big Data

La arquitectura de Big Data está compuesta por cinco bloques:

- ✓ Recolección de datos
- ✓ Almacenamiento
- ✓ Procesamiento de datos
- ✓ Visualización

ILUSTRACIÓN N.1

ARQUITECTURA BIG DATA



Elaboración: Fátima Vizueta Quinto
Fuente: Datos del tema de investigación

Recolección de datos

Esta etapa vamos a obtener los datos que los fabricamos directa e indirectamente segundo tras segundo. Estos datos pueden originarse de diversas fuentes como Web y Social Media (redes sociales), M2M, *Transacciones de datos,* Biométrica, Generados por las personas. Esta información también puede originarse *de* sistemas de lotes (batch) o de sistemas en tiempo real (Streaming) (Sevilla & Salas, s.f.).

Es necesario extraer la información de interés de forma que pueda ser fácilmente indexada y explorada más adelante. En esta fase de recolección se envía los datos a la etapa de almacenamiento, donde se guardarán los datos que han sido recolectados.

Almacenamiento

En esta capa de almacenamiento tenemos dos elementos básicos: Sistemas de ficheros y Bases de datos.
Los métodos de almacenamiento como las Bases de Datos relacionales se han quedado cortos a la hora de tratar esta información conocida como Big Data. Pero con el paso del tiempo han surgido nuevas maneras de guardar información como por ejemplo las bases de datos no relacionales. Debido a que Big Data busca la mayor variedad posible los sistemas de fichero han cobrado mayor importancia porque presentan flexibilidad a la otra de almacenar información que contiene características como la variabilidad, velocidad de generación, volumen de datos, etc (Sevilla & Salas, s.f.).

Procesamiento y Análisis

Hoy en día, contamos con herramientas muy poderosas que permiten procesar esta información que muchas veces presenta diferentes tipos de formato y orígenes como las bases de datos NoSQL o los sistemas de ficheros. Es la fase de mayor importancia a la hora de hablar de Big Data ya que una vez que se tienen almacenados los datos, se busca adquirir conocimiento o valor por medio del procesamiento y análisis de toda esta información almacenada. La necesidad de crear nuevas aplicaciones y que éstas ya estén adaptadas a los sistemas de almacenamiento más recientes ha promovido la aparición del paradigma de procesamiento como MapReduce (Sevilla & Salas, s.f.).

Visualización

Esta capa de Visualización muestra el resultado del almacenamiento y procesamiento de la información que da como resultado la producción de conocimiento. Debemos saber visualizar estos datos, para entenderlos mejor y así sacar provecho de ellos. Este conocimiento es de vital importancia para los diferentes sectores de la industria ya que presenta la oportunidad de innovación y desarrollo de nuevas líneas de negocios para sus organizaciones. No se trata

sólo de contar con la tecnología para obtener y analizar los datos, sino de ser capaces de darle significado (Sevilla & Salas, s.f.).

HERRAMIENTA PARA TRATAMIENTO DEL BIG DATA

BASE DE DATOS NOSQL

Definición

En Informática las bases de datos NoSQL también llamadas No Solo SQL aparecen con la llegada de la web 2.0, ya que hasta ese momento sólo subían contenido a la red aquellas empresas que tenían un portal, pero con la llegada de aplicaciones como Facebook, Twitter o YouTube, cualquier usuario podía subir contenido, provocando así un crecimiento exponencial de los datos (ACENS, 2014).

Podemos definir las bases de datos NoSQL como sistemas de almacenamiento de información que no cumplen con el esquema entidad-relación al que todos nos acostumbramos desde las primeras asignaturas de bases de datos en las carreras de informática. Mientras tanto que las tradicionales bases de datos relacionales basan su funcionamiento en tablas, joins y transacciones ACID, las bases de datos NoSQL no imponen una estructura de datos en forma de tablas y relaciones entre ellas; en ese sentido son más flexibles, ya que suelen permitir almacenar información en otros formatos como clave-valor, Mapeo de Columnas, Documentos o Grafos (Pereira, 2011).

Entonces, base de datos NoSql debe entenderse como la herramienta de almacenamiento de información en aquellas situaciones en las que las bases de datos relacionales generan ciertas dificultades debido principalmente a problemas de escalabilidad y rendimiento de las bases de datos relacionales donde se dan cita miles de usuarios concurrentes y con millones de consultas diarias (ACENS, 2014).

Características

Las características de las bases de datos NoSQL suelen ser las siguientes:

Ausencia De Esquema

Esta primera característica nos dice que los datos no tienen una definición de atributos fija, es decir: Cada registro (o documento, como se les suele llamar en estos casos) puede contener una información con diferente forma cada vez, pudiendo así almacenar sólo los atributos que interesen en cada uno de ellos, facilitando el polimorfismo de datos bajo una misma colección de información. En un mismo documento se pueden almacenar estructuras de datos complejas, como por ejemplo almacenar la información sobre una publicación de un blog junto a los comentarios y etiquetas vertidos sobre el mismo, todo en un único registró. Hacerlo así amplía la claridad (al tener todos los datos relacionados en un mismo bloque de información) y el rendimiento (no hay que hacer un JOIN para obtener los datos relacionados, pues éstos se encuentran directamente en el mismo documento) (Paramio, 2011).

Escalabilidad Horizontal

En esta cualidad se amplía la posibilidad de aumentar el rendimiento del sistema simplemente aumentando más nodos, sin necesidad en muchos casos de realizar ninguna otra operación más que indicar al sistema cuáles son los nodos disponibles. Además de lo comentado anteriormente, muchos de estos sistemas NoSQL permiten utilizar consultas del tipo **Map-Reduce**, las cuales pueden ejecutarse en todos los nodos a la vez (cada uno operando sobre una porción de los datos) y reunir luego los resultados antes de devolverlos al cliente. La gran mayoría permite también indicar otras cosas como el número de réplicas en que se hará una operación de escritura, para garantizar la disponibilidad (Paramio, 2011).

Velocidad

Además de la carencia de un esquema predeterminado, la principal característica de las bases de datos NoSQL es que están pensadas para manipular enormes cantidades de información de manera muy rápida.

Muchos de estos sistemas realizan operaciones directamente en memoria, y sólo vuelcan los datos a disco cada cierto tiempo. Esto permite que las operaciones de escritura sean realmente rápidas. Por supuesto, trabajar de este modo puede sacrificar fácilmente la durabilidad de los datos, y en caso de cuelgue o apagón se podrían perder operaciones de escritura o perder la consistencia. Normalmente, esto lo resuelven permitiendo que una operación de escritura haya de realizarse en más de un nodo antes de darla por válida, o disminuyendo el tiempo entre volcado y volcado de datos a disco. Pero claro, aun así, existe ese riesgo (Paramio, 2011).

Consultas Simples

Son sistemas simples que no tienen un sistema de consulta complejo ni con capacidad declarativa para en una sola línea realizar una cantidad interna de operaciones desorbitada (Antonio, 2011).

Modelo De Concurrencia Débil

El modelo de concurrencia enseñan a no garantizar las propiedades del modelo ACID, las cuales son necesarias para que una serie de instrucciones puedan ser consideradas una transacción, por tal motivo NoSql implementa BASE (Rodriguez & Conesa, 2015).

Ventajas y Desventajas Nosql

Ventajas

Veamos algunas de las ventajas que proveen este tipo de base de datos:

- ✓ Las bases de datos NoSQL utilizan sobre todo el uso de memoria en vez del disco como la principal ubicación de escritura.

- ✓ Permiten manejar volúmenes enormes de información con una velocidad de acceso rápida.

- ✓ Poseen estructuras de datos flexibles, añadir campos (columnas en las base de datos relacionales de forma dinámica). Permiten adaptarse a necesidades de Proyectos mucho más fácilmente que los modelos de Entidad Relación.

- ✓ Fácilmente Escalable. Con las bases de datos NoSQL, solo se tendrán que mover a la nube en entornos virtualizados, esto contrasta con la inversión de hardware.
- ✓ La gran mayoría son open-source y lo único que cobran es el asesoramiento o alguna herramienta especifica.

- ✓ El software del servidor puede ejecutarse en multitud de sistemas operativos.

- ✓ Es la base de datos con más orientación hacía Internet.

- ✓ Diferentes BBDD NoSQL para diferentes Proyectos.

- ✓ Cada tipo de base de dato NoSQL tiene sus propias características.
- ✓ Se pueden ejecutar en clúster en máquinas no muy potentes y baratas

- ✓ Se pueden hacer cambios de los esquemas sin tener que parar bases de datos.
- ✓ Simple de instalar.

Además hay que tener en cuenta también que hay más ventajas que las mencionadas aquí, pero las mismas no se aplican a todas las base de datos (TERASWAP, 2014).

Desventajas

A pesar de los beneficios de las bases NoSQL también tiene algunas desventajas que detallamos a continuación:

- **El código abierto puede significar una "sombra" en el soporte para las empresas**

  Mientras los proveedores de RDBMS tales como Oracle, IBM y Sybase ofrecen buenos soportes desde pequeñas a grandes empresas y típicamente start-ups, los vendedores de código abierto esperan ofrecer un soporte semejante con excepción de un puñado de clientes blue-chip. El hecho de ser código abierto puede ocasionar desconfianza en muchas empresas con respecto al soporte en comparación con empresas globales Generalmente un vendedor de código abierto no tiene el alcance global, servicios de soporte, y la credibilidad de Oracle o IBM (linuxParty, 2011).

- **No está lo suficientemente maduro para algunas empresas**

  Porque a pesar de sus puestas en prácticas en grandes empresas, aun se enfrentan al problema de credibilidad importante con muchas empresas. A diferencia que los distintos servicios que ofrecen madurez, estabilidad y una gran funcionalidad, eso en los sistemas NoSQL, todavía está por ver (linuxParty, 2011).

- **Falta de experiencia**

  La novedad de NoSQL es un gran problema y es que no existen desarrolladores y administradores que conocen la tecnología y la pueden gestionar, por lo que en caso de querer implantar un sistema NoSQL será más complicado encontrar personal capacitado (linuxParty, 2011).

- **Falta de estándares**

  En NoSQL como Open Source, muchas veces el desarrollador se convierte en el “optimizador de las consultas de datos”, lo que sería un retroceso de 20 o 30 años en la manipulación de datos a gran escala.

  Hay muchas bases de datos NoSQL y aún no hay un estándar como si lo hay en las bases de datos relacionales. Al tratarse de una tecnología emergente donde no hay estándares y las alternativas son múltiples, existe un grado importante de no compatibilidad entre herramientas (Ambriz, 2013).

- **Dificultad en el uso de transacciones**

  Las bases de datos no relacionales tienen pobre soporte a transacciones, a lo mucho, las soportan a nivel de "registro" , tampoco soportan transacciones que involucren múltiples entidades. Esto es un problema en el caso de aplicaciones que necesiten mantener la consistencia a lo largo de grupos de registros y múltiples entidades (Ambriz, 2013).

- **Punto débil: Fiabilidad**

  Mientras ofrecen la gestión de un mayor volumen y variedad de datos con mayor agilidad, no necesariamente brindan garantías de fiabilidad de la Data. El principal inconveniente se presenta en que al priorizar por la disponibilidad y la inmediatez del dato para el usuario, trabajan con datos distribuidos y a veces replicados. Y si el dato es muy volátil, puede originar inconsistencia o distintas versiones, algo que puede complicar a las aplicaciones transaccionales (GERENCIA, 2016)

Teorema Cap

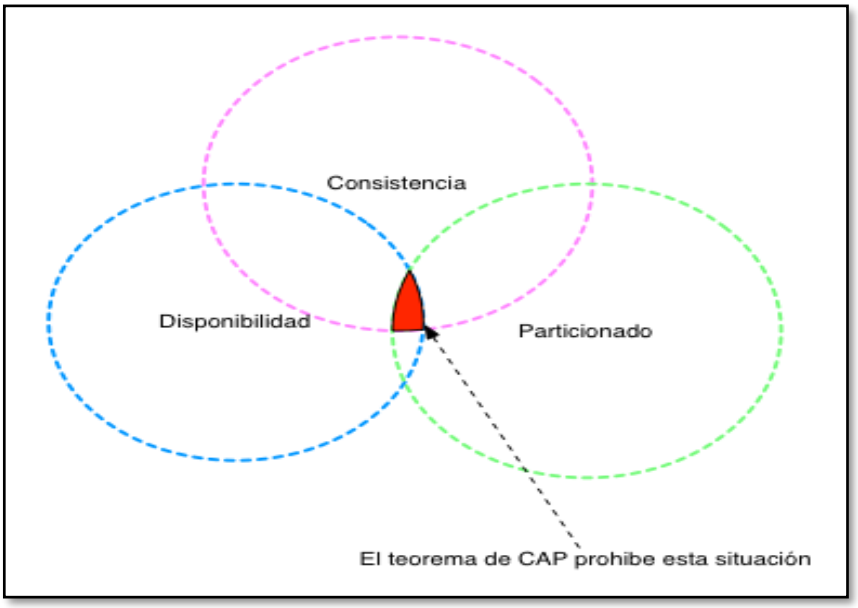
Al tratarse del Proyecto sobre el Big Data, es necesario conocer un poco más a fondo aspectos teóricos para poder elegir la base de datos NoSql que mejor se adapte a nuestro trabajo , eso lo proporciona en gran medida el Teorema CAP.

Como ya vimos en el anterior post, existen muchas bases de datos NoSQL, que a su vez nos proveen distintas ventajas necesarias para implementar soluciones altamente distribuidas y escalables. Según el teorema CAP, también llamado Teorema de Brewer dice que **un** sistema de datos distribuido pude asegurar dos de estas tres propiedades (Strappazzon, 2016) **:**

- ✓ Consistencia (Consistency),
- ✓ Disponibilidad (Availability),
- ✓ Tolerancia al Particionado (Partition Tolerance),

Visualmente esto se representa así:

ILUSTRACIÓN N.2
TEOREMA CAP



Elaboración: *Cecilio Álvarez Caules*
Fuente: http://www.cantabriatic.com/nosql-y-el-teorema-de-cap/

Consistencia (Consistency)

Todos los nodos deben garantizar la misma información al mismo tiempo, si insertamos datos (todos los nodos deben insertar los mismos datos), si actualizamos datos (todos los nodos deben aplicar la misma actualización a todos los datos) y si consultamos datos (todos los nodos deben devolver los mismos datos).Para tal efecto la comunicación entre los nodos debe ser de suma importancia, esto se le llama consistencia atómica, y se consigue replicando la información en todos los nodos (Strappazzon, 2016).

Disponibilidad (Availability)

Independientemente si uno de los nodos se ha caído o a dejado de emitir respuestas, el sistema debe seguir en funcionamiento y aceptar peticiones tanto de escritura como de lectura. Una vez que se pierde comunicación con un nodo, el sistema automáticamente debe tener la capacidad de seguir operando

mientras éste se restablece y una vez lo hace, se debe sincronizar con los demás. (Strappazzon, 2016).

Tolerancia al Particionado (Partition Tolerance)

El sistema debe estar funcionando así existan problemas de comunicación entre los nodos, cortes de red que dificulten su comunicación o cualquier otro aspecto que genere su particionamiento (Strappazzon, 2016).

Una vez que sabemos entonces qué es el teorema de CAP y cuáles son las características más importantes que lo definen, es momento entonces que veamos los posibles escenarios o combinaciones permitidas para cumplir lo anterior:

CP (Consistency–Partition Tolerance)

Los cambios se aplicaran de forma consistente y aunque se pierda la comunicación entre nodos ocasionando el particionado, no se asegura que haya disponibilidad (Strappazzon, 2016).

AP_(Availability–Partition_Tolerance)

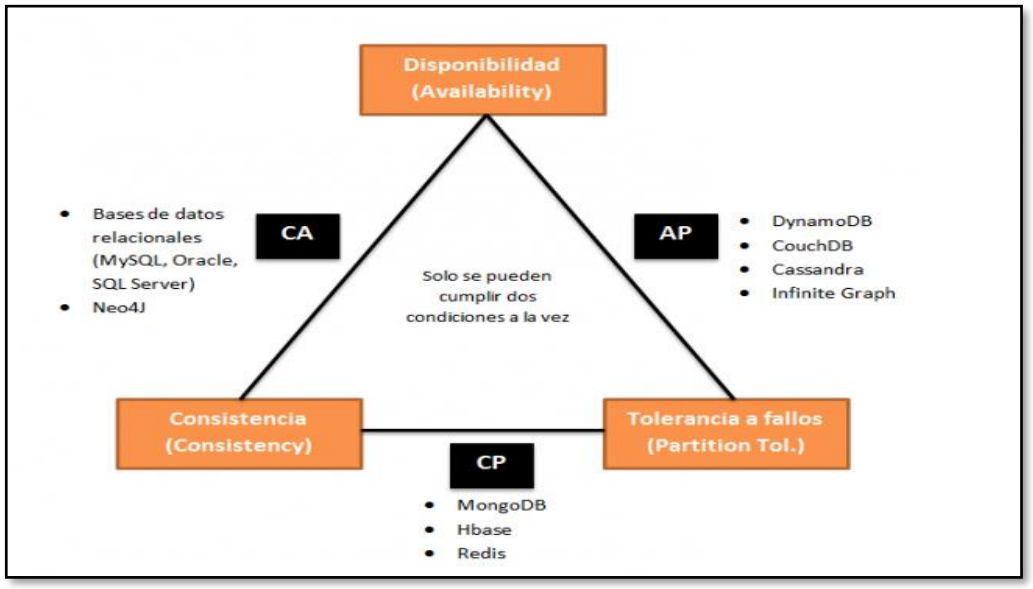
Aquí se garantiza la disponibilidad y tolerancia a particiones, pero no la consistencia, al menos de forma total. Este sistema siempre estará disponible a las peticiones aunque se pierda la comunicación entre los nodos ocasionando el particionado, y en consecuencia por la pérdida de comunicación existirá inconsistencia porque no todos los nodos serán iguales **(Strappazzon, 2016)**.

CA (Consistency – Availability)

En CA garantizan consistencia y disponibilidad, es decir el sistema siempre estará disponible respondiendo las peticiones y los datos procesados serán consistentes**.** En este caso no se puede permitir el particionado (Strappazzon, 2016).

ILUSTRACIÓN N.3

CLASIFICACIÓN DE BASE DE DATOS NOSQL SEGÚN EL TEOREMA CAP



Elaboración: Rubén Fernández
Fuente: http://www.genbetadev.com/bases-de-datos/nosql-clasificacion-de-las-bases-de-datos-segun-el-teorema-cap

Estas son entonces algunas de las clasificaciones que podemos tener con respecto a la base de datos *NoSQL* en la actualidad.

TIPOS DE BASE DE DATOS NOSQL

Las tecnologías NoSQL han evolucionado para dar respuesta a distintos problemas, y aunque tienen muchos aspectos en común, también son muy diferentes entre sí.

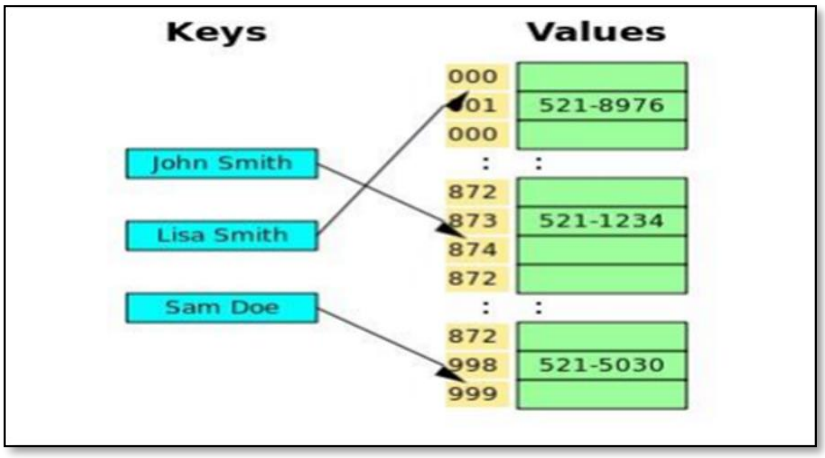
Dependiendo de la forma en la que almacenen la información, nos podemos encontrar varios tipos distintos de BBDD no relacionales. Dado que actualmente hay más de 225 bases de datos NoSql, veamos los tipos más utilizados:

- ✓ Orientadas a clave-valor
- ✓ Orientadas a columnas

- ✓ Orientadas a documentos
- ✓ Orientadas a Grafos

Orientadas a clave-valor

ILUSTRACIÓN N.4
BASES DE DATOS CLAVE – VALOR



Elaboración: Acens
Fuente: https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf

Esta es la base de dato NoSQL más popular, además de ser la más sencillas de entender. En este tipo de sistema, cada elemento está identificado por una clave única, lo que permite la recuperación de la información de forma muy rápida, Esta información que suele almacenarse como un objeto binario (BLOB) , pero la diferencia de este tipo de base de dato es que no necesita definir un esquema (columnas, tipos de datos) para almacenar la información. De esta forma el tipo de contenido no es importante para la base de dato, solo la clave y el valor que tiene asociado. Se caracterizan por ser muy eficientes tanto para las lecturas como para las escrituras, además de que pueden escalar fácilmente particionando los valores de acuerdo a su clave (ACENS, 2014).

El consumidor de esta información es responsable de conocer la estructura de la información almacenada. Están diseñadas para escalar masivamente manteniendo un tiempo de respuesta muy rápido y disponibilidad total. Se suelen usar para almacenar información de sesión, preferencias o perfiles de usuario, carritos de la compra y en general como cachés de cualquier conjunto de información que se pueda recuperar por una clave. Cabe acotar que los almacenes *key / value* son extremadamente rápidos pero no permiten consultas complejas más allá de buscar por su clave. Por lo que si tu aplicación necesita consultas complejas, este tipo de base de datos no se ajusta a tus necesidades (Ayestarán, s.f.).

Algunas BBDD: Redis , Riak y Dynamo.

Orientadas a columnas

Las bases de datos en columnas están pensadas para realizar consultas y agregaciones sobre grandes cantidades de datos, cuando éstas se pueden determinar previamente y no cambian con frecuencia. Funcionan de una forma parecida a las bases de datos relacionales, pero este tipo de base de datos están optimizadas para leer y escribir columnas de datos en lugar de filas así mismo guardan los valores en columnas en lugar de filas (Trigoso, 2016)

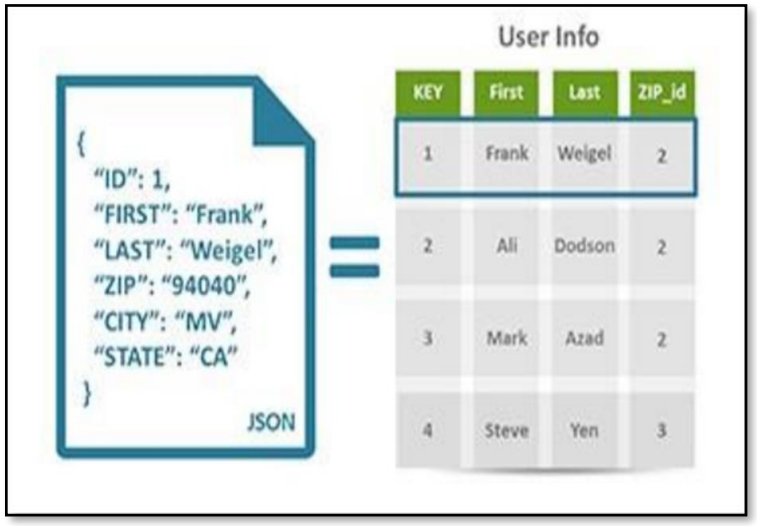
Con este cambio ganamos mucha velocidad en lecturas, ya que si se requiere consultar un número reducido de columnas, es muy rápido hacerlo pero no es eficiente para realizar escrituras. Por ello este tipo de soluciones es utilizado en entornos donde hay poca escritura y existe la necesidad de acceder a varias columnas de muchas filas. Acotando que son muy útiles en procesamiento y análisis de eventos, gestión de contenido y en análisis de datos (Javier, 2015).

Algunas BBDD: Hbase, Cassandra, Hypertable.

Orientadas a Documentos

ILUSTRACIÓN N.5

BASES DE DATOS DOCUMENTALES



Elaboración: Acens
Fuente: https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf

Este tipo de BBDD están diseñadas para almacenar datos semiestructurados como documentos; cabe destacar que no debe confundirse el término *documento* con ficheros Excel, Word, etc. El término *documento* se refiere a una estructura de datos (clave, valor) representada en formato *XML, JSON o BSON* y donde se utiliza una clave única para cada registro (Camacho, s.f.) .Es similar a las bases de datos Key-value, pero con la diferencia que el valor es un fichero que puede ser entendido. Si el servidor entiende los datos, puede hacer operaciones con ellos. De hecho varias de las implementaciones de este tipo de base de dato permiten consultas muy avanzadas sobre los datos, e incluso establecer relaciones entre ellos, aunque siguen sin permitir joins (Camacho, s.f.).

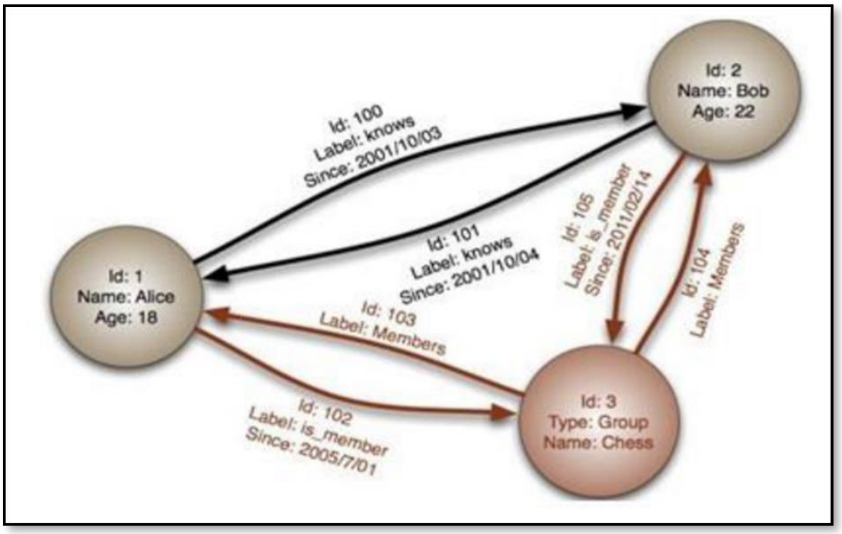
A diferencia de las bases de datos relacionales tradicionales, el esquema de cada documento no relacional (NoSQL) puede variar, lo que ofrece más flexibilidad al organizar y almacenar datos de aplicaciones y permite reducir el almacenamiento necesario para los valores opcionales. Son las bases de datos

más versátiles. Se pueden utilizar en gran cantidad de Proyectos, incluyendo muchos que tradicionalmente funcionarían sobre bases de datos relacionales (ACENS, 2014). Algunas BBDD: Mongo DB, CouchDB y RavenDB

Orientada a Grafos

ILUSTRACIÓN N.6

BASES DE DATOS EN GRAFO



Elaboración: Acens
Fuente: https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf

El almacenamiento en este tipo de bases de datos está optimizado para ejecutar la navegación entre nodos; en donde los nodos son entidades y las aristas son relaciones. Contienen información con uso a menudo de tablas hash distribuidas y ofrecen estructuras de datos sencillas como arrays asociativos o almacenes de pares claves valor (Perez S. , 2015) .

Obviamente, este tipo de bases de datos son muy útiles para guardar información en modelos con muchas relaciones, como redes y conexiones sociales. Algo que ocurre mucho en redes sociales o sistemas de recomendación de productos, donde además se tiene la ventaja de poder aplicar algoritmos estadísticos para determinar recomendaciones que se basan en recorrer grafos (Camacho, s.f.) .

Por este motivo, las bases de datos orientadas a grafo son eficientes para realizar consultas en las que existan relaciones de proximidad entre datos, y no para ejecutar consultas globales. Además para sacar el máximo rendimiento a este tipo de bases de datos, su estructura debe estar totalmente normalizada, de forma que cada tabla tenga una sola columna y cada relación dos (Perez S. , 2015).

Algunas BBDD: Neo4j, HyperGraphDB.

¿Cuándo Usar Nosql?

Si pretendemos desarrollar una aplicación que requiera la lectura/escritura de cantidades enormes de datos y pueda dar servicio a millones de usuarios sin perder rendimiento, entonces debemos plantearnos el uso de una base de datos NoSQL. Las grandes redes sociales como Facebook, Instagram y Twitter o el propio Google las utilizan como medio fundamental de almacenamiento de información. Estas compañías se dieron cuenta que no era tan necesaria la lógica en la organización de sus datos, por esto buscaron un nuevo enfoque para tratar gran cantidad de datos. Sin duda alguna, con esto podemos saber, si deseamos usar un gestor de datos para una aplicación transaccional, que mejor que las viejas bases de datos relacionales. Pero si lo que necesitamos es manejar un sin número de información, debemos de inclinarnos por las Bases de Datos NoSQL (Mireles, 2016).Yo considero que las tecnologías NoSQL pueden ser usadas en los siguientes ámbitos:

- ✓ **Redes sociales:** Es obligatorio. Gracias a las redes sociales, ésta tecnología comenzó a despegar y mostrar su gran utilidad en el campo de la informática y la estadística (Zapata, 2013) .

- ✓ **Desarrollo Web:** Considero más oportuno el uso de éstas tecnologías en ésta área, debido a la poca uniformidad de la información que encontramos en Internet, sin embargo también es posible realizar éstos desarrollos con SQL (Zapata, 2013).

- ✓ **Desarrollo Móvil** En éstos momentos, las empresas están lidiando con un problema grande conocido como **Bring your own device** (trae tu propio dispositivo), abreviado **BYOD–** en realidad no es un problema, es un fenómeno social -, por lo que la información que se recolecte siempre será diferente por más que uno desee estructurarla y mantenerla estática (Zapata, 2013).

- ✓ **BigData:** debido a la administración de grandísimas cantidades de información y su evidente diversidad hace de NoSQL un excelente candidato (Zapata, 2013).

- ✓ **Cloud (XaaS**): "Everything as a service"; NoSQL puede adaptarse casi a cualquier necesidad del cliente, que evidentemente son heterogéneos (Zapata, 2013).

NEO4J

Neo4j pertenece al tipo de base de datos NOSQL orientada a grafos esta implementado en java, almacena los datos de nuestra aplicación en términos de nodos, relaciones y propiedades. El lenguaje de consulta base en Neo4J es Cypher y se basa en la navegación entre nodos (Sánchez, 2014). Se desarrolló en el 2003 y disponible a los usuarios desde el año 2007, el código fuente se encuentra disponible en GitHub (Fernandez J. , 2015) .

¿Qué es un grafo?

Se lo define como una colección de nodos y las relaciones que los conectan. Los grafos nos sirven para componer toda clase de escenarios (Solis, 2013).

- ✓ Los grafos tienen las siguientes características:

- ✓ Están formados por nodos y relaciones

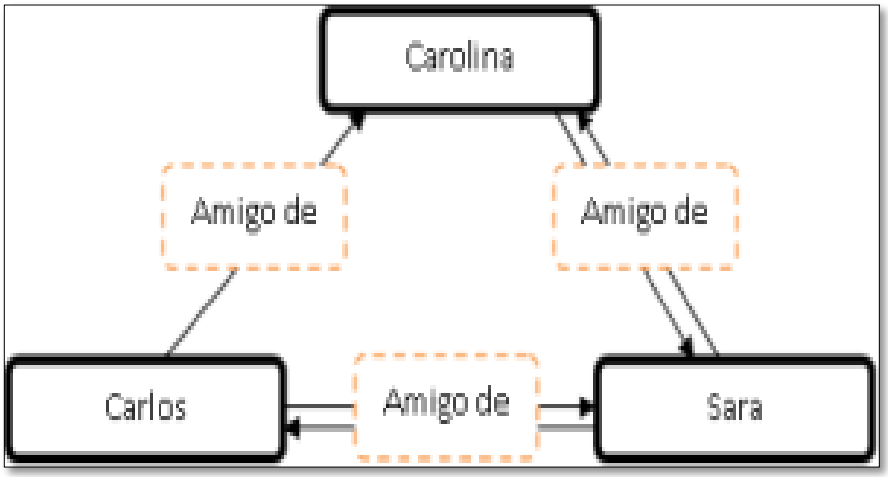
- ✓ Los nodos contienen propiedades (del tipo clave-valor)

- ✓ Las relaciones tienen un nombre y siempre parte de un nodo de inicio a otro de destino.
- ✓ Las relaciones también pueden tener propiedades.

Ejemplo en Facebook, y cómo podría ser su modelo de datos utilizando grafos para una pequeña red de amigos (Solis, 2013).

ILUSTRACIÓN N.7

GRAFO SENCILLO DE RED AMIGOS



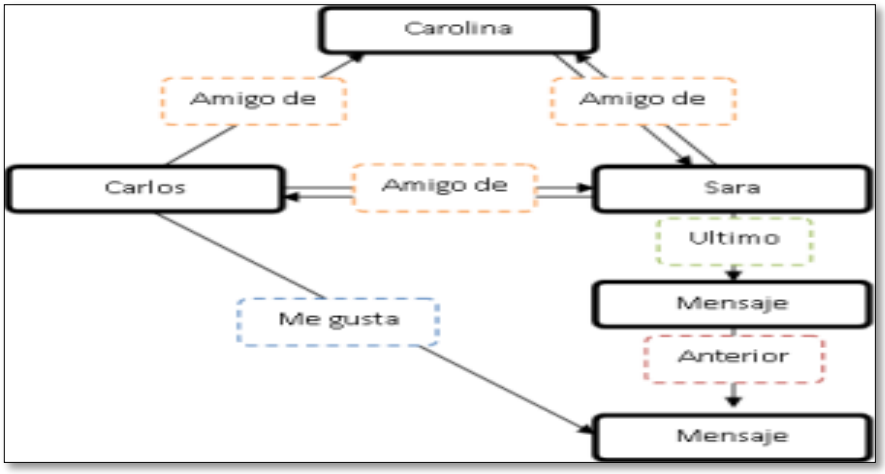
Elaboración: Santiago Márquez Solís
Fuente:http://www.santiagomarquezsolis.com/herramientas-para-Big-Data-neo4j-parte-1/

Detallamos lo que observamos:

Carlos y Sara se consideran amigos mutuos, del mismo modo que Carolina y Sara. Carlos ha establecido que Carolina es su amiga, ésta aún no ha respondido recíprocamente a esta amistad.
Ahora Sara ha dado una serie de mensajes que actualizan su estado, y que de estos mensajes el último se enlaza con Sara mediante una relación “último” y los mensajes anteriores mediante una relación “anterior”, establecer la relación “me gusta” sobre un mensaje es algo tan sencillo como:

ILUSTRACIÓN N.8

GRAFO SENCILLO DE RED AMIGOS-MENSAJE



Elaboración: Santiago Márquez Solís
Fuente:http://www.santiagomarquezsolis.com/herramientas-para-Big-Data-neo4j-parte-1/

Aquí vemos que a Carlos le gusta un mensaje particular de su amiga, pero no ha señalado ninguna opinión sobre el resto de estados emitidos por ella. Una base de datos basada en grafos debe permitir tanto el modelado de un escenario como el anteriormente establecido, como proporcionar las herramientas necesarias para su consulta y análisis.

Características

- ✓ No hay esquema
- ✓ Puede contener billones de nodos y relaciones
- ✓ Rápido recorriendo relaciones, este tipo de queries se conoce como transversals
- ✓ Lenguaje de query propio, Cypher

- ✓ Alta disponibilidad, instalación en diferentes maquinas con balanceador de carga
- ✓ Multilenguaje, proporciona una Api Rest pudiendo utilizarse desde cualquier lenguaje.

Además de las características mencionadas también es open-source (Sánchez, 2014).

Ventajas

- ✓ Es muy fácil de representar los datos conectados.
- ✓ Es muy fácil y rápido para recuperar / recorrido / navegación de los datos.
- ✓ Representa los datos semi-estructurados con mucha facilidad.
- ✓ comandos del lenguaje de consulta Neo4j CQL están en formato legible humano y muy fácil de aprender (Sánchez, 2014).

¿Por Que Usar Neo4j?

- ✓ La mejor y la primera base de datos de grafos del mundo.
- ✓ La comunidad de grafos más grande y activa del planeta.
- ✓ Muy fácil de aprender.
- ✓ Fácil de usar
- ✓ Sólida confiabilidad para aplicaciones de producción de misión crítica
- ✓ Más fácil que nunca es cargar sus datos en Neo4j

- ✓ Modelar como se piensa simplifica mucho el ciclo de desarrollo
- ✓ Alto rendimiento gracias al procesamiento y almacenamiento nativo de grafos.

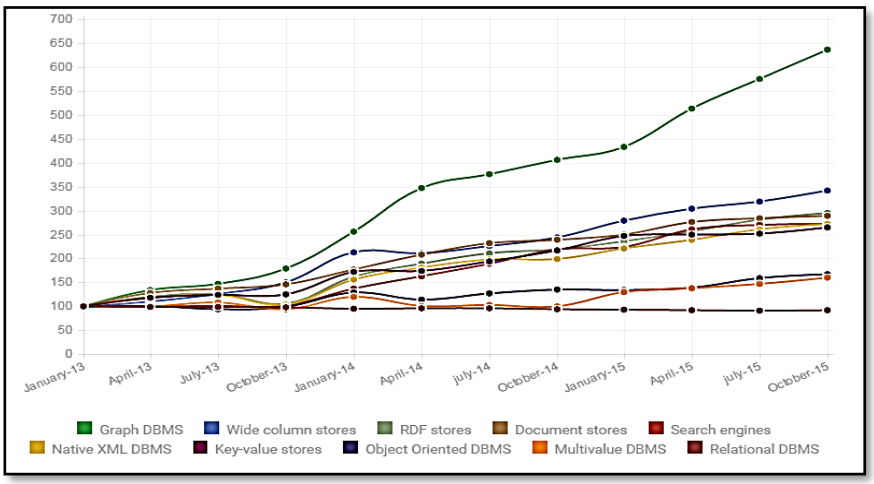
Además Neo4j entrega un alto rendimiento tanto en lecturas como escrituras, sin comprometer la integridad de sus datos (Neo4j, 2016)

Índice de popularidad

El área de las bases de datos es bastante grande, es interesante ver que particularmente las de grafos están tomando un fragmento cada vez mayor de la atención. La capacidad este tipo de base de datos para representar y procesar una variedad de diferentes objetos y conexiones complejas entre estos es una manera muy natural de modelar muchos sistemas (GRAPHEVERYWHERE, s.f.) . A continuación vemos las tendencias en la popularidad de cada tecnología.

GRAFICO N.5

Índice de popularidad por categoría



Elaboración: grapheverywhere
Fuente: http://www.grapheverywhere.com /language/es/jesus-neo4j-2/

Lenguaje Cypher

El lenguaje para realizar consultas en Neo4j es Cypher o también llamado CQL que significa Cypher Query Language . Este lenguaje nos permite afirmar lo que queremos seleccionar, insertar, actualizar o eliminar de nuestros datos basado en grafos sin necesidad de nosotros para detallar cómo hacerlo (Neo4j, 2016). Al igual que Oracle SQL, Neo4j CQL posee comandos para realizar procedimientos de base de datos como Where , Order By, etc. Vamos a emplear este lenguaje en un ejemplo.

Primero procedemos con la instalación que solamente consiste en hacer la descargar de la versión gratuita (“community edition”) desde su web oficial https://neo4j.com.

Función CREATE

Utilizamos CREATE para crear un nodo del grafo, el mismo que va entre paréntesis con etiqueta ‘Movie’ que se guarda en la variable ‘TheMatrix’ con las propiedades ‘title’, ‘released’ y ‘tagline’ con sus respectivos valores en formato JSON. Un trozo del grafo para la película Matrix con sus actores principales, directores y productor.

ILUSTRACIÓN N. 9

FUNCIÓN CREATE

```
CREATE (TheMatrix:Movie {title:'The Matrix', released:1999, tagline:'Welcome to the Real World'})
CREATE (Keanu:Person {name:'Keanu Reeves', born:1964})
CREATE (Carrie:Person {name:'Carrie-Anne Moss', born:1967})
CREATE (Laurence:Person {name:'Laurence Fishburne', born:1961})
CREATE (Hugo:Person {name:'Hugo Weaving', born:1960})
CREATE (AndyW:Person {name:'Andy Wachowski', born:1967})
CREATE (LanaW:Person {name:'Lana Wachowski', born:1965})
CREATE (JoelS:Person {name:'Joel Silver', born:1952})
CREATE
  (Keanu)-[:ACTED_IN {roles:['Neo']}]->(TheMatrix),
  (Carrie)-[:ACTED_IN {roles:['Trinity']}]->(TheMatrix),
  (Laurence)-[:ACTED_IN {roles:['Morpheus']}]->(TheMatrix),
  (Hugo)-[:ACTED_IN {roles:['Agent Smith']}]->(TheMatrix),
  (AndyW)-[:DIRECTED]->(TheMatrix),
  (LanaW)-[:DIRECTED]->(TheMatrix),
  (JoelS)-[:PRODUCED]->(TheMatrix)
```

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Después nos aparecerá el siguiente mensaje que nos dice “Añadido 8 etiquetas, creado 8 nodos, establece 21 propiedades, creado 7 relaciones, sentencia ejecutada en 2896 ms”

ILUSTRACIÓN N. 10

SALIDA FUNCIÓN CREATE

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Crearemos unos cuantos nodos más etiquetados con ‘Person’ y las propiedades ‘name’ y ‘born’ de forma similar al anterior. Estos nodos de momento no forman un grafo, al menos un grafo que tenga cierto sentido para nosotros, por ello es necesario crear las relaciones. Las relaciones se crean de forma muy similar a los nodos, mediante **CREATE** indicamos a través de las variables que identifican los nodos las relaciones con otros nodos. La relación puede tener nombre y propiedades.

ILUSTRACIÓN N. 11

FUNCIÓN CREATE

```
CREATE (TheMatrix:Movie {title:'The Matrix', released:1999, tagline:'Welcome to the Real World'})
CREATE (Keanu:Person {name:'Keanu Reeves', born:1964})
CREATE (Keanu)-[:ACTED_IN {roles:['Neo']}]->(TheMatrix)
```

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

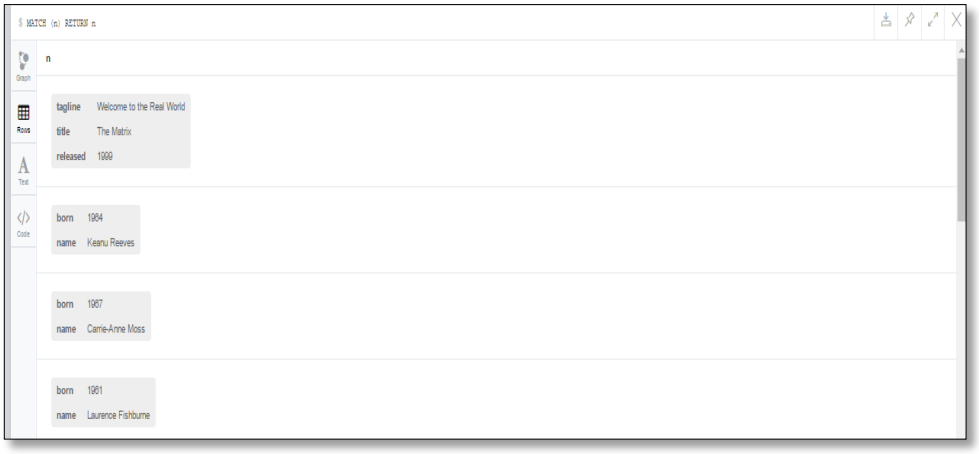
Función MATCH (Vamos a proceder a realizar consultas sobre el grafo.)
Para consultar todo el grafo.

```
MATCH (n)
RETURN n
```

En forma tabular:

ILUSTRACIÓN N. 12

FUNCIÓN MATCH

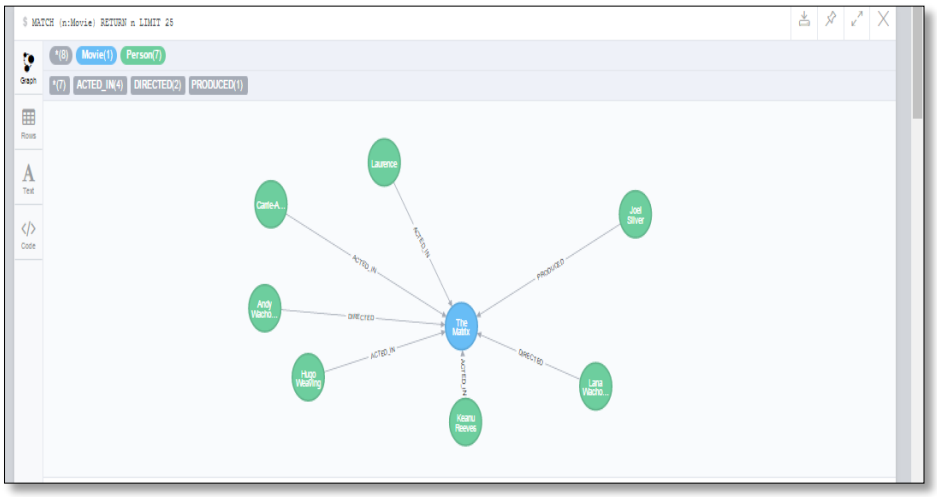


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

En grafo:

ILUSTRACIÓN N. 13

SALIDA FUNCIÓN MATCH



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Para consultar todas las películas en las que ha actuado un actor:

```
MATCH (n {name: "Keanu Reeves"})-[:ACTS_IN]->(d)
RETURN n, COUNT(d)
```

Función DELETE

Esta función es similar a la función sql, y con MATCH podemos hacer de filtro para saber que queremos eliminar.

```
MATCH (n)
DELETE n
```

Ejercicio Propuesto

En este caso se muestra el código programado de las variables como, los datos de la persona, empresa y las relaciones que existen entre ellos.

- Creación de variables y atributos.
- Relaciones de las variables.

Creando Atributos:

```
create (Rosa:Persona {Nombre: "Rosa Vera", Profesion:"Enfermera", Edad: "50"})
create (Bryan:Persona {Nombre: "Bryan Garcia", Profesion:"Licenciado en Enfermeria", Edad: "22"})
create (Cinthya:Persona {Nombre: "cinthya Vizueta", Profesion:"Ingeniera en Sistemas Computacionales", Edad: "24"})
create (Carla:Persona {Nombre: "Carla Moran", Profesion:"Ingeniera en Networking", Edad: "27"})
create (Pedro:Persona {Nombre: "Pedro Valencia", Profesion:"Ingeniero Comercial", Edad: "60"})
create (Virginia:Persona {Nombre: "Virginia Zamora", Profesion:"Ingeniera Industrial", Edad: "40"})
create (Ruben:Persona {Nombre: "Ruben Moran", Profesion:"Ingeniero Civil", Edad: "48"})
create (Cristhian:Persona {Nombre: "Cristhian", Profesion:"Ingeniero en teleinformatica", Edad: "25"})
```

```
create (Centro:Empresa {Nombre: "centro de Salud Camino al Sol", Cargo:"Enfermera", Experiencia: "3 años"})
create (Hospital:Empresa {Nombre: "Hospital del Niño", Cargo: "Licenciado en Enfermeria", Experiencia: "1 año"})
create (Solsa:Empresa {Nombre: "Solsa", Cargo: "Asesor de Software", Experiencia: "2 años"})
```

```
create (Cnt:Empresa {Nombre:"CNT", Cargo:"ANALISTA DE
TELECOMUNICACIONES NOC", Experiencia: "5 años"})
create (MasMusica:Empresa {Nombre: "MasMusica", Cargo:"Asesor Comercial",
Experiencia:"4 años"})
```

Creando la dependencia o enlace:

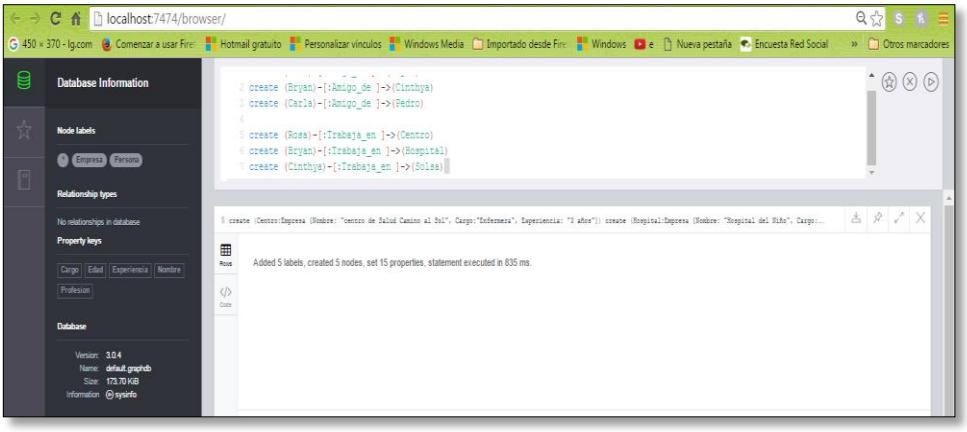
```
create (Rosa)-[:Amigo_de ]->(Bryan)
create (Bryan)-[:Amigo_de ]->(Cinthya)
create (cinthya)-[:Amigo_de ]->(Carla)
create (Carla)-[:Amigo_de ]->(Pedro)
create (Pedro)-[:Amigo_de ]->(Rosa)
create (Pedro)-[:Amigo_de ]->(Virginia)
create (Virginia)-[:Amigo_de ]->(Ruben)
create (Ruben)-[:Amigo_de ]->(cristhian)

create (Rosa)-[:Trabaja_en ]->(Centro)
create (Bryan)-[:Trabaja_en ]->(Hospital)
create (Cinthya)-[:Trabaja_en ]->(Solsa)
create (Carla)-[:Trabaja_en ]->(Cnt)
create (Pedro)-[:Trabaja_en ]->(MasMusica)
```

Una vez terminada los scripts se procede a copiar y pegarlos en la BD Neo4j.

En la siguiente imagen se muestra que han sido creadas las variables con sus atributos y relaciones. En el menú que se encuentra a la izquierda se muestra los atributos que se han creado automáticamente.

ILUSTRACIÓN N. 14
SALIDA FUNCIÓN CREATE



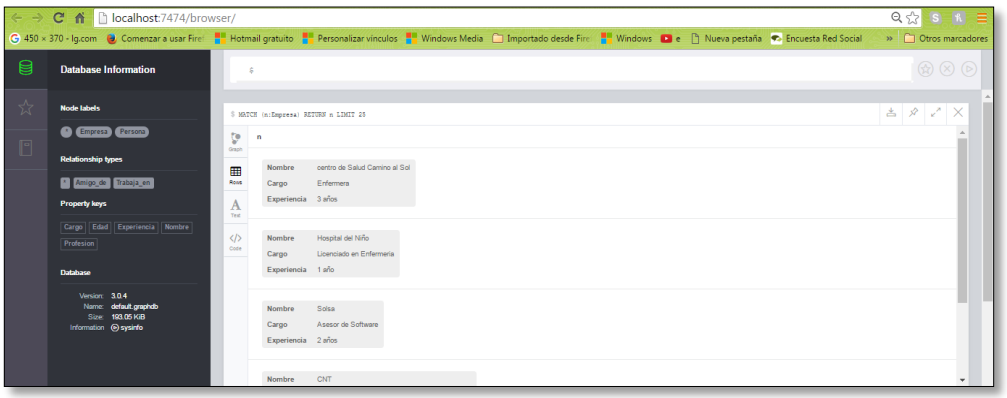
Elaboración: Fátima Cynthia Vizueta Quinto

Fuente: Datos de la Investigación

En la siguiente ventana se refleja los atributos con los datos creados.

ILUSTRACIÓN N. 15

VISUALIZACION DE ATRIBUTOS EN ROW



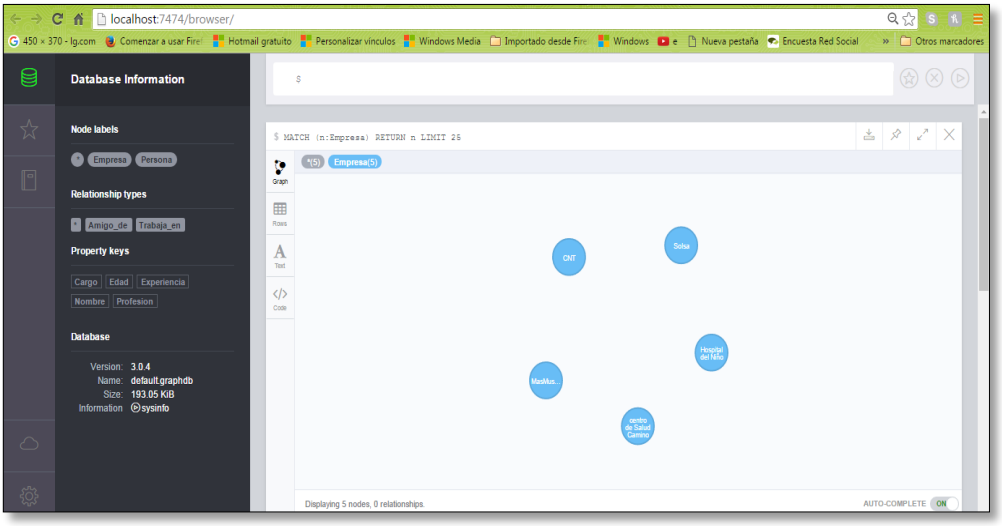
Elaboración: Fátima Cynthia Vizueta Quinto

Fuente: Datos de la Investigación

En la ventana siguiente se muestran las variables de la Empresa en forma de nodos.

ILUSTRACIÓN N. 16

VISUALIZACION DE NODO EMPRESA



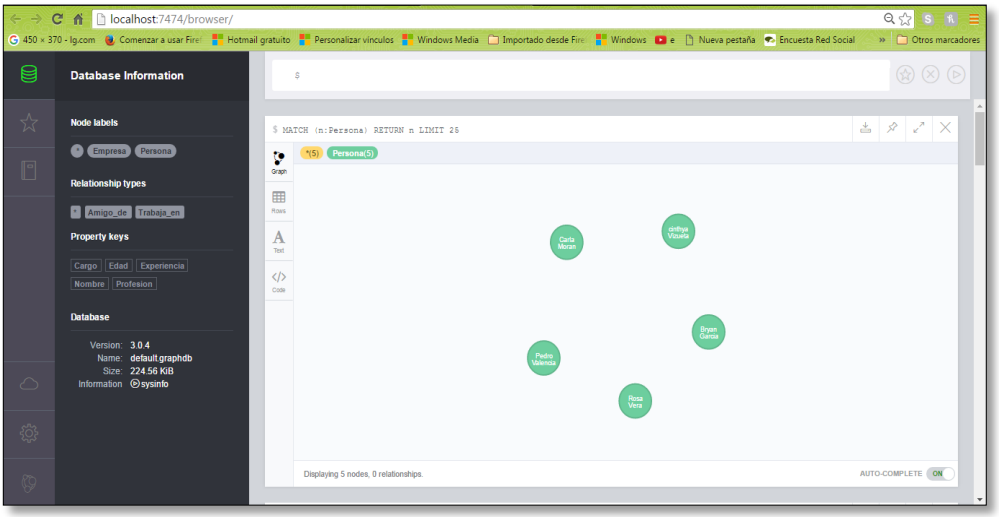
Elaboración: Fátima Cynthia Vizueta Quinto

Fuente: Datos de la Investigación

En la ventana siguiente se muestran las variables de la Persona en forma de nodos.

ILUSTRACIÓN N. 17

VISUALIZACION DE NODO PERSONA

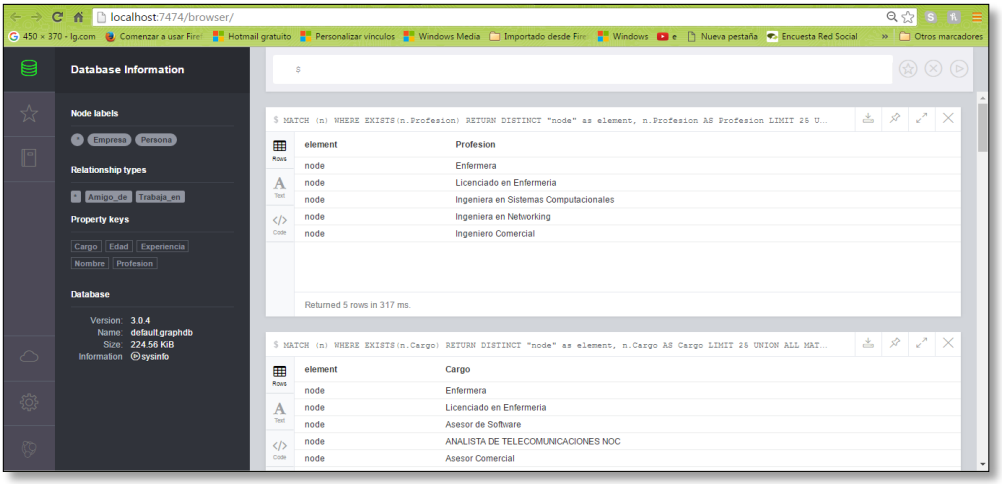


Elaboración: Fátima Cynthia Vizueta Quinto

Fuente: Datos de la Investigación

En esta ventana se refleja el detalle de los atributos de cada nodo.

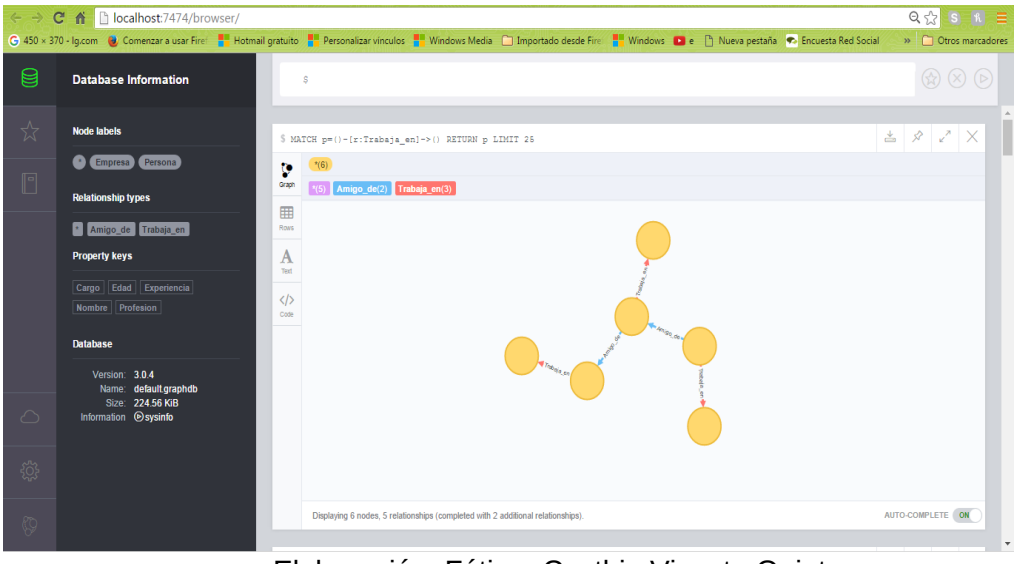
ILUSTRACIÓN N. 18
ATRIBUTOS DE CADA NODO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Representación de Grafos.

ILUSTRACIÓN N. 19
VISUALIZACION DEL GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Características de una implementación efectiva de Big Data

Abierto

Las tecnologías y cargas de trabajo de Big Data cambian continuamente. Cree su arquitectura con bloques modulares interoperables de modo que sus productos de Big Data funcionen con otras soluciones de su centro de datos (Redhat, s.f.).

Ágil

Una TI ágil permite responder a las amenazas competitivas, a los cambios en las tendencias del sector y al comportamiento de cliente. Elija soluciones que se puedan escalar sin problemas y que ofrezcan flexibilidad de implementación y ampliación para ayudarle a crear aplicaciones e integrar orígenes de datos con extrema agilidad (Redhat, s.f.).

Seguro

Los productos de Big Data seguros son compatibles con los estándares estipulados por la TI de su empresa. Las soluciones deben ayudar a normalizar los modelos de seguridad utilizados en su centro de datos, de modo que le aporte una visión sencilla pero completa de la seguridad de datos y aplicaciones basadas en roles (Redhat, s.f.).

Rentable

Reducir el costo de las soluciones de Big Data libera recursos útiles que contribuirán a que se centre en las tareas más necesarias y ayudarán a la empresa detectando mejores modos de fidelizar a los clientes (Redhat, s.f.).

Seguridad para Big Data

Ya tenemos una noción de todo lo relacionado al Big Data, pero hay un aspecto que no hemos considerado que es la seguridad de la información. Cuanta más información se esté procesando, mayor riesgo existe de incumplir normativas asociadas al tema de Privacidad de la información (incluyendo datos personales) especialmente en sectores normados. En cuanto a las implementaciones propiamente dichas, existen múltiples aspectos a evaluar tales como:

Sistemas/procesamiento distribuido

(Guirado, 2014) afirma:

> La mayor parte de las soluciones actuales se basan en sistemas básicamente centralizados es decir cuando el procesamiento se realiza en forma distribuida (MapReduce, etc.), la seguridad de cada "nodo", de todos los canales de comunicación y de la solución general de coordinación de tareas pasa a ser fundamental. Esto se dificulta aún más cuando se utilizan soluciones tipo "nube". (p.12)

Nuevas tecnologías

(Guirado, 2014) afirma:

> Cada uno de los componentes de la solución (Hadoop, bases de datos NoSQL, sistemas de visualización, etc.) pueden tener requerimientos de seguridad específicos y no necesariamente contarse con la experiencia para evaluar y administrar adecuadamente los mismos.Como con cualquier tecnología, a medida que se desarrolle su uso también se desarrollarán ataques sobre las mismas por lo que es es importante recordar que usualmente la seguridad suele "venir atrás" de las nuevas funcionalidades. (p.13)

Aspectos tradicionales

Autenticación

No debe corresponder solo a los usuarios, sino de todos los elementos que formen parte del ambiente (Guirado, 2014).

Integridad de los datos

Hay mayor problema por la distribución de los datos, la utilización de herramientas tecnológicas diferentes de las usuales para manejarlos, la variedad de los mismos y un “ciclo de vida” mucho más corto (Guirado, 2014).

Perfiles de acceso

Al obtener acceso a mayor cantidad de información y tener mecanismos más poderosos para procesar la misma, el contar con mecanismos con una granularidad adecuada para definir los accesos es crítico (Guirado, 2014).

Infraestructura

Naturalmente que en sistemas con un alto grado de distribución la administración de seguridad de la propia infraestructura se dificulta significativamente. La utilización de herramientas centralizadas de configuración/control y la utilización de estándares pasarán de ser una buena práctica a un requerimiento ineludible (Guirado, 2014).

Telecomunicaciones

Del mismo modo que en lo referente a infraestructura, todos los datos en vía deberán contar con una protección adecuada que además hoy en día estas situaciones son relativamente rara en redes internas (Guirado, 2014).

En que sectores se puede aplicar Big Data

Actualmente, la mayoría de negocios está experimentando una transformación digital gracias al Big Data. Por lo tanto, aquellos negocios que lo aprovechen mediante un análisis inteligente de datos a gran escala podrán conocer las pautas y las tendencias de los consumidores a los que dirigen sus productos y servicios.

Estos son algunos de los sectores que ya están aplicando la utilización del Big Data:

Seguros Calcular con éxito la cuota del seguro que deben pagar los usuarios en función de los datos obtenidos según la forma de conducir (LLAVINIA, 2016).

Farmacias: Monitorizar en tiempo real el efecto de medicamentos sobre enfermedades y evaluar su grado de efectividad (LLAVINIA, 2016).

Banca: Conocer el riesgo de los mercados en tiempo real en función de los datos obtenidos de las transacciones y cotizaciones. Conocer mejor a los clientes para ofrecer un trato personalizado (LLAVINIA, 2016).

Transporte: Conseguir la optimización de rutas en tiempo real a partir de datos de tráfico, meteorológicos y paradas obligatorias de ruta (LLAVINIA, 2016).

Educación: Aprenden de las acciones de los usuarios con ejercicios de formación y son capaces de proponernos el siguiente ejercicio a realizar para maximizar el desarrollo cognitivo del usuario (LLAVINIA, 2016).

Comercio: Las empresas usan Big Data para insertar publicidad de forma segmentada para llegar mejor al cliente, personalizar sitios web y tomar decisiones clave más eficientes sobre sus clientes (ElEconomista, 2016).

Deporte: Aquí el Big Data se está utilizando para decidir el mejor precio a pagar a un nuevo futbolista o incluso para predecir posibles lesiones de jugadores y sus momentos más óptimos para el descanso (ElEconomista, 2016).

Medio Ambiente: El procesado de millones de datos en tiempo real y su modelización permite vaticinar catástrofes y otros fenómenos naturales, según los expertos (ElEconomista, 2016).

Empresas que Aplican Big Data

El mercado de Big Data está creciendo constantemente, pues su uso permite obtener resultados exactos sobre el consumidor. Esto incita que las marcas realicen mejores estrategias basadas en aquellos resultados que obtienen a través de su uso. A continuación algunas de las empresas exitosas hacen uso del Big Data

Facebook

Facebook dirige anuncios específicos a sus usuarios en base a las montañas de datos personales que recopila sobre ellos (Redhat, s.f.).

Macy's

Es uno de los comercios minoristas más importantes de los Estados Unidos, que destaca por su e-commerce. Esta compañía ajusta el precio casi en tiempo real para 73 millones de artículos en función de la demanda y disponibilidad del artículo (Redhat, s.f.).

Netflix

La empresa norteamericana Netflix gracias a esta tecnología de análisis de datos consiguió recabar los gustos de sus usuarios y con ellos creó la serie House of Cards (Nubit, 2016).

Amazon

Esta compañia explota tan bien el BigData que llega a averiguar con gran rapidez no sólo los libros que decimos que nos gustan, sino los que realmente nos gustan (Nubit, 2016).

Zara

Esta empresa ha logrado rentabilizar el gran volumen de datos extraído de su actividad comercial gracias a un excelente uso del Big Data. Gracias a este tipo de herramientas consiguen por ejemplo, averiguar en tiempo real qué tiendas a nivel mundial necesitan reposición de stock (Nubit, 2016).

Walmart

Walmart.com ha añadido la búsqueda semántica, mejorando la compleción de las compras en línea de un 10 % a un 15 %, un incremento de miles de millones de dólares para la empresa (Redhat, s.f.).

Linkedin

Usa Big Data y científicos de datos para desarrollar un amplio espectro de productos y características, como “Gente que podrías conocer”, “Grupos que te podrían gustar”, “Trabajos que te podrían interesar”, “¿Quién ha visto tu perfil?”, y otros. Lo que le ha hecho ganar millones de usuarios (Nubit, 2016).

American Express

La entidad está analizando las operaciones que hacen sus clientes VIP con su tarjeta para ofrecerles productos relacionados con su estilo de vida. Por ejemplo, si hay una persona que gasta mucho dinero en gastronomía, puede invitarle a reservar en restaurantes concretos (Nubit, 2016).

Nike

En fabricación de equipos deportivos tenemos que mencionar a la mas popular Nike, está compañía se está convirtiendo en algo más que eso. Actualmente ya es una firma de alta tecnología y software, es una de las pioneras en el terreno de los wearables, los dispositivos que permiten añadir una capa de inteligencia a elementos que antes no lo eran, como puede ser el caso de relojes o pulsómetros (Nubit, 2016).

FUNDAMENTACIÓN LEGAL

El problema planteado se fundamenta en leyes y reglamentos señalados a continuación:

Ley 67 Del Comercio Electrónico, Firmas Electrónicas, Mensaje De Datos, de 2002

Art. 2.- Reconocimiento jurídico de los mensajes de datos:

Los mensajes de datos tendrán igual valor jurídico que los documentos escritos. Su eficacia, valoración y efectos se someterá al cumplimiento de lo establecido en esta Ley y su reglamento.

Decreto Ejecutivo No. 3496, R.O.735 Del 31 De Diciembre Del 2002**. Reglamento de la Ley de Comercio Electrónico, Firmas Electrónicas, Mensaje De Datos**

Art. 10.- Elementos de la infraestructura de firma electrónica:

La firma electrónica es aceptada bajo el principio de neutralidad tecnológica. Las disposiciones contenidas en la Ley 67 y el presente reglamento no restringen la autonomía privada para el uso de otras firmas electrónicas generadas fuera de la infraestructura de llave pública, ni afecta los pactos que acuerden las partes sobre validez y eficacia jurídica de la firma electrónica conforme a lo establecido en la ley y este reglamento.

Ley 83 De La Propiedad Intelectual, de 1998

Art. 1.- El Estado reconoce, regula y garantiza la propiedad intelectual adquirida de conformidad con la ley, las decisiones de la Comisión de la Comunidad Andina y los convenios internacionales vigentes en el Ecuador.

La propiedad intelectual comprende:

1. Los derechos de autor y derechos conexos.

2. La propiedad industrial, que abarca, entre otros elementos, los siguientes:

a. Las invenciones.

b. Los dibujos y modelos industriales.

c. Los esquemas de trazado (topografías) de circuitos integrados.

d. La información no divulgada y los secretos comerciales e industriales.

e. Las marcas de fábrica, de comercio, de servicios y los lemas comerciales.

f. Las apariencias distintivas de los negocios y establecimientos de comercio.

g. Los nombres comerciales.

h. Las indicaciones geográficas.

i. Cualquier otra creación intelectual que se destine a un uso agrícola, industrial o comercial.

3. Las obtenciones vegetales.

Artículo 10.- El derecho de autor protege también la forma de expresión mediante la cual las ideas del autor son descritas, explicadas, ilustradas o incorporadas a las obras.

No son objeto de protección:

a.) Las ideas contenidas en las obras, los procedimientos, métodos de operación o conceptos matemáticos en sí; los sistemas o el contenido ideológico o técnico de las obras científicas, ni su aprovechamiento industrial o comercial; y,

b.) Las disposiciones legales y reglamentarias, las resoluciones judiciales y los actos, acuerdos, deliberaciones y dictámenes de los organismos públicos, así como sus traducciones oficiales.

En una editorial, publicada en el sitio web http://www.derechoecuador.com/ recalca el acceso a la información mediante el uso de la tecnología, donde dice lo siguiente:

La Tecnología Informática está al alcance de todos, la actividad del ser humano se desarrolla y se desenvuelve por medios automáticos, por lo que se ha determinado una serie de ventajas y desventajas, que en lo principal pueden afectar a los derechos fundamentales de las personas, como por ejemplo lesionar la "intimidad" y la "privacidad" de los "datos" que se procesan electrónicamente. Por ello la necesidad de precautelar, "regular", proteger, controlar y sancionar los actos y hechos que afecten negativamente a los sujetos sin que estos hayan tenido conocimiento de lo que ha sucedido con los datos y mensajes proporcionados electrónicamente. Esta tarea le compete al Derecho

Informático como una nueva rama del derecho de las nuevas tecnologías de la información.

Protección de datos

La doctrina utiliza la expresión "protección de datos" en lo referente a la protección jurídica de la persona frente a la tecnología que automatiza sus datos. Pero que es lo que se protege, al respecto la mayoría de autores coinciden en los siguientes aspectos:

a) Proteger al individuo ante el "manejo o manipulación, no autorizada, de sus datos personales" que se encuentren en medios o formas electrónicas.

b) Los resultados de procesamientos informáticos, "deben ser identificable con el titular de los mismos" puesto que es muy fácil conocer características de la personalidad y de la intimidad de las personas.

c) Y, por último, el consentimiento no autorizado del uso de los datos, para fines en los que el titular no autorizo o fue obligado a darlos

En el 2008, el Presidente del Ecuador, firmó el decreto 1014 con el cual el Software Libre pasa a ser una política de Estado para ser adoptado por todas las entidades.

Art. 1.-Establecer como política pública para las entidades de la administración Publica Central la utilización del Software libre en sus sistemas y equipamientos informáticos.

Art. 2.-Se entiende por Software Libre, a los programas de computación que se pueden utilizar y distribuir sin restricción alguna, que permita su acceso a los códigos fuentes y que sus aplicaciones puedan ser mejoradas.

Estos programas de computación tienen las siguientes libertades:

a. Utilización del programa con cualquier propósito de uso común.
b. Distribución de copias sin restricción alguna.
c. Estudio y modificación del programa (Requisito: código fuente disponible).
d. Publicación del programa mejorado (Requisito: código fuente disponible).

Según la Constitución De La República Del Ecuador:

Título VII, Capítulo Primero, Sección primera (Educación).

De acuerdo a la constitución del Ecuador, Titulo VII (Régimen del Buen Vivir), Capítulo Primero (Inclusión y Equidad), Sección Primera (Educación).

Art. 350.-El sistema de educación superior tiene como finalidad la formación académica y profesional con visión científica y humanista; la investigación científica y tecnológica; la innovación, promoción, desarrollo y difusión de los saberes y las culturas; la construcción de soluciones para los problemas del país, en relación con los objetivos del régimen de desarrollo.

Título VII, Capítulo Primero, Sección octava

De acuerdo a la constitución del Ecuador, Titulo VII (Régimen del Buen Vivir), Capítulo Primero (Inclusión y Equidad), Sección octava (Ciencia, tecnología, innovación y saberes ancestrales).

Art. 387.- Será responsabilidad del Estado:

1. Facilitar e impulsar la incorporación a la sociedad del conocimiento para alcanzar los objetivos del régimen de desarrollo.

2. Promover la generación y producción de conocimiento, fomentar la investigación científica y tecnológica, y potenciar los saberes ancestrales, para así contribuir a la realización del buen vivir, al sumak kausay.

3. Asegurar la difusión y el acceso a los conocimientos científicos y tecnológicos, el usufructo de sus descubrimientos y hallazgos en el marco de lo establecido en la Constitución y la Ley.

4. Garantizar la libertad de creación e investigación en el marco del respeto a la ética, la naturaleza, el ambiente, y el rescate de los conocimientos ancestrales.

5. Reconocer la condición de investigador de acuerdo con la Ley.

Según el Reglamento de la Investigación Científica y Tecnológica de la Universidad e Guayaquil:

Título Preliminar, Disposiciones Fundamentales Objetivo de la Investigación Científica y Tecnológica

Art. 1.- Los objetivos de la investigación en la Universidad de Guayaquil están concebidos como parte de un proceso de enseñanza único, de carácter docente investigativo, orientado según norma el Estatuto Orgánico, para permitir el conocimiento de la realidad nacional y la creación de ciencia y tecnología, capaces de dar solución a los problemas del país. Las investigaciones dirigidas a la comunidad tienen por finalidad estimular las manifestaciones de la cultura popular, mejorar las condiciones intelectuales de los sectores que no han tenido acceso a la educación superior; la orientación del pueblo frente a los problemas que lo afectan; y la prestación de servicios, asesoría técnica y colaboración en los planes y Proyectos destinados a mejorar las condiciones de vida de la comunidad.

Capítulo IV, Coordinación de Investigación de las Unidades Academias

Art 14.- las unidades académicas son responsables de la labor investigativa de sus profesores (as) e investigadores (as) y trabajaran por lograr la mayor integración posible de los Proyectos de investigación a las necesidades del desarrollo científico y metodológico del pregrado y el posgrado, y a los fines de la formación integral y profesional de sus docentes y alumnos.

Según la Ley de Educación Superior:

Capítulo I

De la constitución, fines y objetivos del sistema nacional de educación Superior

Art 3.- Las instituciones del sistema nacional de educación superior ecuatoriano, en sus diferentes niveles, tienen los siguientes objetivos y estrategias fundamentales:

a) Desarrollar sus actividades de investigación científica en armonía con la legislación nacional de ciencias y tecnología y la ley de la propiedad intelectual.

Sección Novena

De la Ciencia Y Tecnología

Art 80.- El Estado fomentará la ciencia y la tecnología, especialmente en todos los niveles educativos, dirigidas a mejorar la productividad, la competitividad, el manejo sustentable de los recursos naturales, y a satisfacer las necesidades básicas de la población. La investigación científica y tecnológica se llevará a cabo en las universidades, escuelas politécnicas, institutos superiores técnicos y tecnológicos y centros de investigación científica, en coordinación con los sectores productivos cuando sea pertinente, y con el organismo público que establezca la ley, la que regulará también el estatuto del investigador científico

HIPÓTESIS A CONTESTARSE

¿Qué ha propiciado la aparición de Big Data?

¿Qué se entiende por Big Data?

¿Cuál es el objetivo de Big Data?

¿Cuáles son sus principales beneficios?

¿Existe conciencia en las compañías sobre la relevancia de contar con Big Data?

VARIABLES DE LA INVESTIGACIÓN

Variable Independiente

Estudio de la tecnología Big Data

En la presente investigación se pretende enseñar nuevas alternativas de almacenamiento debido a las ingentes cantidades de datos, en donde esta tecnología actúa almacenando y procesando todos estos datos para generar información útil y a tiempo real.

Variable Dependiente

Presentación de una guía con el prototipo para demostrar la investigación realizada.

Es una herramienta útil para conseguir mejores resultados en el aprendizaje, en el mismo que contiene información adecuada y comprensiva acerca de la aplicación de la tecnología de almacenamiento, se la ha realizado para los usuarios que posean poco conocimiento de este tema.

DEFINICIONES CONCEPTUALES

Nosql

Definimos a NoSQL como una multitud de bases de datos que intentan solventar las limitaciones que el modelo relacional se encuentra en entornos de almacenamiento masivo de datos, y concretamente en las que tiene en el momento de escalar, donde es necesario disponer de servidores muy potentes y de balanceo de carga.

Data-Warehouse

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

Interfaz Web

La interfaz web son elementos gráficos que permiten al usuario acceder a los contenidos, navegar e interactuar. Para lograr que un usuario se quede y vuelva, el diseño de la interfaz es importante.

Big Data

Big Data conocido en español como *datos masivos* lo denominamos como a la gestión y análisis de colosales volúmenes de datos que no pueden ser tratados de manera convencional, ya que superan los límites y capacidades de las herramientas de software.

Cluster

Lo podemos puntualizar como un sistema de procesamiento paralelo o distribuido. Consta de un conjunto de computadoras independientes, interconectadas entre sí, de tal manera que funcionan como un solo recurso y en donde cada uno de los elementos del *cluster* se le conoce como nodo**.**

OpenSource

Open Source, también llamado “Código Abierto” es un término que se utiliza para indicar a cierto tipo de software que se distribuye mediante una licencia que le permite al usuario final el acceso a su código de programación para estudiarlo, modificarlo y realizar mejoras en el mismo, pudiendo incluso hasta redistribuirlo.

RDBMS

Un RDBMS (Relational Database Management System) es un Sistema Gestor de Bases de Datos Relacionales. Se trata de software capaz de producir, manipular y gestionar bases de datos de tipo relacional

WEB 2.0

La Web 2.0 o Web social se refieren a una segunda generación en la historia de los sitios web. Su denominador común es que están basados en el modelo de una comunidad de usuarios. Abarca una amplia variedad de redes sociales, blogs, wikis y servicios multimedia interconectados cuyo propósito es el intercambio ágil de información entre los usuarios y la colaboración en la producción de contenidos.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

Diseño de la Investigación

Modalidad de la Investigación

El presente Proyecto: "**Análisis de la importancia del Big Data Storage sobre problemáticas reales de gran volumen de almacenamiento**" tiene como modalidad investigación de campo un 30% e investigación bibliográfica un 70%. Será de tipo bibliográfica, al llevarse a cabo por medio de fuentes bibliográficas tales como libros, revistas, videos, consultas en internet. El porcentaje de campo corresponde a las encuestas y pruebas del tema propuesto.

Investigación Bibliográfica

Se puede entender como la búsqueda de información en documentos para determinar cuál es el conocimiento existente en un área particular, un factor importante en este tipo de investigación la utilización de la biblioteca y realizar pesquisas bibliográficas. La habilidad del investigador se demostrara en la cuidadosa indagación de un tema, de la habilidad para escoger y evaluar materiales, de tomar notas claras bien documentadas y depende además de la presentación y el orden del desarrollo en consonancia con los propósitos del documento. Una idea que ayuda a entender este punto es que no debe de existir ningún investigador que inicie su trabajo, hasta que no haya explorado la literatura existente en la materia de su trabajo. Una investigación bibliográfica se busca en textos, tratados, monografías, revistas y anuarios. (Salazar, 2009)

Investigación de Campo

Este tipo de investigación es entendida como el análisis sistemático de problemas en la realidad, con el propósito bien sea de describirlos, interpretarlos, entender su naturaleza y factores constituyentes, manifestar sus causas, y efectos, o augurar su ocurrencia, haciendo uso de métodos característicos de cualquiera de los paradigmas de investigación conocidos o en desarrollo. (Perez J. , 2012). Otra característica de la investigación de campo es que los datos de interés son recogidos en forma directa de la realidad.

Tipo De Investigación

Por los objetivos

Investigación pura o básica

Este tipo de investigación básica también se conoce como investigación pura o fundamental, la misma que se apoya dentro de un contexto teórico con el propósito de desarrollar nuevas teorías la cual busca aumentar y profundizar cada vez los conocimientos ya existidos en la realidad.

En este Proyecto se estudiara la importancia que tiene en la actualidad el Big Data para el manejo de grandes cantidades de información la cual esta tecnología actúa almacenando y procesando todos estos datos para generar información útil y a tiempo real.

Por el lugar

De campo

Este tipo de investigación se apoya en informaciones que provienen de otras como son las entrevistas, los cuestionarios, las encuestas y observaciones. En el presente Proyecto se usara la encuesta, la misma que es una técnica de la investigación de campo por medio de la cual el investigador recibe de la encuestada información pertinente para los fines que persigue. Es la recolección de testimonios orales y escritos de personas vivas.

En nuestra investigación la mayor parte de información viene en donde está ubicado el problema en nuestro caso seria las web 2.0 por las ingentes cantidades de datos que hay cada minuto; los mismo que se manifiestan de diversas naturaleza. En nuestro medio el conocimiento del Big Data no es muy avanzado y para comprobar con exactitud la falta de este tema en la actualidad, se realizó una encuesta dirigida a dos cursos de la carrera de Ingeniería en Sistemas con el objetivo de medir sus conocimientos.

Por la Naturaleza

Para la toma de decisiones

Porque me permite comparar que tipo de tecnología utilizar, mediante un análisis previo, para ver cual nos beneficia. En este caso se implementara Neo4j que es una base de datos de grafos diseñada y construida para ser resistente y optimizada para grafos en vez de tabla.

Por el alcance

Histórico

A medida que transcurren los años los datos se incrementan de una manera desmesurada, donde ya la web ha pasado de gigabyte a terabyte de información por tal motivo todos estos datos distintos necesitan un tratamiento diferente.Actualmente vivimos en un mundo conectados constantemente a una gran cantidad de información, ya sea a través del ordenador, del teléfono móvil, la Tablet, etc. Tanto en la industria, como en el negocio e incluso en nuestra sociedad se maneja grandes cantidades de información en donde es necesario pensar en la tecnología Big Data la cual actúa almacenando y procesando todos estos datos para generar información útil y a tiempo real.

Por la factibilidad

Proyecto factible

En esta investigación se dará a conocer nuevas alternativas de base de datos NoSQL por su gran ventaja de no seguir el modelo relacional y ser Open Source. También se analizara las causas por las que las empresas necesitan la

tecnología Big Data, el uso de la misma, y que empresas han implementado este tipo de soluciones.

POBLACIÓN Y MUESTRA

Población

Se habla de población como el conjunto de cosas, personas, animales o situaciones que tiene una o varias características o atributos comunes.

Nuestra población estará definida por los estudiantes de la carrera de Ingeniería en Sistemas Computacionales para evaluar sus conocimientos en Big Data, con el propósito de que en un futuro exista personal calificado para trabajar con este tipo de tecnología que manejan grande volúmenes de información. La población la comprenden dos cursos de quinto semestre de la carrera de Ingeniería en Sistemas Computacionales (72 estudiantes).

Muestra

Una muestra es una parte extraída de un población, representativa de un todo, que es usada para llevarla a conocimiento público o para analizarla.

Para el presente Proyecto se considera como muestra de estudio a los alumnos de 2 cursos de 5to semestres de la carrera de Ingeniería en Sistemas Computacionales. Se consideró a los alumnos para conocer si tienen comprensión del tema.

CUADRO N. 2
POBLACIÓN

POBLACIÓN DE ESTUDIANTES	NÚMERO DE ELEMENTOS
Estudiantes de S5D	27
Estudiantes de S5I	45
TOTAL	72

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

De una población de 72 personas se selecciona una muestra de 57 personas.

EL TAMAÑO DE LA MUESTRA

Dentro de nuestro trabajo, se tomaron un total de 57 personas. Para determinar el tamaño del estrato de la muestra representativa de la población se utilizó la siguiente fórmula estadística:

$$n = \frac{P.Q.N}{(N-1)E^2/K^2 + P.Q}$$

P = Probabilidad de éxito (0.50)
Q = Probabilidad de fracaso (0.50)
N= Tamaño de la población (72)
E= error de estimación (6%)
K= # de desviac. Típicas "Z" (1: 68%, 2: 95,5%, 3: 99.7%)
n = Tamaño de la muestra (57)

$$n = \frac{0.50x0.50x72}{(72-1)0.06^2/2^2 + 0.50x0.50}$$

$$n = \frac{18}{(71)(0.0036)/4 + 0.25}$$

$$n = \frac{18}{(71)(0.0009) + 0.25}$$

$$n = \frac{18}{0.0639 + 0.25}$$

$$n = \frac{18}{0.3139}$$

$$n = 57$$

Cálculo de la fracción muestral:

$$f = \frac{n}{N} = \frac{57}{72} = 0.7916$$

Estrato	Población	Muestra
Medio	22	17
Bajo	50	40
Total	**72**	**57**

El resultado de nuestra formula nos indica que, del total de la población de 72 estudiantes de la carrera de Ingeniería en Sistemas Computacionales se debe considerar una muestra de 57 personas a las cuales se les desarrollara una encuesta.

OPERACIONALIZACIÓN DE VARIABLES

CUADRO N. 3

OPERACIONALIZACIÓN DE VARIABLES

Variables	Dimensiones	Indicadores	Tecnicas y/o Instrumentos
Variable Independiente Estudio de la tecnología Big Data Hoy en día todos los procesos relacionados con la humanidad y aquellos fenómenos que a ésta le corresponden están generando grandes volúmenes de datos de una manera cada vez más desproporcionada y sin control, que requieren un tratamiento adecuado.	Estudio desarrollado: los dos cursos de quinto semestre la carrera de Ingeniería en Sistemas Computacionales.	Los resultados de la encuesta realizada para verificar el nivel de conocimiento sobre el tema Big Data.	La investigación y la encuesta realizada.
	Conceptos , beneficios e inconvenientes de Big Data orientados a su utilización en las empresas.	Exposición de cada beneficio que presenta Big Data para las empresas.	Referencias bibliográficas Estudios e informes realizados. Blogs y foros
Variable Dependiente Presentación de una guía con el prototipo para demostrar la investigación realizada. La guía es una documentación muy útil por su información oportuna y comprensiva acerca de como se debe implementar Big Data.	Después de analizar la información sobre big data se realizo un prototipo para la demostración de la investigación realizada.	Las pruebas realizadas al prototipo una vez hecha su implementación para ver su funcionamiento.	Investigación y consulta a personas con conocimiento superior. Bibliografía de como estructurar un ambiente Big Data.
	Guía	Información del trabajo de investigación.	

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

INSTRUMENTOS DE RECOLECCIÓN DE DATOS

En toda investigación es importante la recolección de datos, de esta manera este es un paso fundamental para tener éxito en nuestros resultados. La recolección de datos es la actividad que consiste en la recopilación de información que consiste en trabajar con lo recolectado para convertirlo en conocimiento útil. Las herramientas que pueden ser utilizadas por el investigador para desarrollar sus Proyectos pueden ser las entrevistas, la encuesta, el cuestionario, la observación, etc.

La Técnica

La técnica es un conjunto de conocimientos prácticos o procedimientos para obtener el resultado deseado. En cuanto a las técnicas de investigación, se estudiarán dos formas generales: Técnica Documental y Técnica de Campo
La técnica documental permite la recopilación de información para articular las teorías que sustentan el estudio de los fenómenos y procesos. La técnica de campo permite la observación en contacto directo con el objeto de estudio, y la recolección de testimonios que permitan confrontar la teoría con la práctica (Rivas, 2012).

En el Proyecto

Los instrumentos de recolección de datos para la investigación que se está desarrollando son:
Técnicas documentales: Lectura científica y Análisis de contenido
Técnicas de campo: la encuesta y la entrevista

Lectura Científica

Para el Proyecto investigativo utilizamos la lectura científica, esta técnica es el primer paso de toda investigación. Con esta técnica analizamos y extraemos

con facilidad las ideas principales de un escrito, fue utilizada para entender el marco teórico y se utilizó los materiales de estudio como libros, revistas, diagramas, tablas, cronogramas, etc.

Análisis de contenido

También se empleó el análisis de contenido el cual nos sirvió para realizar la mayor parte de la recolección. Es una técnica para estudiar y analizar la comunicación objetiva, sistemática y cuantitativa, por tanto el análisis textual de distintas fuentes que se ha consultado, depende del entendimiento del problema.

Encuesta

También se utilizó la encuesta, que es una técnica de recogida de datos mediante la aplicación de un cuestionario a una muestra de individuos; a través de esta técnica se pudo conocer el nivel de conocimiento de los estudiantes de dos cursos de 5to semestre de la carrera Ingeniería en Sistemas Computacionales. La encuesta fue elaborada en la web mediante Google Docs, el cuestionario puede ser observado en el siguiente enlace: https://docs.google.com/forms/d/1SJ4bbThfsuLsAlP4_Fkx2GjQytOi3WiG2Cf8DD_effM

Los Instrumentos

Un instrumento es cualquier objeto que se usa como un medio o recurso, para arribar a lo que se desea conseguir. Lo que permite manipular a la técnica es el instrumento de investigación. Para el presente Proyecto se utilizó la técnica de la encuesta bajo la modalidad de cuestionario como instrumento (Rivas, 2012).

Instrumentos de Investigación

Los instrumentos de Investigación son los recursos de que puede valerse el investigador para acercarse a los problemas y extraer de ellos la información.Los instrumentos empleados para esta investigacion son muy importante porque me va premitir obtener informacion específica para cumplir con los objetivos propuestos.
De acuerdo a las tecnicas antes mencionadas , los instrumentos empleados para el Proyecto son:

El Cuestionario

El cuestionario consiste de 8 preguntas que tocaban aspectos del tema del Proyecto, siendo estas revisadas con un profesional con el proposito de autorizar el instrumento, despues se procedio a enviarlo via correo electronico a los respectivos cursos de 5to semestre para medir sus conocimientos del tema.

Las Fuentes Bibliograficas

Las fuentes bibliograficas consultadas la comprenden articulos de algun libro citado en internet y conceptos actuales escritos por profesionales.

Internet

El internet es otro instrumento de investigacion en el presente Proyecto, puesto que la mayor parte de informacion se debe a el.

Tutor

Por medio de los conocimientos del tutor se pudo establecer que bibliografias eran utiles para el Proyecto y que temas se debia investigar.

La Encuesta y el Cuestionario

La encuesta se la realizo con el objetivo de ver en que nivel estan los conocimientos de los alumno de 5to semestre de la carrera de Ingenieria en Sistemas sobre el tema Big Data, el cuestionario contiene preguntas sumamente sencillas para facilitar la contestacion.

Cuestionario de la Encuesta Realizada

A continuacion presento las preguntas que se elaboraron para la encuesta sobre Analisis de la importancia del Big Data en problematicas reales de gran volumen de almacenamiento: ver ANEXO No. 3

Validación

En la elaboración del Proyecto de investigación, uno de los aspectos que debe cuidarse es el perteneciente con la validez, buscando con ese criterio que el Proyecto tenga la calidad necesaria. La validez del instrumento de recolección de datos que se aplicó, se determinó mediante el juicio de un experto con experiencia en el tema con el objetivo de ver el nivel de conocimiento de Big Data.

Mientras tanto los instrumentos para conseguir la informacion del Proyecto , se tomo en base a dos criterios esenciales que no han de faltar para la construccion del instrumento que son: validez y confiabilidad.

Validez

Para empezar se habla de validez de un instrumento, como el hecho de que una prueba sea de tal manera concebida, elaborada y aplicada y que mida lo que se plantea medir. Con la validez se crea la revisión de la presentación del

contenido, el contraste de los indicadores con los ítems que miden las variables correspondientes.

Una de las formas de determinar la validez es a través del juicio de expertos que es el conjunto de opiniones , valoraciones y recomendaciones apoyados sobre su experiencia y conocimiento del tema. En este caso sera mediante el juicio de un experto.

La confiabilidad

Se detalla la confiabilidad de un instrumento de medición cuando permite determinar que el mismo, mide lo que se quiere medir, y aplicado varias veces, muestre el mismo resultado. Mientras más reactivos tenemos en el instrumento, mayor debe ser la confiabilidad, por lo cual existe un momento donde no importa cuántos reactivos tengas, la confiabilidad del instrumento se va a mantener estable.

Procedimiento de la Investigación

A continuación se detallan en orden los pasos que permitieron desarrollar mi tesis, desde la concepción del problema hasta la elaboración del informe del Proyecto son:

El problema:

- Planteamiento del problema
- Ubicación del problema en un contexto
- Situación conflicto nudos críticos
- Causas y consecuencias del problema
- Delimitación del problema
- Formulación del problema
- Evaluación del problema
- Objetivo general
- Objetivo especifico
- Alcances del problema

- Justificación o importancia

Marco teórico:

- Antecedentes del estudio
- Fundamentación Teórica
- Fundamentación legal
- Preguntas científicas a contestarse
- Variables de la investigación
- Definiciones conceptuales

Metodología:

- Modalidad de la Investigación
- Tipo de investigación
- Población y Muestra
- Operacionalización de variables
- Instrumentos de recolección de datos
- La técnica
- Instrumentos de investigación
- La encuesta y el cuestionario
- Validación
- Procedimiento de la Investigación
- Recolección de información
- Procesamiento y Análisis
- Criterios para la elaboración de la propuesta

Resultados conclusiones y recomendaciones:

- Resultados
- Conclusiones
- Recomendaciones

Recolección de Información

Para el procesamiento de recolección de información se utilizaron técnicas de campo como es la encuesta y las documentales empleadas fueron la lectura científica y análisis de contenido, pues son las que mejor se adaptan al entorno de la investigación y además serán agregadas con las herramientas tales como sitios web sobre Big Data, libros, revistas, blogs, y reuniones con persona experta en el tema para revisión y retroalimentación del trabajo investigativo.

En nuestro cuestionario se escogieron preguntas con el fin de conocer el nivel de compresión del tema del Proyecto. La población a evaluar fueron dos cursos de 5to semestre de la CISC, donde se selecciona una muestra de 57 personas a los cuales se les envió un cuestionario por correos vía web.

Procesamiento y Análisis

Para comenzar el análisis de los datos se puede precisar como la ciencia que examina los datos en bruto con el propósito de sacar conclusiones sobre la información obtenida.

Una vez ejecutada la recolección de información a través de los cuestionarios, comienza una etapa fundamental para toda la investigación que es el realizar el tratamiento adecuado para el procesamiento y análisis de datos obtenidos de la encuesta.

Técnicas para el Procesamiento y Análisis de Datos

La técnica utilizada fue la encuesta y el instrumento que se aplico fue el cuestionario que consta de 8 preguntas las cuales van dirigidas a 57 estudiantes de 5to semestre de la CISC. La información de las encuestas fue obtenida de forma digital (Google Docs.) y exportada directamente en una hoja de Excel.

Una vez que las encuestas fueron tomadas a los estudiantes, el resultado de cada una de las preguntas esta expresado mediante porcentajes para su fácil comprensión, una vez realizada esta operación se empezó a realizar gráficos para dar nuestra opinión oportuna de acuerdo a los resultados obtenidos.

A continuación se muestran los resultados de cada una de las preguntas del cuestionario y el objetivo de la misma:

ANÁLISIS DE LA ENCUESTA

Pregunta No 1.- ¿ha escuchado usted acerca del tema Big Data?

CUADRO N. 4

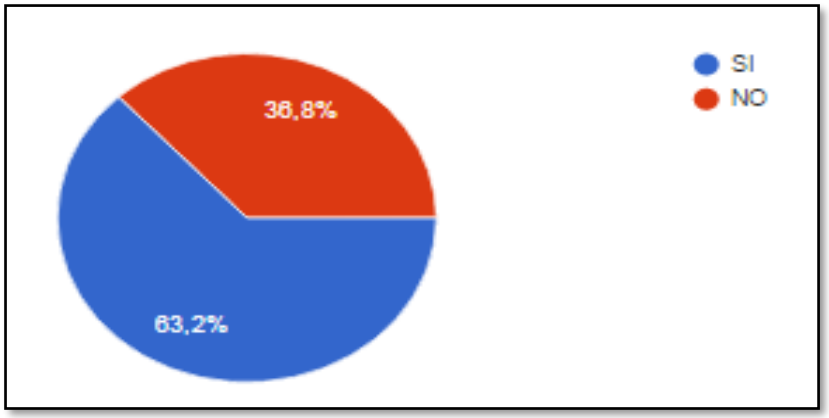
PREGUNTA 1 DE LA ENCUESTA

Alternativas	frecuencia	porcentaje
Si	36	63,20%
No	21	36,80%
TOTAL	57	100,00%

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

GRÁFICO N. 6

ESTADÍSTICA DE LA PREGUNTA 1 DE LA ENCUESTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Análisis de Resultado

Una buena parte de los encuestados han oido sobre el tema Big Data , no todos pero si gran parte de la muestra que es el 63.20% ; mientras que el 36.80 opina que no lo conocen.

Pregunta No 2-¿sabe usted cuál de estas opciones no es una herramienta para tratar con Big Data?

CUADRO N. 5

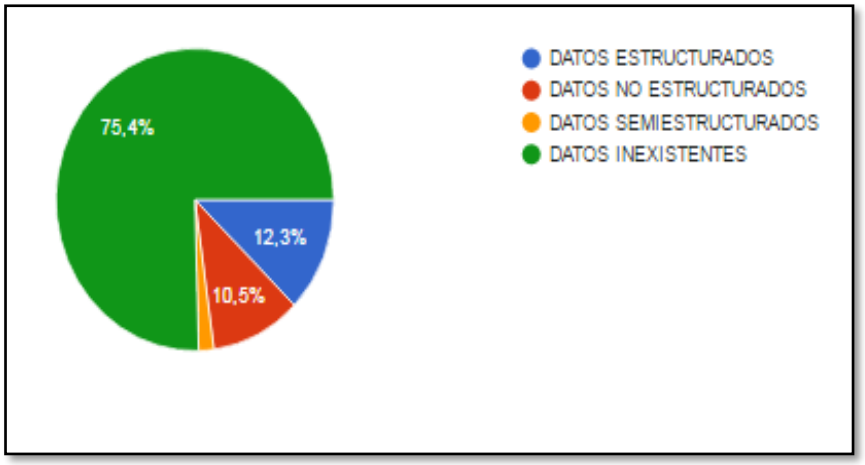
PREGUNTA 2 DE LA ENCUESTA

Alternativas	frecuencia	porcentaje
Datos estructurados	7	12,30%
Datos no estructurados	6	10,50%
Datos semiestructurados	1	1,80%
Datos inexistentes	43	75,40%
TOTAL	57	100,00%

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

GRÁFICO N. 7

ESTADÍSTICA DE LA PREGUNTA 2 DE LA ENCUESTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Análisis de Resultado

El 75.40% de los encuestados seleccionaron la opcion de “DATOS INEXISTENTES” mientras que el 1.80% seleccionaron “DATOS SEMIESTRUCTURADOS” con estos resultados podemos concluir que gran parte de la mayoría no se equivoco y se denota que tienen leves conocimientos de esta tecnología.

Pregunta No 3.- ¿sabe usted cuál es un tipo de captura de información en Big Data**?**

CUADRO N. 6

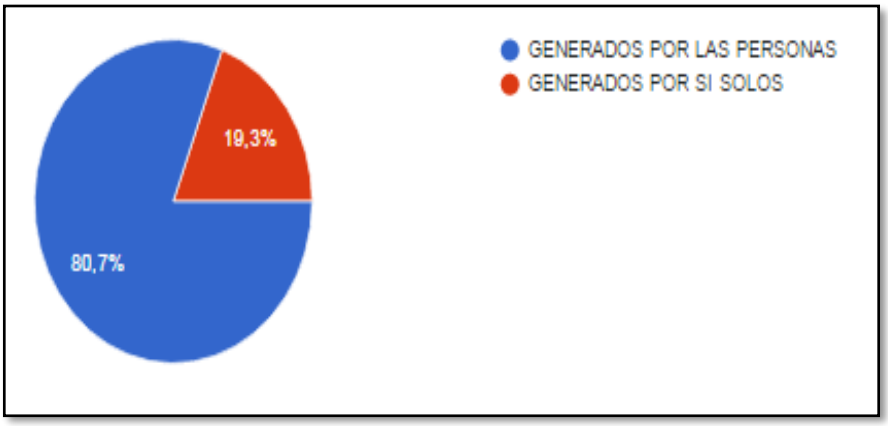
PREGUNTA 3 DE LA ENCUESTA

Alternativas	frecuencia	porcentaje
Generados por las personas	46	80,70%
Generados por si solos	11	19,30%
TOTAL	57	100,00%

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

GRÁFICO N. 8

ESTADÍSTICA DE LA PREGUNTA 3 DE LA ENCUESTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Análisis de Resultado

Según los resultados obtenidos de la muestra seleccionada que equivale a 57 estudiantes , podemos decir que un 80.70% consideran que el tipo de captura de Big Data es generado por las personas quedando un 19.30% que piensa que son generados por si solos.

Pregunta No 4.- ha escuchado sobre el volumen de los datos masivos

CUADRO N. 7

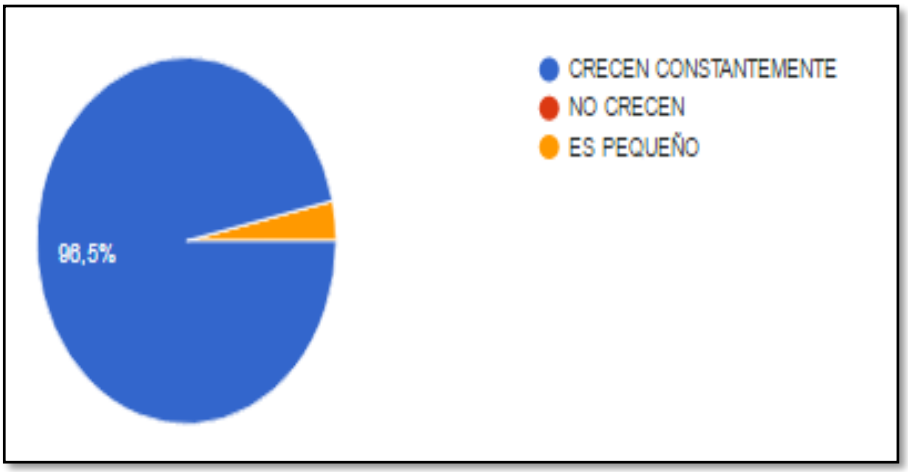
PREGUNTA 4 DE LA ENCUESTA

Alternativas	frecuencia	porcentaje
Crecen constantemente	55	96,50%
No crecen	0	0,00%
Es pequeño	2	3,50%
TOTAL	57	100,00%

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

GRÁFICO N. 9

ESTADÍSTICA DE LA PREGUNTA 4 DE LA ENCUESTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Análisis de Resultado

El 96.50% de las personas encuestadas opinan que los datos crecen constantemente mientras que solo un 3.50 % lo consideran como que es pequeño.

Pregunta No 5.- ¿sabe usted en qué ámbitos tiene utilidad Big Data?

CUADRO N. 8

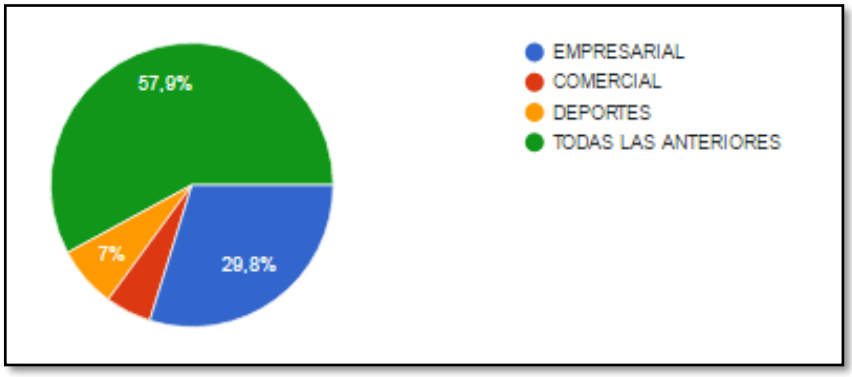
PREGUNTA 5 DE LA ENCUESTA

Alternativas	frecuencia	porcentaje
Empresarial	17	29,80%
Comercial	3	5,30%
Deportes	4	7,00%
Todas las anteriores	33	57,90%
TOTAL	57	100,00%

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

GRÁFICO N. 10

ESTADÍSTICA DE LA PREGUNTA 5 DE LA ENCUESTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Análisis de Resultado

Como podemos observar en las preguntas anteriores con respecto al conocimiento de Big Data , las personas encuestadas tienen un cierto nivel de comprensión pero un 42.10% está confundido con respecto a que ámbitos tiene utilidad.

Pregunta No 6.- ¿sabe usted como se manifiestan los datos semiestructurados?

CUADRO N. 9

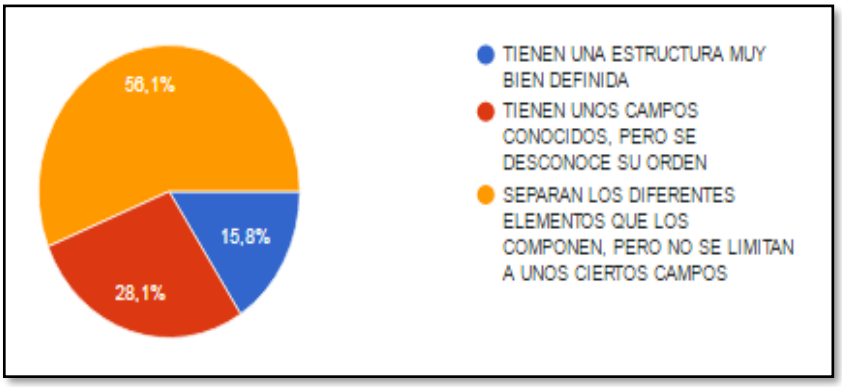
PREGUNTA 6 DE LA ENCUESTA

Alternativas	frecuencia	porcentaje
Tienen una estructura muy bien definida	9	15,80%
Tienen unos campos conocidos, pero se desconoce su orden	16	28,10%
Separan los diferentes elementos que lo componen, pero no se limitan a unos ciertos campos	32	56,10%
TOTAL	57	100,00%

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

GRÁFICO N. 11

ESTADÍSTICA DE LA PREGUNTA 6 DE LA ENCUESTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Análisis de Resultado

De 57 estudiantes encuestados sobre que piensan de los datos semiestructurados, 32 estudiantes expresan que se separan los diferentes elementos que los componen, pero no se limitan a unos ciertos campos, el 16 creen que tienen unos campos conocidos pero se desconoce su orden y por último los 9 piensan que tienen una estructura bien definida.

Pregunta No 7.- ¿sabe usted qué permite la asociación en análisis de datos?

CUADRO N. 10

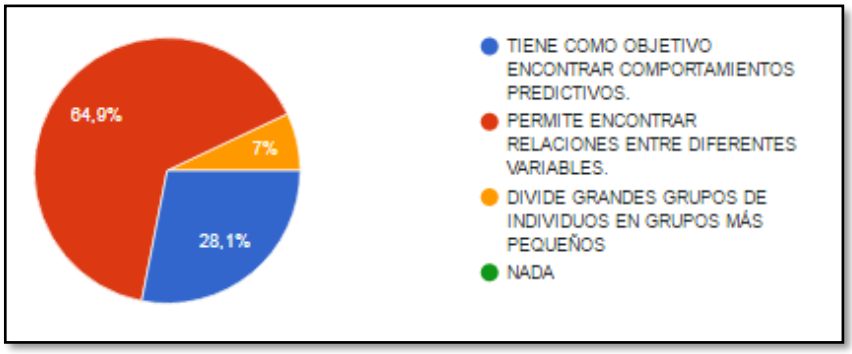
PREGUNTA 7 DE LA ENCUESTA

Alternativas	frecuencia	porcentaje
Tiene como objetivo encontrar comportamientos predictivos.	16	28,10%
Permite encontrar relaciones entre diferentes variables.	37	64,90%
Divide grandes grupos de individuos en grupos mas pequeños	4	7,00%
nada	0	0,00%
TOTAL	57	100,00%

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

GRÁFICO N. 12

ESTADÍSTICA DE LA PREGUNTA 7 DE LA ENCUESTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Análisis de Resultado

El 64.90% contesto que la asociación de análisis de datos permite encontrar relaciones entre diferentes variables. El 28.10% cree que el objetivo es encontrar comportamientos predictivos mientras que el 7% dice que se divide grandes grupos de individuos en grupos mas pequeños.

Pregunta No 8.- ¿Ha escuchado a qué se refiere la herramienta NoSql?

CUADRO N. 11

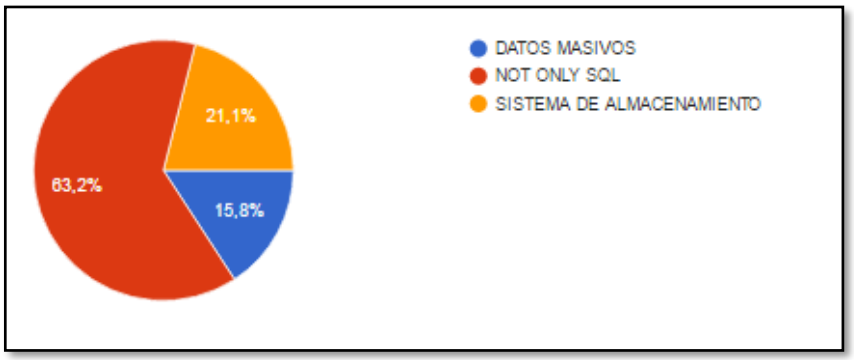
PREGUNTA 8 DE LA ENCUESTA

Alternativas	frecuencia	porcentaje
Datos masivos	9	15,79%
Not only sql	36	63,16%
Sistema de almacenamiento	12	21,05%
TOTAL	57	100,00%

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

GRÁFICO N. 13

ESTADÍSTICA DE LA PREGUNTA 8 DE LA ENCUESTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Análisis de Resultado

Según los resultados obtenidos de la muestra seleccionada que equivalen a 57 estudiantes, podemos decir que el 63.16% conocen a que se refiere NoSql, quedando un 15.79% que piensan que son datos masivos y el 21.05% que creen que es sistema de almacenamiento.

CRITERIOS PARA LA ELABORACIÓN DE LA PROPUESTA

El tema de investigación “Análisis de la importancia del Big Data Storage sobre problemáticas reales de gran volumen de almacenamiento” tiene varias partes en su contenido, pero la más importante es el paso a paso de la herramienta Neo4j.

Para la validación del tema, primero se consultara su historia así como su importancia actualmente, cuáles son sus ventajas e inconvenientes y por consiguiente que beneficios aporta a las empresas que lo empleen. Como conclusión se implementara Neo4j para ver su comportamiento con ingentes cantidades de datos desde la carga de los datos hasta su visualización en forma de grafos.

Se utilizó la tecnología Neo4j por sus múltiples ventajas entre ellas su lenguaje de consulta Cypher que está en formato legible humano que es fácil de aprender donde también su sintaxis le resultara muy familiar al MySql. Además esta tecnología permite visualizar como un mapa estas relaciones, proporcionando una capacidad de visualización poco comparable con otras soluciones.

Otra superioridad de esta tecnología es que al importar los datos que necesitemos, no habrá errores al cargar los datos de los nodos por lo que no se cancelaria la carga y la información de ellos en el fichero bad.log que se crea en la carpeta de la base datos. Podemos afirmar que la importación de datos es sumamente rápida pero su desventaja es que hay borrar la BBDD anteriores para nuevas inserciones. Cypher ofrece una gama muy avanzada de consultas por lo que su estudio debe ser muy avanzado para hacer cualquier tipo de consulta que necesitemos.

En los anexos 3 y 4 se expondrá los procedimientos a seguir para implementar Neo4j.

CAPÍTULO IV

RESULTADOS CONCLUSIONES Y RECOMENDACIONES

Al considerar los diferentes aspectos relacionados con la realización y ejecución del presente trabajo de investigación, y luego del estudio de la problemática y el marco teórico se expone el resultado obtenido y se definen las siguientes conclusiones y recomendaciones:

RESULTADOS

En este módulo se expone brevemente la tecnología escogida para emplear el ambiente Big Data. En esta guía se optó por utilizar Neo4j como Base de Datos basada en grafos que son básicamente diseñadas para modelar la información a partir de las relaciones entre las entidades, y para procesar y explorar de manera rápida y eficaz esta marea de datos usando Cypher (Lyon & Hambre, 2016).

Tras el proceso del escenario, se observan grafos que representan la información obtenida, por lo que gracias a esta herramienta se podrán realizar análisis de información procedente de los Panamá Papers que nos permita sacar conclusiones de gran interés. Al emplear esta herramienta convertimos el **Big Data** a **Big Información**, y por ultimo a **Conocimiento**.

A continuación se mostraran los resultados obtenidos de las pruebas realizadas con la plataforma Neo4j. Se procederá a ver el tiempo de importación de datos como su tiempo de respuestas a las consultas, para esto se adjuntan algunos ejemplos de consultas que se extraen de esta plataforma. Los resultados obtenidos corresponden a famosos involucrados, empresas offshore, intermediarios, beneficiarios, etc.

El primer análisis hace referencia al tiempo de importación en el que ha finalizado la carga de información en la base de datos.

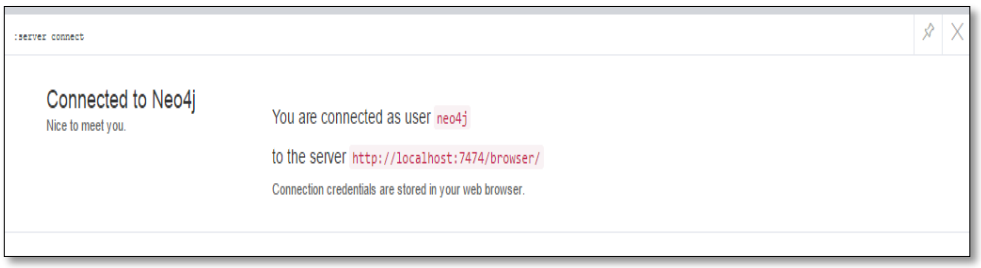
Para desarrollar esta prueba hemos tomado de la página del Consorcio Internacional de Periodistas de Investigación (ICIJ) archivos en formato CSV de empresas, personas, direcciones, intermediarios que forman parte de los Panamá Papers, que son una de las mayores filtraciones periodísticas de la historia, donde podemos encontrar a personas de todo el mundo aprovechando los paraísos fiscales para esconder su dinero del pago de impuestos en sus respectivos países (ICIJ, s.f.).

Para empezar vamos a comprobar si estamos conectados, abrimos Neo4j y hacemos la siguiente consulta en el editor:

: Server connect

A continuación podemos visualizar que estamos conectados como usuario Neo4j al servidor http://localhost:7474/browser. Conexión

ILUSTRACIÓN N. 20
ESTADO DE CONEXIÓN NEO4J



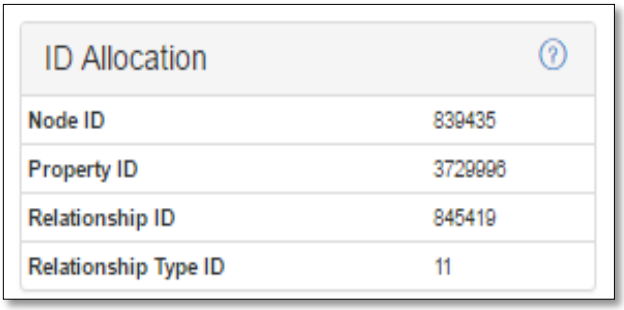
Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Ahora procedemos a ver la cantidad de nodos, propiedades, relaciones y la cantidad de relaciones creadas digitalizando el siguiente comando:

: play sysinfo

ILUSTRACIÓN N. 21

INFORMACIÓN SISTEMA NEO4J



Elaboración: Fátima Cynthia Vizueta Quinto

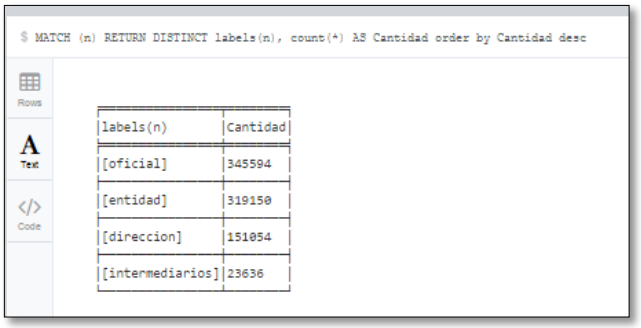
Fuente: Datos de la Investigación

Por consiguiente vamos a consultar los nodos que existen en nuestra base de datos y su cantidad importada.

```
MATCH (n)
RETURN
DISTINCT labels (n),
Count (*) AS Samp
Order by Samp desc
```

ILUSTRACIÓN N. 22

CONSULTA DE NODOS



Elaboración: Fátima Cynthia Vizueta Quinto

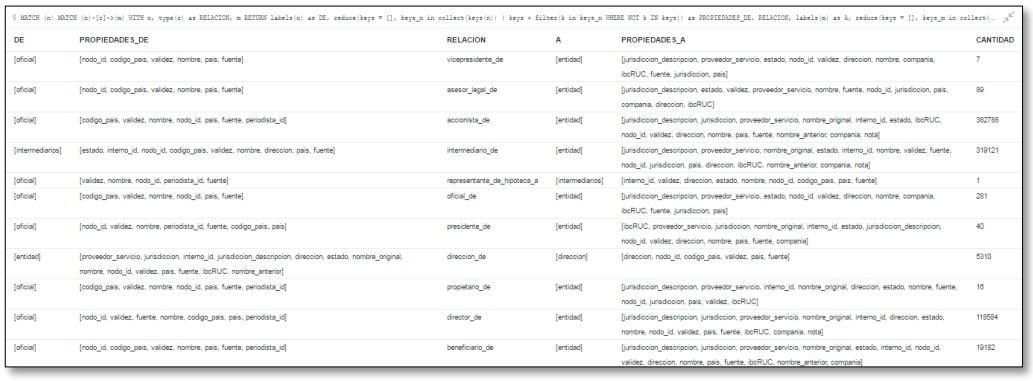
Fuente: Datos de la Investigación

Visualizamos las relaciones que se han creado con sus propiedades y la cantidad de conexiones que tiene cada una.

```
MATCH (n)
MATCH (n)-[r]->(m)
WITH n, type(r) as RELACION, m
RETURN labels(n) as DE,
reduce(keys = [], keys_n in collect(keys(n)) | keys + filter(k in keys_n
WHERE NOT k IN keys)) as PROPIEDADES_DE,
RELACION,
labels(m) as A,
reduce(keys = [], keys_m in collect(keys(m)) | keys + filter(k in keys_m
WHERE NOT k IN keys)) as PROPIEDADES_A,
count(*) as CANTIDAD
```

ILUSTRACIÓN N. 23

CONSULTA DE RELACIONES



DE	PROPIEDADES_DE	RELACION	A	PROPIEDADES_A	CANTIDAD
[oficial]	[nodo_id, codigo_pais, validez, nombre, pais, fuente]	vicepresidente_de	[entidad]	[jurisdiccion_descripcion, proveedor_servicio, estado, nodo_id, validez, direccion, nombre, compania, ibcRUC, fuente, jurisdiccion, pais]	7
[oficial]	[nodo_id, codigo_pais, validez, nombre, pais, fuente]	asesor_legal_de	[entidad]	[jurisdiccion_descripcion, estado, validez, proveedor_servicio, nombre, fuente, nodo_id, jurisdiccion, pais, compania, direccion, ibcRUC]	89
[oficial]	[codigo_pais, validez, nombre, nodo_id, pais, fuente, periodista_id]	accionista_de	[entidad]	[jurisdiccion_descripcion, jurisdiccion, proveedor_servicio, nombre_original, interno_id, estado, ibcRUC, nodo_id, validez, direccion, nombre, pais, fuente, nombre_anterior, compania, nota]	392788
[intermediarios]	[estado, interno_id, nodo_id, codigo_pais, validez, nombre, direccion, pais, fuente]	intermediario_de	[entidad]	[jurisdiccion_descripcion, proveedor_servicio, nombre_original, estado, interno_id, nombre, validez, fuente, nodo_id, jurisdiccion, pais, direccion, ibcRUC, nombre_anterior, compania, nota]	319121
[oficial]	[validez, nombre, nodo_id, periodista_id, fuente]	representante_de_hipoteca_a	[intermediarios]	[interno_id, validez, direccion, estado, nombre, nodo_id, codigo_pais, pais, fuente]	1
[oficial]	[codigo_pais, validez, nombre, nodo_id, pais, fuente]	oficial_de	[entidad]	[jurisdiccion_descripcion, proveedor_servicio, estado, nodo_id, validez, direccion, nombre, compania, ibcRUC, fuente, jurisdiccion, pais]	281
[oficial]	[nodo_id, validez, nombre, periodista_id, fuente, codigo_pais, pais]	presidente_de	[entidad]	[ibcRUC, proveedor_servicio, jurisdiccion, nombre_original, interno_id, estado, jurisdiccion_descripcion, nodo_id, validez, direccion, nombre, pais, fuente, compania]	40
[entidad]	[proveedor_servicio, jurisdiccion, interno_id, jurisdiccion_descripcion, direccion, estado, nombre_original, nombre, nodo_id, validez, pais, fuente, ibcRUC, nombre_anterior]	direccion_de	[direccion]	[direccion, nodo_id, codigo_pais, validez, pais, fuente]	5310
[oficial]	[codigo_pais, validez, nombre, nodo_id, pais, fuente, periodista_id]	propietario_de	[entidad]	[jurisdiccion_descripcion, proveedor_servicio, interno_id, nombre_original, direccion, estado, nombre, fuente, nodo_id, jurisdiccion, pais, validez, ibcRUC]	16
[oficial]	[nodo_id, validez, fuente, nombre, codigo_pais, pais, periodista_id]	director_de	[entidad]	[jurisdiccion_descripcion, jurisdiccion, proveedor_servicio, nombre_original, interno_id, direccion, estado, nombre, nodo_id, validez, pais, fuente, ibcRUC, compania, nota]	118584
[oficial]	[nodo_id, codigo_pais, validez, nombre, pais, fuente, periodista_id]	beneficiario_de	[entidad]	[jurisdiccion_descripcion, jurisdiccion, proveedor_servicio, nombre_original, estado, interno_id, nodo_id, validez, direccion, nombre, pais, fuente, ibcRUC, nombre_anterior, compania]	19182

Elaboración: Fátima Cynthia Vizueta Quinto

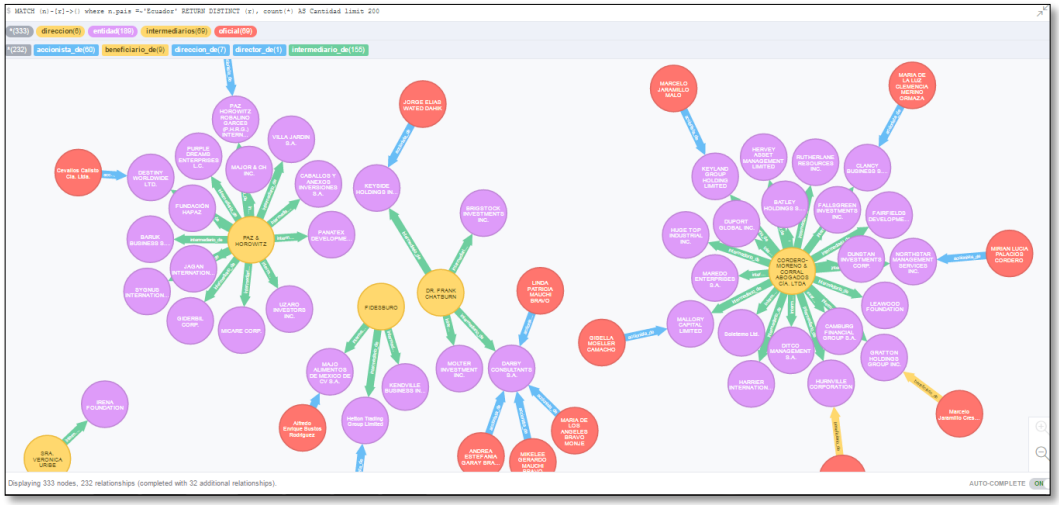
Fuente: Datos de la Investigación

A continuación se mostrara un ejemplo de obtener datos de los nodos y sus conexiones filtrando solo a Ecuador con un límite de 200 relaciones.

```
MATCH (n)-[r]->() where n.pais =~'Ecuador'
RETURN DISTINCT (r),
count(*) AS Cantidad
limit 200
```

ILUSTRACIÓN N. 24

CONSULTA DE OBTENCIÓN DE DATOS

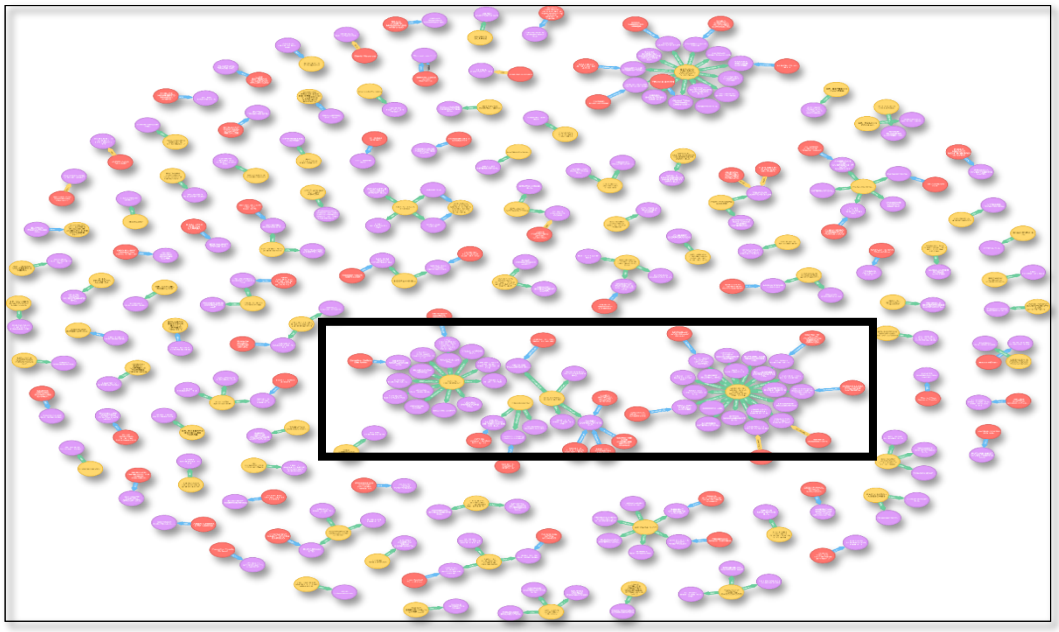


Elaboración: Fátima Cynthia Vizueta Quinto

Fuente: Datos de la Investigación

ILUSTRACIÓN N. 25

VISUALIZACIÓN COMPLETA DE LA OBTENCIÓN DE DATOS



Elaboración: Fátima Cynthia Vizueta Quinto

Fuente: Datos de la Investigación

En el anexo 2 podemos visualizar las consultas que se han empleado a Neo4j con su lenguaje Cypher, donde comprobamos que este entorno de trabajo acelera el acceso a la información y a la toma de resultados.

CONCLUSIONES

Para comenzar debo acotar que al desarrollar este Proyecto de Titulación me he dado cuenta de lo amplio que es este campo, esto es solo un manto de lo que queda por investigar, espero que los próximos estudiantes sigan esta línea de investigación y que este Proyecto les sea útil para ampliar nuevas ideas.

Con este Proyecto se trata de promover el almacenamiento de grandes datos, realizamos una introducción muy atractiva sobre Big Data mostrando sus características, ventajas y la herramienta que se va a emplear para el desarrollo de la guía sobre el ambiente Big Data.

Big Data es una de las tendencias actuales para el manejo de grandes volúmenes de datos a través del cual podemos obtener valiosas informaciones, de forma que pueda ser de ayuda para detectar problemas y plantear una toma de decisiones y la más importante de predecir futuras situaciones.

Debido al gran volumen de datos que se manejan en el Big Data, estos no pueden ser procesados por la base de datos tradicionales; en comparación con las bases de datos relacionales, los sistemas NoSQL son más escalables y ofrecen un mayor rendimiento; de igual manera, su modelo de datos aborda varios requerimientos para los que el modelo relacional no está diseñado; entre ellos gestión de grandes volúmenes de datos estructurados, semi-estructurados y no estructurados. De eso trata el Big Data de afronta la labor de almacenar, clasificar, analizar y compartir ese cúmulo pesado de información, es decir, lidiar con las denominadas “5 V´s” del Big Data: Volumen, Velocidad, Variedad, Veracidad, Valor.

Por otra parte en la implementación practica se ha desarrollado un ambiente Big Data con la tecnología Neo4j perteneciente a la base de datos NoSql orientado a

grafo por sus múltiples ventajas entre ellas su lenguaje de consulta Cypher que está en formato legible humano que es fácil de aprender. Además esta tecnología permite visualizar como un mapa estas relaciones, proporcionando una capacidad de visualización poco comparable con otras soluciones, dibuja las relaciones entre personas y cosas.

Posteriormente, se ha considerado los aspectos tecnológicos y metodológicos necesarios para llevar a cabo el desarrollo de este Proyecto completando así la investigación teórica, y contando con el tutor con los conocimientos necesarios para desarrollar una solución al problema planteado.

Después del estudio de esta tecnología podemos concluir que es de gran utilidad en la actualidad, por lo que cada vez hay más datos y se necesita extraer el valor de ellos.

Para finalizar puedo manifestar que ha sido una magnifica experiencia explorar en un terreno desconocido para mí y poco documentado ; donde este Proyecto ya materializado ha alcanzado los objetivos generales tanto como los específicos planteados en este trabajo de investigación.

RECOMENDACIONES

Cada vez son más las empresas que sin importar su tamaño están implementando Big Data. A continuación se detallan algunas recomendaciones:

Las empresas u organizaciones deben tomar el tiempo necesario para detectar en que departamentos de las empresas no bastan los medios de análisis tradicionales y para ello es necesario implantar métodos más sofisticados como el Big Data.

Es importante que las compañías tengan bien claros los objetivos de negocio, fijar metas concretas como mejorar la calidad del servicio, como crear más ingresos; es decir cuestionarse qué tipo de pregunta es la que se quiere resolver;

ya que las organizaciones que han hecho uso del Big Data han aumentado sus ingresos. Un reto muy importante para las empresas es identificar la herramienta que se va a utilizar por lo que muchas de estas soluciones de almacenamiento son de código libre y se pueden manipular para ajustarla a nuestros requerimientos, además de que existen un amplio soporte por la comunidad open source.

Además que visualizar e interpretar datos puede llegar a ser complicado, motivo por el cual se recomienda requerir a personas expertas que sepan del tema para entender dónde y cómo puede ser aplicado el Big Data. Las fuentes de datos y las tecnologías de Big Data están en constante progreso por lo cual las empresas deben aprender a estar alerta y reaccionar con rapidez para aprovechar las oportunidades que ofrecen las nuevas tecnologías.

Por último, cada cierto mes la herramienta que se utiliza para el desarrollo de este Proyecto aumenta su versión añadiendo nuevas funcionalidades, por lo que se recomienda estar al tanto de las nuevas versiones para realizar prácticas sobre ellas.

ANEXOS

ANEXO 1

GUÍA PARA LA CREACIÓN DE UN AMBIENTE BIG DATA

Instalación de neo4j

En este apartado se puntualizan los pasos necesarios para la creación de un ambiente Big Data, utilizando la herramienta Neo4j la cual se escogió por su interesante manejo que se le dan a los datos. Esta tecnología facilita una visión de grafos que nos permite visualizar los que podría ser el contenido de una base de datos entidad relación y recorrerla en forma de grafos.

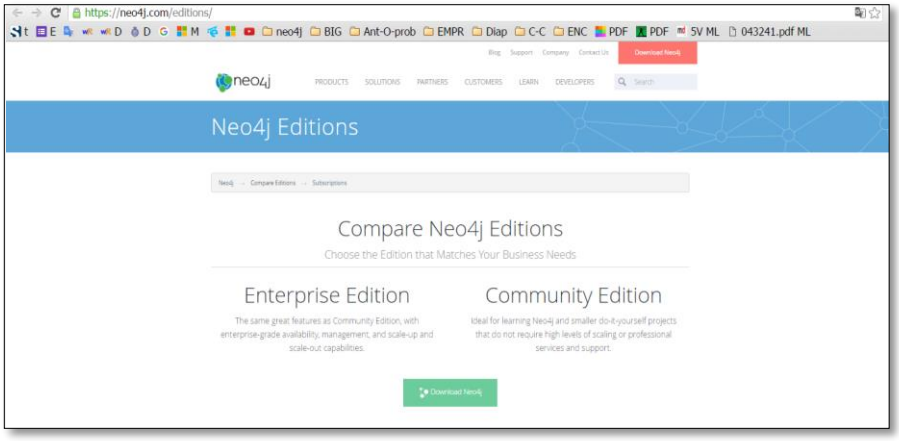
La instalación que voy a explicar es sobre Windows. Lo primero que tenemos que hacer es ir a la página de Neo4j, y pinchar en el enlace de descarga de la versión community, hay varias versiones, pero esta es la gratuita con licencia Pública General de GNU (GPL) (Sanchez J. , 2014).

La versión actualmente de Neo4j es 3.04 que fue publicada el 29/07/2016.

Pasos De Instalación

Ingresar a la página https://neo4j.com/download/

ILUSTRACIÓN N. 26
DIRECCION PAGINA NEO4J



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Descargar la opción Community Edition

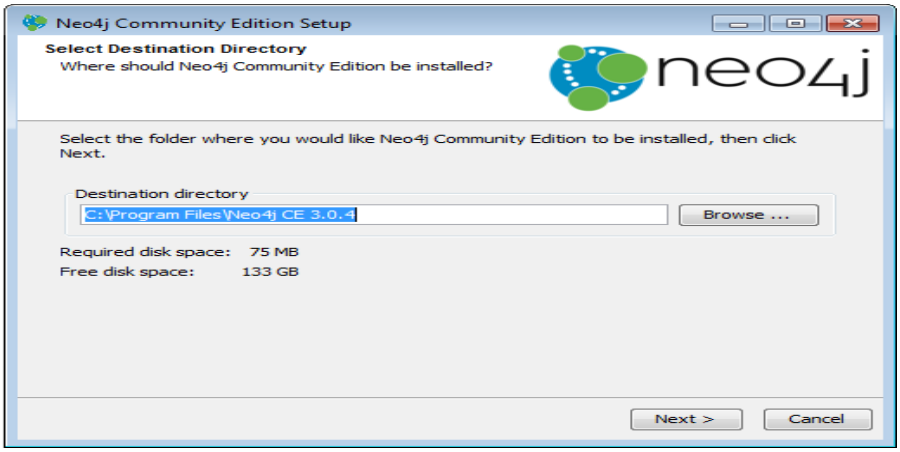
ILUSTRACIÓN N. 27
DESCARGA NEO4J COMMUNITY

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Ejecutar el instalador de Neo4j

Una vez que se ejecuta el instalador Neo4j, se muestra una ventana con la opción de escoger el directorio destino donde se instalara el software.

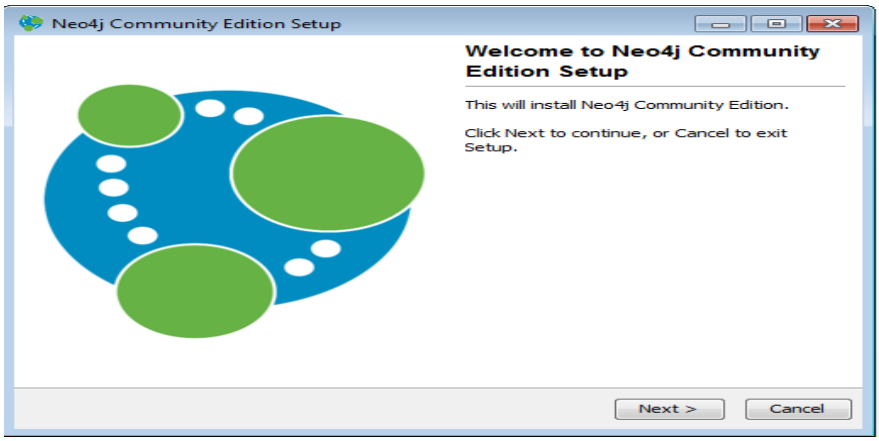
ILUSTRACIÓN N. 28
DIRECTORIO DE NEO4J



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Seleccionamos el botón siguiente y se abrirá una ventana que se muestra a continuación:

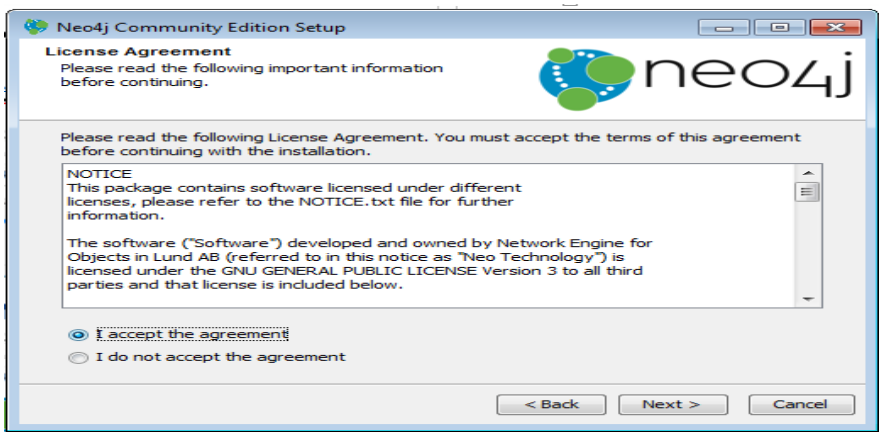
ILUSTRACIÓN N. 29
PAGINA DE BIENVENIDA DE NEO4J



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Una vez ejecutado el asistente para instalación se acepta el acuerdo de licencia, y damos clic en el botón siguiente.

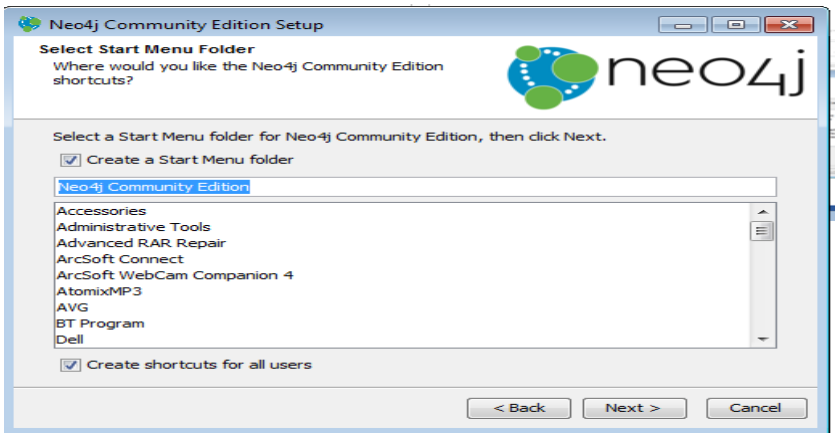
ILUSTRACIÓN N. 30
ACEPTA LICENCIA DE NEO4J



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

La ventana siguiente nos indicara que debemos activar la opción para crear una carpeta con el nombre del instalador y es recomendable dejar el mismo nombre por defecto.

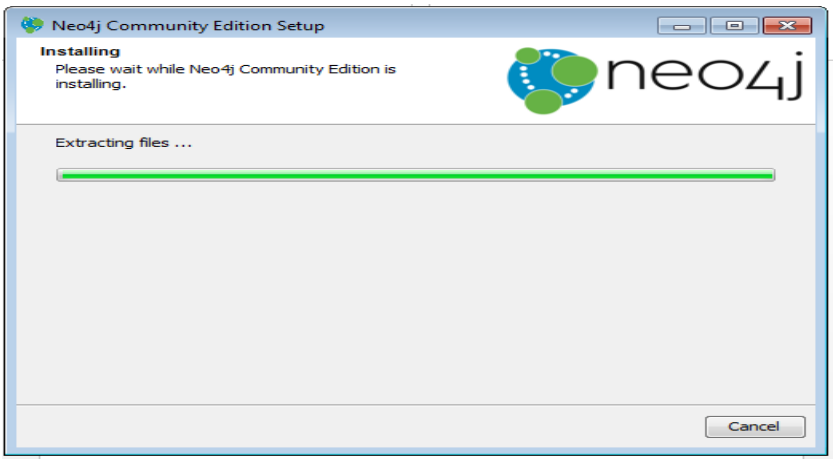
ILUSTRACIÓN N. 31
CREACION DE CARPETA DE INICIO DE MENU



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Una vez más damos clic en el botón siguiente y seguido esto se mostrara otra ventana mostrando la extracción de los archivos que compone el instalador de neo4j.

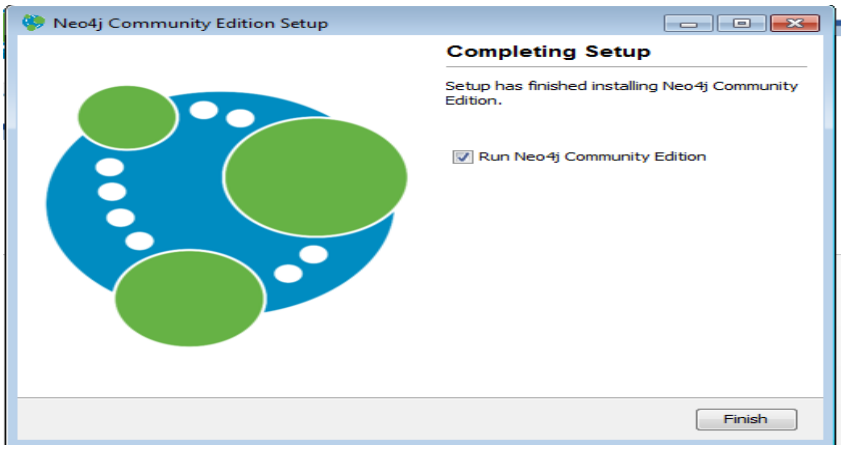
ILUSTRACIÓN N. 32
EXTRACCION DE CARPETA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

En el momento que se completa la extracción del software se procede a dar clic en el botón cancelar y automáticamente se generara otra ventana que indicara que se ha completado exitosamente la instalación.

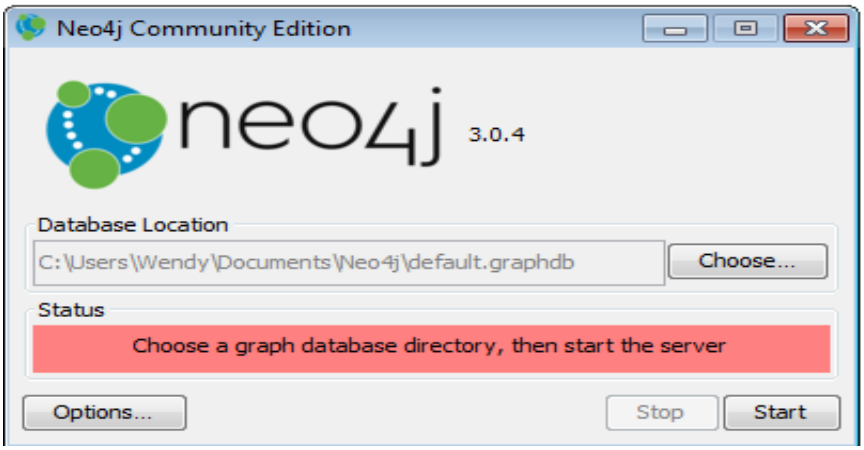
ILUSTRACIÓN N. 33
CONFIGURACION COMPLETA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

A continuación de haber completado la instalación se genera la siguiente ventana en donde choose permitirá escoger la ruta en donde se guardaran los archivos de la base de datos y escogemos el botón start para iniciar el servicio de neo4j.

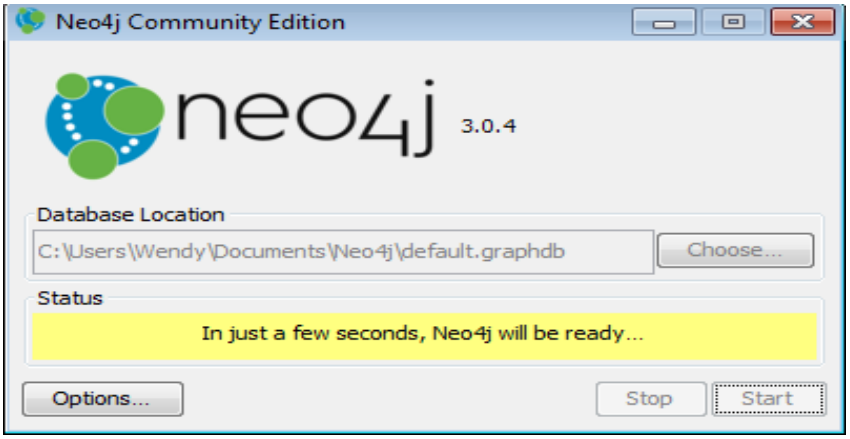
ILUSTRACIÓN N. 34
DIRECCIÓN DE LA BASE DE DATOS



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Al iniciar este ejecutable nos aparece una ventana donde podemos ver la ruta de la base de datos, que es C: \Users\wendy\Documents\Neo4j\default.graphdb, y seguidamente aparecerá una ventana que muestra que se está cargando el servicio.

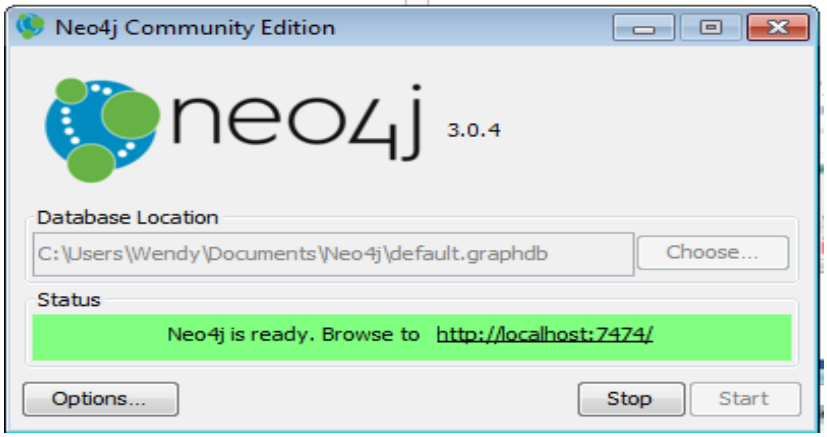
ILUSTRACIÓN N. 35
PANTALLA DE INICIO DE SESION EN CARGA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

En la ventana siguiente aparecerá el link al browser que por defecto es **http://localhost:7474/.**

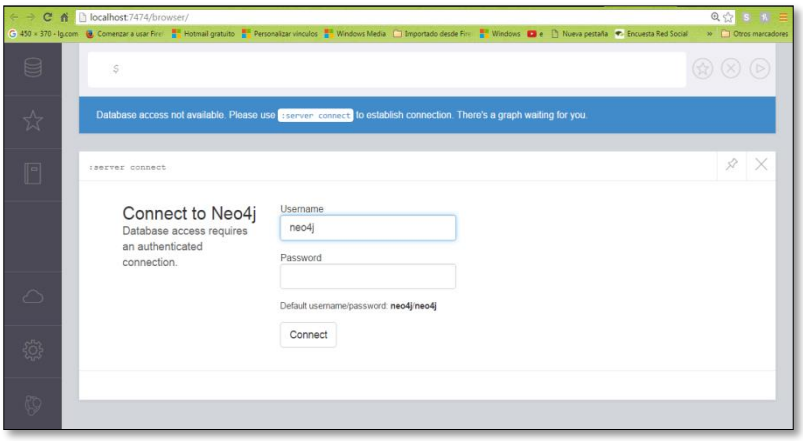
ILUSTRACIÓN N. 36
PANTALLA DE INICIO DE SESION



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Una vez seleccionado el link del servicio neo4j se abrirá el browser. Ingresamos neo4j como password y damos clic en Start.

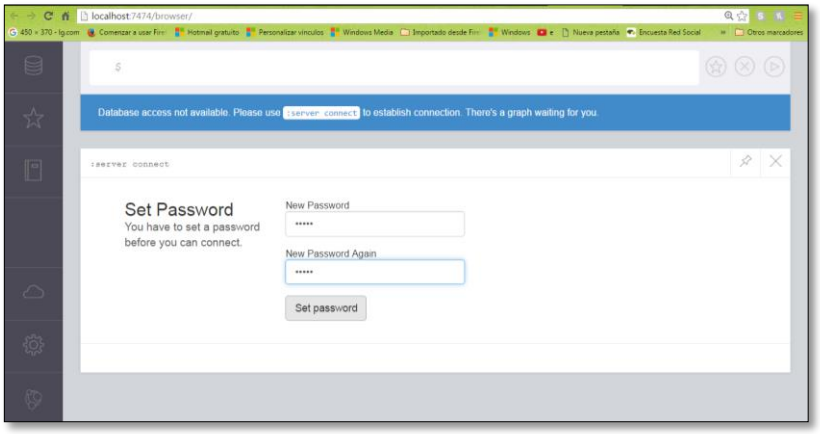
ILUSTRACIÓN N. 37
PANTALLA DE INICIO DE BROWSER



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Registramos una nueva clave y confirmamos la clave.

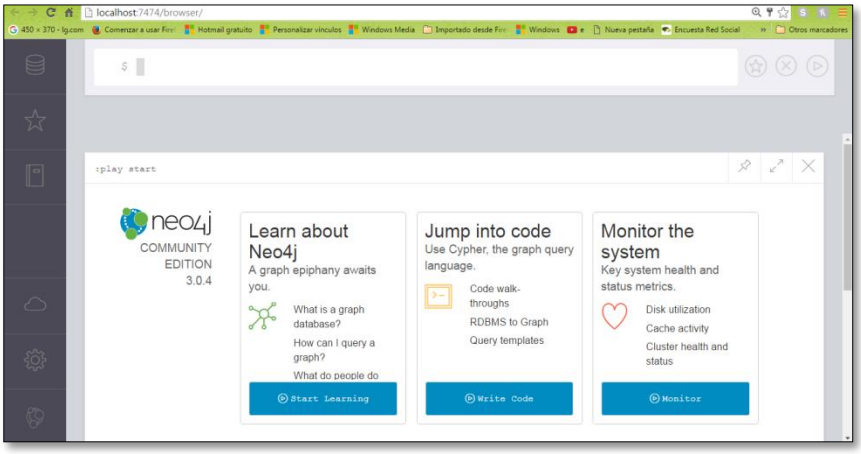
ILUSTRACIÓN N. 38
PANTALLA DE REGISTRO DE CLAVE



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Una vez registrada la clave, se generara automáticamente la siguiente ventana. En la parte se puede escribir las consultas Cypher.

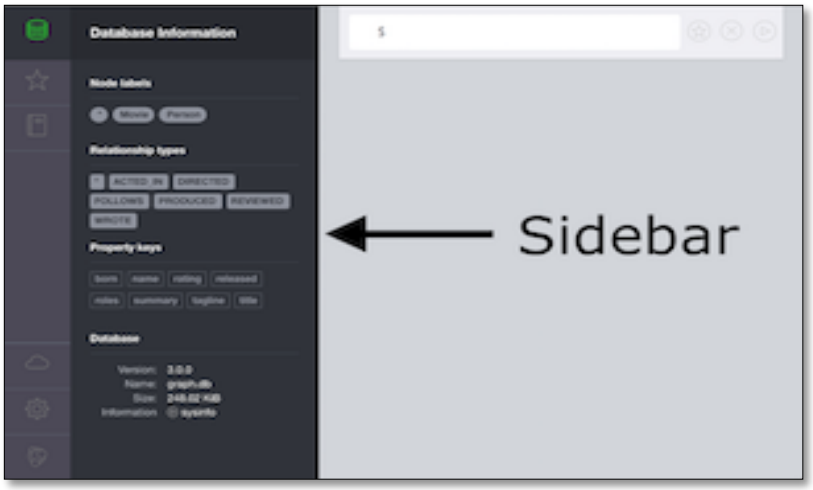
ILUSTRACIÓN N. 39
PANTALLA DE CONSULTA



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

También tiene un Sidebar donde hay 6 menús. La barra lateral se expande para mostrar diferentes paneles prácticos para consultas e información comunes. En el menú base de datos podemos ver los tipos de nodos que tenemos, las relaciones y las propiedades. Al hacer clic en cada uno de estos elementos que hemos creado se genera un Query de cónsula y un frame de resultado.

ILUSTRACIÓN N. 40
PANTALLA DE INFORMACION DE SIDEBAR



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

ANEXO 2

GUÍA PARA LA IMPLEMENTACIÓN

A continuación, se presentara la parte práctica correspondiente al trabajo de fin de grado. Para realizar esta prueba hemos tomado de la página del Consorcio Internacional de Periodistas de Investigación cinco archivos en formato CSV de empresas, personas, direcciones, intermediarios que fueron mencionados en los documentos de los Panamá Papers por su evasión de impuestos, fraudes, lavado de dinero, corrupción, etc.

A partir de estos ficheros hemos creados grafos para facilitar la compresión de las interacciones entre sujetos distintos en esta red, el punto es analizar estos datos con Neo4j.

Cuya información es la siguiente:

Oficial	:	*345.594 nodos*
Intermediario	:	23.636 *nodos*
Entidad	:	*319.150 nodos*
Dirección	:	*151.054 nodos*
tipo_rel		*: 1.048.576 nodos*

Características De La Maquina

Tipo S.O.	: Windows 7 Home Premium
S.O.	: 64 bits
Memoria RAM	: 4 GB
Procesador	: Intel ® Core ™ 2.40 GHz

Ingesta De Datos

Al momento de crear un nodo o una relación, el servidor de la base de datos de Neo4j asignará un número interno “Id”. Procederemos a evaluar la complejidad de los datos, el método de ingesta es por ETL. Neo4j-Import; Load Create.

Los nodos elegidos para realizar esta prueba han sido:

ILUSTRACIÓN N. 41
ARCHIVOS EN FORMATO CSV

all_edges	21/06/2016 13:24	Archivo de valores separados por comas...	41.821 KB
direccion	21/06/2016 13:24	Archivo de valores separados por comas...	25.509 KB
entidad	21/06/2016 13:24	Archivo de valores separados por comas...	104.016 KB
Intermediarios	21/06/2016 13:24	Archivo de valores separados por comas...	3.547 KB
oficial	21/06/2016 13:24	Archivo de valores separados por comas...	41.888 KB

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Estructura

Los datos de cada línea del archivo son csv (comma separeted value). La estructura de los ficheros a importar son:

Entidad

Una empresa, fideicomiso o fondo creado en una baja de impuestos.

Contiene: *319.150 nodos*

*Atributos:*Nombre,Nombre_original,Nombre_antiguo,Jurisdicción,Jurisdicción_descripcion,Compañía,Dirección,Interno_id,Fecha_incorporacion,Fecha_inactiva,Fecha_baja,Fecha_residencia,Estado,Proveedor_servicio,IbcRUC,Código_país,País,Nota,Validez,Nodo_id,Fuente

Oficial

Una persona o empresa que juega un papel en una entidad.

Contiene: 345.594 nodos

Atributos: Nombre, Periodista_id, Validez, Código_país, País, Nodo_id, Fuente

Intermediario

Un intermediario para alguien que busca una sociedad offshore y un proveedor de servicios offshore; por lo general es un despacho de abogados o un

intermediario que pide a un proveedor de servicios para crear una firma para un cliente.

Contiene: 23.636 nodos

Atributos: Nombre, Interno_id, Dirección, Validez, Código_país, País, Estado, Nodo_id, Fuente

Dirección

Dirección que figura en la base de datos originales obtenidos por ICIJ.

Contiene: 151.054 nodos

Atributos: Dirección, Periodista_id, Validez, Código_país, País, Nodo_id, Fuente

Neo4j-import

La herramienta Neo4j nos permite importar ficheros CSV con datos grandes a Neo4j. Se ha realizado una prueba con 5 nodos para poder estudiar mejor el procedimiento. Una de las desventajas que se ha encontrado es que la inserción debe ser base de datos inexistente, apuntando a que no se puede insertar si primero no se borra los trabajos anteriores. Por consiguiente no permite la actualización de nodos, ya que es necesario un borrado previo de la base de datos para volver a insertar en esta misma.

Carga masiva de datos

Lo primero que debemos hacer para construir nuestro modelo graficó es cargar todos los datos, donde estos mismos se convierten en nodos en el gráfico.

Los ficheros con los datos se encuentran en las siguientes direcciones:

C:\Users\Owner\Documents\Neo4j\ProyectoTitulación.graphdb\import\oficial.csv
C:\Users\Owner\Documents\Neo4j\ProyectoTitulación.graphdb\import\entidad.csv
C:\Users\Owner\Documents\Neo4j\ProyectoTitulación.graphdb\import\intermediario.csv
C:\Users\Owner\Documents\Neo4j\ProyectoTitulación.graphdb\import\direccion.csv

En el navegador ejecutamos los comandos CQL y así mismo visualizamos su salida. Ejecutar los comandos donde se visualice el indicador de dólares: “$” y después clic en “play” para ejecutar sus órdenes.

ILUSTRACIÓN N. 42
VISUALIZACIÓN DE BUTTON EXECUTE

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

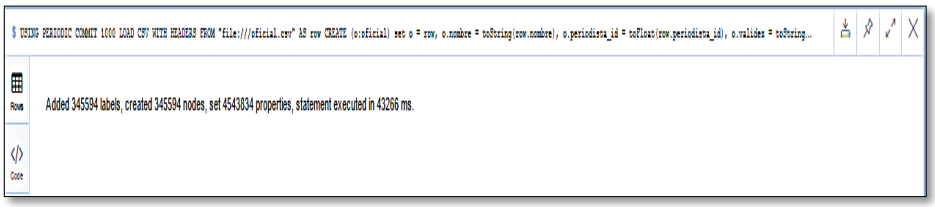
Recordemos que solo se permite la carga de un fichero CSV en cada ejecución.

Nodo: oficial

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM "file:///oficial.csv" AS row
CREATE (o: oficial)
Set o = row,
o.nombre = toString (row.nombre),
o.periodista_id = toFloat (row.periodista_id),
o.validez = toString (row.validez),
o.codigo_pais = toString (row.codigo_pais),
o.pais = toString (row.pais),
o.nodo_id = toFloat (row.nodo_id),
o.fuente = toString (row.fuente)
```

La primera línea nos indica que se efectué un envió periódico de los datos para ir creando los nodos. La segunda línea LOAD CSV WITH HEADERS nos indica que se carga el csv con cabecera que se encuentra en el directorio donde cada línea se llamara “row”. Con estas instrucciones ya tenemos el csv cargado, pero falta que aquellas líneas se conviertan en nodos. La instrucción CREATE crea el nodo, además se añaden las propiedades que se le indique. Cada propiedad es asociada a una parte de la línea procedente del CSV.
La salida que se obtiene en el explorador cuando la inserción ha funcionado correctamente será:

ILUSTRACIÓN N. 43
CREACIÓN DE NODO OFICIAL



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

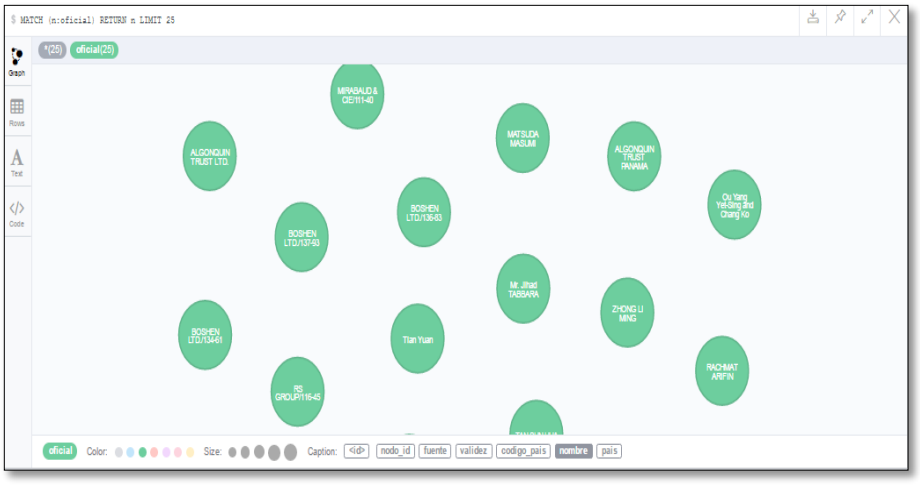
Neo4j nos está indicando que ha creado 345.594 nodos con 345.594 etiquetas y se ejecutó en 43266 milisegundos.

Después procederemos a ejecutar la siguiente consulta para cargar los datos, en la cual podemos comprobar que los nodos se han creado, indicando un límite de 25 nodos.

```
MATCH (n: oficial)
RETURN n LIMIT 25
```

El resultado podemos visualizarlo como grafo.

ILUSTRACIÓN N. 44
NODO OFICIAL VISUALIZADO EN GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

O en forma tabular:

ILUSTRACIÓN N. 45
NODO OFICIAL VISUALIZADO EN FORMA TABULAR

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Procederemos a efectuar lo mismo con los otros archivos CSV:

Nodo: entidad

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM "file:///entidad.csv" AS row
CREATE (e: entidad)
Set e = row,
e.nombre = toString (row.nombre),
e.nombre_original = toString (row.nombre_original),
e.nombre_anterior = toString (row.nombre_anterior),
e.jurisdiccion = toString (row.jurisdiccion),
e.jurisdiccion_descripcion = toString (row.jurisdiccion_descripcion),
e.compania = toString (row.compania),
e.direccion = toString (row.direccion),
e.interno_id = toString (row.interno_id),
e.fecha_incorporacion = toFloat (row.fecha_incorporacion),
e.fecha_inactivacion = toFloat (row.fecha_inactivacion),
e.fecha_baja = toFloat (row.fecha_baja),
e.fecha_residencia = toFloat (row.fecha_residencia),
e.estado = toString (row.estado),
e.proveedor_servicio = toString (row.proveedor_servicio),
e.ibcRUC = toFloat (row.ibcRUC),
e.codigo_pais = toInt (row.codigo_pais),
e.pais = toString (row.pais),
e.nota = toString (row.nota),
e.validez = toString (row.validez),
```

```
e.nodo_id = toFloat (row.nodo_id),
e.fuente = toString (row.fuente)
```

La salida al ejecutar será:

ILUSTRACIÓN N. 46
CREACIÓN DE NODO ENTIDAD

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Ejecutamos la siguiente consulta para visualizar los datos:

```
MATCH (n: entidad)
RETURN n LIMIT 25
```

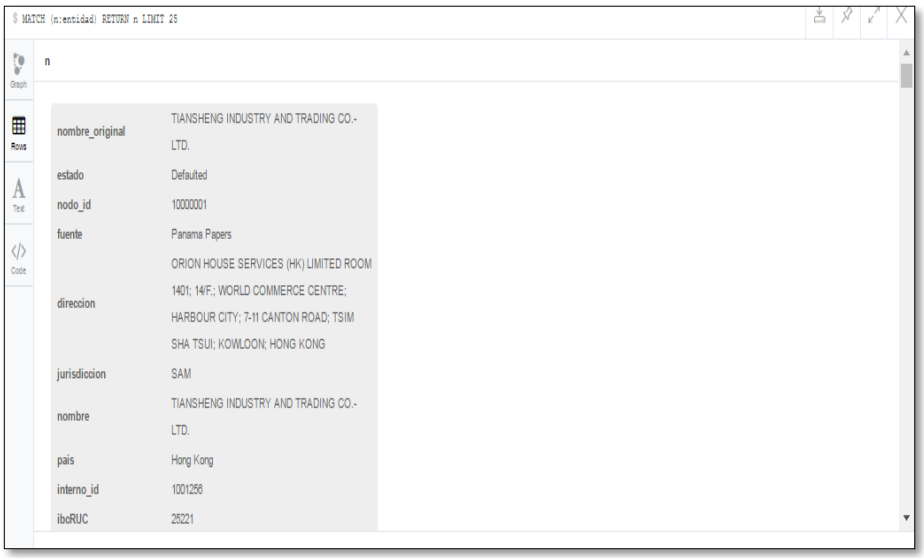
El resultado como grafo:

ILUSTRACIÓN N. 47
NODO ENTIDAD VISUALIZADO EN GRAFO

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

O en forma tabular:

ILUSTRACIÓN N. 48
NODO ENTIDAD VISUALIZADO EN FORMA TABULAR

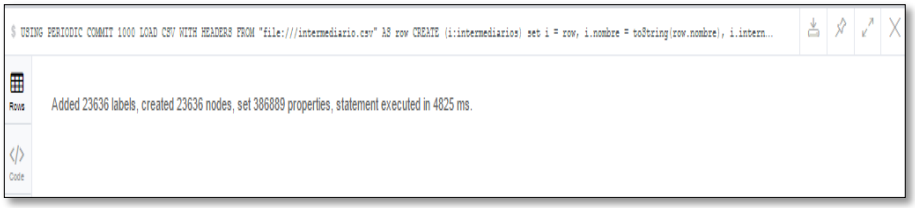


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Nodo: intermediaries

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM "file:///intermediario.csv" AS row
CREATE (i: intermediarios)
Set i = row,
i.nombre = toString (row.nombre),
i.interno_id = toFloat (row.interno_id),
i.direccion = toString (row.direccion),
i.validez = toString (row.validez),
i.codigo_pais = toString (row.codigo_pais),
i.pais = toString (row.pais),
i.estado = toString (row.estado),
i.nodo_id = toFloat (row.nodo_id),
i.fuente = toString (row.fuente)
```

ILUSTRACIÓN N. 49
CREACIÓN DE NODO INTERMEDIARIO

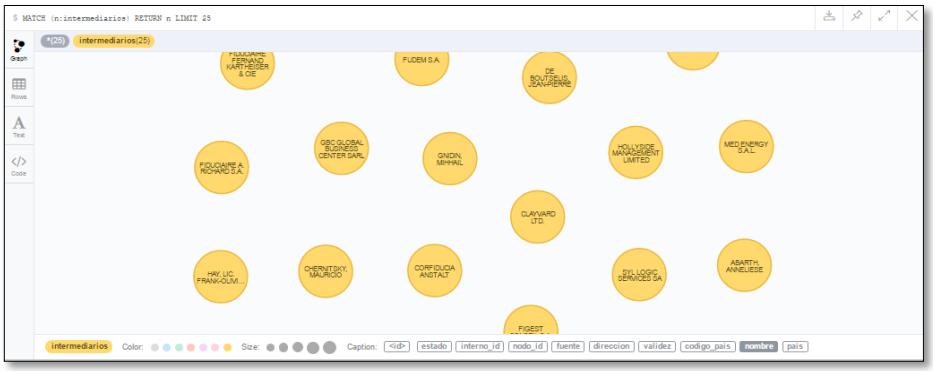


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Ejecutamos la siguiente consulta para visualizar los datos:

```
MATCH (n: intermediarios)
RETURN n LIMIT 25
```

ILUSTRACIÓN N. 50
NODO INTERMEDIARIO VISUALIZADO EN GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

En forma tabular:

ILUSTRACIÓN N. 51
NODO INTERMEDIARIO VISUALIZADO EN FORMA TABULAR

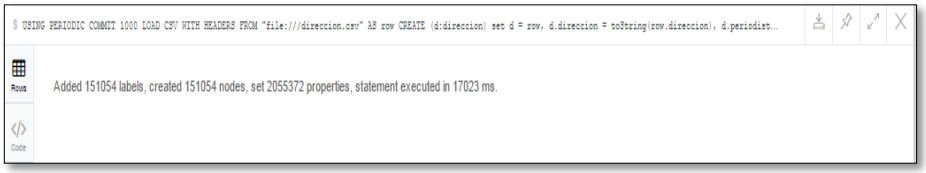
Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Nodo: direccion

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM "file:///direccion.csv" AS row
CREATE (d: direccion) Set d = row,
d.direccion = toString (row.direccion),
d.periodista_id = toFloat (row.periodista_id),
d.validez = toString (row.validez),
d.codigo_pais = toString (row.codigo_pais),
d.pais = toString (row.pais),
d.nodo_id = toFloat (row.nodo_id),
d.fuente = toString (row.fuente)
```

La salida al ejecutar será:

ILUSTRACIÓN N. 52
CREACIÓN DE NODO DIRECCIÓN



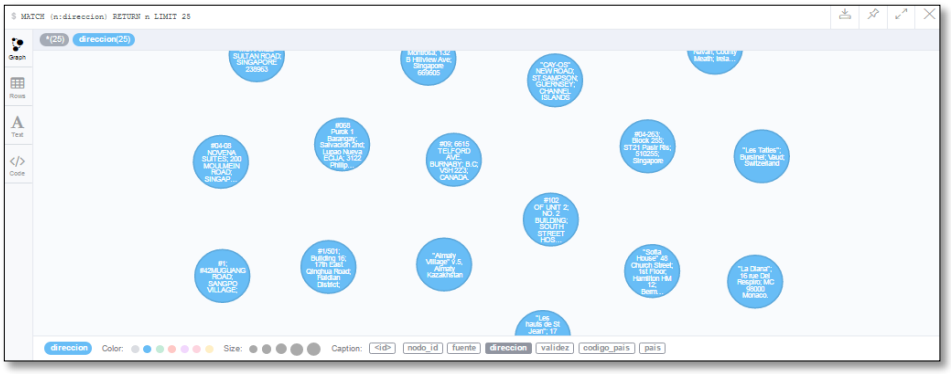
Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Ejecutamos la siguiente consulta para visualizar los datos:

```
MATCH (n: direccion)
RETURN n LIMIT 25
```

El resultado visualizarlo como grafo:

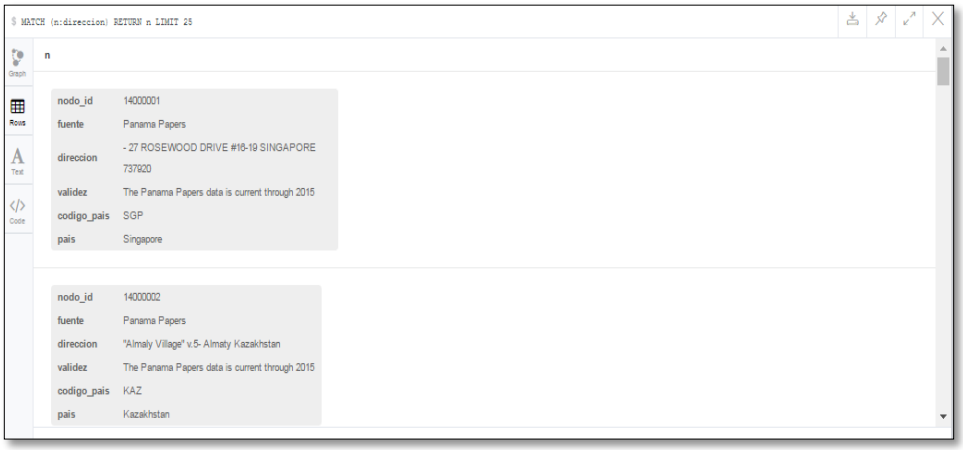
ILUSTRACIÓN N. 53
NODO DIRECCIÓN VISUALIZADO EN GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

O en forma tabular:

ILUSTRACIÓN N. 54
NODO DIRECCIÓN VISUALIZADO EN FORMA TABULAR



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Creación de índices

La creación de estos índices permite a la aplicación de la base de datos encontrar más rápidamente el dato. Neo4j no permite la crear índices antes de la inserción. Para crear los índices miremos el siguiente modelo:
CREATE INDEX ON: Nombre_de_Etiqueta (clave_de_la_propiedad).

```
CREATE INDEX ON: direccion (nodo_id)
CREATE INDEX ON: intermediarios (nodo_id)
CREATE INDEX ON: entidad (nodo_id)
CREATE INDEX ON: oficial (nodo_id)
```

Creación de relaciones

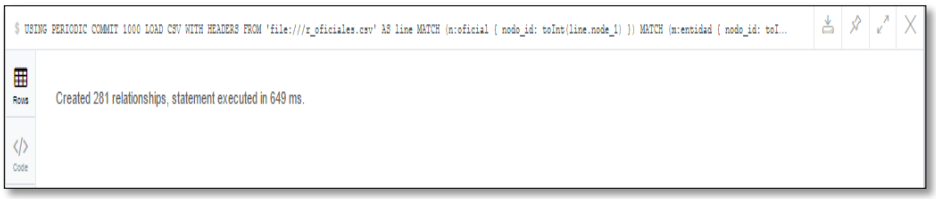
Las relaciones se derivan de los nodos creados. Por ejemplo creamos la relación ***oficial_de***, modelando el nodo oficial relacionado directo con el nodo entidad.

Relación: oficial_de

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM 'file: ///r_oficiales.csv' AS line
MATCH (n: oficial { nodo_id: toInt(line.node_1) })
MATCH (m: entidad { nodo_id: toInt(line.node_2) })
CREATE (n)-[:oficial_de] -> (m)
```

La salida al ejecutar será:

ILUSTRACIÓN N. 55
CREACIÓN DE RELACIÓN OFICIAL_DE

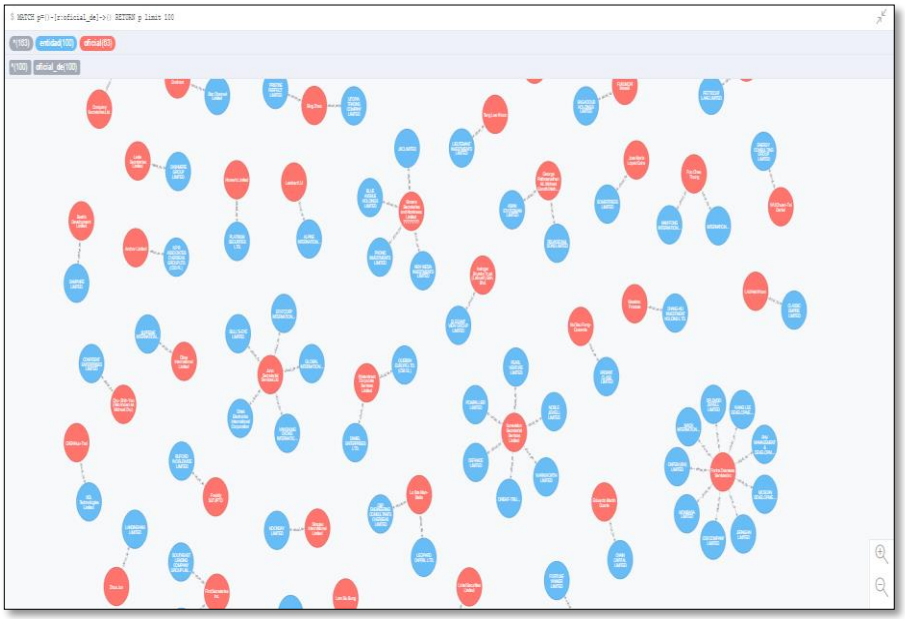


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Ejecutamos la siguiente consulta para visualizar las relaciones:

```
MATCH p=()-[r :oficial_de]->()
RETURN p
Limit 100
```

ILUSTRACIÓN N. 56
RELACIÓN OFICIAL_DE VISUALIZADO EN GRAFO



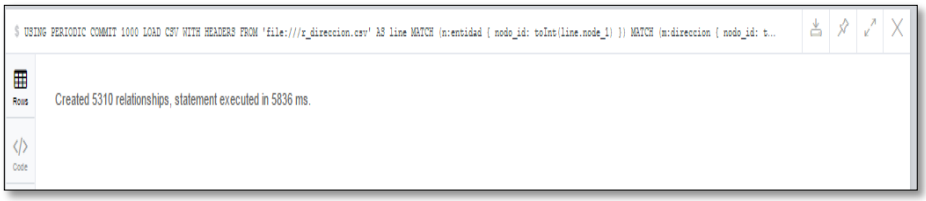
Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Relación: direccion_de

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM 'file: ///r_direccion.csv' AS line
MATCH (n: entidad { nodo_id: toInt(line.node_1) })
MATCH (m: direccion { nodo_id: toInt(line.node_2) })
CREATE (n)-[: direccion_de ] ->(m)
```

La salida al ejecutar será:

ILUSTRACIÓN N. 57
CREACIÓN DE RELACIÓN DIRECCIÓN_DE

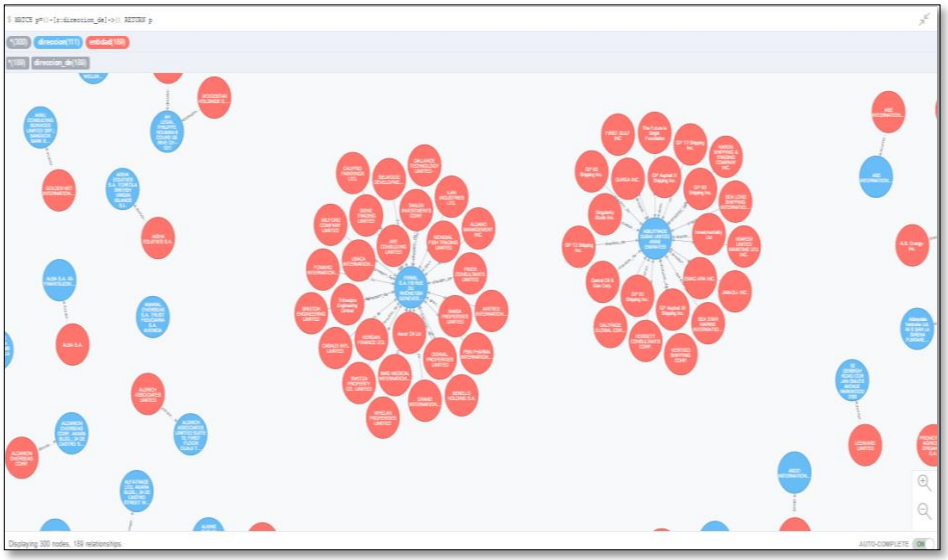


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Ejecutamos la siguiente consulta para visualizar las relaciones:

```
MATCH p=()-[r :direccion_de]->()
RETURN p
```

ILUSTRACIÓN N. 58
RELACIÓN DIRECCIÓN_DE VISUALIZADO EN GRAFO



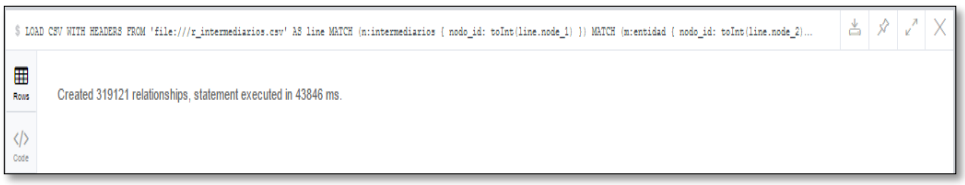
Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Relación: intermediario_de

```
LOAD CSV WITH HEADERS FROM 'file:///r_intermediarios.csv' AS line
MATCH (n: intermediarios { nodo_id: toInt(line.node_1) })
MATCH (m: entidad { nodo_id: toInt(line.node_2) })
CREATE (n)-[:INTERMEDIARIO_DE]->(m)
```

La salida al ejecutar será:

ILUSTRACIÓN N. 59
CREACIÓN DE RELACIÓN INTERMEDIARIO_DE

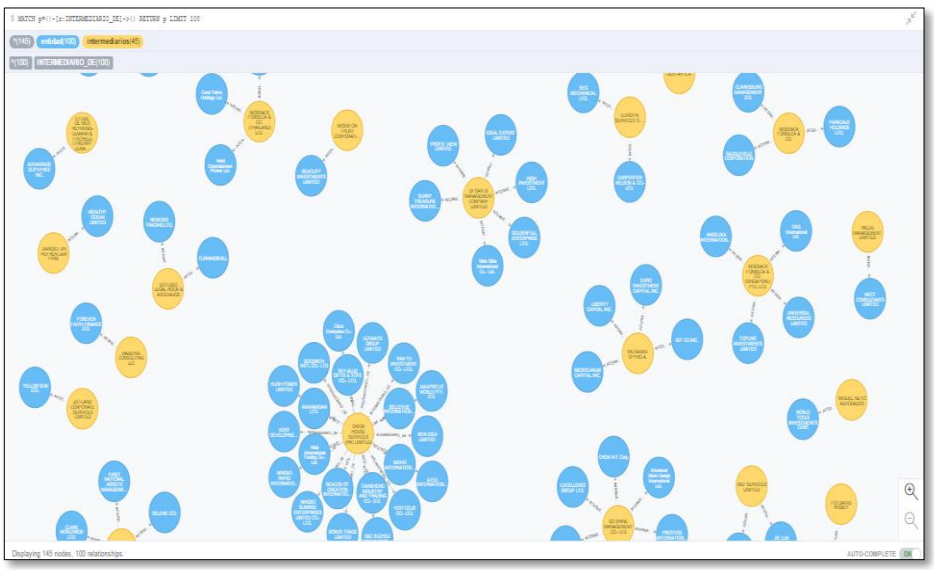


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Ejecutamos la siguiente consulta para visualizar las relaciones:

```
MATCH p=()-[r: INTERMEDIARIO_DE]->()
RETURN p
Limit 100
```

ILUSTRACIÓN N. 60
RELACIÓN INTERMEDIARIO_DE VISUALIZADO EN GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

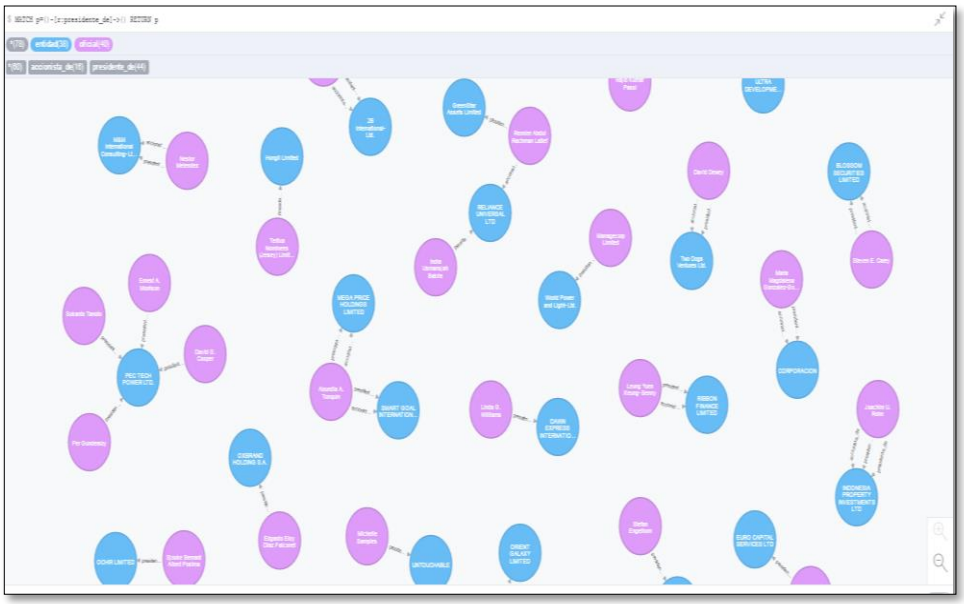
Relación: presidente_de

```
LOAD CSV WITH HEADERS FROM 'file: ///r_presidente.csv' AS line
MATCH (n: oficial { nodo_id: toInt(line.node_1) })
MATCH (m: entidad { nodo_id: toInt(line.node_2) })
CREATE (n)-[: presidente_de] -> (m)
```

Ejecutamos la siguiente consulta para visualizar las relaciones:

```
MATCH p = ()-[r :presidente_de]->()
RETURN p
```

ILUSTRACIÓN N. 61
RELACIÓN PRESIDENTE_DE VISUALIZADO EN GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

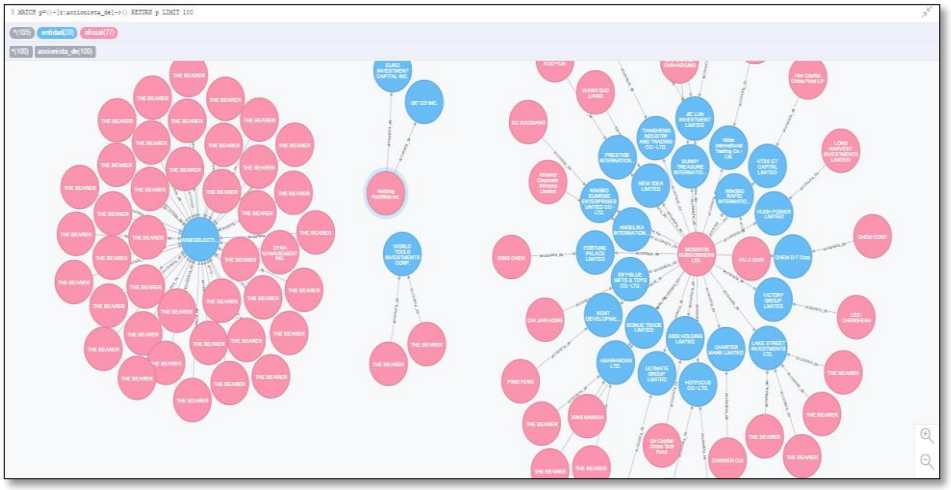
Relación: accionista_de

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM 'file:///r_accionista.csv' AS line
MATCH (n: oficial { nodo_id: toInt(line.node_1) })
MATCH (m: entidad { nodo_id: toInt(line.node_2) })
CREATE (n)-[:accionista_de]->(m)
```

Ejecutamos la siguiente consulta para visualizar las relaciones:

```
MATCH p=()-[r :accionista_de]->()
RETURN p  Limit 100
```

ILUSTRACIÓN N. 62
RELACIÓN ACCIONISTA_DE VISUALIZADO EN GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

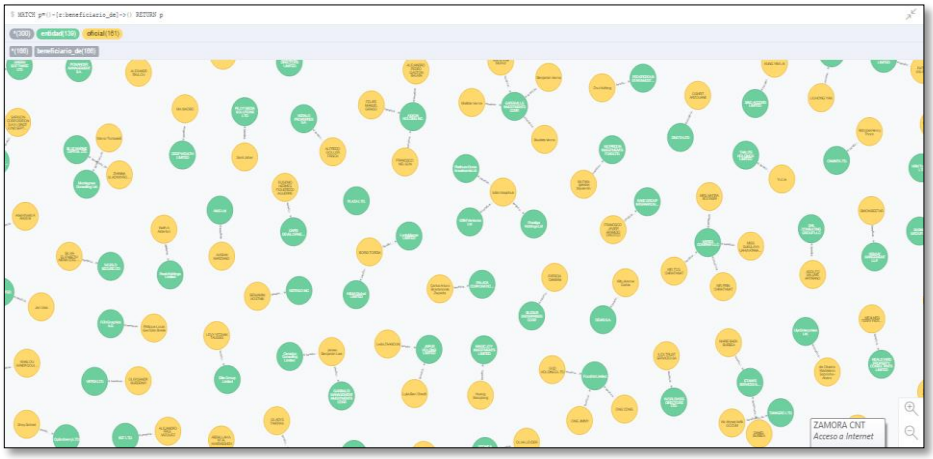
Relación: beneficiario_de

```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM 'file: ///r_beneficiario.csv' AS line
MATCH (n: oficial { nodo_id: toInt(line.node_1) })
MATCH (m: entidad { nodo_id: toInt(line.node_2) })
CREATE (n)-[: beneficiario_de ] -> (m)
```

Ejecutamos la siguiente consulta para visualizar las relaciones:

```
MATCH p=()-[r : beneficiario_de]->() RETURN p
```

ILUSTRACIÓN N. 63
RELACIÓN BENEFICIARIO_DE VISUALIZADO EN GRAFO



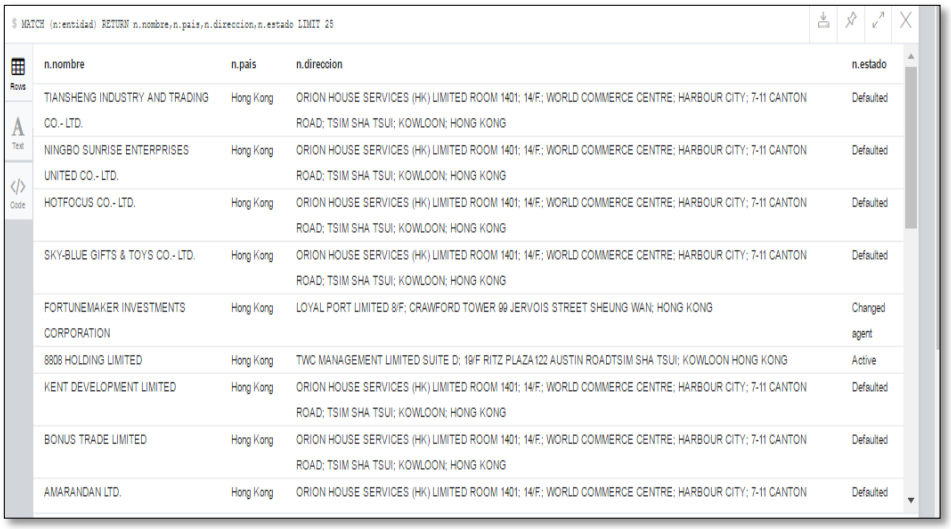
Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

CONSULTAS

1.-Consulta que devuelve los nombres de las empresas offshore, su país, dirección y su estado

```
MATCH (e: entidad)
RETURN e.nombre as EMPRESA, e.pais as PAIS, e.direccion as DIRECCION, e.estado as ESTADO
```

ILUSTRACIÓN N. 64
CONSULTA VISUALIZADA EN FORMA TABULAR



```
$ MATCH (n:entidad) RETURN n.nombre,n.pais,n.direccion,n.estado LIMIT 25
```

Rows | Text | Code

n.nombre	n.pais	n.direccion	n.estado
TIANSHENG INDUSTRY AND TRADING CO.- LTD.	Hong Kong	ORION HOUSE SERVICES (HK) LIMITED ROOM 1401; 14/F.; WORLD COMMERCE CENTRE; HARBOUR CITY; 7-11 CANTON ROAD; TSIM SHA TSUI; KOWLOON; HONG KONG	Defaulted
NINGBO SUNRISE ENTERPRISES UNITED CO.- LTD.	Hong Kong	ORION HOUSE SERVICES (HK) LIMITED ROOM 1401; 14/F.; WORLD COMMERCE CENTRE; HARBOUR CITY; 7-11 CANTON ROAD; TSIM SHA TSUI; KOWLOON; HONG KONG	Defaulted
HOTFOCUS CO.- LTD.	Hong Kong	ORION HOUSE SERVICES (HK) LIMITED ROOM 1401; 14/F.; WORLD COMMERCE CENTRE; HARBOUR CITY; 7-11 CANTON ROAD; TSIM SHA TSUI; KOWLOON; HONG KONG	Defaulted
SKY-BLUE GIFTS & TOYS CO.- LTD.	Hong Kong	ORION HOUSE SERVICES (HK) LIMITED ROOM 1401; 14/F.; WORLD COMMERCE CENTRE; HARBOUR CITY; 7-11 CANTON ROAD; TSIM SHA TSUI; KOWLOON; HONG KONG	Defaulted
FORTUNEMAKER INVESTMENTS CORPORATION	Hong Kong	LOYAL PORT LIMITED 8/F; CRAWFORD TOWER 99 JERVOIS STREET SHEUNG WAN; HONG KONG	Changed agent
8808 HOLDING LIMITED	Hong Kong	TWC MANAGEMENT LIMITED SUITE D; 19/F RITZ PLAZA 122 AUSTIN ROADTSIM SHA TSUI; KOWLOON HONG KONG	Active
KENT DEVELOPMENT LIMITED	Hong Kong	ORION HOUSE SERVICES (HK) LIMITED ROOM 1401; 14/F.; WORLD COMMERCE CENTRE; HARBOUR CITY; 7-11 CANTON ROAD; TSIM SHA TSUI; KOWLOON; HONG KONG	Defaulted
BONUS TRADE LIMITED	Hong Kong	ORION HOUSE SERVICES (HK) LIMITED ROOM 1401; 14/F.; WORLD COMMERCE CENTRE; HARBOUR CITY; 7-11 CANTON ROAD; TSIM SHA TSUI; KOWLOON; HONG KONG	Defaulted
AMARANDAN LTD.	Hong Kong	ORION HOUSE SERVICES (HK) LIMITED ROOM 1401; 14/F.; WORLD COMMERCE CENTRE; HARBOUR CITY; 7-11 CANTON	Defaulted

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

2.-Consultar la cantidad y los nombres de las empresas que tiene cada pais en paraísos fiscales.

```
MATCH (n: entidad)-[r: direccion_de] -> (m: direccion)
WITH m.pais as PAIS, COLLECT (n.nombre) as EMPRESA, COUNT (n.nombre) as CANTIDAD
RETURN PAIS, CANTIDAD, EMPRESA
```

ILUSTRACIÓN N. 65
CONSULTA VISUALIZADA EN FORMA TABULAR

```
$ MATCH (n: entidad)-[r: direccion_de] -> (m: direccion) WITH m.pais as PAIS, COLLECT (n.nombre) as EMPRESA, COUNT (n.nombre) as CANTIDAD RETURN PAIS, CANTIDAD, EMPRESA
```

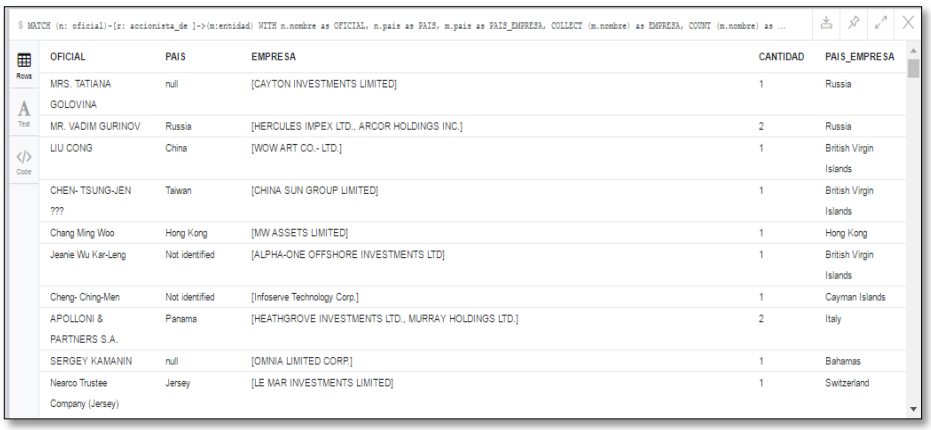
PAIS	CANTIDAD	EMPRESA
Azerbaijan	2	[CANVIL MANAGEMENT LTD., JOHN BROWN LTD.]
Saint Kitts and Nevis	1	[DISVASARI INTERNATIONAL INC.]
Saudi Arabia	7	[DORETTE CORPORATION, AMG DEVELOPMENTS LIMITED, CHIPSTEAD PROPERTIES CORP. S.A., NP ESTATE S.A., GLOBAL MEDICAL GROUP CORP., HEMPEL FINANCE LTD., DAIVEL INTERNATIONAL LTD.]
Egypt	5	[GALERIE GABY S.A., BRIDGEPORT CORPORATION LTD., ORCHARD ENTERPRISES S.A., SUN TRAVEL CORP., SURREY TRADING S.A.]
British Virgin Islands	694	[BELVO INVESTMENTS LIMITED, EMILY INVESTMENTS LIMITED, JOLIMA INVESTMENTS LIMITED, 3DK Consultant Corp., 4TH APRIL INVESTMENT INC., ABBEYFIELD INVESTMENTS LIMITED, ABDO INTERNATIONAL CORP., A.B. Energy- Inc., ABINA MARKETING LIMITED, Abson Investments Limited, Action Equity Group Limited, ADVINTRA SERVICES INC., AEDON NYASA LTD, Africa Energy Ltd, Aftract Limited, AGUILAR BUSINESS CORP., AISHA EQUITIES S.A., AKER FINANCE LIMITED, ALBIREO HOLDINGS INC., ALCARION OVERSEAS CORP., ALDAN FINANCE INVESTMENT INC., ALDERLY PROPERTIES GROUP CORP., ALDRIDGE CONSULTANTS LTD., ALFATRADE LTD., ALKINS SUGAR TRADING LIMITED, ALLROUND TRADING S.A., ALMAR GLOBAL CORP., ALTIUS GROUP INC., SALUKI LIMITED, GAIN SUN DEVELOPMENT LIMITED, ANATOL FINANCIAL ASSETS LIMITED, ANCORA MANAGEMENT INC., ANJUNA LIMITED, ANTARES CONSULTANCY INC., APOLLO FINANCIAL MANAGEMENT S.A., APPLE INVESTMENTS SERVICES LIMITED, ARABIAN INVESTORS GROUP S.A., ARAMIX DEVELOPMENT PROPERTIES LTD., ARAMIX MARKETING CORP., ARBUTUS BLUE LIMITED, ARCHER DEVELOPMENTS LIMITED, ARKWOOD INVESTMENTS LIMITED, ARLON ASSET HOLDINGS LTD., ARMENGAR TRADING CORP., Arnold Estates Limited, ARPS FINANCIAL INC., ARTIQUE INTERNATIONAL LTD., ARVEST CORPORATION, ASHLAWN PROPERTY HOLDINGS LIMITED, Asian World Investment Group- Inc., ASTRO STAR GROUP LTD., HIREWARD INVESTMENTS LIMITED, IT STAR HOLDINGS LIMITED, ATINON MANAGEMENT CORP., ATLANTIC WAREHOUSE LTD., ATTIJA GROUP LTD., AURELINE HOLDING LTD., AVE ENERGY SERVICES S.A., AXBRIDGE GROUP INC., AZTEC MARITIME ASSETS

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

3.-Consultar el nombre de la persona , a que pais pertenece , las empresas offshore creadas y en que pais estan registradas.

```
MATCH (n: oficial)-[r: accionista_de ]->(m:entidad)
WITH n.nombre as OFICIAL, n.pais as PAIS, m.pais as PAIS_EMPRESA, COLLECT
(m.nombre) as EMPRESA, COUNT (m.nombre) as CANTIDAD
RETURN OFICIAL, PAIS, EMPRESA, CANTIDAD, PAIS_EMPRESA
```

ILUSTRACIÓN N. 66
CONSULTA VISUALIZADA EN FORMA TABULAR



```
$ MATCH (n: oficial)-[r: accionista_de ]->(m:entidad) WITH n.nombre as OFICIAL, n.pais as PAIS, m.pais as PAIS_EMPRESA, COLLECT (m.nombre) as EMPRESA, COUNT (m.nombre) as ...
```

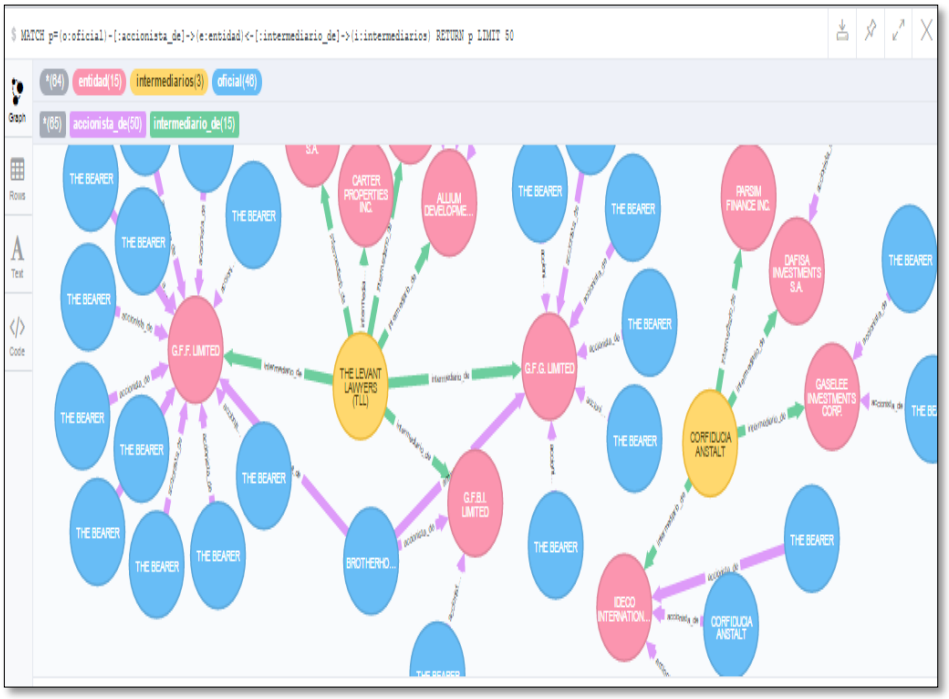
OFICIAL	PAIS	EMPRESA	CANTIDAD	PAIS_EMPRESA
MRS. TATIANA GOLOVINA	null	[CAYTON INVESTMENTS LIMITED]	1	Russia
MR. VADIM GURINOV	Russia	[HERCULES IMPEX LTD., ARCOR HOLDINGS INC.]	2	Russia
LIU CONG	China	[WOW ART CO.- LTD.]	1	British Virgin Islands
CHEN- TSUNG-JEN ???	Taiwan	[CHINA SUN GROUP LIMITED]	1	British Virgin Islands
Chang Ming Woo	Hong Kong	[MW ASSETS LIMITED]	1	Hong Kong
Jeanie Wu Kar-Leng	Not identified	[ALPHA-ONE OFFSHORE INVESTMENTS LTD]	1	British Virgin Islands
Cheng- Ching-Men	Not identified	[Infoserve Technology Corp.]	1	Cayman Islands
APOLLONI & PARTNERS S.A.	Panama	[HEATHGROVE INVESTMENTS LTD., MURRAY HOLDINGS LTD.]	2	Italy
SERGEY KAMANIN	null	[OMNIA LIMITED CORP.]	1	Bahamas
Nearco Trustee Company (Jersey)	Jersey	[LE MAR INVESTMENTS LIMITED]	1	Switzerland

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

4.- Consultar quienes son los accionistas y los intermediarios que ayudaron a crear las empresas offshore.

```
MATCH p=(o:oficial)-[:accionista_de]->(e:entidad)<-[:intermediario_de]-
>(i:intermediarios)
RETURN p
```

ILUSTRACIÓN N. 67
CONSULTA VISUALIZADA EN GRAFO

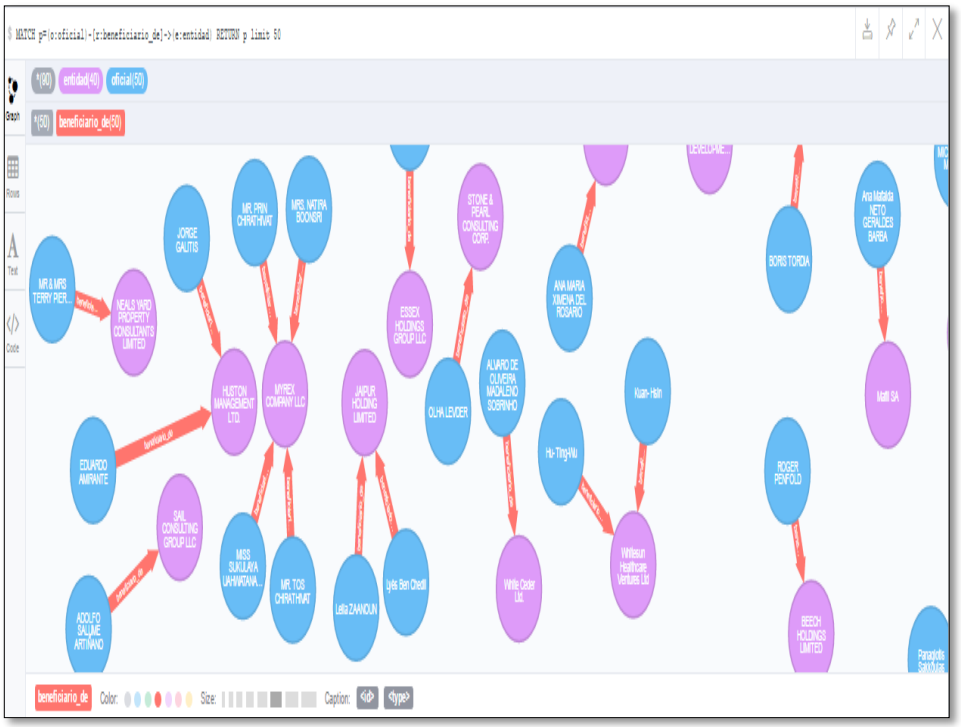


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

5.- Visualizar quienes son los beneficiarios finales de cada compañía.

```
MATCH p=(o:oficial)-[r:beneficiario_de]->(e:entidad)
RETURN p
```

ILUSTRACIÓN N. 68
CONSULTA VISUALIZADA EN GRAFO

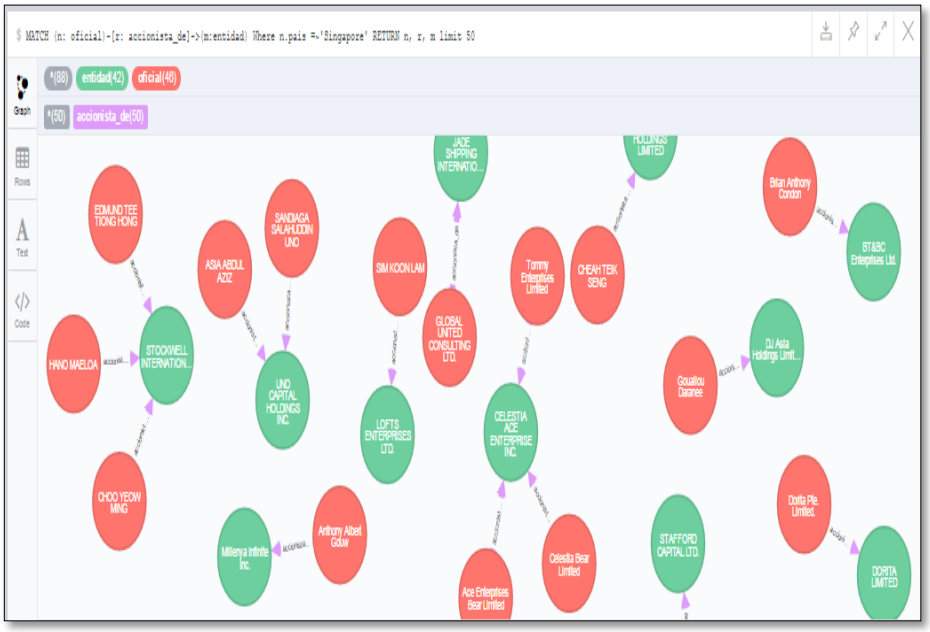


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

6.**-Consultar los nombre de los accionistas de Singapore y sus empresas registradas.**

```
MATCH (n: oficial)-[r: accionista_de]->(m:entidad)
Where n.pais =~'Singapore'
RETURN n, r, m
```

ILUSTRACIÓN N. 69
CONSULTA VISUALIZADA EN GRAFO

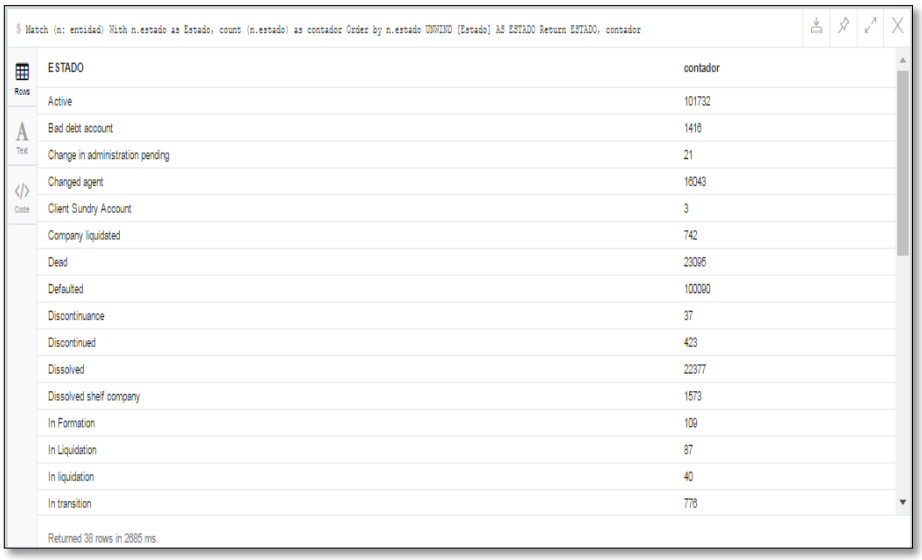


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

7.- Consultar los estados en orden alfabético y el total de ellos. Los mismos que podemos exportarlos en formato JSON o CSV.

```
Match (n: entidad)
With n.estado as Estado, count (n.estado) as contador
Order by n.estado
UNWIND [Estado] AS ESTADO  Return ESTADO, contador
```

ILUSTRACIÓN N. 70
CONSULTA VISUALIZADA EN FORMA TABULAR



```
$ Match (n: entidad) With n.estado as Estado, count (n.estado) as contador Order by n.estado UNWIND [Estado] AS ESTADO Return ESTADO, contador
```

ESTADO	contador
Active	101732
Bad debt account	1416
Change in administration pending	21
Changed agent	16043
Client Sundry Account	3
Company liquidated	742
Dead	23095
Defaulted	100080
Discontinuance	37
Discontinued	423
Dissolved	22377
Dissolved shelf company	1573
In Formation	109
In Liquidation	87
In liquidation	40
In transition	778

Returned 38 rows in 2885 ms.

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Exportar Datos

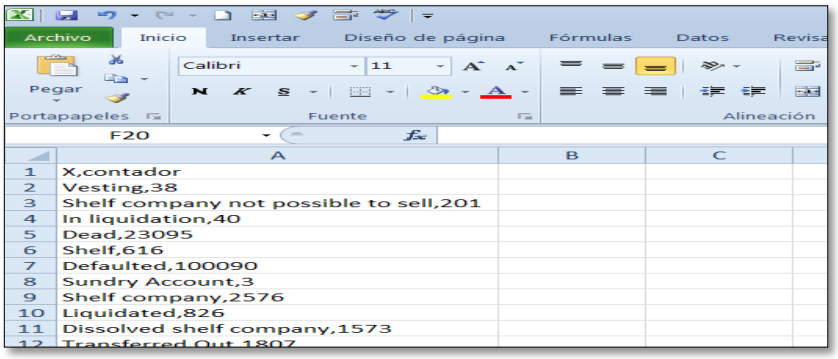
También nos da las opciones de exportar sus datos en formato CSV y JSON.

ILUSTRACIÓN N. 71
EXPORTAR EN JSON-CSV

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Hacer clic en "Exportar CSV" para exportar los resultados en formato de archivo CSV

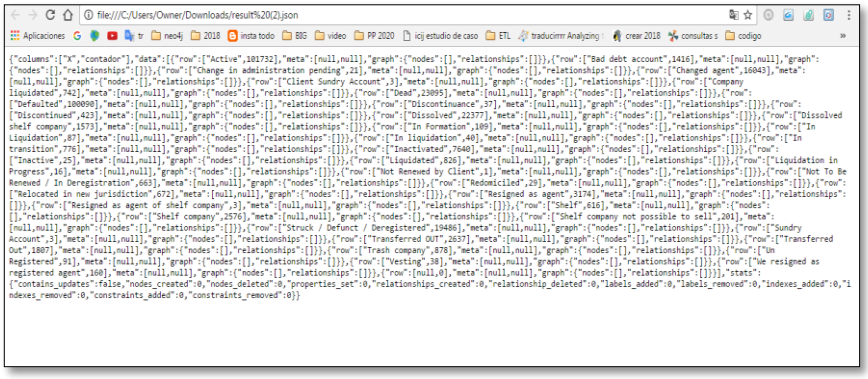
ILUSTRACIÓN N. 72
DATO EXPORTADO EN CSV



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Hacer clic en "Exportar JSON" para exportar los resultados en formato de archivo JSON.

ILUSTRACIÓN N. 73
DATO EXPORTADO EN JSON

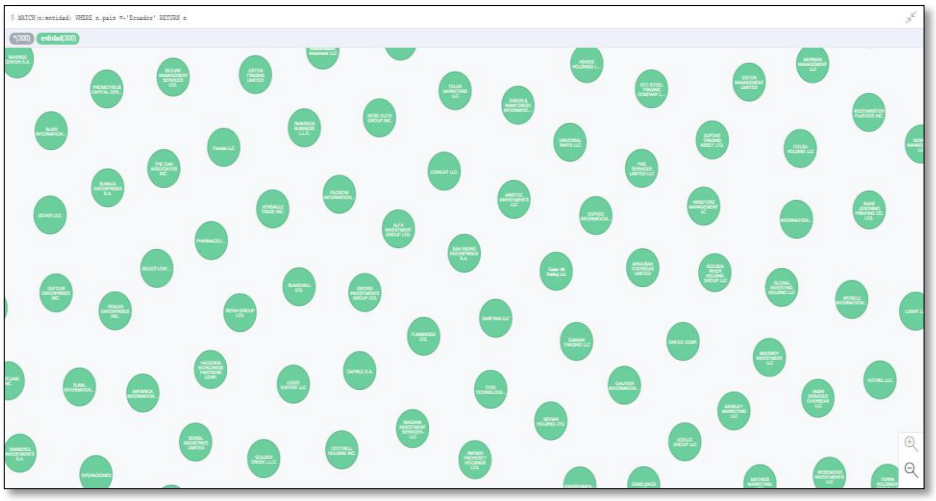


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

8.- Consulta que devuelve solo las empresas creadas en Ecuador. Usaremos Where para filtrar los resultados.

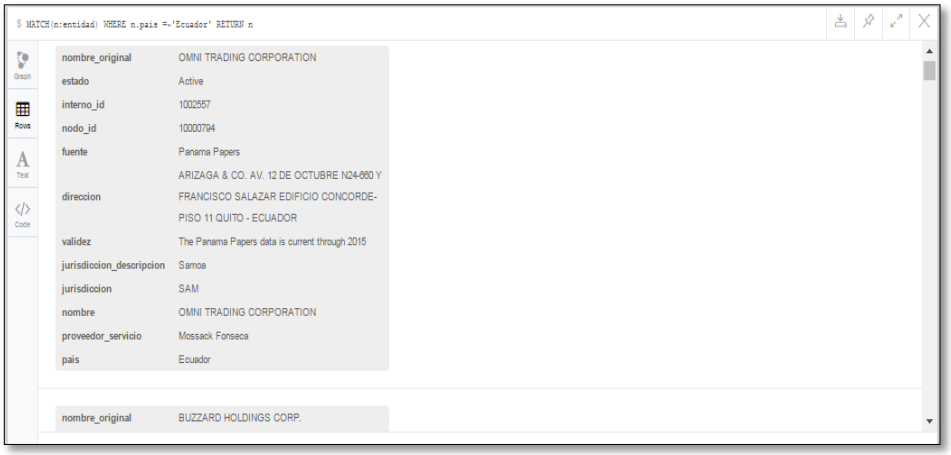
```
MATCH(n: entidad)
WHERE n.pais =~'Ecuador'
RETURN n
```

ILUSTRACIÓN N. 74
CONSULTA VISUALIZADA EN GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

ILUSTRACIÓN N. 75
CONSULTA VISUALIZADA EN FORMA TABULAR



nombre_original	OMNI TRADING CORPORATION
estado	Active
interno_id	1002557
nodo_id	10000794
fuente	Panama Papers
direccion	ARIZAGA & CO. AV. 12 DE OCTUBRE N24-660 Y FRANCISCO SALAZAR EDIFICIO CONCORDE-PISO 11 QUITO - ECUADOR
validez	The Panama Papers data is current through 2015
jurisdiccion_descripcion	Samoa
jurisdiccion	SAM
nombre	OMNI TRADING CORPORATION
proveedor_servicio	Mossack Fonseca
pais	Ecuador

nombre_original	BUZZARD HOLDINGS CORP.

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

9.- Consulta que devuelve el total de las empresas offshore creadas en Ecuador.

```
MATCH (n: entidad)//-[r: direccion_de] -> (m: direccion)
WITH n.pais as PAIS, COLLECT (n.nombre) as EMPRESA, COUNT (n.nombre) as CANTIDAD
Where PAIS =~'Ecuador'
RETURN PAIS, EMPRESA, CANTIDAD
```

ILUSTRACIÓN N. 76
CONSULTA VISUALIZADA EN FORMA TABULAR

PAIS	EMPRESA	CANTIDAD
Ecuador	[UNIGLOBAL CORPORATION LIMITED, OMNI TRADING CORPORATION, BUZZARD HOLDINGS CORP., Cedar Hill Trading Ltd., Century Enterprise Corporation, FULASA Corp., Internacional de Inversiones INVERLATIN S.A., MONAR HOLDING LTD., ENSENADA CORP., MEMORIL INVESTMENT CORP., ROSEBELLA FARM'S CORPORATION, SYBEST OVERSEAS LIMITED, GROSELLA CORP., ORFEO CORP., GROENE OIL COMPANY LLP, OBLIGACORP REPRESENTATIVE S.A., GEBEKO OVERSEAS LIMITED, DARFRAN LLC, Business & Migration Holding Corporation, ALGOYA LLC, JACKLIN INVESTMENTS LLC, PANGEA INVESTMENTS LLP, OMATE CONSULTANTS LIMITED, OLVEX CAPITAL LTD., ENVIRONMENTAL BIO TECHNOLOGIES LTD., VADAL GROUP CORPORATION, FAIRCENT CORPORATION, TERRAFERTIL US LLC, RENALCO LLP, CAMELBACK LLC., LYNTON TRADING LTD., ARKADIAN OVERSEAS LIMITED, MEDICALTECH INVESTMENTS LLC, KELBURN ENTERPRISES LC, INTERGRAMA LIMITED, RICHMO-BILIARIA LLC, TERRA INVESTMENTS LLC, EXTRA ASSET MANAGEMENT LLC, TIDEFORD MANAGEMENT LLP, UCELLO GROUP LLC, JAGAN INTERNATIONAL LTD., BARKLEY MARKETING LLC, GRYSEL FINANCE S.A., TERRA HOLDINGS GROUP LLC, BEESROY INVESTMENT LLC, VERNON GROUP ASSETS LLC, JIMSON HOLDING LTD., UNIVERSAL PARTS LLC, GOLFIN FOOD CORP. LLC, ROSEMOOR INVESTMENTS LLC, GOLDEN RIVER HOLDING GROUP LLC, SUNLIGHT INVESTMENTS GROUP LLC, HASHVILLE INTERNATIONAL LLC, ARISDALE TRADING LTD., NATRAL INVESTMENTS LLC, LOBALAT LLC., MOBATA INTERNATIONAL LTD., OCCIDENTAL TRADING SERVICES LLC, GLOBAL INVESTING HOLDING LLC, FALER CORP. LLC, VADIN SERVICES OVERSEAS LLC, MEDICAL ASSISTANCE REPRESENTATIVES LLC, LUGNY LLC, ARISTEC INVESTMENTS LLC, FLEXPRO LIMITED, ITELSA HOLDING LLC, THERSANDER LLC, MONELLI INTERNATIONAL L.C., ARKELEY DEVELOPMENT PORTFOLIO LLC, ALFA INVESTMENT GROUP LTD., XTREME MARKETING GROUP LLC, AZORINE CONSULTANTS LTD., KINGFORD MANAGEMENT LC, BRALUM LLC, FRINVESSMENT LLC., VOTREL LLC., LATINAMERICAN GROUP INVESTMENT LLC, WORLDWIDE REAL ESTATE GROUP LLC, INTERNATIONAL REAL ESTATE GROUP LLC, MUTUAL INVESTMENTS LLC, KENNON MANAGEMENT LLC, STC STEEL TRADING COMPANY LLC, ESTOR MANAGEMENT LIMITED, ARISIA HOLDING LTD., TELSIS TRADING GROUP LLC, GLOBAL RECYCLING COMPANY- GRC RECYCLERS LLC, PORTFOLIO INVESTMENT LLC, HEWES HOLDINGS LLC, TRENKO CORPORATION LLP, ARISE REGENT SERVICES LLC, SAINT JERONIMO PRINTING CO. LTD., WESTMINSTER PLASTICS INC. LC, BARTA MANAGEMENT LLC, PNS SERVICES LIMITED LLC, MERIBAN MANAGEMENT LLC,	1862

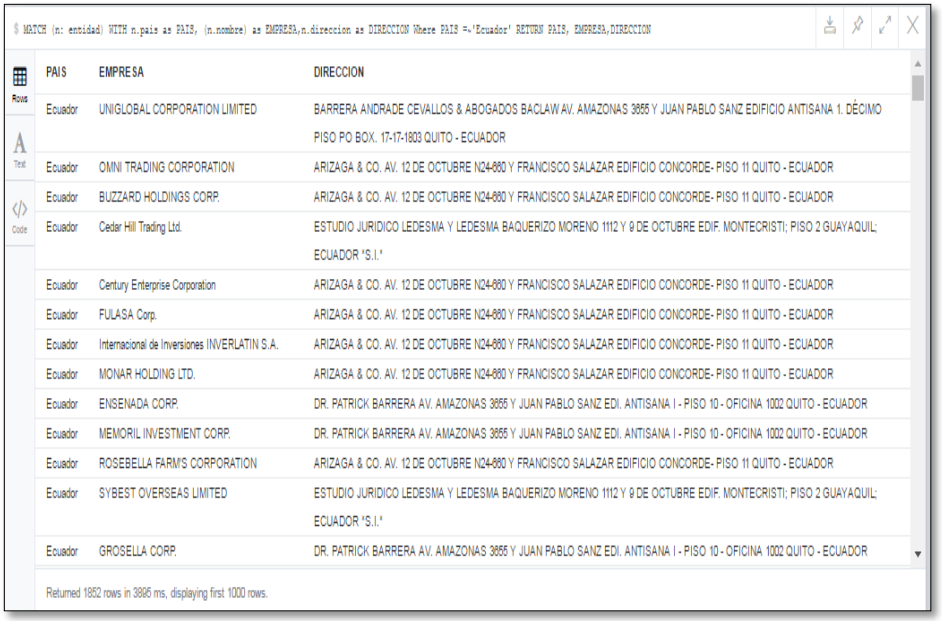
Returned 1 row in 4783 ms.

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

10.- Consulta que devuelve las empresas offshore creadas en Ecuador y sus direcciones registradas.

```
MATCH (n: entidad)
WITH n.pais as PAIS, (n.nombre) as EMPRESA, n.direccion as DIRECCION
Where PAIS =~'Ecuador'
RETURN PAIS, EMPRESA,DIRECCION
```

ILUSTRACIÓN N. 77
CONSULTA VISUALIZADA EN FORMA TABULAR



```
$ MATCH (n: entidad) WITH n.pais as PAIS, (n.nombre) as EMPRESA,n.direccion as DIRECCION Where PAIS =~'Ecuador' RETURN PAIS, EMPRESA,DIRECCION
```

Rows | Text | Code

PAIS	EMPRESA	DIRECCION
Ecuador	UNIGLOBAL CORPORATION LIMITED	BARRERA ANDRADE CEVALLOS & ABOGADOS BACLAW AV. AMAZONAS 3855 Y JUAN PABLO SANZ EDIFICIO ANTISANA 1. DÉCIMO PISO PO BOX. 17-17-1803 QUITO - ECUADOR
Ecuador	OMNI TRADING CORPORATION	ARIZAGA & CO. AV. 12 DE OCTUBRE N24-860 Y FRANCISCO SALAZAR EDIFICIO CONCORDE- PISO 11 QUITO - ECUADOR
Ecuador	BUZZARD HOLDINGS CORP.	ARIZAGA & CO. AV. 12 DE OCTUBRE N24-860 Y FRANCISCO SALAZAR EDIFICIO CONCORDE- PISO 11 QUITO - ECUADOR
Ecuador	Cedar Hill Trading Ltd.	ESTUDIO JURIDICO LEDESMA Y LEDESMA BAQUERIZO MORENO 1112 Y 9 DE OCTUBRE EDIF. MONTECRISTI; PISO 2 GUAYAQUIL: ECUADOR 'S.I.'
Ecuador	Century Enterprise Corporation	ARIZAGA & CO. AV. 12 DE OCTUBRE N24-860 Y FRANCISCO SALAZAR EDIFICIO CONCORDE- PISO 11 QUITO - ECUADOR
Ecuador	FULASA Corp.	ARIZAGA & CO. AV. 12 DE OCTUBRE N24-860 Y FRANCISCO SALAZAR EDIFICIO CONCORDE- PISO 11 QUITO - ECUADOR
Ecuador	Internacional de Inversiones INVERLATIN S.A.	ARIZAGA & CO. AV. 12 DE OCTUBRE N24-860 Y FRANCISCO SALAZAR EDIFICIO CONCORDE- PISO 11 QUITO - ECUADOR
Ecuador	MONAR HOLDING LTD.	ARIZAGA & CO. AV. 12 DE OCTUBRE N24-860 Y FRANCISCO SALAZAR EDIFICIO CONCORDE- PISO 11 QUITO - ECUADOR
Ecuador	ENSENADA CORP.	DR. PATRICK BARRERA AV. AMAZONAS 3855 Y JUAN PABLO SANZ EDI. ANTISANA I - PISO 10 - OFICINA 1002 QUITO - ECUADOR
Ecuador	MEMORIL INVESTMENT CORP.	DR. PATRICK BARRERA AV. AMAZONAS 3855 Y JUAN PABLO SANZ EDI. ANTISANA I - PISO 10 - OFICINA 1002 QUITO - ECUADOR
Ecuador	ROSEBELLA FARM'S CORPORATION	ARIZAGA & CO. AV. 12 DE OCTUBRE N24-860 Y FRANCISCO SALAZAR EDIFICIO CONCORDE- PISO 11 QUITO - ECUADOR
Ecuador	SYBEST OVERSEAS LIMITED	ESTUDIO JURIDICO LEDESMA Y LEDESMA BAQUERIZO MORENO 1112 Y 9 DE OCTUBRE EDIF. MONTECRISTI; PISO 2 GUAYAQUIL: ECUADOR 'S.I.'
Ecuador	GROSELLA CORP.	DR. PATRICK BARRERA AV. AMAZONAS 3855 Y JUAN PABLO SANZ EDI. ANTISANA I - PISO 10 - OFICINA 1002 QUITO - ECUADOR

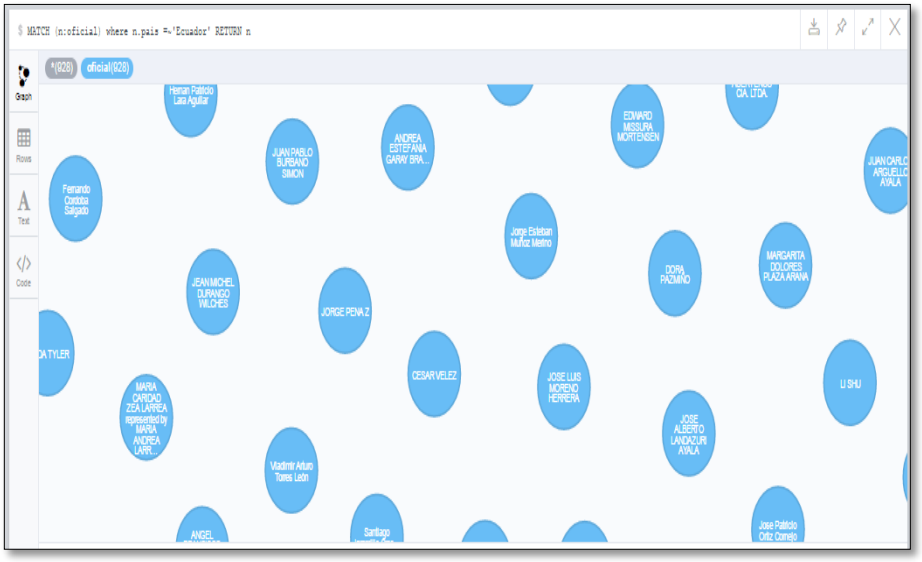
Returned 1852 rows in 3895 ms, displaying first 1000 rows.

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

11.-Consulta que devuelve los nombres de las personas de Ecuador que tienen un papel importante en una empresa.

```
MATCH (o:oficial) where o.pais =~'Ecuador'
RETURN o
```

ILUSTRACIÓN N. 78
CONSULTA VISUALIZADA EN GRAFO

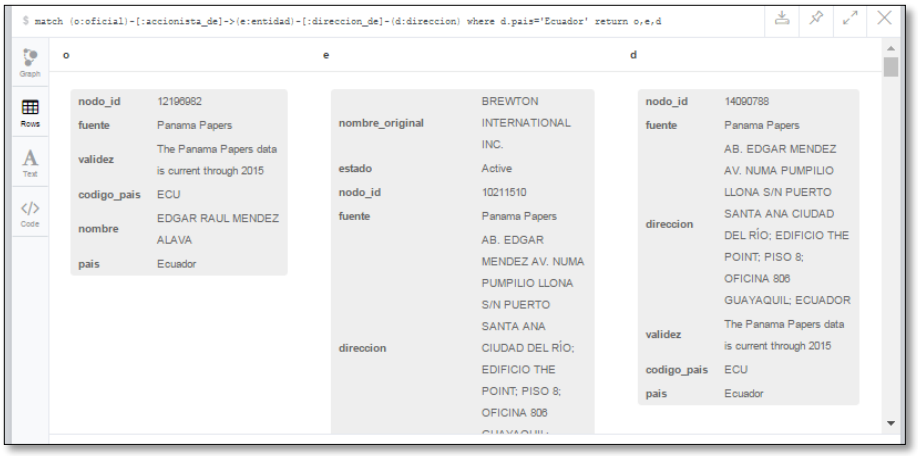


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

12.- Visualizar los accionistas Ecuatorianos de las empresas offshore , sus respectivas direcciones registradas.

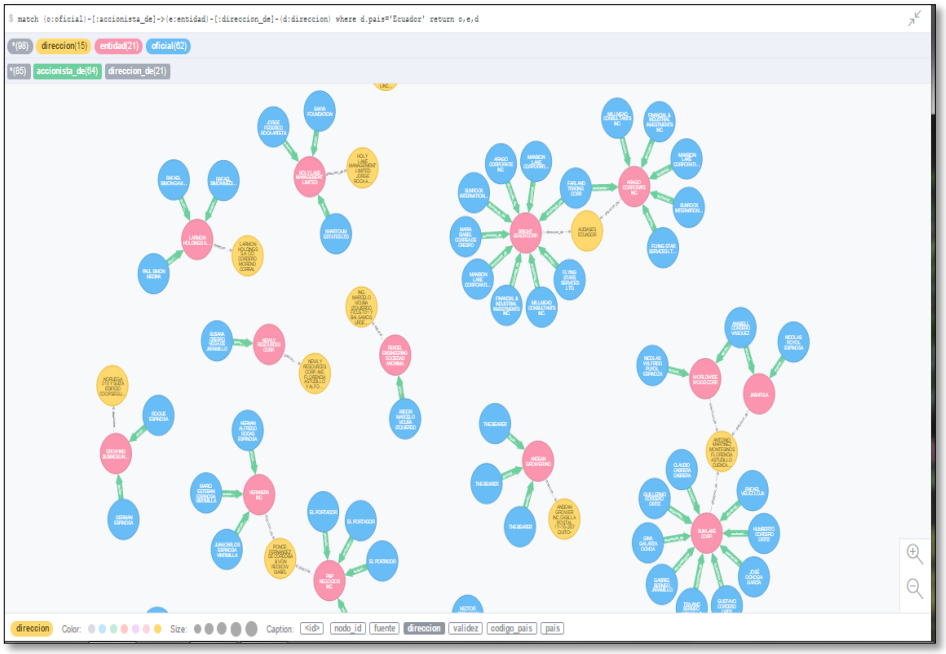
```
match p=(o:oficial)-[:accionista_de]->(e:entidad)-[:direccion_de]-(d:direccion)
where d.pais='Ecuador' return p
```

ILUSTRACIÓN N. 79
CONSULTA VISUALIZADA EN FORMA TABULAR



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

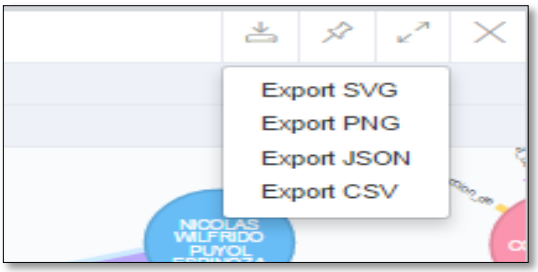
ILUSTRACIÓN N. 80
CONSULTA VISUALIZADA EN GRAFO



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

La visualización gráfica tambien se puede exportar como PNG y SVG.

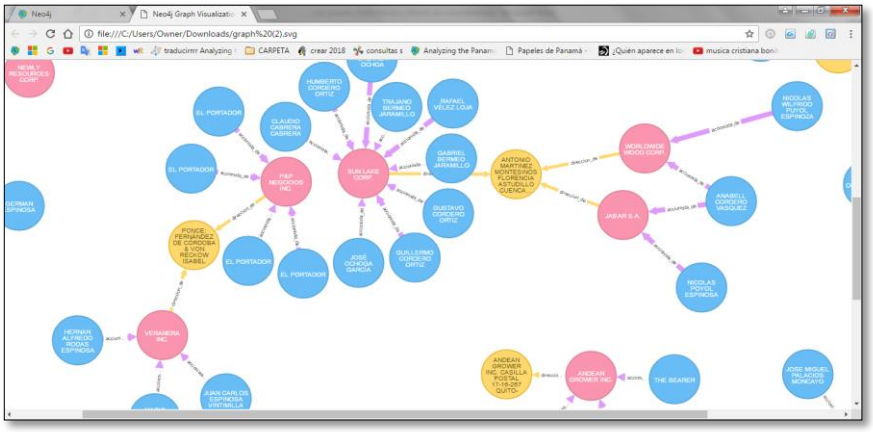
ILUSTRACIÓN N. 81
EXPORTAR EN SVG-PNG



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Hacer clic en "Exportar SVG" para exportar los resultados en formato de archivo SVG.

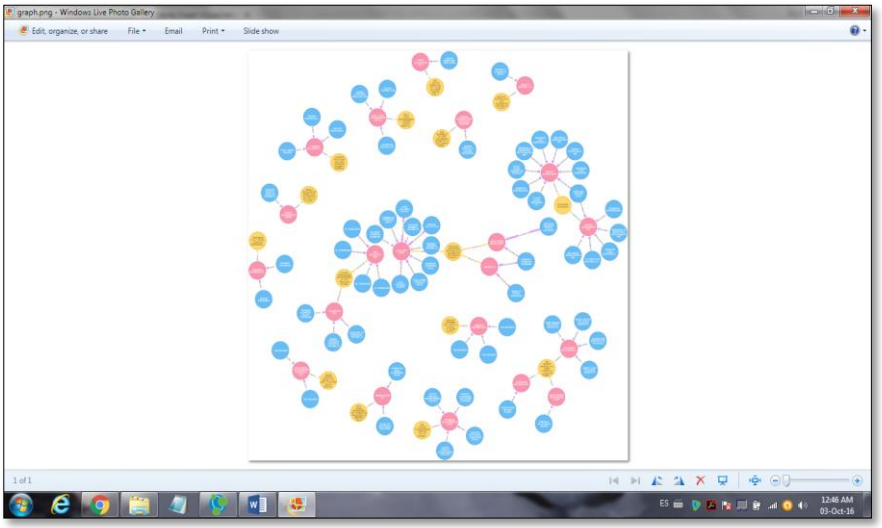
ILUSTRACIÓN N. 82
DATO EXPORTADO EN SVG



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Hacer clic en "Exportar PNG" para exportar los resultados en formato de archivo JSON.

ILUSTRACIÓN N. 83
DATO EXPORTADO EN PNG

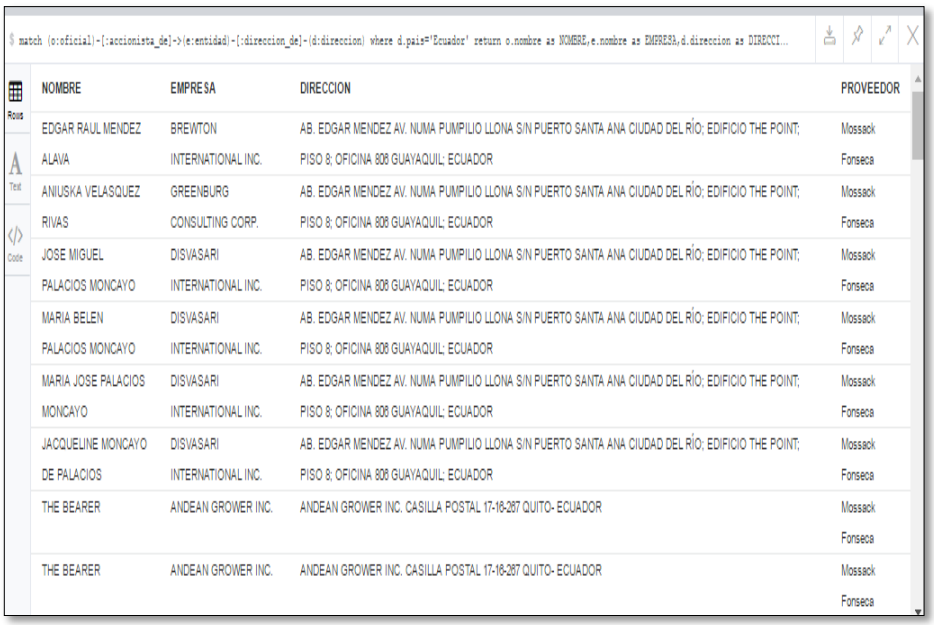


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

13.- Visualizar los accionistas Ecuatorianos, las empresas offshore, sus respectivas direcciones y quien es su proveedor.

```
match (o:oficial)-[:accionista_de]->(e:entidad)-[:direccion_de]-(d:direccion)
where d.pais='Ecuador'
return o.nombre as NOMBRE,e.nombre as EMPRESA,d.direccion as
DIRECCION,e.proveedor_servicio as PROVEEDOR
```

ILUSTRACIÓN N. 84
CONSULTA VISUALIZADA EN FORMA TABULAR



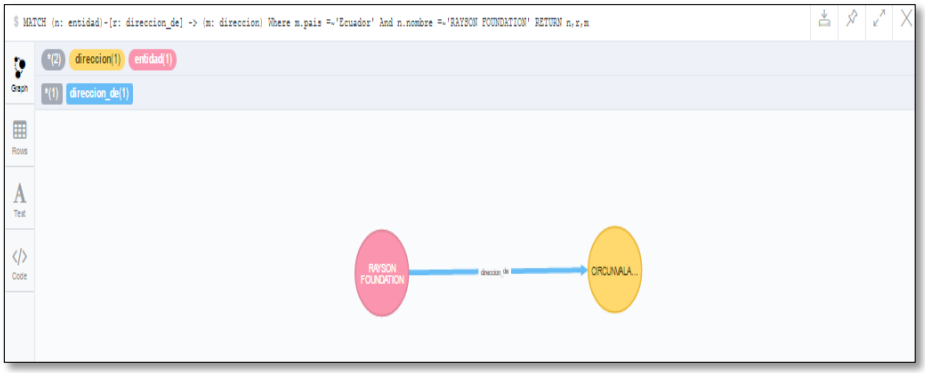
NOMBRE	EMPRESA	DIRECCION	PROVEEDOR
EDGAR RAUL MENDEZ	BREWTON	AB. EDGAR MENDEZ AV. NUMA PUMPILIO LLONA S/N PUERTO SANTA ANA CIUDAD DEL RÍO; EDIFICIO THE POINT;	Mossack
ALAVA	INTERNATIONAL INC.	PISO 8; OFICINA 808 GUAYAQUIL; ECUADOR	Fonseca
ANIUSKA VELASQUEZ	GREENBURG	AB. EDGAR MENDEZ AV. NUMA PUMPILIO LLONA S/N PUERTO SANTA ANA CIUDAD DEL RÍO; EDIFICIO THE POINT;	Mossack
RIVAS	CONSULTING CORP.	PISO 8; OFICINA 808 GUAYAQUIL; ECUADOR	Fonseca
JOSE MIGUEL	DISVASARI	AB. EDGAR MENDEZ AV. NUMA PUMPILIO LLONA S/N PUERTO SANTA ANA CIUDAD DEL RÍO; EDIFICIO THE POINT;	Mossack
PALACIOS MONCAYO	INTERNATIONAL INC.	PISO 8; OFICINA 808 GUAYAQUIL; ECUADOR	Fonseca
MARIA BELEN	DISVASARI	AB. EDGAR MENDEZ AV. NUMA PUMPILIO LLONA S/N PUERTO SANTA ANA CIUDAD DEL RÍO; EDIFICIO THE POINT;	Mossack
PALACIOS MONCAYO	INTERNATIONAL INC.	PISO 8; OFICINA 808 GUAYAQUIL; ECUADOR	Fonseca
MARIA JOSE PALACIOS	DISVASARI	AB. EDGAR MENDEZ AV. NUMA PUMPILIO LLONA S/N PUERTO SANTA ANA CIUDAD DEL RÍO; EDIFICIO THE POINT;	Mossack
MONCAYO	INTERNATIONAL INC.	PISO 8; OFICINA 808 GUAYAQUIL; ECUADOR	Fonseca
JACQUELINE MONCAYO	DISVASARI	AB. EDGAR MENDEZ AV. NUMA PUMPILIO LLONA S/N PUERTO SANTA ANA CIUDAD DEL RÍO; EDIFICIO THE POINT;	Mossack
DE PALACIOS	INTERNATIONAL INC.	PISO 8; OFICINA 808 GUAYAQUIL; ECUADOR	Fonseca
THE BEARER	ANDEAN GROWER INC.	ANDEAN GROWER INC. CASILLA POSTAL 17-16-267 QUITO- ECUADOR	Mossack
			Fonseca
THE BEARER	ANDEAN GROWER INC.	ANDEAN GROWER INC. CASILLA POSTAL 17-16-267 QUITO- ECUADOR	Mossack
			Fonseca

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

14.-Consultar si la empresa 'RAYSON FOUNDATION' pertenece a Ecuador y retornar si es el caso su direccion.

```
MATCH (n: entidad)-[r: direccion_de] -> (m: direccion)
Where m.pais =~'Ecuador'
And n.nombre =~'RAYSON FOUNDATION'
RETURN n,r,m
```

ILUSTRACIÓN N. 85
CONSULTA VISUALIZADA EN GRAFO

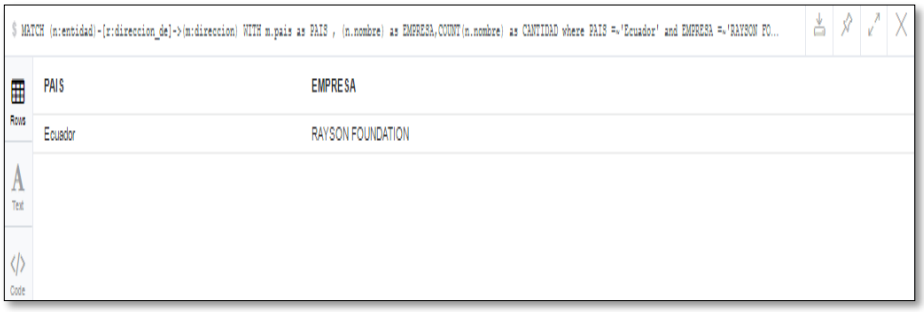


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

17.- Consultar si la empresa 'RAYSON FOUNDATION' pertenece a Ecuador y retornar si es el caso.

```
MATCH (n: entidad)-[r: direccion_de] -> (m: direccion)
WITH m.pais as PAIS, (n.nombre) as EMPRESA Where PAIS =~'Ecuador'
And EMPRESA =~'RAYSON FOUNDATION'  RETURN PAIS, EMPRESA
```

ILUSTRACIÓN N. 86
CONSULTA VISUALIZADA EN FORMA TABULAR

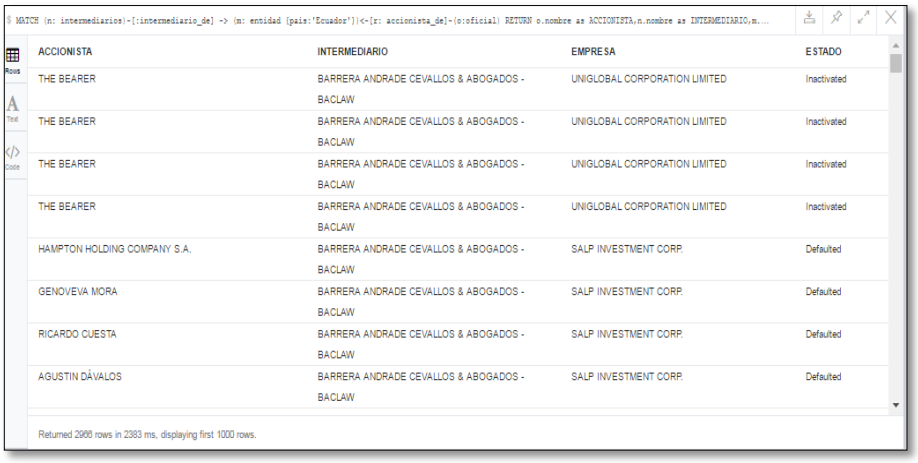


Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

18.- Consultar los estados actuales de las empresas offshore de Ecuador con su intermediario y accionista.

```
MATCH (n: intermediarios)-[:intermediario_de] -> (m: entidad {pais:'Ecuador'})<-[r:
accionista_de]-(o:oficial)
RETURN o.nombre as ACCIONISTA,n.nombre as INTERMEDIARIO,m.nombre as
EMPRESA, m.estado AS ESTADO
```

ILUSTRACIÓN N. 87
CONSULTA VISUALIZADA EN FORMA TABULAR



ACCIONISTA	INTERMEDIARIO	EMPRESA	ESTADO
THE BEARER	BARRERA ANDRADE CEVALLOS & ABOGADOS - BACLAW	UNIGLOBAL CORPORATION LIMITED	Inactivated
THE BEARER	BARRERA ANDRADE CEVALLOS & ABOGADOS - BACLAW	UNIGLOBAL CORPORATION LIMITED	Inactivated
THE BEARER	BARRERA ANDRADE CEVALLOS & ABOGADOS - BACLAW	UNIGLOBAL CORPORATION LIMITED	Inactivated
THE BEARER	BARRERA ANDRADE CEVALLOS & ABOGADOS - BACLAW	UNIGLOBAL CORPORATION LIMITED	Inactivated
HAMPTON HOLDING COMPANY S.A.	BARRERA ANDRADE CEVALLOS & ABOGADOS - BACLAW	SALP INVESTMENT CORP.	Defaulted
GENOVEVA MORA	BARRERA ANDRADE CEVALLOS & ABOGADOS - BACLAW	SALP INVESTMENT CORP.	Defaulted
RICARDO CUESTA	BARRERA ANDRADE CEVALLOS & ABOGADOS - BACLAW	SALP INVESTMENT CORP.	Defaulted
AGUSTIN DÁVALOS	BARRERA ANDRADE CEVALLOS & ABOGADOS - BACLAW	SALP INVESTMENT CORP.	Defaulted

Returned 2966 rows in 2383 ms, displaying first 1000 rows.

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

Hemos observado algunas consultas dentro de un ambiente Big Data para comprobar como Neo4j nos muestra los resultados de manera rápida y comprensible transformando todos estos datos en información y por consiguiente a conocimiento.

ANEXO 3

LA ENCUESTA

1.- ¿Ha escuchado usted acerca del tema Big Data?

Si	
No	

2¿Sabe usted cuál de estas opciones no es una herramienta para tratar con Big Data?

Datos estructurados	
Datos no estructurados	
Datos semiestructurados	
Datos inexistentes	

3. ¿Sabe usted cuál es un tipo de captura de información en Big Data**?**

Generados por las personas	
Generados por si solos	

4. **-** Ha escuchado sobre el volumen de los datos masivos

Crecen constantemente	
No crecen	
Es pequeño	

5¿Sabe usted en qué ámbitos tiene utilidad Big Data?

Empresarial	
Comercial	
Deportes	
Todas las anteriores	

6.- ¿Sabe usted cómo se manifiestan los datos semiestructurados?

Tienen una estructura muy bien definida	
Tienen unos campos conocidos, pero se desconoce su orden	
Separan los diferentes elementos que los componen, pero no se limitan a unos ciertos campos	

7.- ¿Sabe usted qué permite la asociación en análisis de datos?

Tiene como objetivo encontrar comportamientos predictivos.	
Permite encontrar relaciones entre diferentes variables.	
Divide grandes grupos de individuos en grupos más pequeños	
nada	

8.- ¿Ha escuchado a qué se refiere la herramienta NoSql?

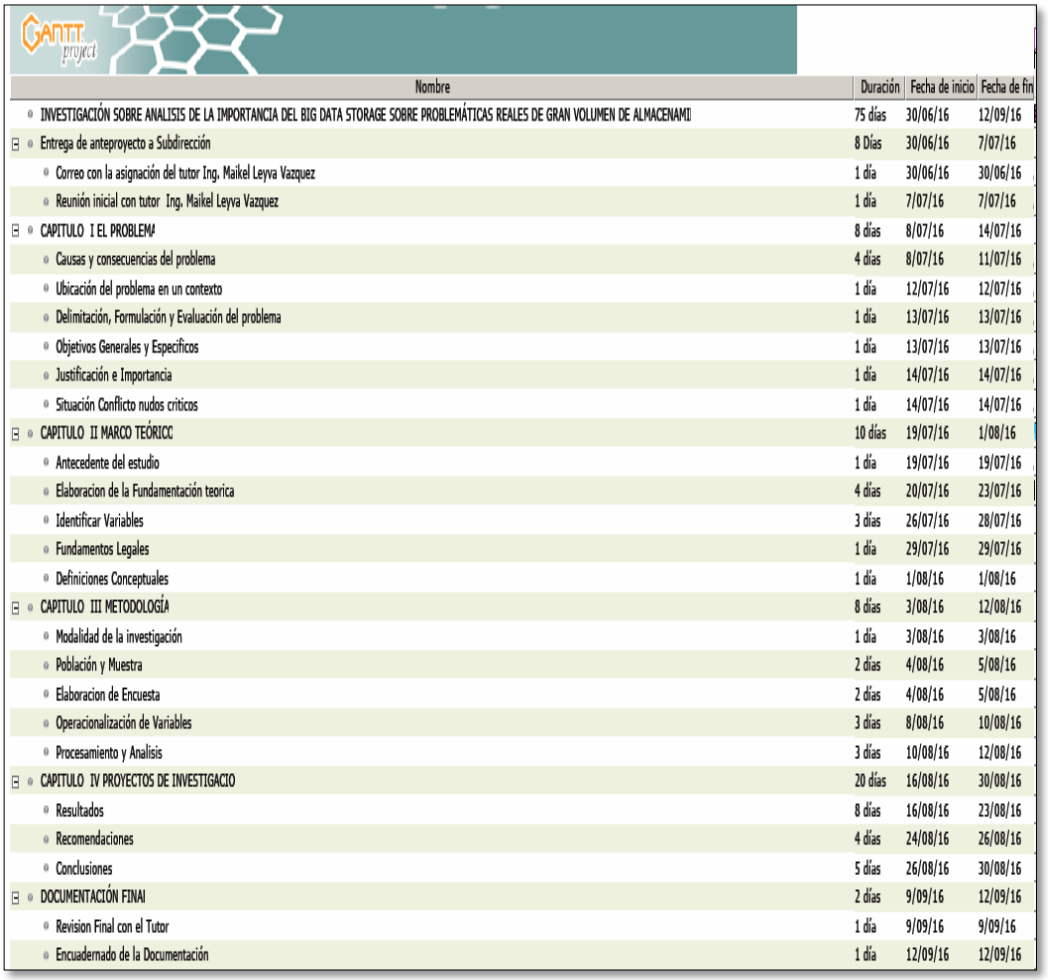
Datos masivos	
Not only sql	
Sistema de almacenamiento	

ANEXO 4

CRONOGRAMA

En el presente cronograma se detallan las actividades realizadas en el Proyecto. Se calcula que el estudio se efectuará en el transcurso de tres meses y se estructura en diferentes capítulos. La ilustración muestra las tareas que se llevaran a cabo, indicando así mismo su duración.

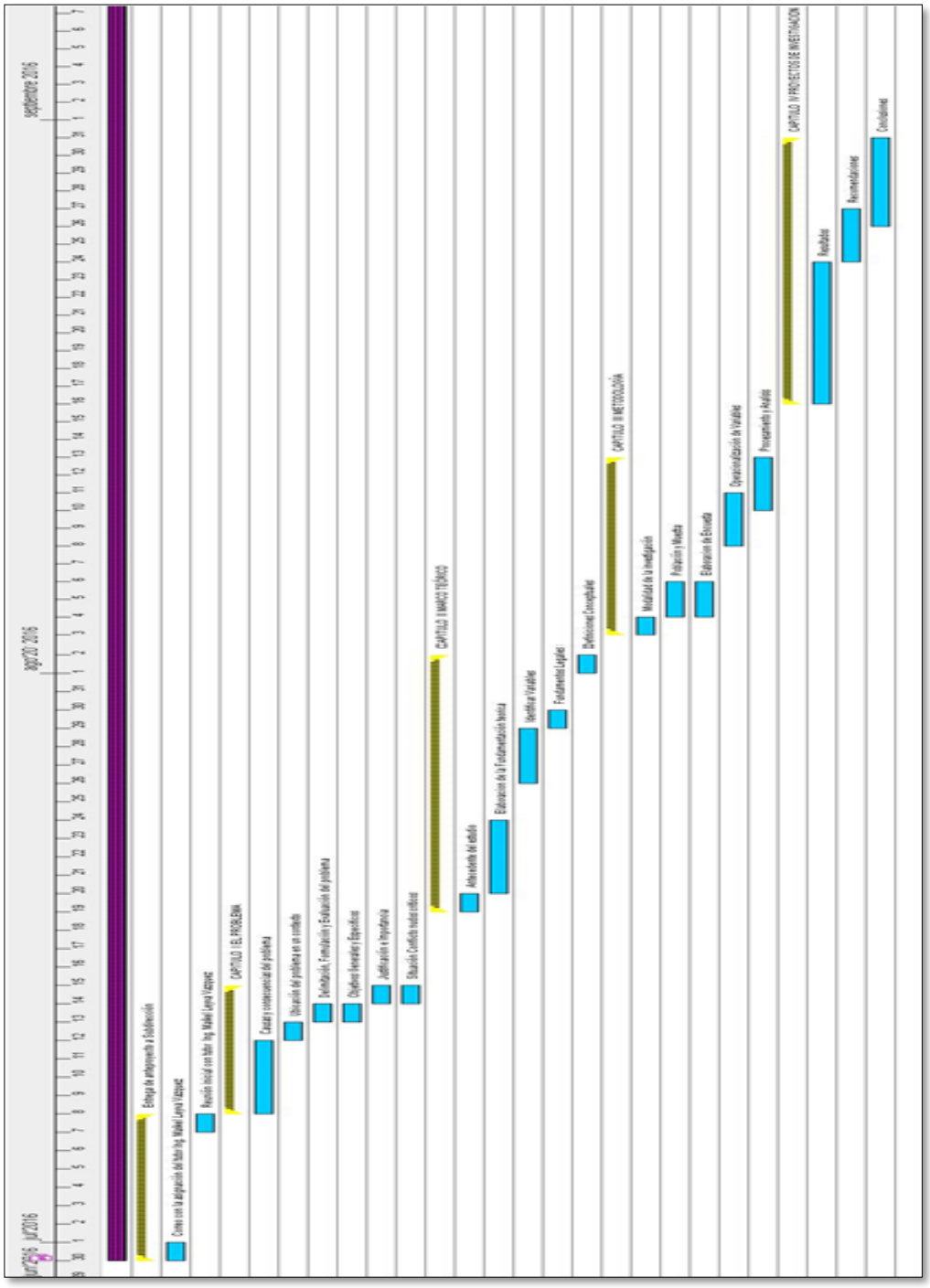
ILUSTRACIÓN. 88
DIAGRAMA DE GANTT DEL PROYECTO



Nombre	Duración	Fecha de inicio	Fecha de fin
INVESTIGACIÓN SOBRE ANALISIS DE LA IMPORTANCIA DEL BIG DATA STORAGE SOBRE PROBLEMÁTICAS REALES DE GRAN VOLUMEN DE ALMACENAMI	75 días	30/06/16	12/09/16
Entrega de anteproyecto a Subdirección	8 Días	30/06/16	7/07/16
Correo con la asignación del tutor Ing. Maikel Leyva Vazquez	1 día	30/06/16	30/06/16
Reunión inicial con tutor Ing. Maikel Leyva Vazquez	1 día	7/07/16	7/07/16
CAPITULO I EL PROBLEMA	8 días	8/07/16	14/07/16
Causas y consecuencias del problema	4 días	8/07/16	11/07/16
Ubicación del problema en un contexto	1 día	12/07/16	12/07/16
Delimitación, Formulación y Evaluación del problema	1 día	13/07/16	13/07/16
Objetivos Generales y Especificos	1 día	13/07/16	13/07/16
Justificación e Importancia	1 día	14/07/16	14/07/16
Situación Conflicto nudos criticos	1 día	14/07/16	14/07/16
CAPITULO II MARCO TEÓRICC	10 días	19/07/16	1/08/16
Antecedente del estudio	1 día	19/07/16	19/07/16
Elaboracion de la Fundamentación teorica	4 días	20/07/16	23/07/16
Identificar Variables	3 días	26/07/16	28/07/16
Fundamentos Legales	1 día	29/07/16	29/07/16
Definiciones Conceptuales	1 día	1/08/16	1/08/16
CAPITULO III METODOLOGÍA	8 días	3/08/16	12/08/16
Modalidad de la investigación	1 día	3/08/16	3/08/16
Población y Muestra	2 días	4/08/16	5/08/16
Elaboracion de Encuesta	2 días	4/08/16	5/08/16
Operacionalización de Variables	3 días	8/08/16	10/08/16
Procesamiento y Analisis	3 días	10/08/16	12/08/16
CAPITULO IV PROYECTOS DE INVESTIGACIO	20 días	16/08/16	30/08/16
Resultados	8 días	16/08/16	23/08/16
Recomendaciones	4 días	24/08/16	26/08/16
Conclusiones	5 días	26/08/16	30/08/16
DOCUMENTACIÓN FINAI	2 días	9/09/16	12/09/16
Revision Final con el Tutor	1 día	9/09/16	9/09/16
Encuadernado de la Documentación	1 día	12/09/16	12/09/16

Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de la Investigación

DIAGRAMA DE GANTT DEL PROYECTO
ILUSTRACIÓN. 89



Elaboración: Fátima Cynthia Vizueta Quinto
Fuente: Datos de Investigación

BIBLIOGRAFÍA

ACENS. (11 de 02 de 2014). bbdd-nosql-wp-acens Bases de datos NOSQL Qué son y tipos que nos podemos encontrar. Recuperado el 11 de 07 de 2016, de Cloud Server: https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf

Alcazar, J. (01 de 02 de 2016). Estadísticas Facebook Ecuador. Recuperado el 02 de 08 de 2016, de Formacion Gerencial: http://blog.formaciongerencial.com/2016/02/01/estadisticasfacebookecuador/

Ambriz, A. (30 de 09 de 2013). Datos estructurados en Big Data. Recuperado el 18 de 07 de 2016, de Almacenamiento de datos estructurados con Big Data: https://www.ibm.com/developerworks/ssa/library/bd-almacenamiento-datos/

Antonio. (16 de 11 de 2011). ¿Qué es NoSQL? Recuperado el 15 de 07 de 2016, de Codecriticon: http://codecriticon.com/que-es-nosql/

Ariely, D. (26 de 07 de 2016). Big Data. Lo que fue, lo que es y lo que será. Recuperado el 02 de 08 de 2016, de Big Data: http://www.intelygenz.es/big-data-lo-que-fue-lo-que-es-y-lo-que-sera/

Ayestarán, V. (s.f.). Trabajamos con todas las bases de datos NoSQL. Recuperado el 22 de 07 de 2016, de NoSQL we care for data: https://www.paradigmadigital.com/lineas-servicio/nosql/

Barranco, R. (18 de 06 de 2012). Todos formamos parte de ese gran crecimiento de datos. Recuperado el 27 de 08 de 2016, de ¿Qué es Big Data?: https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/

Camacho, E. (s.f.). NOSQL. Recuperado el 22 de 07 de 2016, de NoSQL la evolución de las bases de datos: http://sg.com.mx/revista/42/nosql-la-evolucion-las-bases-datos#.V7a271vhDDd

Demassieux, N. (9 de 11 de 2015). Retos del Big Data. Recuperado el 30 de 07 de 2016, de Los 6 retos del Big Data: http://www.orange.com/es/noticias/2015/novembre/Los-6-retos-del-Big-Data

Diaz, W. (27 de 05 de 2013). BASES DE DATOS NOSQL: LLEGARON PARA QUEDARSE. Recuperado el 20 de 07 de 2016, de Bases de datos NOSQL: http://basesdedatosnosql.blogspot.com/2013/05/bases-de-datos-nosql-llegaron-para.html

Dominguez, E. (17 de 01 de 2013). MapReduce: Ejemplo teórico. Recuperado el 03 de 08 de 2016, de Hadoop On The Road:

http://hadoopontheroad.blogspot.com/2013/02/mapreduce-ejemplo-teorico.html

ElEconomista. (28 de 03 de 2016). Cómo sacar partido del Big Data en el negocio. Recuperado el 10 de 08 de 2016, de El Economista: http://www.eleconomista.es/tecnologia/noticias/7448343/03/16/Como-sacar-partido-del-Big-Data-en-el-negocio.html

Fernandez, J. (17 de 04 de 2015). Neo4j - A Graph Database. Recuperado el 28 de 08 de 2016, de A Graph Database: http://es.slideshare.net/JavierdelaRosaFernan/neo4j-47124820

Fernandez, S. (13 de 07 de 2016). La importancia del analista de datos para las empresas. Recuperado el 28 de 07 de 2016, de El Economista: http://www.eleconomista.es/gestion-franquicias/noticias/7701528/07/16/La-importancia-del-analista-de-datos-para-las-empresas.html

Fernández, T. (22 de 07 de 2013). hadoop-taller Introducción a Hadoop. Recuperado el 04 de 08 de 2016, de Páxinas de Eventos: http://eventos.citius.usc.es/bigdata/workshops/hadoop-taller.pdf

Forbes Staff. (19 de 04 de 2016). Esto es lo que pasa en un minuto en internet - Interesante: : http://scl.io/ksU4PC-X#gs.YI93Q=w. Recuperado el 01 de 08 de 2016, de Ultimas Noticias: http://www.forbes.com.mx/lo-pasa-minuto-internet/#gs.YI93Q=w

GERENCIA. (19 de 08 de 2016). ¿El fin del reinado de las bases de datos relacionales? Recuperado el 20 de 07 de 2016, de NoSQL: http://www.emb.cl/gerencia/articulo.mvc?xid=105

González, I., & Carrillo, L. (03 de 08 de 2016). El Dark Data, el lado oscuro del Big Data. Recuperado el 11 de 07 de 2016, de Blogthinkbig.com: http://blogthinkbig.com/el-dark-data-el-lado-oscuro-del-big-data/

GRAPHEVERYWHERE. (s.f. de s.f. de s.f.). Las bases de datos de grafos han venido para quedarse. Recuperado el 28 de 08 de 2016, de No te pierdas este avion: http://www.grapheverywhere.com/language/es/jesus-neo4j-2/

Guirado, R. (10 de 09 de 2014). CIGRAS2014%20-%20Seguridad%20en%20tiempos%20de%20Big%20Dat Seguridad en tiempos de Big Data. Recuperado el 05 de 08 de 2016, de ISACA: http://www.isaca.org/chapters8/Montevideo/cigras/Documents/CIGRAS2014%20-%20Seguridad%20en%20tiempos%20de%20Big%20Data.pdf

ICIJ. (s.f. de s.f. de s.f.). How to download this database. Recuperado el 05 de 08 de 2016, de Offshore Leaks Database: https://offshoreleaks.icij.org/pages/database

INEC. (02 de 08 de 2016). Presentacion_TIC_2015 Tecnologías de la Información y la Comunicación 2015. Recuperado el 10 de 08 de 2016,

de Ecuador ama la vida: http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/2015/Presentacion_TIC_2015.pdf

Instituto DataKey. (20 de 05 de 2015). ¿QUÉ ES BIG DATA? Recuperado el 24 de 07 de 2016, de Archivo de la categoría: Uncategorized: http://www.instituto-datakey.com/category/uncategorized/page/2/

Javier. (18 de 11 de 2015). Tipos de bases de datos y las mejores bases de datos del 2016. Recuperado el 24 de 07 de 2016, de Bases de datos Monitorización: http://blog.pandorafms.org/es/tipos-de-bases-de-datos-y-las-mejores-bases-de-datos-del-2016/

jgaribay. (26 de 05 de 2014). Ejemplo MapReduce Hadoop 1. Recuperado el 04 de 08 de 2016, de Java Mexico: http://www.javamexico.org/blogs/jgaribay/ejemplo_mapreduce_hadoop_1

Kodak. (12 de 07 de 2016). La ventajas del Big Data en los proyectos de servicios públicos digitales. Recuperado el 29 de 07 de 2016, de Soluciones de modernizacion: http://www.kodak.com/ek/kodakgcg/mx/es/modernization-solutions/blog/blog_post/?contentid=4294997771

linuxParty. (09 de 01 de 2011). 5 Pros y 5 contras de cinco bases de datos NoSQL. Recuperado el 18 de 07 de 2016, de linuxParty: http://www.linux-party.com/index.php/6599-5-pros-y-5-contras-de-cinco-bases-de-datos-nosql

LLAVINIA, X. (03 de 02 de 2016). Jordi Griful: "El Big Data generará miles de puestos de trabajo". Recuperado el 09 de 08 de 2016, de Talento Digital: http://www.elperiodico.com/es/noticias/entrevistas-talento-digital/entrevista-jordi-griful-big-data-generara-miles-puestos-trabajo-4867709

Lyon, W., & Hambre, M. (13 de 04 de 2016). Analyzing the Panama Papers With Neo4j: Data Models, Queries, and More. Recuperado el 01 de 09 de 2016, de DZone Spotlight: https://dzone.com/articles/analyzing-the-panama-papers-with-neo4j-data-models

Mireles, r. (24 de 05 de 2016). NoSQL. Recuperado el 22 de 07 de 2016, de Tecnologia en el siglo XXI: http://rosauramireles.jimdo.com/2016/05/24/nosql/

Natividad, P. (02 de 08 de 2016). La importancia del Big Data. Recuperado el 28 de 07 de 2016, de Conexion: http://www.esan.edu.pe/conexion/actualidad/2016/08/02/la-importancia-del-big-data/

Neo4j. (19 de 08 de 2016). 5 Reasons to Choose Neo4j. Recuperado el 27 de 08 de 2016, de Neo4j: https://neo4j.com/

Nishikawa, A. (20 de 09 de 2014). El teorema CAP de Brewer. Recuperado el 22 de 07 de 2016, de Tema 1: Descripción del problema a resolver: http://redmine.nishilua.com/projects/goraexplorer/wiki/Tema_1_Descripci%C3%B3n_del_problema_a_resolver

Nubit. (30 de 03 de 2016). CÓMO USAR BIGDATA Y ANALÍTICA DE NEGOCIO PARA OPTIMIZAR TU EMPRESA. Recuperado el 12 de 08 de 2016, de Nubit: http://www.nubit.es/como-usar-bigdata-y-analitica-de-negocio-para-optimizar-tu-empresa/

Olmedo, J. (20 de 08 de 2012). ¿Qué es MapReduce? Recuperado el 03 de 08 de 2016, de Big Data: http://blogs.solidq.com/es/big-data/que-es-mapreduce/

Ortoll, E. (10 de 2014). Big Data se escribe con V. Recuperado el 03 de 08 de 2016, de Esrudios de cincias de la Informacion y Comunicacion: http://www.uoc.edu/divulgacio/comein/es/numero37/articles/Article-Eva-Ortoll.html

Paramio, C. (26 de 04 de 2011). Pero, ¿en qué se diferencian exactamente? Recuperado el 14 de 07 de 2016, de El concepto NoSQL, o cómo almacenar tus datos en una base de datos no relacional: http://www.genbetadev.com/bases-de-datos/el-concepto-nosql-o-como-almacenar-tus-datos-en-una-base-de-datos-no-relacional

Pereira, M. (15 de 06 de 2011). ¿Qué son las bases de datos NOSQL? Recuperado el 14 de 07 de 2016, de ¿Qué son las bases de datos NoSQL?: http://manuelpereiragonzalez.blogspot.com/2011/05/que-son-las-bases-de-datos-nosql.html

Perez, J. (26 de 06 de 2012). LOS TIPOS DE INVESTIGACIÓN. Recuperado el 25 de 08 de 2016, de ASESORIA DE TESIS Y TRABAJOS DE GRADO: http://asesoriatesis1960.blogspot.com/2012/06/los-tipos-de-investigacion.html?m=0

Perez, S. (30 de 03 de 2015). CLAVES PARA ELEGIR TU BASE DE DATOS NOSQL. Recuperado el 24 de 07 de 2016, de NOSQL: https://www.ondho.com/claves-para-elegir-tu-base-de-datos-nosql/

Perez, V. (29 de 09 de 2011). ¿Qué es el Big Data? Recuperado el 26 de 07 de 2016, de ANALITICA WEB: http://www.analiticaweb.es/que-es-big-data/

Power Data. (16 de 04 de 2014). C-mo-afecta-el-Big-Data-a-tu-d-a-a-d-a El valor de la gestión de datos. Recuperado el 18 de 07 de 2016, de El valor de la gestión de datos: http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/381766/C-mo-afecta-el-Big-Data-a-tu-d-a-a-d-a

PowerData. (12 de 09 de 2013). ¿Qué es el Big Data y por qué es tan importante? Recuperado el 28 de 07 de 2016, de El valor de la gestión de datos: http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/328884/Qu-es-el-Big-Data-y-por-qu-es-tan-importante

Rayo, A. (21 de 07 de 2016). La seguridad: parte central de cualquier proyecto. Recuperado el 10 de 08 de 2016, de Principales retos de seguriidad para proyecto Big Data: http://www.netmind.es/knowledge-center/principales-retos-seguridad-big-data/

Real-Time, T. I. (2016). The Internet in Real-Time. Obtenido de interactive-gif-internet-in-real-time-what-happens-every-second-on-web-infographi: http://3.bp.blogspot.com/-voc74Tzs4XU/U4OS8NOYTXI/AAAAAAAAegs/9v6MyBrJbPc/s1600/interactive-gif-internet-in-real-time-what-happens-every-second-on-web-infographic.gif

Redhat. (s.f.). 5 características de una implementación efectiva de big data. Recuperado el 06 de 08 de 2016, de PERCEPCIONES E IDEAS: https://www.redhat.com/es/insights/big-data

Rivas, A. (29 de 11 de 2012). Técnica, herramienta, método y metodología. Recuperado el 25 de 08 de 2016, de MUNDO INFORMATICO: http://mundoinformatico321.blogspot.com/2012/11/tecnica-herramienta-metodo-y-metodologia.html

Rivero, S. (08 de 08 de 2016). Descubre cómo tu empresa puede ser más efectiva gracias al Big Data. Obtenido de Universia: http://noticias.universia.es/practicas-empleo/noticia/2016/08/08/1142481/descubre-como-empresa-puede-efectiva-gracias-big-data.html

Rodriguez, E., & Conesa, J. (17 de 12 de 2015). M4-NoSQL-Caracteristicas Bases de datos NoSQL Caracteristicas. Recuperado el 16 de 07 de 2016, de Miriadax: https://miriadax.net/documents/54270034/56288160/M4-NoSQL-Caracteristicas.pdf

Salazar, H. (09 de 11 de 2009). Investigacion Bibliografica. Recuperado el 23 de 08 de 2016, de Investigacion Bibliografica: http://es.slideshare.net/HernanSalazar/investigacin-bibliogrfica-2463165

Salvador, F. (20 de 03 de 2014). Big%20Data%20ESP%207 Big data: ¿La ruta o el Destino ? Recuperado el 14 de 07 de 2016, de International Excellence: http://www.ie.edu/fundacion_ie/Comun/Publicaciones/Publicaciones/Big%20Data%20ESP%207.pdf

Sanchez , J. (02 de 07 de 2015). ¿QUÉ ES BIG DATA? Recuperado el 27 de 08 de 2016, de ¿QUÉ ES BIG DATA?: http://www.t2omedia.com/ideas/actualidad/que-es-big-data-2/

Sanchez, J. (04 de 04 de 2014). Instalación de Neo4j y un primer vistazo al Neo4j Browser. Recuperado el 02 de 09 de 2016, de Xurxo Developer: http://xurxodeveloper.blogspot.com/2014/04/neo4j-instalacion-y-neo4j-browser.html

Sánchez, J. (30 de 03 de 2014). Neo4j una base de datos NoSQL orientada a grafos. Recuperado el 28 de 07 de 2016, de Xurxo Developer: http://xurxodeveloper.blogspot.com/2014/03/neo4j-una-base-de-datos-nosql-orientada.html

Santacruz, I. (13 de 08 de 2013). BigData y su ecosistema de bases de datos: NoSQL y RDBMS. Recuperado el 11 de 07 de 2016, de Tecnología e informática: https://ismasantacruz.wordpress.com/2013/08/13/bigdata-y-su-nuevo-ecosistema-de-bases-de-datos-nosql-y-rdbms/

SAT. (13 de 08 de 2014). ¿Qué es el Big Data? Recuperado el 01 de 08 de 2016, de SAT: http://www.fundacionctic.org/sat/articulo-que-es-el-big-data

Sevilla, E., & Salas, D. (s.f.). Diseño+de+una+arquitectura+para+Big+Data Dise˜no de una arquitectura para Big Data. Recuperado el 03 de 08 de 2016, de Base de Datos de Business Intelligence: https://lpi-businessintelligencedb.wikispaces.com/file/view/Dise%C3%B1o+de+una+arquitectura+para+Big+Data.pdf

Smith, C. (s.f.). Big Data. Recuperado el 29 de 07 de 2016, de Big Data: http://www.academia.edu/9252520/Big_Data

Solis, S. (23 de 04 de 2013). Herramientas para Big Data. Neo4J (Parte 1). Recuperado el 28 de 08 de 2016, de Santiago Márquez Solís: http://www.santiagomarquezsolis.com/herramientas-para-big-data-neo4j-parte-1/

Strappazzon, N. (12 de 06 de 2016). El teorema CAP en base de datos. Recuperado el 21 de 07 de 2016, de El teorema CAP: https://www.swapbytes.com/teorema-cap-base-datos/

TERASWAP. (12 de 05 de 2014). NoSql: Una nueva forma de base de datos – Parte I. Recuperado el 18 de 07 de 2016, de NOSQL: http://www.blog.teraswap.com/nosql-introduccion-parte-i/

THEBIGGESTDATA. (19 de 04 de 2016). Tipos de Big Data: diferentes clasificaciones. Recuperado el 03 de 08 de 2016, de Big Data de los numeros a la informacion: https://thebiggestdata.wordpress.com/2016/04/19/tipos-de-big-data-diferentes-clasificaciones/

Trazada. (07 de 05 de 2015). Los cinco grandes retos del Big Data. Recuperado el 01 de 09 de 2016, de TRAZADA: http://trazada.com/cinco-retos-big-data/

Trigoso, H. (05 de 2016). Bases de datos NoSQL. Recuperado el 22 de 07 de 2016, de Kawcode: http://kawcode.com/novedades/bases-de-datos-nosql/

Vera, R. (01 de 01 de 2016). Big Data. Recuperado el 26 de 08 de 2016, de Big Data: http://es.calameo.com/read/0047891339338242aabb5

YERO, Y. (04 de 12 de 2013). MapReduce en el procesamiento distribuido de grandes volúmenes de información. Recuperado el 03 de 08 de 2016, de Revista vinculando: http://vinculando.org/beta/revision-modelo-programacion-mapreduce.html

Zapata, E. (04 de 04 de 2013). SQL vs. NoSQL ¿Cuál es el mejor? Obtenido de SQL vs. NoSQL: https://estebanz01.wordpress.com/2013/04/04/sql-vs-nosql-cual-es-el-mejor/