Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría de Tecnologías de la Información y Comunicación

ANÁLISIS DE BIG DATA BASADO EN REPRESENTACIONES DE DIAGRAMAS GRÁFICOS EN UN AMBIENTE DE REALIDAD VIRTUAL

Ing. Otto Efraín Anaya López
Asesorado por el Ing. Msc. Edwin Estuardo Zapeta Gómez

Guatemala, noviembre de 2020

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

ANÁLISIS DE BIG DATA BASADO EN REPRESENTACIONES DE DIAGRAMAS GRÁFICOS EN UN AMBIENTE DE REALIDAD VIRTUAL

TRABAJO DE GRADUACIÓN

PRESENTADO A JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ING. OTTO EFRAÍN ANAYA LÓPEZ
ASESORADO POR EL ING. MSC. EDWIN ESTUARDO ZAPETA GÓMEZ

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN

GUATEMALA, NOVIEMBRE DE 2020

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Cristian Moisés de la Cruz Leal
VOCAL V	Br. Kevin Armando Cruz Lorente
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Ing. Edgar Darío Álvarez Cotí
EXAMINADOR	Ing. Marlon Antonio Pérez Türk
EXAMINADORA	Inga. Gabriela María Díaz Domínguez
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

ANÁLISIS DE BIG DATA BASADO EN REPRESENTACIONES DE DIAGRAMAS GRÁFICOS EN UN AMBIENTE DE REALIDAD VIRTUAL

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 30 de marzo de 2019.

Ing. Otto Efraín Anaya López

USAC
TRICENTENARIA
Universidad de San Carlos de Guatemala

Decanato
Facultad de Ingeniería
24189101 - 24189102

DTG. 390.2020.

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Postgrado, al Trabajo de Graduación titulado: **ANÁLISIS DE BIG DATA BASADO EN REPRESENTACIONES DE DIAGRAMAS GRÁFICOS EN UN AMBIENTE DE REALIDAD VIRTUAL,** presentado por el Ingeniero **Otto Efraín Anaya López,** estudiante de la **Maestría en Tecnologías de la Información y Comunicación** y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
DECANA
FACULTAD DE INGENIERÍA

Inga. Anabela Cordova Estrada
Decana

Guatemala, noviembre de 2020.

AACE/asga

ESCUELA DE ESTUDIOS DE
POSTGRADO
FACULTAD DE INGENIERÍA

Guatemala, Noviembre de 2020

EEPFI-1504-2020

En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen y verificar la aprobación del Revisor y la aprobación del Área de Lingüística al Trabajo de Graduación titulado: **"ANÁLISIS DE BIG DATA BASADO EN REPRESENTACIONES DE DIAGRAMAS GRÁFICOS EN UN AMBIENTE DE REALIDAD VIRTUAL"** presentado por el Ingeniero **Otto Efraín Anaya López** quien se identifica con Carné **201020696** correspondiente al programa de **Maestría en Artes en Tecnologías de la Información y la Comunicación** apruebo y autorizo el mismo.

Atentamente,

"Id y Enseñad a Todos"

DIRECCIÓN

Mtro. Ing. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Universidad de San Carlos de Guatemala

Guatemala, Noviembre de 2020

Como Coordinador de la **Maestría en Artes en Tecnologías de la Información y Comunicación** doy el aval correspondiente para la aprobación del Trabajo de Graduación titulado: **"ANÁLISIS DE BIG DATA BASADO EN REPRESENTACIONES DE DIAGRAMAS GRÁFICOS EN UN AMBIENTE DE REALIDAD VIRTUAL"** presentado por el Ingeniero **Otto Efraín Anaya López** quien se identifica con Carné **201020696.**

Atentamente,

"Id y Enseñad a Todos"

Mtro. Ing. Marlon Antonio Pérez Türk
Coordinador de Maestría
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Universidad de San Carlos de Guatemala

Guatemala, 18 de Noviembre de 2020

EEPFI-1505-2020

En mi calidad como Asesor del Ingeniero **Otto Efraín Anaya López** quien se identifica con Carné **201020696** procedo a dar el aval correspondiente para la aprobación del Trabajo de Graduación titulado: **"ANÁLISIS DE BIG DATA BASADO EN REPRESENTACIONES DE DIAGRAMAS GRÁFICOS EN UN AMBIENTE DE REALIDAD VIRTUAL"** quien se encuentra en el programa de **Maestría en Artes en Tecnologías de la Información y la Comunicación** en la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala.

Atentamente,

"Id y Enseñad a Todos"

MSc. Ing. Edwin Estuardo Zapeta Gómez
Asesor

ACTO QUE DEDICO A:

Dios Por las bendiciones recibidas en todo momento, sé que ha estado ahí desde siempre y guía mi caminar todos los días desde que salgo de casa hasta que regreso.

Mis padres Mi padre, Otto Anaya Gallardo (q.e.p.d), por enseñarme que trabajando duro se puede alcanzar el éxito y que la honestidad y la honradez son herramientas para llegar más alto. Mi madre, Mirna López, que me ha acompañado toda la vida y siempre me ha aconsejado, no se ha apartado de mí y siempre está ahí para apoyarme en nuevas metas creyendo que las puedo alcanzar.

Mi abuela María Delfina López, por siempre aconsejarme a que siga el camino del bien y estar ahí apoyándome y siempre orando por mí, uno de los motores por quien sigo adelante.

Mi hermano José Anaya, por su apoyo incondicional, a pesar de que cada uno tiene caminos separados para hacer las cosas, coincidimos en los valores que nos fueron enseñados.

Mi familia A cada uno que me ha apoyado no solo en esta etapa de mi carrera profesional, sino desde mucho antes.

Familia Rubio Vidal Que me enseñaron que una familia no está formada solo por los lazos de sangre, sino por todo aquel que te acompaña en los buenos momentos y en los no tan buenos.

Mis amigos Con quienes perseveramos y fuimos fuertes para afrontar los cursos de maestría, sabiendo que, a pesar de las arduas labores siempre compartimos el objetivo de culminar esta etapa.

AGRADECIMIENTOS A:

Dios Porque a Él debo mi existir y cada día me da la fuerza, la sabiduría y el entendimiento para salir adelante a pesar de las circunstancias me hace ver el camino correcto.

Mi asesor Por compartir sus conocimientos sin egoísmo, guiarme y acceder al asesoramiento de esta tesis, agradezco al Ingeniero Estuardo Zapeta.

Mis catedráticos Por la dedicación que pusieron en cada curso para transmitir su sabiduría y lecciones aprendidas, además de mostrarnos que el estudiar no solo es leer y memorizar, sino aprender y aplicar en nuestras labores el conocimiento adquirido.

Universidad de San Carlos de Guatemala Por brindarme una nueva oportunidad de crecer académicamente. A la Facultad de Ingeniería por albergarme durante los cursos y a la unidad de postgrado que aceptó mi petición de ampliar mis conocimientos y criterios con esta maestría.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

FIGURAS

TABLAS

LISTA DE SÍMBOLOS

Símbolo	Significado
2D	Dos dimensiones
3D	Tres dimensiones
HDFS	*Hadoop Distributed File System*
KB	*Kilobytes*
MB	*Megabytes*
ORC	*Optimized Row Columnar*
SOAP	*Simple Object Access Protocol*
SQL	*Structured Query Language*
RCFile	*Record Columnar File*
REST	*Representational State Transfer*
RV	Realidad virtual
TV	Televisión
UDF	*User Defined File*

GLOSARIO

API	*Application Programming Interface.*
Arquitectura	En informática se considera como el diseño inicial con componentes definidos y descritos para la construcción de un sistema de información.
Big Data	Conjunto de datos que tienen un volumen, variabilidad y velocidad de crecimiento a grandes escalas, pueden ser estructurados y no estructurados.
CISCO	*CISCO Systems* es una empresa con sede en San José, California, dedicada a la manufactura, venta, consultoría y mantenimiento de telecomunicaciones.
Cardboard	Gafas de realidad virtual hechas de cartón y con dos lentes de distancia focal, que se combinan con un teléfono móvil para crear un ambiente de realidad virtual.
Docker	Proyecto de código abierto que permite el despliegue de aplicaciones en contenedores o ambientes separados según un lenguaje o

	tecnología.
Exabyte	Su símbolo es EB y se estima que equivale a 10^{18} *bytes*.
FODA	Herramienta que sirve para estudiar las situaciones de una empresa, proyecto, producto en la cual se analizan sus fortalezas, oportunidades, debilidades y amenazas.
Google	Motor de búsqueda más utilizado por los navegadores e internautas para la búsqueda de información, en ella se indexan la mayor parte de información de páginas web.
Joystick	Dispositivo periférico generalmente utilizado para videojuegos, en el caso de RV utilizado para la navegación dentro del ambiente.
no-SQL	Sistema de base de datos que no sigue una conducta relacional ni tiene estructuras fijas como tablas bien definidas; por el contrario, puede contener datos no estructurados o semiestructurados.
Oculus Go	Sistema de RV conformado por un par de gafas y otros dispositivos periféricos para el control del usuario dentro del ambiente, tiene diferentes aplicaciones como educación y entretenimiento, además que posee un sistema configurable y

	programable para un usuario desarrollador.
Petabyte	Su símbolo es PB y se estima que equivale a 10^{15} *bytes*.
Prototipo	Primera versión de un producto con características fundamentales que muestra cual será la aproximación al mismo terminado.
Realidad virtual	Entorno de escenas y objetos que parecen reales, pero no lo son, estos están generados a través de tecnología informática.
SDK	Software Development Kit, es un conjunto de herramientas de software que tiene la finalidad de facilitar al desarrollador la creación de aplicaciones.
Terabyte	Su símbolo es TB y se estima que equivale a 10^{12} *bytes*.
Treemap	Estructura de datos semejante a un árbol en la cual la información se deriva en sus ramas y hojas para organizar mejor la información que contiene.
UNESCO	*United Nations Educational, Scientific and Cultural Organization*

RESUMEN

El gráfico, ya sea científico, estadístico o de otra clase, representa un trozo de información del cual se desprenden muchas interpretaciones y apelaciones a las cuales un expositor puede expresar al público con cientos de palabras, o bien la persona quien lo observa, puede de una forma percibir los conceptos y hacer comparaciones para extraer conclusiones.

Los diagramas gráficos ayudan a extraer y manipular datos utilizando el sistema de análisis visual y permiten correlacionar información para finalmente obtener ese conocimiento y emitir una conclusión sobre los datos analizados. A este punto, se sabe que los diagramas gráficos pueden contener errores o pueden ser manipulados de forma errónea, en algún momento puede ser convincente y brindar la seguridad que la información reflejada en ellos es real.

La exploración de los gráficos y su estructura puede beneficiar al usuario que se encuentra en un ambiente semi o completamente inmerso ya que existe una región en el cerebro que conecta los datos con los gráficos abstractos de la habitación virtual en la que se encuentra (Bellgardt *et al*., 2017). El problema radica en la falta de agudeza en la percepción visual del ser humano y lograr la distinción en un gráfico en 2D de datos muy grandes y otros muy pequeños, así también la saturación de un gráfico cuando este excede un cierto límite de categorías, el usuario no es capaz de tomar todos los detalles o tiende a confundir las cantidades porcentuales asignadas a cada diagrama.

Por ello, el objetivo general de este proyecto es diseñar e implementar una arquitectura capaz de generar diagramas gráficos en un ambiente de realidad virtual (RV), los datos son extraídos de la tecnología *Big Data* y analizados a través de sus diferentes herramientas. Los resultados son consumidos a través de

servicios RESTful montado a través de microservicios, estos llegan al dispositivo de RV inmersivo y genera 10 diferentes gráficos, como pie, barras, anillo y dispersión. El ambiente de RV es manejado a través de un control remoto que utiliza *Bluetooth* para conectarse al dispositivo móvil que renderiza los gráficos.

En sí, el experimento involucra a 25 voluntarios de pruebas, los cuales fueron sometidos a dos pruebas: la primera consiste en que los voluntarios leen una serie de 10 gráficas en 2D e indican al evaluador las cantidades aproximadas que cada gráfico tiene, la segunda parte consta de las mismas 10 graficas solo que en el ambiente inmersivo de RV, en donde le usuario tiene la oportunidad de explorar los gráficos desde diferentes ángulos mediante el control remoto, acercarse y alejarse y tomar la perspectiva que mejor le convenga y le permita asociar los conceptos que ve.

Según el experimento, se encontró que la interacción entre las personas y los gráficos mejora en algunos aspectos la toma de decisiones, pero también se requiere experiencia para manejarse dentro de un mundo virtual. Así pues, los resultados lanzaron que los niveles de confianza y exactitud del análisis 3D se mantienen en algunos casos, pero pueden aumentar. Para este estudio se debe mejorar la navegabilidad del usuario dentro del mundo virtual, que le permita tener nuevas perspectivas y dominio sobre los gráficos.

PLANTEAMIENTO DEL PROBLEMA Y FORMULACIÓN DE PREGUNTAS ORIENTADORAS

Miles de millones de datos son registrados en todo momento del día, cada 60 segundos se manejan alrededor de 2 millones de búsquedas en Google, se generan 571 páginas, se envían 204 millones de correos electrónicos. Según datos de Cisco diariamente existen *Exabytes* de información transmitida por internet, lo cual representa un volumen enorme de datos a los que se recurre para convertirla en información.

Para dicha extracción y análisis de información la solución es *Big Data* que es un conjunto de datos estructurados y no estructurados con variabilidad, difíciles de manipular cuando no existe una organización básica, por ejemplo, datos generados por geolocalización, sensores de temperatura, información generada desde nuestros dispositivos móviles. Es considerado como Big Data un volumen desde los 50 *Terabytes* hasta varios *Petabytes*.

De *big data* surge información que es utilizada por compañías para ser analizada y utilizada para estrategias de venta, pero por la gran cantidad de datos es casi imposible localizar uno en específico y es tedioso leer largas listas de estos para concluir en un punto y tomar una decisión con base a ese dato. Por lo tanto, para estos casos existe la técnica llamada visualización, la cual utiliza representaciones gráficas sobre los datos previamente analizados y que a través de tablas, imágenes, diagramas y gráficos fáciles de leer simplifica el proceso de entendimiento de estos en la cual el usuario puede realizar comparaciones de diferentes datos mediante la capacidad de percepción visual.

Dentro de los gráficos existe una serie que con frecuencia es más utilizada para representar datos, entre los cuales se puede mencionar: *treemap*, coordenadas polares, diagramas circulares, diagramas de barras, gráficas de flujos de datos, diagramas de pie.

Los anteriores mencionados son comúnmente utilizados en sitios *web*, presentaciones para la gerencia, documentos, entre otros, pero algunas veces al comparar datos muy grandes con otros muy pequeños, se tiende a perder la perceptibilidad del gráfico siendo este un problema para comprender los datos y perder esos pequeños detalles, hasta que se hace un acercamiento a los mismos y no todas las representaciones planas en papel o pantallas ofrecen esta funcionalidad.

Otro problema común sucede cuando los datos tienen diferentes series, es decir varias clasificaciones o categorías, la cantidad de barras o segmentos de las gráficas tienden a crecer, la visibilidad de los datos comienza a perderse y nuestra capacidad humana perceptual sobre la imagen disminuye en cuanto aumentan los datos desplegados, por lo tanto, la comprensión de lo que se analiza baja. A continuación, se muestra un análisis sobre las herramientas actuales de visualización para *Big Data.*

Tabla I. **Análisis FODA de las herramientas de visualización de *Big Data***

Fortalezas	**Oportunidades**
• Capacidad de interacción con los datos por parte del usuario. • Capacidad de visualización de diferentes tipos de datos. • Variedad de representaciones gráficas para análisis de datos.	• Visualización inmersiva con realidad virtual resulta en una mejor percepción geométrica de datos y más intuitivo entendimiento de datos. • El intrínseco patrón humano de reconocimiento (descubrimiento visual) esta habilidad puede ser maximizada.
Debilidades	**Amenazas**
• Hay espacios para mejorar grandes cantidades de datos con las tres V (volumen, velocidad y variedad). • Existen capacidades de mejorar la experiencia del usuario a través de análisis de datos y gráficos.	• No es posible visualizar adecuadamente muchos datos generados a través de Big Data.

Fuente: elaboración propia.

Si se aplica la técnica de visualización de datos y la combinamos con un área poco estudiada, pero con gran potencial, que es la inmersión por realidad virtual (RV) que permitiría representar datos de forma geométrica en lugar de la forma tradicional. La visualización inmersiva podría llevar a las personas que analizan *big data* a tener un mejor entendimiento de los datos que están utilizando para un análisis.

Las áreas de aplicación de realidad virtual son varias: arte, al permitir la proyección de este a través de dispositivos periféricos como *cardboard* o dispositivos especializados para celular. Educación, en donde se imparte enseñanza a través de la creación de mundos virtuales con lecciones de aprendizaje. Entretenimiento, entre estos los videojuegos, que al combinarse con controles como un *joystick* permiten crear una experiencia de usuario mayor a la que se tiene con el videojuego en una televisión. La medicina también se ha visto beneficiada de esta tecnología en donde se utiliza para dar entrenamiento en cirugías, en el tratamiento de pacientes que necesitan rehabilitación física y terapias para enfermedades mentales.

Teniendo en cuenta las capacidades, ventajas y el gran potencial en otras áreas que ofrece la RV y la combinamos con tecnologías como *Big Data* y las técnicas de visualización podríamos ser capaces de mejorar la experiencia del usuario al analizar muestras muy grandes de datos, así como muy pequeñas y realizar comparaciones entre las mismas.

La pregunta central es: ¿qué sistema de información permite la proyección de datos generados por *big data* de forma geométrica e interactiva en RV? Se hace una búsqueda de software que permita el mejor análisis a través de gráficos geométricos en 3D.

La siguiente pregunta es: ¿qué tipos de tecnología existen para la generación de representaciones gráficas interactivas de *Big Data* a través de RV?, si en verdad existe esta tecnología o tiene alguna oportunidad de mercado al desarrollarse uno, si podemos hacer más interactivo un análisis o no y si es comprensible al igual que una gráfica en 2D.

La otra pregunta es: ¿qué método permite generar una representación gráfica con RV interactiva de percepción geométrica en *Big Data*?, responde a si existe un método de RV que realice esta función que se quiere evaluar o algún indicio de esta, de lo contrario crear una solución inicial.

La pregunta final es: ¿cuál es la arquitectura de sistemas para combinar RV y *big data* para generar una representación gráfica interactiva de manera eficiente?, esta pregunta responde a si existe una solución desarrollada de lo contrario desarrollar la misma en una etapa inicial, que se pueda construir sin muchos recursos o bien que simule el funcionamiento de un producto ya terminado.

OBJETIVOS

General

Diseñar e implementar una aplicación que permita la proyección de análisis de datos creados por *Big Data* en un ambiente inmersivo de realidad virtual.

Específicos

- Integrar tecnologías inmersivas de RV como *Oculus* con análisis de datos generados por *Big Data*.

- Generar gráficas geométricas con RV a partir de datos generados por *Big Data*.

- Diseñar e implementar una arquitectura que permita la conexión entre tecnología de RV y *Big Data* para representación geométrica de análisis de datos.

MARCO METODOLÓGICO

- Tipo de investigación

Se realiza una investigación de tipo mixto, es decir, en este se evalúan variables cuantitativas y cualitativas. La solución propuesta busca mejorar la forma en la que el usuario logra percibir los diagramas gráficos no solamente en 2 dimensiones, sino la influencia que tiene el sentido de profundidad en la percepción e interpretación de datos. El punto final y uno de los objetivos es también probar que se pueden crear nuevos artefactos, innovadores que sean capaces de no solo entretener, sino ayudar en temas tan importantes como la toma de decisiones con base en diagramas gráficos provenientes de datos.

Este tipo de investigación tecnológica busca no solo la creación y aplicación de RV, sino también probar que ciertas teorías, como la mejora de la percepción visual en un ambiente en tres dimensiones y que el entender los datos e interpretación de estos también mejora en este mundo virtual.

- Diseño de investigación

Para este trabajo se diseñó un modelo experimental, debido a que existen solamente registros de aplicaciones en áreas como medicina, educación y entretenimiento, entre otros; se tienen muy pocas referencias que realidad virtual haya sido utilizado con anterioridad para la presentación de diagramas gráficos para el análisis de información.

La segunda parte, se centra en la confirmación que, en efecto, el colocar al cerebro en un mundo virtual y su interacción con el mismo permite que el cerebro logre analizar de mejor forma la información que recibe a través de la vista. Por otro lado, también se necesita comprobar que la teoría es aplicable, esto a través de experimentos con sujetos que estén dispuestos a participar de un programa de utilización de prototipos y una serie de cuestionarios que permitirán dar un veredicto al problema planteado y si este es solucionado, a través de técnicas de medición para evaluar que tanto puede los de pruebas interpretar diagramas en un ambiente virtual respecto a diagramas en dos dimensiones.

- Procedimiento metodológico

A continuación, se definen cada una de las fases involucradas para la elaboración de esta investigación, desde sus inicios en planeación hasta la ejecución y pruebas realizadas:

  - Fase I. Observación documental

Se hizo una observación técnica de diferentes áreas como gerencias y personal que practicaron un análisis de datos mediante gráficos, así como su comportamiento con respecto a diversos tipos de gráficos, por ejemplo, pie, barras, gráficos de pila, entre otros; además de comprobar qué sucede cuando los diagramas son difíciles de analizar cuando despliegan varias categorías, que es el principal problema por tratar por la investigación. Esta primera fase tuvo una duración aproximada de 3 semanas.

- Fase II. Revisión documental

Una vez analizada la fase de observación y extraídos los datos se procedió con la investigación de sistemas que sean capaces de generar gráficos en dos y tres dimensiones, y así se construyó una herramienta de recolección de datos, por ejemplo, si las personas conocen el tipo de tecnología de realidad virtual, o bien si tienen alguna pista de este tipo de dispositivos. En esta fase se definieron las herramientas a utilizar, para el desarrollo del proyecto; es decir, hardware y software para el desarrollo del mundo de realidad virtual, métodos de comunicación entre los dispositivos periféricos y el servidor de análisis de datos. La segunda fase tuvo una duración de 3 semanas.

- Fase III. Definición de servicios

Se utilizaron dos técnicas para esta investigación: la primera es la entrevista, esta nos dirá qué tanto saben las personas acerca de las tecnologías de realidad virtual y sus beneficios, esto para sondear si es conveniente o factible utilizar este tipo de tecnología. Esto sucede antes de la implementación del proyecto, es decir en fases previas a realizar los experimentos en una población de muestra. Se estima que la tercera fase duro alrededor de una a una semana y media.

La segunda técnica es un cuestionario, el cual se aplicó a los sujetos de prueba que se sometieron al estudio, es decir refleja que tan bien o mal pueden comprender los datos reflejados en un gráfico en dos dimensiones y en un gráfico en tres dimensiones. Los cuestionarios tienen preguntas con escalas numéricas y preguntas acerca de la experiencia que se tuvo con el prototipo, esto para medir si la comprensión de los sujetos aumenta, se mantiene o disminuye utilizando es prototipo de realidad virtual.

- Fase IV. Desarrollo del sistema

El desarrollo del sistema se basó en la creación del módulo de realidad virtual capaz de representar diagramas gráficos en tres dimensiones, que tenga cierto ángulo de navegación para que el usuario pueda explorar los datos. Un servidor con datos recolectados tendrá un análisis previo de los mismos, es decir datos condensado listos para ser consumidos y reflejados en una gráfica.

El dispositivo de realidad virtual (RV) se conecta a este servidor a través de servicios REST que retornan la información al periférico y este lo genera el diagrama graficó para el usuario. De esta forma, el prototipo refleja un mundo virtual con diagramas gráficos en tres dimensiones capaces de ser manipulados por el usuario de forma limitada, para que le sea más fácil navegar entre ellos y tener diferentes ángulos y perspectivas de visualización. Para la implementación de esta fase se tomaron 26 semanas, esto por la complejidad del desarrollo y la investigación para llevarla a cabo.

- Fase V. Experimento

Para el experimento se ocupó una muestra de 25 personas que se sometieron a pruebas en el sistema para ver diferentes diagramas gráficos en tres dimensiones. También utilizaron un sistema tradicional en gráficos en dos dimensiones, de esta forma se construyó una comparativa entre los dos sistemas y cual es más eficiente. Esta fase tuvo una duración de 3 semanas para la evaluación de los casos.

- Fase VI. Análisis y comparación de resultados

Una vez recolectados los datos del experimento, se procedió a realizar la comparativa de resultados obtenidos, así como dar un veredicto de lo desarrollado en el sistema y si este cumple con las expectativas definidas en un inicio. En esta última fase, se analizaron los resultados en aproximadamente una semana y media.

- Instrumentos de recolección de información

Para reunir información es fundamental la utilización de diversas fuentes, en este caso primarias y secundarias que permitan la obtención de los datos necesarios para la implementación del proyecto.

- Fuentes primarias

  - Búsqueda de datos a analizar mediante páginas que brindan datos abiertos.
  - Equipo de computadora que permite la renderización del mundo virtual.
  - Dispositivos periféricos que generan los diagramas gráficos en el mundo virtual y proyectan los resultados de los análisis.

- Fuentes secundarias

  - Libros sobre temas de realidad virtual y *Big Data*.
  - Publicaciones científicas y artículos en revistas que respalden la investigación.
  - Tesis de maestría y doctorado.
  - Sitios web de documentación para el desarrollo del prototipo.

- Videos e instructivos para implementación de prototipo.

INTRODUCCIÓN

Este trabajo sigue una línea de investigación: dispositivos y sistemas para ampliar la tecnología 3D en imágenes, a través de métodos de visualización aplicados en realidad virtual (RV). Esto mediante un sistema que sea capaz de generar representaciones geométricas sobre gráficas basadas en datos analizados a través de *Big Data* y presentadas en un dispositivo de inmersión de RV.

Como bien es sabido RV tiene muchas áreas de aplicación hoy en día, por ejemplo, educación, en la medicina y entretenimiento con la creación de mundos virtuales aplicados a las categorías mencionadas. Por otro lado, *Big Data*, transforma el paradigma de uso de modelos de base de datos tradicionales a uso de bases no-SQL y modelos híbridos, permite la masificación de datos estructurados, semiestructurados y no estructurados para luego convertirlos en información valiosa para el usuario final en un tiempo prudencial en el que sean útiles.

Dadas las premisas y objetivos anteriores, este proyecto busca combinar dos tecnologías como *Big Data* y RV para generar gráficos de forma geométrica en tres dimensiones el cual permita al usuario navegar a través de estos de forma interactiva. Esta nueva alternativa permitirá al analista tener una mejor perspectiva de lo que está analizando.

Debido a las limitaciones que tienen las técnicas de visualización actuales, entre las que se menciona el crecimiento de las categorías o clasificación de datos a graficar y analizar, los criterios aplicables disminuyen, esto se debe a la

gran cantidad de datos mostrados en un solo gráfico. Otra debilidad de la visualización es que el cerebro humano tiende a perder la percepción de los datos cuando se realizan comparaciones entre dos series una con datos muy grandes y otra con datos muy pequeños, es como comparar a una distancia muy corta la altura entre un edificio de treinta pisos y una casa de tan solo dos, esta diferencia es abismal.

Al ampliar la cantidad de puntos de vista, es posible también ampliar el criterio bajo el cual un usuario evalúa un diagrama gráfico, de esta forma le será más fácil la interpretación de la información que analiza y de igual forma le será más fácil y segura la toma de decisiones.

Las decisiones tomadas a partir la información son extraídas de datos tabulados y gráficas que en ocasiones son difíciles de analizar afectadas por los problemas anteriormente expuestos, este proyecto pretende la combinación de dos tecnologías como *Big Data* para la recolección y análisis de datos, aplicando los beneficios que ofrece la RV a través de sus dispositivos que permiten la inmersión del usuario en un mundo virtual en el cual podrá navegar y obtener una perspectiva geométrica de los datos, en comparación a un diagrama en dos dimensiones, aumentando así la capacidad de aplicar más criterios a las gráficas y obtener mejores decisiones a partir de ellas.

Experimentos realizados por la Universidad de New Brunswick, encontraron que las personas prefieren una representación gráfica en tres dimensiones (Ware y Franck, 1994), como por ejemplo doctores que necesitan hacer cirugías ortopédicas, en lugar de solamente diagramas o representaciones en dos dimensiones, su conclusión fue que más del 40 % de los sujetos del experimento pudieron entender 3 veces más en un gráfico en 3D que las personas que solamente observaron el gráfico en 2D.

En el mundo de las finanzas, la toma de decisiones se realiza con diferentes técnicas y de diferentes fuentes de datos, entre ellos gráficos en 2D que han probado ser efectivos al momento de realizar comparaciones y extraer información y conclusiones (Cheng, 2014). Según el estudio realizado en la Universidad Utrecht por Cheng Xuan, la mezcla de la lectura de gráficos en 2D con un espacio virtual en 3D consigue que las personas encuentren la información de una forma más rápida, además de la velocidad, se mide la exactitud y el nivel de confianza con la que alguien toma una decisión con base a los gráficos.

El trabajo se divide de la siguiente forma: el planteamiento del problema expone cuales son las principales debilidades de la percepción visual durante los análisis gráficos que realizan los usuarios. Los objetivos son uno general y tres específicos relacionados al primero, sobre la arquitectura e implementación, además de cómo están interrelacionadas las 3 tecnologías que conforman el prototipo desarrollado.

El marco metodológico muestra cuales fueron los pasos a seguir para realizar la investigación y el experimento del prototipo, las personas que involucra para las pruebas y cuál fue el *roadmap* a grandes rasgos y cada etapa ejecutada para obtener los resultados finales. Durante la introducción se describe información general bajo la cual se constituyó el proyecto, las generalidades sobre percepción visual y diagramas gráficos, de las cuales arrancó este documento. Los antecedentes por su parte hablan sobre las investigaciones previas de *Big Data* y RV, así como de los diagramas gráficos y cuáles son sus aplicaciones en la vida diaria entre otras cosas.

La justificación explaya los motivos que apoyan a esta investigación y ejecución del experimento, habla sobre la forma en la que este trabajo aporta a investigaciones previas sobre percepción visual y los dispositivos y sistemas para ampliar la tecnología 3D en imágenes. Los alcances describen los resultados esperados del desarrollo e implementación de la arquitectura y prototipo, las áreas técnicas que abarca y los alcances investigativos del mismo y una selección de la mejor forma de implementación.

El marco teórico abarca toda la información recopilada a través de otras investigaciones, artículos, fuentes primarias y secundarias que explican datos sobre *Big Data*, RV y como estas aportan valor a las tareas diarias. Adicional explica las bases de la percepción visual y como esta se relaciona con la realidad virtual y cuál es su correlación en el proyecto. La presentación de resultados es un condensado de datos sobre lo que se obtuvo en el proyecto, desde los diagramas de arquitectura y las especificaciones del software utilizado para la implementación, así como los resultados del experimento realizado con los 25 voluntarios que probaron el prototipo.

Al final del escrito se encuentran la discusión de resultados, las conclusiones y las recomendaciones. La discusión por su parte habla sobre los hallazgos durante el experimento, si se alcanzaron los objetivos y cuál es el veredicto acerca de los resultados y el comportamiento respecto a las expectativas. Las conclusiones responden a cada uno de los objetivos que se escriben en el apartado de objetivos, mencionando si estos fueron alcanzados o no. Y Para terminar las recomendaciones, que como su nombre lo indica, son enunciados que ayudarán a orientar a futuros investigadores sobre qué camino tomar si se decide incursionar en esta rama de la informática y siguientes pasos para complementar el experimento.

1. ANTECEDENTES

Los diagramas gráficos ayudan a extraer y manipular datos utilizando el sistema de análisis visual y permiten correlacionar información para finalmente obtener ese conocimiento y emitir una conclusión sobre los datos analizados. A este punto se sabe que los diagramas gráficos pueden contener errores o manipulaciones erróneas, en algún momento puede ser convincente y brindar la seguridad que la información reflejada en ellos es real.

La exploración de los gráficos y su estructura puede beneficiar al usuario que se encuentra en un ambiente semi o completamente inmerso ya que existe una región en el cerebro que conecta los datos con los gráficos abstractos de la habitación virtual en la que se encuentra (Bellgardt *et al*, 2017), en el mismo camino, la inmersión en el mundo virtual crea en el usuario la sensación de encontrarse dentro de los datos. Los datos pueden ser comprendidos por la persona que los visualiza, siempre y cuando exista una conexión y una pertenencia entre los anteriores mencionados. Esta investigación tiene como objetivo, una forma más eficiente de relacionar datos con el usuario y que le permita tomar mejores decisiones.

Relacionado con los gráficos estadísticos y matemáticos está la educación, según un taller y estudio llevado a cabo por la Universidad de Taiwán (Hwan, Su, Huang y Dong, 2009), estudiantes que utilizaron un sistema virtual de manipulación de objetos en tres dimensiones, fueron capaces de escoger puntos de vista apropiados y resolver un problema geométrico con mayor facilidad, que las personas que únicamente tenían acceso a los modelos en pizarras. En conclusión, un sistema con representaciones geométricas en 3D

puede ser beneficioso no solo para las áreas de educación, sino para toda aquella aplicación de diagramas gráficos en diferentes ámbitos.

En ocasiones los sistemas de estructuras en 2D necesitan mostrar diferentes tipos de asociaciones a los datos para que estos puedan ser comprendidos (Parker, Franck y Ware, 2000), por lo que es necesario distorsionar y ajustar la vista desde diferentes puntos y ángulos para que pueda ser comprendido. El acercar y alejar de forma rápida un gráfico en 3D puede ayudar al usuario a relacionar tamaños de gráficos, nodos y barras y relacionar de forma más sencilla diferentes series de datos, para comprender y concluir en ideas a ser aplicadas.

2. JUSTIFICACIÓN

Este trabajo sigue una línea de investigación: dispositivos y sistemas para ampliar la tecnología 3D en imágenes, a través de métodos de visualización aplicados en realidad virtual (RV). Esto mediante un sistema que sea capaz de generar representaciones geométricas sobre gráficas basadas en datos analizados a través de *Big Data* y presentadas en un dispositivo de inmersión de RV.

Como bien es sabido RV tiene muchas áreas de aplicación hoy en día, por ejemplo, educación, en la medicina y entretenimiento con la creación de mundos virtuales aplicados a las categorías mencionadas. Por otro lado, *Big Data*, transforma el paradigma de uso de modelos de base de datos tradicionales a uso de bases no-SQL y modelos híbridos, permite la masificación de datos estructurados, semiestructurados y no estructurados para luego convertirlos en información valiosa para el usuario final en un tiempo prudencial en el que sean útiles.

Dadas las premisas y objetivos anteriores, este proyecto busca combinar dos tecnologías como *Big Data* y RV para generar gráficos de forma geométrica en tres dimensiones el cual permita al usuario navegar a través de estos de forma interactiva. Esta nueva alternativa permitirá al analista tener una mejor perspectiva de lo que está analizando.

Debido a las limitaciones que tienen las técnicas de visualización actuales, entre las que se menciona el crecimiento de las categorías o clasificación de datos a graficar y analizar, los criterios aplicables disminuyen,

esto se debe a la gran cantidad de datos mostrados en un solo gráfico. Otra debilidad de la visualización es que el cerebro humano tiende a perder la percepción de los datos cuando se realizan comparaciones entre dos series una con datos muy grandes y otra con datos muy pequeños, es como comparar a una distancia muy corta la altura entre un edificio de treinta pisos y una casa de tan solo dos, esta diferencia es abismal.

Al ampliar la cantidad de puntos de vista, es posible también ampliar el criterio bajo el cual un usuario evalúa un diagrama gráfico, de esta forma le será más fácil la interpretación de la información que analiza y de igual forma le será más fácil y segura la toma de decisiones.

Las decisiones tomadas a partir la información son extraídas de datos tabulados y gráficas que en ocasiones son difíciles de analizar afectadas por los problemas anteriormente expuestos, este proyecto pretende la combinación de dos tecnologías como *Big Data* para la recolección y análisis de datos, aplicando los beneficios que ofrece la RV a través de sus dispositivos que permiten la inmersión del usuario en un mundo virtual en el cual podrá navegar y obtener una perspectiva geométrica de los datos, en comparación a un diagrama en dos dimensiones, que aumenta así la capacidad de aplicar más criterios a las gráficas y obtener mejores decisiones a partir de estas.

La técnica de visualización de datos sería la principal beneficiada por la tecnología emergente de RV que ya tiene más aplicaciones en diferentes ramos como: educación, entretenimiento y medicina. RV tiene el potencial de resolver las debilidades de las técnicas de visualización de datos que las herramientas para graficar en dos dimensiones implementan.

Si se aprovecha esta ’relativa’ nueva tecnología y se aplica de la forma adecuada como en los campos de aplicación que se mencionan anteriormente, es posible mejorar el área de análisis de datos y no limitarlo a solo un punto de vista plano sino, a distintos puntos con una ancho, alto y profundidad.

3. ALCANCES

3.1. Resultados

Un prototipo funcional completo con la integración de los módulos de *Big Data*, un dispositivo de realidad virtual y un módulo de microservicios que los interconecte. El prototipo contará con las siguientes funcionalidades:

- Carga de datos al módulo de *Big Data.*
- Módulo de procesamiento y análisis de datos.
- Visualización de gráficos en 3D en el dispositivo de realidad virtual.
- Navegación del usuario sobre diagramas gráficos generados a partir de gráficos dentro del mundo de realidad virtual.

Los resultados que se esperan son:

- Creación de una aplicación que pueda reproducir diagramas gráficos navegables en un mundo virtual de inmersión completa a través de dispositivos de RV y controles.

- Diseño e implementación de un módulo integral, completamente funcional interconectado que combine RV y análisis de *Big Data*.

- Comprobar que la interpretación de datos por parte de un usuario del módulo es mayor en un ambiente de diagramas gráficos en 3D que una gráfica de 2D.

- Implementación de un sistema completamente integrado de análisis de datos con *Big Data* y un dispositivo de realidad virtual a través de arquitectura de microservicios.

3.2. Técnicos

Los alcances técnicos que fueron generados durante el desarrollo de este proyecto son:

- Crear un módulo de generación de diagramas gráficos geométricos en un mundo virtual de inmersión completa.

- Desarrollar un API capaz de conectar un dispositivo de realidad virtual y un banco de procesamiento de *Big Data*, con la definición microservicios a través de servicios REST y SOAP para la extracción de información y consumo de estos desde un dispositivo de realidad virtual.

- Crear un ambiente apto para análisis de datos que aumente el nivel de interpretación de datos a través de la percepción visual en un ambiente 3D.

3.3. Investigativos

A continuación, se presentan los alcances investigativos generados por el diseño e implementación del proyecto y que buscan una finalidad de mejorar la percepción visual y criterio del usuario:

- Investigar, evaluar y clasificar tecnologías de realidad virtual y *Big Data* que logren acoplarse adecuadamente según los requerimientos del proyecto.

- Evaluar el método que permita hacer eficiente la construcción de gráficos con orígenes de datos masivos.

- Evaluar y seleccionar la tecnología de realidad virtual y *Big Data* adecuada en función de la arquitectura que permita la construcción geométrica de diagramas.

4. MARCO TEÓRICO

4.1. Percepción visual y su influencia en el aprendizaje

Existen tantas definiciones de qué es la percepción, una de ellas según Groffman (2006) “la percepción es el proceso activo de localizar y extraer información del ambiente” (p.241). También, es uno de los medios principales de adquisición del conocimiento mediante el aprendizaje.

De la misma forma, el aprendizaje, complementa y facilita el proceso perceptual a través de la experiencia y almacenamiento de la información. Posterior al aprendizaje y el conocimiento se combinan para extraer nueva información, aplicando criterios previamente conocidos, la relación entre procesos se complementan una a la otra, convirtiéndose en un ciclo autosuficiente que permite aumentar las habilidades con base a experiencias adquiridas previamente.

Tomando en cuenta la teoría anterior de Groffman, se asegura que la percepción visual influye mucho durante el ciclo del aprendizaje y por ende en la extracción e interpretación de la información. Según Garzía (1996), la percepción también involucra el entendimiento que llevará a la adquisición de conocimiento, pero para que dicha percepción funcione involucra tres diferentes sistemas: un sistema visoespacial, un sistema de análisis visual y un sistema visomotor.

En resumen, la percepción cubre el concepto de buscar, localizar, centrar y extraer datos los cuales son organizados y procesados para obtener información de los datos obtenidos a través de la vista. Según la pirámide del conocimiento de los datos se extrae información, de esta viene el conocimiento y la acumulación de este desemboca en la sabiduría.

Por lo tanto, un sistema de percepción visual óptimo permite la extracción de datos de mejor manera, haciendo más fácil la comprensión y generación de información por parte del cerebro durante la asociación de conceptos. En otras palabras, si la percepción visual mejora también la organización y procesamiento de información lo harán.

4.2. Problemas de percepción visual frente a diagramas gráficos

Los diagramas gráficos son comunes para la extracción de información ya que son una fuente de datos visual que puede ser analizada. Dentro de los gráficos más comunes utilizados normalmente en periódicos, internet y conferencias encontramos las coordenadas polares, diagramas de barra, diagramas de pie y circulares, así como flujos de datos en vivo o en un lapso.

Pero todos los diagramas anteriores son inútiles al existir un gran conjunto de clasificaciones las cuales no hacen posible a la vista hacer una distinción exacta o extraer información de un gráfico de proporciones muy grandes o pequeñas. Las comparaciones entre datos de gran tamaño frente a datos de poca proporción hacen que la información interpretada por el cerebro humano no sea la correcta.

Otro de los problemas surge cuando existe una diversidad grande de datos que se tienen que comparar en un solo gráfico y que no es posible al ojo

humano interpretar. Pero si aplicamos técnicas de visualización avanzada, por ejemplo, gráficos en tres dimensiones, aumentaremos las probabilidades de los ojos y el cerebro de concentrar, localizar y extraer información de los diagramas gráficos.

4.3. ¿Qué es *Big Data*?

En internet, cada momento del día se está generando información de diferentes índoles, desde registros de nuevos usuarios a páginas *web*, sitios publicados, videos en conocidas páginas de *streaming*, publicaciones en redes sociales. Cisco *Systems* informa que el registro de datos en internet es de varios *Exabytes*, lo cual representa un gran volumen que se podría convertir en información valiosa.

La solución para tratar con esta gran cantidad de datos es *Big Data,* ya que es capaz de manejar diferentes tipos, entre datos estructurados y no estructurados, es decir variables en su organización y forma. Sin embargo, no basta únicamente con almacenar y manipular estos datos, tenemos que obtener un beneficio de ellos y convertirlos en información, la cual a su vez se convierte en sabiduría según la jerarquía del conocimiento.

En sí, no hay una definición de lo que es *Big Data*, anteriormente se decía que era un volumen grande de datos que no era posible de analizar en memoria o incluso imposible de analizar. Una de las ideas de lo que es (*Big Data*) hoy en día es:

> Los *big data*, datos masivos se refieren a cosas que se pueden hacer a gran escala, pero no a una escala inferior, para extraer nuevas percepciones o crear nuevas formas de valor, de tal forma que se

> transforman los mercados, las organizaciones, las relaciones entre los ciudadanos y los gobiernos (Mayer-Schönberger y Cukier, 2013, p.17).

Es decir, en la actualidad este conjunto de datos, al ser analizados, proporcionan información que puede cambiar la forma de vida.

Adicional a ello, si se quiere definir qué es *Big Data*, se debe aclarar que no tiene un software como tal, sino que es la combinación de viejas y nuevas tecnologías que ayudan a las compañías a ampliar su visión respecto a los datos que posee (Hurwitz, *et al*., 2013). Posee la capacidad de manejar un gran volumen de datos a una alta velocidad y realizar análisis en tiempo real.

Big Data se utiliza por diversas empresas para la toma de decisiones, por ejemplo: una empresa dedicada a la venta de productos varios, como Walmart, posee la capacidad de indagar entre sus datos de ventas registradas, formar patrones y tendencias, ya sea para abastecerse de un producto que se sabe es muy vendido en la época, o bien para el aumento/disminución de precios según su conveniencia. Ahora bien, al imaginar la cantidad de ventas que hace esta empresa diaria, mensual y anualmente se puede tener una idea de la cantidad de *petabytes* de información existente. Examinar esta información con una base de datos relacional normal no sería posible, por lo cual se necesita software especializado para manejar este volumen de datos, a una velocidad razonable.

Y esta es precisamente una de las características más notables de *Big Data,* o las llamadas 3V:

- Volumen
- Velocidad
- Variedad´

4.3.1. Volumen

Se refiere a la cantidad de datos que se está generando y almacenando desde aplicaciones y sitios web asociados a empresas. Por ejemplo, sitios en línea como Facebook, Google, cada una de las búsquedas que hacen los usuarios queda registrada, almacenada y posteriormente analizada por parte de estas empresas. Videos, imágenes, *inbox* enviados, generan un gran volumen de datos, a pesar de ello están siempre disponibles y pueden ser accedidos de forma frecuente y concurrente.

4.3.2. Velocidad

La segunda característica de *Big Data* apunta hacia la rapidez en la generación de datos y acceso durante un rango de tiempo. Además, existen diferentes tipos de velocidad. Una de ellas es el tiempo de generación de datos en tiempo real, la velocidad en la que se genera un volumen de datos a partir de datos en línea y fuera de línea, velocidad de transmisión de datos al ser consultados.

4.3.3. Variabilidad

O variedad de datos, se refiere la forma que pueden llegar a tener los datos: estructurados y no estructurados. Los primeros son datos que se encuentran bajo un estándar y siguen un formato definido, por ejemplo, datos obtenidos de tablas de una base de datos relacional, sabemos que tienen una cantidad de campos asignados, estén llenos o no con información, estos siempre tendrán la misma estructura.

Por otro lado, los datos no estructurados son todos aquellos que no poseen una forma fija, en otras palabras, los campos de su estructura son variables. En esta categoría entran textos planos, imágenes, datos obtenidos de redes sociales como *tweets*, publicaciones, correos electrónicos (contenido), audio, videos, entre otros. Toda esta información no tiene una estructura fija por lo tanto no puede ser categorizada bajo una entidad a analizar, sino que antes debe ser tratada mediante otros medios.

4.4. Áreas en las que se utiliza *Big Data*

Big data puede aplicarse en diferentes modelos de negocio: seguros, venta de productos, telecomunicaciones, gobierno, entre otros, Las decisiones tomadas generalmente son de jefaturas como *marketing*, ventas, gerencias y altos mandos.

Una de las aplicaciones más utilizadas por *Big Data* es la segmentación y clasificación de clientes. La obtención de una visión general o de 360 grados sobre el cliente al cual está dirigido nicho de negocio. Si las empresas logran comprender a su cliente, sus comportamientos, necesidades, preferencias. La vista que tenga una empresa de sus clientes es muy importante y de ello depende de cuál será su estrategia para con él y convertirlo de un prospecto de venta a un cliente real de la compañía.

Al continuar analizando, se observa que también las instituciones utilizan *Big Data* para analizar la eficiencia en el desempeño que tiene su personal de acuerdo con datos recolectados de los *logs* en sus computadoras, tareas realizadas y mediciones de software en los cuales está registrado el usuario en su puesto de trabajo.

Otra aplicación de *Big Data* se encuentra en el análisis realizado a los datos generados por los dispositivos que se utilizan actualmente, maquinaria y equipo que poseen sensores inteligentes que miden todo tipo de reacción cuantificable, que genera un volumen de datos con el cual se podrán hacer análisis de diferente índole, como cambios de piezas a la máquina, sondeo de errores medidos, cambios de comportamiento en el rendimiento, entre otros. De esta forma es posible prevenir daños a dispositivos, mejora y optimización de rendimiento e incluso un alargamiento en la vida de este.

En deportes, dispositivos móviles y *gadgets* específicamente, generan volúmenes grandes de datos, como cantidad de pasos, ritmo cardiaco, toma de oxigenación y presión sanguínea, inactividad del usuario, horas de sueño, consumo y quema de calorías, entre otros. De esta forma, empresas que diseñan los dispositivos pueden crear un análisis al usuario de su forma de vida, diseñar controles de hábitos saludables para la persona, programación de ejercicios que estén a su nivel y que mejoren la condición física o bien que se adapten a mejorar la salud de la persona.

Algunos gobiernos utilizan datos para modernizar las ciudades, a través de análisis realizados a cámaras de la ciudad, dispositivos inteligentes que miden la cantidad de personas que entran y salen de un edificio, conteo de vehículos que transitan por una avenida, sensores de temperatura, entre otros. Con base a análisis generados por este gran conjunto de datos, el gobierno puede crear planes para disminución de tránsito, mejora en el acceso a lugares públicos, mejora del transporte y señalización públicos.

Estos solamente son algunos ejemplos de aplicación de *Big Data*, aunque hay muchos más como análisis de datos de entidades públicas de salud,

educación. Las áreas financieras como *trading* y la ciencia también se han visto beneficiadas por este conjunto de tecnologías que les han ayudado a mejorar procesos y a descubrir nuevas formas de optimización.

4.5. *Big Data* como herramienta para análisis de datos

Las herramientas de *Big Data* no se basan únicamente en tener un acceso rápido a los datos, un gran volumen de almacenamiento y una variedad de los mismos, si estos no se les da un uso, simplemente es data almacenada que ocupa un espacio y genera costos el hacerlo.

Con la caja de herramientas adecuada se le puede dar un buen uso, convertir los datos en información y posteriormente en sabiduría. Con ellas se puede hacer el reconocimiento de un nicho de mercado, el segmento de cliente al cual va dirigido un producto, un análisis profundo de *marketing* y la toma de decisiones correctas con base a datos reales y medibles. Dichas decisiones se harán a una mayor velocidad que si tan solo se hiciera con base a presunciones y a investigaciones encontradas en internet, publicadas por otras empresas que ya han hecho una búsqueda previa de datos convertida a información.

Los costos pueden ser reducidos al mismo tiempo que se aumentan las ganancias de una empresa, ya que se puede ver como un producto se desempeña en el mercado, cuál es la aceptación de este entre los clientes e incluso si está generando las ganancias proyectadas de acuerdo con la madurez de este.

Los clientes, al ser una parte fundamental de una compañía son el activo más importante que esta tiene, por lo tanto, es necesario fidelizarlos y tener una

perspectiva global de lo que quiere, lo que necesita y lo que se le puede entregar.

Los análisis que se realizan con este tipo de herramientas y las decisiones que se toman influyen mucho a nivel personal o empresarial de acuerdo con los datos a los que se refiera. Los beneficios son altos al tomar una decisión fundamentada y concreta de datos documentados y almacenados, no guías de investigaciones que no tienen base alguna.

4.6. El ecosistema Hadoop

Hadoop es un proyecto *Open source* conformado, por pequeños módulos o subproyectos que trabajan para un mismo fin: el almacenamiento, procesamiento y análisis de datos a gran escala. Entre las funciones más destacables de *Hadoop* tenemos:

- Optimización de almacenamiento de datos.
- Optimización de acceso a datos (*Map Reduce*).
- Análisis de datos.
- Procesamiento de diferentes tipos de datos (estructurados y no estructurados).
- Minería de datos.
- Sistema de datos distribuidos y escalables.
- Sistema de archivos óptimo para almacenamiento y acomodamiento de datos ingresados.
- Alta disponibilidad.

Hadoop nace por la necesidad de procesamiento de grandes volúmenes de datos a través de clústeres. Se fundamenta en el proyecto de Google File

System y las funciones MapReduce, adicional a ello se dice que tiene tres módulos principales y que soportan el amasamiento de datos: MapReduce para el acceso a datos, Hadoop Distributed File system; como soporte fundamental del almacenamiento de las cadenas de datos generadas por *MapReduce* y que brinda un alto rendimiento de lectura y escritura. Y finalmente Hadoop Common, que encierra las utilidades, aplicaciones y scripts necesarios para la construcción de inicio del sistema *Hadoop* en general, coordina el funcionamiento del sistema operativo y sistema de archivos (Ghemawat, *et al*, 2003).

4.6.1. Variabilidad

MapReduce es un paradigma por la tecnología *Hadoop* y es aplicado en el análisis y procesamiento de grandes volúmenes de datos en forma paralela. Este surge en los años 80 de acuerdo con modelos de sistemas distribuidos basados en álgebra lineal. Las ideas surgieron de a partir de cómo Google maneja el llamado *PageRank* con el cual despliegan sus resultados de búsquedas, de primera mano se sabe que *Hadoop* utilizó este sistema para apoyar el proyecto de búsqueda llamado *Nutch*.

MapReduce se basa en dos funciones características: *Map* y *Reduce*. Utiliza búsquedas a través de método llave-valor para búsquedas e inserciones, (Aragundi *et al*, 2009). A continuación, se describe cómo funcionan los métodos de MapReduce.

La primera, *Map*, funciona sobre grandes cantidades de datos, estos son divididos por partes llamados paquetes, por lo general de 64KB con los registros dentro de ellos. Esta función calcula valores intermedios con base a

cada registro, los agrupa de acuerdo al valor de ese criterio, convirtiéndolos en una lista intermedia, posterior los envía a la función Reduce.

Reduce es una función que toma un conjunto de valores intermedio, realiza una especie de 'barajado' de datos a los cuales aplica una reducción al combinar partes que sean iguales en forma, las reordena y devuelve un resultado. De esta forma se obtiene un acceso de salida rápido a los datos.

Las aplicaciones de *MapReduce* no se ven exclusivamente limitadas para *Hadoop*, sino son implementadas en ordenamientos distribuidos, análisis de *logs*, machine *learning* y sistemas operativos como *Unix*, por ejemplo, el comando *grep* capaz de realizar búsquedas de datos a través de una expresión regular.

4.6.2. HDFS

Es la versión *open source* de Google File System utilizado por Hadoop y diseñado para manejar grandes volúmenes de datos, con tamaños de *tera* y *petabytes* de información. Por su capacidad de procesamiento, utiliza funciones de tolerancia a fallos, que indican los bloques de archivos a ser replicados, generalmente son tres, pero este número puede ser cambiado, así se asegura error de pérdida de información. Los datos ingresados en este sistema se dividen en bloques que son almacenados y distribuidos en diferentes nodos que posteriormente son accedidos.

Un solo nodo maestro, llamado *NameNode* es el encargado de manejar el espacio y el acceso a los archivos. Este nodo principal maneja los *DataNodes*, que contienen la información separada por bloques insertadas en ficheros de fácil acceso, siempre y cuando se tenga la dirección correcta de entrada al mismo. El *NameNode* se limita únicamente a contestar peticiones de

lectura y escritura, es un directorio de grandes proporciones que indica la dirección física del nodo a consultar para que el cliente pueda acceder a él (Ghemawat, *et al*, 2003).

Los *DataNodes* poseen en su interior una serie de archivos de 64 a 128 Mb de tamaño, que a su vez se conforman de varios bloques de datos de igual tamaño, siendo desde 64 hasta los 128Mb. Estos bloques de información se encuentran distribuidos en todos los nodos del clúster de forma balanceada por medio de algoritmos que se encargan de reubicarlos cada cierto tiempo y optimizar el espacio, así como las opciones de lectura y escritura.

4.7. Herramientas de Hadoop para análisis de datos

Hadoop como ecosistema, posee una gran variedad de software especializado en diferentes áreas, como *analytics*, machine *learning*, controladores de *streaming* de datos, gestores de funciones de MapReduce, *Zookeeper* encargado de la coordinación de microservicios entre servidores, *HBase* también especializado en análisis en tiempo real, *Pig* que se encarga de traducir lenguaje de alto nivel a funciones de *MapReduce* para recuperación de datos. En puntos posteriores se tratan dos *softwares* especializados que se utilizarán en el trabajo de investigación e implementación de la arquitectura.

4.7.1. Hadoop Hive

Entre la variedad de software que ofrece *Hadoop* se encuentra Hive, que permite la consulta, compresión, agrupación y análisis de datos. Inicialmente fue desarrollado por Facebook Inc. pero en la actualidad la utilizan empresas como *Amazon* en servicios de *Amazon Web Services* que utilizan MapReduce; Netflix es otra beneficiada de este recurso, al utilizarlo para el manejo de datos

de sus servidores. Bancos utilizan este recurso para amasar grandes cantidades de datos sobre transacciones, clientes, *logs*, entre otros.

Hive posee un editor de lenguaje *HiveQL*, que muestra características similares al lenguaje SQL normal y transaccional. A diferencia del SQL tradicional *HiveQL* automáticamente traduce las consultas de alto nivel a consultas que utilizan *MapReduce* para acceder a los datos almacenados en los nodos (Carpiolo, *et al*, 2012). Puede manejar datos en diferentes tipos de almacenamiento como *RCFile* (*Record Columnar File*), en otras palabras, archivos columnares. ORC, similar al formato anterior, únicamente que de forma ordenada.

4.7.2. Hadoop Pig

Es una herramienta parte del ecosistema *Hadoop* que utiliza el lenguaje llamado *Pig Latin*, encargado de traducir programación Java a *MapReduce* (Natkovich, 2008). Además, puede ejecutar funciones UDF o *User Defined Functions* por sus siglas en inglés, que se escriben en lenguaje *Ruby*, *Javascript, Groovy* o *Java* como se menciona inicialmente.

Pig puede analizar grandes sets de datos a través de la programación secuencias y funciones de análisis de datos en lenguajes de alto nivel y que se acopla a una infraestructura *Hadoop* que evalúa estos programas para la extracción de información. Posee tres características fundamentales:

- Fácil de programar: capaz de ejecutar secuencias programadas simples que se traducen a operaciones de *MapReduce*. También, existen secuencias paralelas que permiten la ejecución, entendimiento y mantenimiento de tareas de análisis de datos.

- Oportunidades de optimización: la forma de programación permite que el mismo sistema optimice la ejecución, esto hace que la escritura de programas sea más rápida, el programador se centre más en la semántica del lenguaje que en la eficiencia que el programa tiene.

- Extensibilidad: permite al usuario crear sus propias funciones de procesamiento de datos (Apache Foundation, 2018).

4.8. Realidad virtual

Realidad virtual o RV se refiere a un conjunto de objetos que al unirse generan en el usuario una sensación de apariencia real y que también genera una percepción de encontrarse directamente interactuando con el mundo digital generado a través de gráficos de computadora. Básicamente a través de este entorno se crean en el cerebro señales.

Durante 1980 el término realidad virtual fue acuñado por Jaron Lanier que es uno de los primeros investigadores de esta área. Una definición formal dada por la enciclopedia Britannica (2000) es “el uso del modelado y la simulación por computadora que permite a una persona interactúa con un entorno sensorial tridimensional (3D) artificial u otro entorno sensorial”. Además, menciona que: “las aplicaciones creadas de realidad virtual hacen una inmersión completa del usuario en un entorno creado a través de gráficos de computadora que simula ser real mediante el uso de dispositivos interactivos como lo son gafas, cascos y auriculares que envían y reciben información” (p.1).

La reproducción de estos objetos y entorno con los cuales el usuario convive es una serie de dispositivos creados especialmente para este propósito

y por lo general son gafas o cascos de realidad virtual. Al aislar a la persona que los utiliza es capaz de crear la sensación de estar sumergido o inmerso en este mundo, esto se logra aumentando la atención del sentido de la vista en imágenes en tres dimensiones que dan la sensación al cerebro presencia en el mundo virtual.

La RV puede ser clasificada de diferentes formas, no inmersiva, semiinmersiva e inmersiva. Este tipo de categorización se basa en la forma en la que un usuario se conecta a la realidad virtual. Primero, la RV no inmersiva muestra al usuario un ambiente generado por computadora y navegable a través de una pantalla plana, por ejemplo, una TV o bien una computadora personal navegable a través de dispositivos periféricos como teclado y mouse.

La RV semiinmersiva va un poco más allá y se basa en la creación de un ambiente aislado por al menos cuatro pantallas con formación de cubo, capaces de dar seguimiento a los movimientos que realiza el usuario y trasladándolo a un ambiente navegable de RV. Finalmente, la RV inmersiva sube de nivel al utilizar dispositivos más pequeños como lentes y otros controles con botones o bien guantes y otros artefactos *wearables* capaces de dar seguimiento a los movimientos del cuerpo y moviendo el personaje del usuario, en el caso de los videojuegos, dentro del mundo de RV.

Adicional al casco o gafas, el usuario puede acompañar y navegar en el mundo mediante dispositivos como controles, guantes y otros accesorios especiales como ropa con la cual la experiencia aumenta ya que la sensación de permanencia en el mundo virtual no se limita a ver, sino a interactuar con objetos que cambian de forma, se acercan o se alejan y lo más importante crean una sensación de tres dimensiones o lo más acercado a la realidad que el usuario vive en su diario vivir.

4.8.1. Áreas de aplicación de RV

Las áreas de RV aplican a la mayoría de las edades, desde niños hasta adultos y personas de la tercera edad. Las aplicaciones incluyen: Educación, entretenimiento, medicina, videojuegos y en la actualidad se encuentra en desarrollo en áreas potenciales como la milicia y terapia psicológica.

4.8.1.1. Educación

Es uno de los campos que utiliza la inmersión para tomar la atención del estudiante y lograr una concentración en una tarea, al mismo tiempo que el estudiante aprende a través de conceptos y actividades que el mundo virtual propone a este. La utilización de tecnología para el aprendizaje ha sido uno de los principales puntos reforzados por la Unesco que pretende utilizar los medios digitales para aumentar el aprendizaje de los estudiantes.

RV ofrece diferentes mundos programados a través de aplicaciones proyectadas en gafas especializadas y controles que permiten a los estudiantes moverse a través del mundo virtual. Por ejemplo, visitas a museos, mundos virtuales dentro del cuerpo humano, visitas a otras regiones del mundo. Esto rompe los paradigmas de tener que realizar visitas físicas a estos lugares cuando son completamente accesibles desde el asiento de un escritorio, además de ser interactivos y navegables. Las exploraciones se llevan a cabo en diferentes materias de estudio, desde las ciencias naturales hasta clases como geografía y arte, y no solo en la era actual, sino que RV permite la reconstrucción de antiguas civilizaciones y un salto hacia atrás en el tiempo para revivir lo ya encontrado e investigado en diversas partes del planeta.

4.8.1.2. Medicina

El área más beneficiada de la medicina se encuentra en la cirugía, esto debido a que ahora es más aceptable el enseñar teóricamente a los nuevos doctores, pero también lograr poner la práctica dichos conocimientos. De esta forma se agilizan los procesos de práctica supervisada que complementan las enseñanzas del médico.

Esto se logra mediante el uso de mundos virtuales que permiten al estudiante una mejor vista de lo aprendido, ya que al ser gráficos en 3D son capaces de relacionar mejor los conceptos y de visualizar de mejor forma que solamente en imágenes de libros y vídeos. Además, a través de complementos a los lentes de RV también se ofrece un entrenamiento de cirugía en donde el usuario interactúa directamente en prácticas directamente supervisadas por parte de su tutor, de esta forma adquiere los conocimientos básicos para aplicar.

Las simulaciones ofrecidas no solamente muestran al estudiante objetos tridimensionales sino también son capaces de evaluar al mismo a través de la interacción que este ha tenido y la medición del comportamiento de este de acuerdo con criterios programados y calibrados en los dispositivos usados, pues de esta forma se mide el control psicomotor que el usuario posee y le ayuda a mejorar. El objetivo principal es dar los conocimientos adecuados al futuro cirujano mediante entrenamientos antes de ejecutar una cirugía real en un quirófano.

4.8.1.3. Entretenimiento

En cuestión de entretenimiento, varias empresas han desarrollado desde clips de películas hasta videojuegos. Los vídeos pueden ser generados mediante dispositivos dedicados como cámaras que graban a un ángulo de 360°, por ejemplo, Samsung, Nikon *360*, *GoPro*, entre otras. Y que a su vez son capaces de transmitir al dispositivo móvil que se encargará de reproducir a través de gafas, un ejemplo de ellos es *Oculus* de *Samsung*, el cual permite la conexión del teléfono a las mismas, este se convierte en la pantalla de proyección.

En el ámbito de los videojuegos, los gráficos han evolucionado en calidad durante los últimos años y con las tecnologías necesarias son capaces de permitir al usuario transportarse dentro del mundo virtual a tal punto que las sensaciones brindadas por el dispositivo hacen que el jugador se sienta presente dentro de este mundo generado por computadora.

La interacción con controles, tales como mandos *joystick*, armas de fuego que funcionan como en el mundo virtual hace que la experiencia de usuario sea mejor y la aplicación de sensores en los dispositivos aumentan la funcionalidad de estos, haciendo que el usuario tenga un comportamiento más acertado de sus movimientos durante el juego.

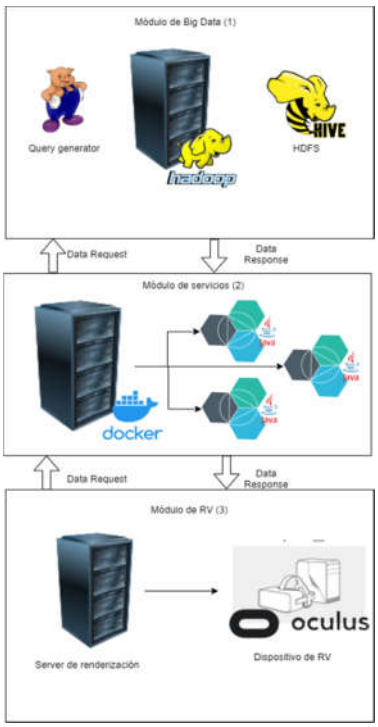
5. PRESENTACIÓN DE RESULTADOS

5.1. Análisis y diseño del sistema de gráficos en RV

La creación de este proyecto está divida en dos partes, la primera es la manipulación de datos a través de software de *Big Data*, y la segunda es la representación gráfica de dichos datos en forma condensada a través de gafas *Oculus Gear VR*. Pero a través de estas dos funcionalidades principales, fue necesaria una capa de transmisión de datos que fue implementada con microservicios.

Por lo tanto, el proyecto fue dividido en tres grandes módulos: el primero es el de *Big Data*, conformado por una máquina virtual de *Cloudera*; ver figura 1 parte (1). El segundo, es un mini *datacenter* de microservicios, que sirvieron como intermediario entre los datos y su representación en VR; ver figura 1 parte (2). Y finalmente, el tercer módulo está conformado por los dispositivos de RV, el *Oculus Gear VR* de Samsung, un *Pro-Controller* de la consola de videojuegos Nintendo Switch y un teléfono Samsung Galaxy *S7*, adaptado al dispositivo de RV; ver figura 1 parte (3).

Figura 1. **Arquitectura general del sistema**



Fuente: elaboración propia utilizando software draw.io.

5.1.1. Módulo de *Big Data*

El módulo de *Big Data* es el primer módulo desarrollado, fue creado con a partir de la máquina virtual basada en un *sandbox* de la fusión de *Cloudera* y *Hortonworks* después de su fusión. El archivo de virtualización completa puede ser descargado de https://www.cloudera.com/downloads/hortonworks-sandbox.html, la descarga consta de un archivo con extensión OVA, que puede ser montado con el software de *Oracle*, *Virtual Box*, el cual hará una exportación de todas las configuraciones como red, disco y tamaño en memoria óptima para el funcionamiento del software interno como Hadoop.

Para iniciar, los datos fueron extraídos de páginas de datos abiertos del Gobierno de Guatemala, de presupuestos ejecutados en años pasados y otros datos abiertos de gobiernos del Reino Unido, obtenidos de https://data.gov.uk/, con datos de distribución demográfica, datos climáticos por zona, emisiones de dióxido de carbono, entre otros, La carga fue aproximadamente de 4 GB, los cuales sirvieron como base para ejecutar instrucciones de HiveQL de tipo *Select*. Los datos fueron ordenados y estructurados por la herramienta de *Hadoop*.

Los datos fueron cargados a través de la herramienta gráfica de *Hadoop* Hive, que mapeo todos los campos contenidos en el JSON, y escogió la estrategia óptima para la carga de datos, en algunas ocasiones si necesita un poco de ayuda para saber la definición de los datos y como se separan, por ejemplo, un archivo CSV puede ser separado y cargado a tablas por un carácter de coma o punto y coma. *Cloudera* es una herramienta muy versátil que permite una carga fácil de los archivos descargados de internet, Hive se encargó de la mayor parte del trabajo.

Hablando concretamente, el módulo de *Big Data* está conformado por un núcleo completo con las herramientas necesarias, construidas para estas tareas, se hizo bastante fácil la manipulación de datos. No fue necesaria su reestructuración, pero si la limpieza de estos, que fue bastante sencilla. Cabe mencionar que las consultas son fáciles de manejar y construir, el lenguaje de *HiveQL* es parecido al SQL y permite generar consultas de *Data Manipulation Language* (DML) de forma como se haría en un *Data Base Management System*.

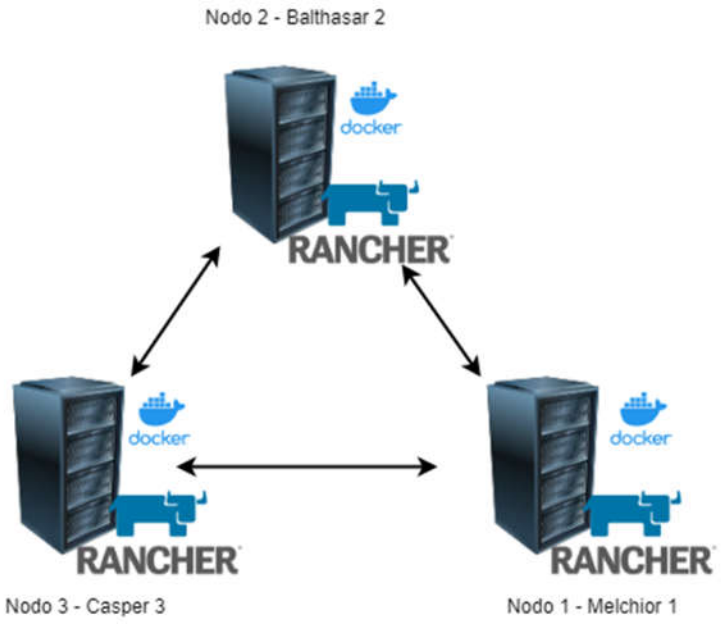
5.1.2. Módulo de microservicios

Para la implementación de microservicios se evaluaron diferentes tecnologías para la administración de contendores *Docker*, desde clúster en *Amazon Web Services (AWS), Microsoft Azure y Google Cloud*. Estos tienen un excelente performance para el manejo de contendores, auto escalamiento, de acuerdo con reglas definidas por el programador, como la carga de red, carga de *CPU* o memoria *RAM*. Desafortunadamente la conexión de datos a estas tecnologías en la nube es impredecible y el cambio de IP pública de una conexión residencial es cambiante, por lo que no fue una solución viable en primera instancia.

Por el problema anterior expuesto la configuración de una conexión estable y segura a través de la máquina virtual de *Cloudera* y los microservicios en la nube fue de baja calidad y con una latencia muy grande que en ocasiones no permitía la devolución de los datos al existir un *timeout* o una respuesta tardía del contenedor. Por lo que se optó por hacer una implementación local en *Rancher*.

. El modelo consta de un clúster de 3 nodos como se puede ver en la figura n. Cada nodo es un *Linux CentOS 7* con 2 GB de memoria de capacidad y con una instalación de *Docker* con versión 18.0.6, siendo la más estable. Cada instalación *Docker* contiene un agente de *Rancher* con el cual se realiza el monitoreo de la salud del servidor, es decir *RAM* disponible, uso del *CPU*, entradas/salidas de datos de la red, de esta forma se asegura al momento de escalar cual es el servidor óptimo para agregar el contendor (ver figura 2).

Figura 2. **Clúster de microservicios**



Fuente: elaboración propia, utilizando software draw.io.

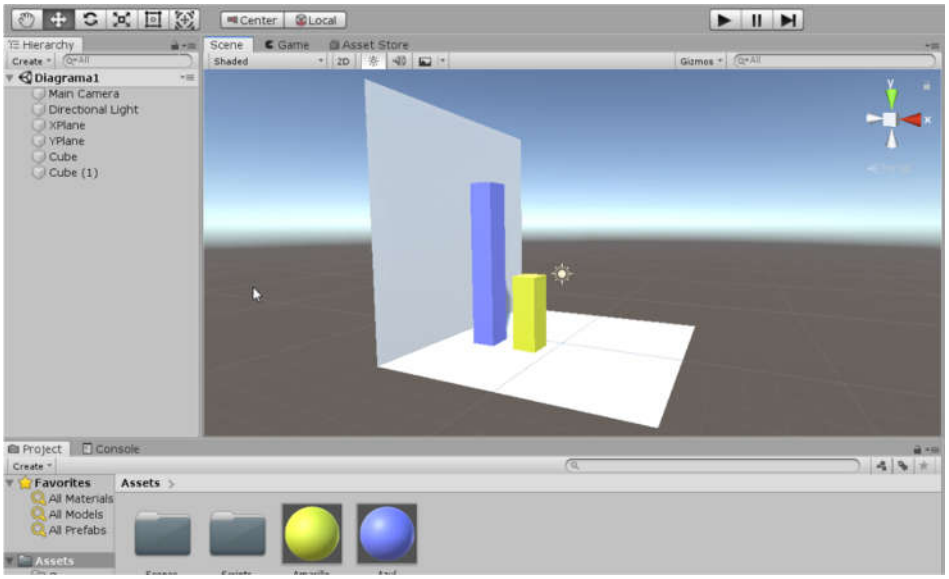
Se creó una aplicación en el lenguaje Java con un módulo Web de servicios *Restful*, capaz de consultar la base de datos de Hive, la herramienta para las librerías es *Maven*, de esta forma se centraliza todo en un archivo POM para configurar *endpoints*, URL de consumo de datos, variables, entre otros, En este caso esta pequeña aplicación montada en un servidor *Glassfish* versión 3.1.2, por supuesto que dentro de un contendor y con las instrucciones de escalamiento para que se distribuyan dentro del clúster la cantidad necesaria para atender la cantidad de peticiones y el volumen de datos que trasmite.

5.1.3. Módulo de realidad virtual (RV)

Para el módulo de RV, se hizo la evaluación de dos motores principales *Unreal Engine* y *Unity 3D*. Por un lado, *Unreal* tiene una forma bastante accesible para generar cualquier tipo de gráfico o escenario, pero la curva de aprendizaje es mucho más grande, además utiliza C++ para como lenguaje para sus *scripts* de programación de acciones.

Unity 3D, por otro lado, la creación de gráficos es un poco más difícil, pero la programación se realiza con el lenguaje *C# de Microsoft*, ayudando un poco más a usuarios no experimentados (ver figura 3). Este último juega como un pro y contra sobre este software, ya que las librerías de este sistema suelen ser un poco más pesadas para arrancar con su funcionamiento, pero al mismo tiempo facilita la manipulación de objetos y su programación ya que contiene todas las librerías necesarias para ello.

Figura 3. **Editor de desarrollo *Unity 3D***



Fuente: elaboración propia, utilizando *Unity 3D.*

La conexión hacia el *web service* de Java fue creado con una de las librerías de C# y la clase WWW, en la cual se realizó la invocación para la obtención de datos y su mapeo en los diagramas gráficos. Ya que la renderización del juego requiere que todos los datos estén presentes, se llamaron secuencialmente a un balanceador de carga, desplegado en el clúster de servidores de microservicios. Este último se encargó de enviar las peticiones hacia la aplicación y hacer la recuperación de los datos para su posterior renderización.

Finalmente, el despliegue del juego se realizó siempre bajo la plataforma de *Unity*, únicamente que fue necesaria la instalación de paquetes adicionales de *Oculus*, disponibles en https://docs.unity3d.com/Manual/VROverview.html, específicamente el paquete de *Oculus XR*. Una vez instalado es posible utilizar las clases necesarias para que los controles sean útiles dentro de la aplicación. El sistema de gafas *Oculus Samsung Gear VR* (ver figuras 4 y 5), fue utilizado para la renderización a través de un teléfono Samsung Galaxy S7 que se adapta a este dispositivo.

Figura 4. **Samsung *Gear VR powered by Oculus*, vista frontal**

Fuente: elaboración propia.

Figura 5. **Samsung *Gear VR powered by Oculus*, vista trasera**

Fuente: elaboración propia.

El control para el movimiento del usuario dentro del ambiente es sincronizado a través de la tecnología *Bluetooth*. Ya que no se tenía disponible el control original del Samsung *Gear VR*, se utiliza un *Pro-controller* perteneciente a la marca Nintendo de la consola *Switch* (ver figura 6). Los botones y palancas necesarias fueron mapeados completamente y programados para que el usuario final se pueda mover en cualquier dirección dentro del ambiente, así como girar sobre su propio eje. Las librerías de *Unity* en conjunto con el *SDK* de *Oculus* permiten realizar este tipo de tareas de forma fácil.

Figura 6. **Nintendo Switch *Pro-Controller***

Fuente: elaboración propia.

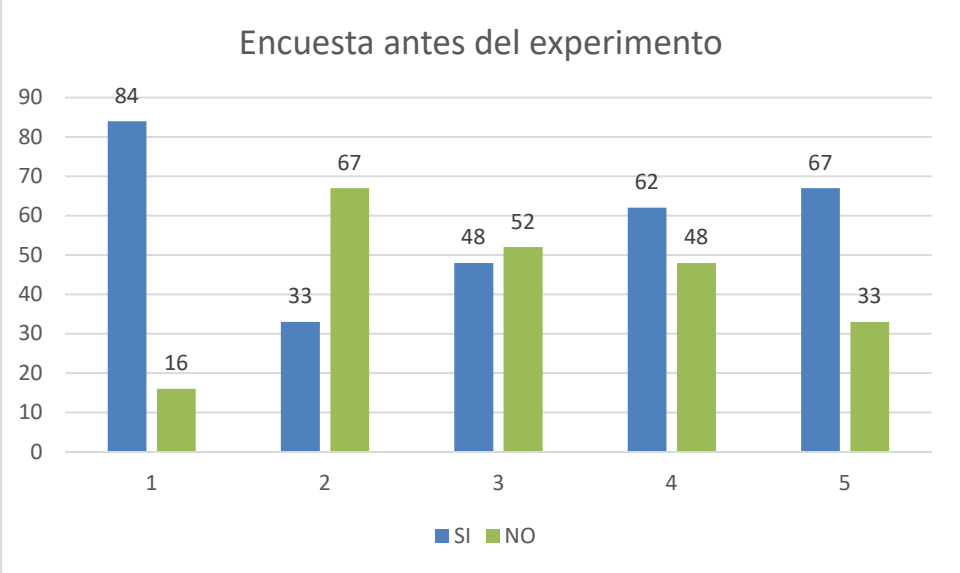
5.2. Conocimiento poblacional de la tecnología de RV

Como se explicó en capítulos anteriores, la tecnología de RV es un área medianamente explorada en diversas ramas como medicina, entretenimiento, entre otros. Pero en general las personas no tienen idea de la existencia de esta o solamente han escuchado hablar de ellas en algunas películas, internet o artículos de revista, pero, para fines académicos se desea medir la cantidad de personas que tienen el conocimiento o bien han experimentado con esta tecnología y es parte de los fines de este trabajo.

De acuerdo con la planificación, este es una de las dos entrevistas que se realizaron a los sujetos de prueba, una antes y el otro después de utilizar la tecnología RV. Para tomar las medidas se han entrevistado a los 25 sujetos de prueba, con preguntas simples de respuesta 'sí' y 'no'. Las preguntas son las siguientes y en el diagrama gráfico de la figura 7 se muestran las respuestas a las mismas expresadas en porcentajes.

- ¿Ha escuchado usted el término realidad virtual?
- ¿De ser así, sabe que aplicaciones tiene RV en la vida diaria?
- ¿Sabe usted la definición de realidad virtual?
- ¿Ha utilizado usted alguna vez algún dispositivo de realidad virtual?
- Si su respuesta anterior es sí, ¿le pareció agradable la experiencia?

Figura 7. **Resultados de la encuesta realizada a los sujetos de prueba**



Fuente: elaboración propia.

Acorde con la información generada, en general se tiene que los porcentajes de personas que han escuchado de RV son altos, pero solamente la definición o que hace RV, pero no tienen el conocimiento o potencial que este tiene, más que solo entretenimiento.

Por otro lado, un hecho importante es que casi dos tercios han utilizado uno de estos dispositivos en alguna ocasión y la mayoría responde que dispositivos *Samsung* que son los más promocionados, además se sienten conforme con la utilización. Por lo que el acceder a utilizar un dispositivo de RV no representó un reto, todos aceptaron colaborar con el experimento.

5.3. Visualización y comprensión de los diagramas gráficos

Los resultados de esta sección se obtuvieron realizando la comparación de los 25 voluntarios que colaboraron con experimento (ver anexo 1), se escogieron estas personas por motivos de cercanía, estos voluntarios se conformaron por vecinos, familiares y compañeros de trabajo del sector informático. El número 25 se escogió porque se consideró un número razonable para realizar la comparación de datos con mayor exactitud. Durante una prueba realizada en el que analizaban 10 diagramas gráficos y de *pie* en un formato plano en un archivo de Excel corriente, en el que solamente podían moverse sobre el gráfico, pero no modificarlo para cambiar sus valores o ver etiquetas ni leyendas de datos.

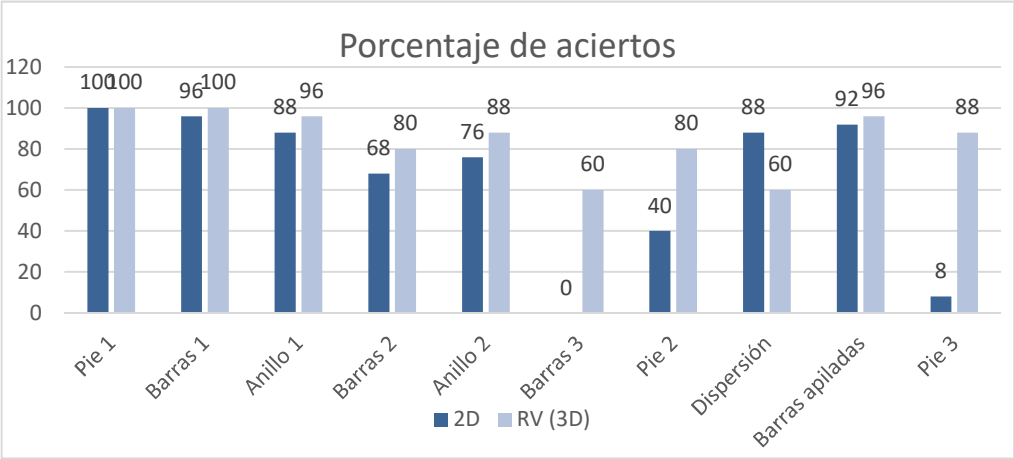
La segunda parte de la prueba está basada en los mismos 10 diagramas gráficos, únicamente que en el ambiente de RV, en el cual el usuario tendría que moverse dentro del ambiente alrededor de los diagramas y dar una respuesta de los valores que cada uno contenía. Se comenzó por definir un “acierto” y es el acercamiento o concordancia a los resultados reales de un

+/- 5 % de exactitud, es decir que al momento que un sujeto de pruebas ve un diagrama gráfico dice a la persona que ejecuta el laboratorio una cantidad de lo que logra observar en los diagramas de pie o de barras o dispersión propuestos.

Por ejemplo, se muestra un diagrama de pie con una sola etiqueta para una porción de este, y el sujeto debe dar datos del resto de particiones, o bien si es de barras debe calcular sin etiqueta aplicada alguna a la barra, su medida y este no debe estar por arriba ni por debajo de unos 5 puntos porcentuales para que se considere como acertado. Los resultados son los siguientes, basados en el análisis de las gráficas y la respuesta de los sujetos de prueba:

Como se observa en la figura 8, los gráficos en 3D superan, para todos los casos, excepto el del gráfico de dispersión, la capacidad de comprensión. Según las entrevistas practicadas a los sujetos de prueba posterior al uso del dispositivo de RV, el hecho de tener navegabilidad a través del diagrama gráfico y visitarlo desde diferentes perspectivas, hace que la comprensión de este aumente en un pequeño porcentaje.

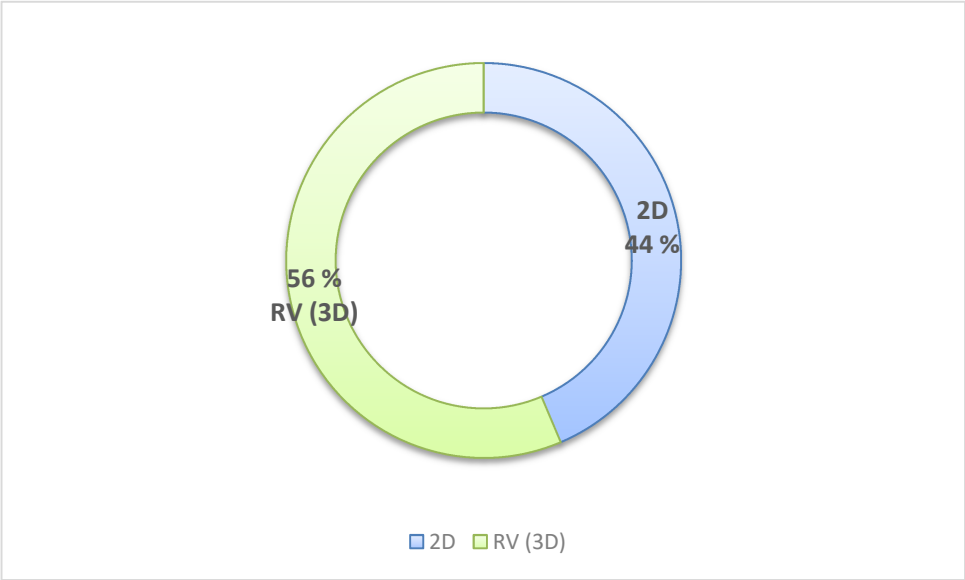
Figura 8. **Condensado del porcentaje de aciertos del experimento**



Fuente: elaboración propia.

Como prueba de lo anterior, se hizo un análisis de los porcentajes de comprensión y aciertos en 2D y en 3D (RV). La mayoría de los casos de gráficos eran de barras, pie y solo uno de dispersión, el cual mostró ser útil únicamente cuando se muestra en 2D, la comprensión no aumenta, sino más bien disminuyó. El porcentaje de aciertos en 3D es 12 % mayor a los gráficos planos (ver figura 9).

Figura 9. **Condensado de porcentaje de comprensión 2D vs. 3D**



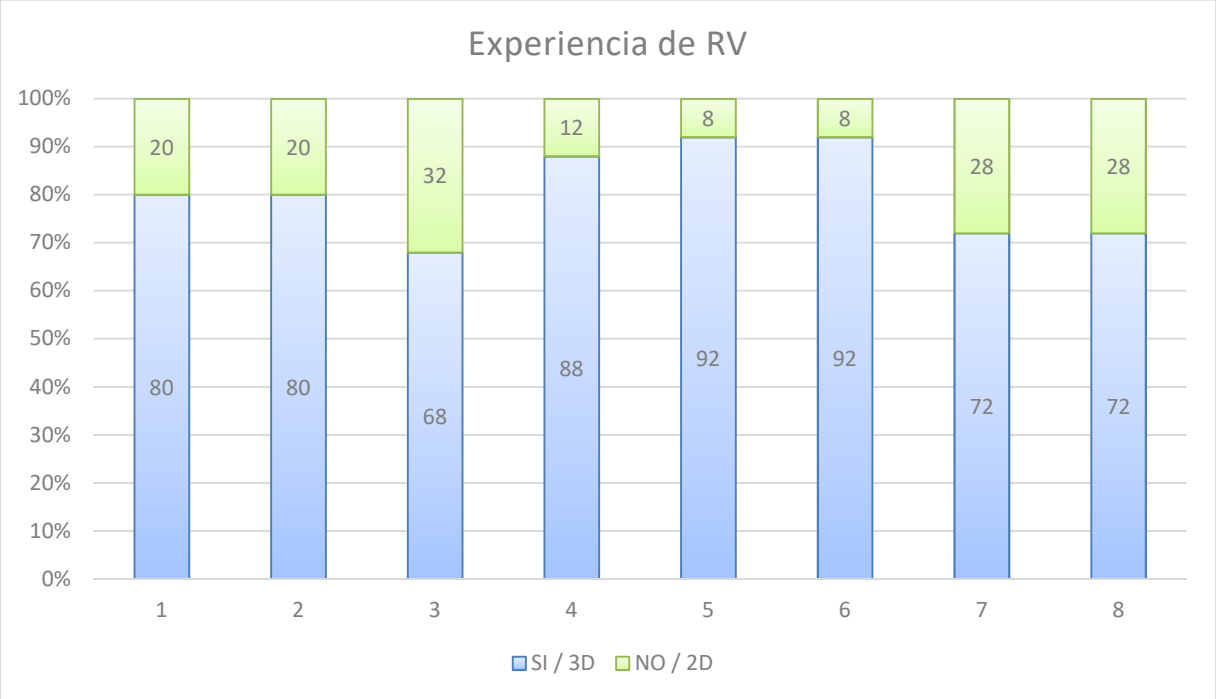
Fuente: elaboración propia.

5.4. Experiencia RV

Finalmente, después del experimento, además de medir los niveles de comprensión de los sujetos de prueba, se encuestaron y entrevistaron nuevamente para validar si su experiencia utilizando esta tecnología fue buena y que efectos secundarios experimentaron durante o posterior al uso del dispositivo RV (ver figura 10 con resultados condensados). Las preguntas elaboradas a los sujetos fueron las siguientes:

- ¿Le pareció buena la experiencia con el dispositivo RV?
- ¿Le gustaría repetirlo nuevamente?
- De acuerdo con los gráficos vistos, ¿qué le pareció más fácil de entender, los gráficos 3D o gráficos planos?
- ¿Le pareció fácil el utilizar el control para movilizarse a través del gráfico?
- ¿Recomendaría esta experiencia a alguien más?
- ¿Cree que los gráficos en 3D son más fáciles de entender?
- ¿Utilizaría esta tecnología para analizar gráficos más complejos
- ¿Recomendaría usted el uso de RV para este tipo de actividades de análisis?

Figura 10. **Resultados de la entrevista de experiencia de RV**



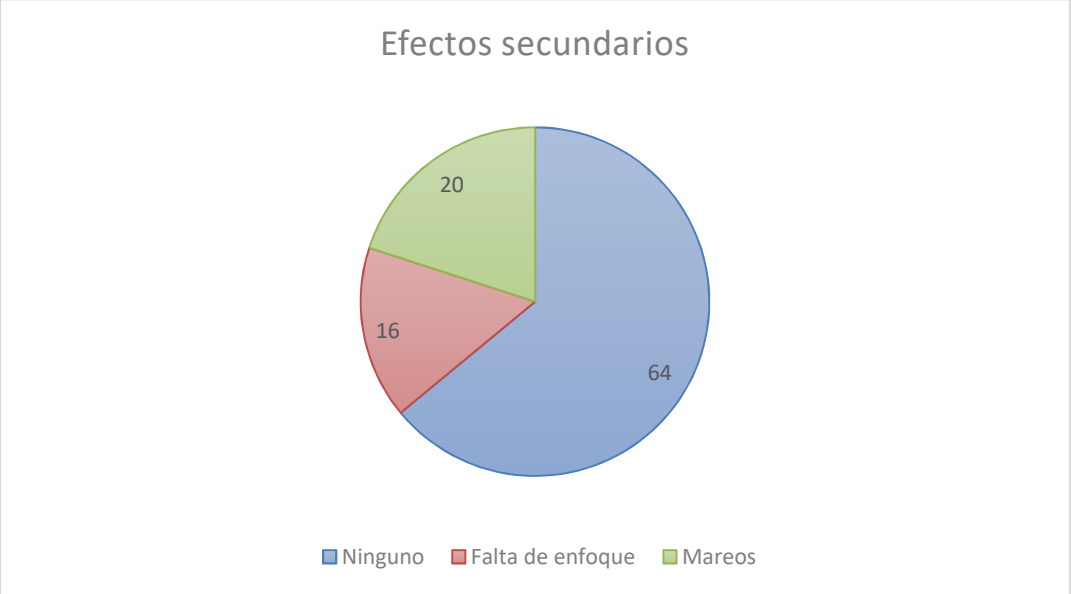
Fuente: elaboración propia.

Del total de preguntas realizadas a los sujetos de pruebas, la mayoría coincide que es mejor la utilización de una tecnología 3D ante una 2D, que tuvieron mejor relación con el gráfico y pudieron entenderlo de una mejor forma. Además, recomiendan el uso de la tecnología. Aunque en una minoría, los sujetos dicen no estar de acuerdo con su uso ya que experimentaron algunos efectos secundarios como se muestra a continuación.

5.4.1. Efectos secundarios

Durante la elaboración de este proyecto, específicamente en el desarrollo del módulo de RV, encontré que este puede tener una especie de “efectos secundarios”, o bien malestares al utilizar mucho esta tecnología, por lo que también decidí preguntar acerca de esto a los sujetos de prueba. Entre estos malestares encontré dos principales, que son: mareos, al final de utilizar las gafas de RV. Igualmente, la falta de enfoque se refiere a que, durante el uso de las gafas, la persona puede perder un poco la capacidad de enfocar o vista borrosa, lo mismo sucede al quitarse estas puede haber una falta de enfoque hasta que la vista se reacomoda a la luz natural, la siguiente figura muestra los porcentajes de los efectos encontrados (ver figura 11).

Figura 11 . **Porcentaje de efectos secundarios experimentados por el uso de RV**



Fuente: elaboración propia.

El gráfico muestra que la mayoría no presentaron ningún efecto secundario, solo una minoría presentó falta de enfoque durante y después del uso del dispositivo de RV, como una especie de cansancio visual, mientras que otros presentaron mareos al final del uso y tuvieron un poco de dificultad al incorporarse después del uso del dispositivo de RV.

6. DISCUSIÓN DE RESULTADOS

Para el desarrollo de este proyecto, y para la disminución de tiempos, se utilizaron herramientas de tipo *community* que están listas para trabajar sin mucha configuración, por ejemplo, *Hortonworks Data Platform* y su máquina virtual en co-creación con *Cloudera*, configurada con las herramientas de *big data*, necesarias para carga, modificación y selección de datos.

Además, por la disponibilidad y una menor curva de aprendizaje se utilizó Unity, que permite crear de forma sencilla, y de buena calidad, escenarios para móviles siendo precisamente el fin de este proyecto. *Unity* trabaja directamente con el lenguaje C# (*C Sharp*), que es fácil de utilizar y basado en objetos, por lo que no es tan complicado de aprender a manejar, especialmente que casi todas las librerías para añadir *joystick* y controlar los objetos dentro del mundo de RV vienen incluidas con la instalación.

6.1. *Big Data* y microservicios

Los datos para este proyecto fueron descargados de diferentes lugares y cargados a *Hadoop*, en sí el proyecto contiene una cantidad variada de datos abiertos, entre ellos presupuestos 2019 ejecutados por el gobierno de Guatemala, obtenidos desde la página https://datos.minfin.gob.gt/**,** sitio del gobierno que provee datos abiertos. También, datos sobre el clima provistos por https://data.gov.uk/ como emisión de dióxido de carbono emitido por el transporte público. Estos sitios y datos obtenidos de ellos fueron los mejores para analizar, ya que reestructurarlos para un análisis fue sencillo por las herramientas utilizadas.

Entre estas herramientas, muy recomendadas también, se utilizó la máquina virtual de *Cloudera* para la carga y selección de datos, mientras que para recuperarlos se utilizaron microservicios de Java en un ambiente *Docker* con *HiveQL*, con datos semi-estructurados. *HiveQL* permite la integración con varios lenguajes, para facilidad de implementación se utilizó *Java* que permitió la recuperación de grandes datos también fue fácil su recuperación.

Por otro lado, del lado de microservicios, existen diferentes plataformas que permiten el *autoscaling* o escala automática para microservicios, dependiendo de la cantidad de tráfico, requerimientos, porcentaje de carga por CPU's y memoria. Entre ellos *AWS, Google Cloud, Kubernetes* y sus administradores como *RancherOS*. Para este se escogieron los últimos dos (*Kubernetes* y *RancherOS*), aunque otras tecnologías también son más fáciles de implementar, pero implican un costo en la nube. Para este experimento no se necesitaba demasiada potencia de procesamiento de datos por lo que fue una implementación exitosa.

6.2. Proyección de diagramas gráficos en RV

El objetivo principal de este proyecto es la proyección de análisis de *big data* en un ambiente inmersivo de realidad virtual, ya que este ámbito ha sido muy poco investigado y existe poca implementación en análisis de diagramas gráficos, es más, esta tecnología ha sido más implementada para entretenimiento, medicina y educación.

Por casi no encontrar información sobre proyecciones de diagramas gráficos en RV, puedo llegar a afirmar que ha sido muy poco investigado e implementado. Este proyecto fue un poco difícil de diseñar y construir, debido a la complejidad de conexiones que se tienen que llevar a cabo entre un servidor

y una tecnología de RV. Además de las limitaciones por el precio que representa un buen dispositivo de gafas y mandos de RV, puede que las empresas o personas tiendan a no invertir en este tipo de tecnología.

Por otro lado, el tiempo de construcción de un gráfico lleva una cantidad considerable de tiempo. La generación de los gráficos del proyecto se hizo de una forma semi-estática, esto quiere decir que no puede generar cualquier gráfico deseado, solamente puede generar barras, *pie*, anillos y diagramas de dispersión. Pero no en el momento, se necesitan hacer ajustes siempre en el proyecto sin renderizar para que funcione correctamente.

6.3. Aciertos en el experimento

La exactitud durante el experimento se midió con un +/- 5 %, en otras palabras, los sujetos de pruebas podían equivocarse con un margen de +/- 5 puntos porcentuales en acertar los valores reales de las pruebas, ya que los diagramas fueron hechos en escala 1:1. Por ello se tomaron en cuenta 10 diagramas gráficos similares de tipo *pie*, barras, anillos y dispersión. Por lo que se puede asegurar que las pruebas fueron suficientes para extraer datos y generar una cantidad de aciertos muy cerca de la realidad.

Entre las pruebas de 2D y 3D de RV, el aumento fue poco significativo de un 12 %. Es decir, la comprensión en RV subió varios puntos porcentuales, aunque sea una pequeña cantidad. Esto no aplica para todos los diagramas, como el de dispersión que fue menos comprendido en 3D que en planos de dos dimensiones. Aunque en porcentaje no se alcanzó el 40 % de comprensión, como se expuso en un estudio mostrado en capítulos anteriores, se puede afirmar que RV mostrando imágenes en 3D aumenta la comprensión del individuo que lo usa al analizar diagramas gráficos.

CONCLUSIONES

1. Se diseñó e implementó un sistema con tres diferentes módulos que permite la proyección de análisis de datos creados a partir de *big data* en un ambiente inmersivo de realidad virtual (RV). Con ello se mejora un 12 % el nivel de comprensión de RV (proyección 3D), sobre los diagramas gráficos planos en 2D, así también se aumenta la navegabilidad del usuario a través del diagrama pudiendo este extraer más detalles.

2. La tecnología RV tiene un gran potencial en diversos campos, no solo medicina, entretenimiento y educación. Prueba de ello es la flexibilidad que tiene para adaptarse a nuevos campos. Durante este experimento se demostró que también es posible aplicar RV al análisis de datos ya que se logró la integración con *big data (Hadoop)* a través de microservicios *(Docker y RancherOS)*, obteniendo una plataforma capaz de generar diagramas gráficos con información brindada por *big data*.

3. Se generaron 10 diferentes gráficos a partir de una serie de datos procesados con *big data* y transmitidos a través de microservicios hacia un dispositivo de RV. Con base a este sistema también se comprueba que RV aumenta la comprensión y amplia el campo visual de una persona al analizar diagramas gráficos que representan información, así pues, tiene la capacidad hacer que el usuario utilice esta tecnología para moverse en diferentes ángulos y perspectivas, dándole un movimiento mayor y la libertad de explorar los diagramas de una forma interactiva y desde un lugar que pueda captar cada detalle.

4. Se diseñó e implementó la arquitectura para generar gráficos en 3D en un ambiente de RV. La integración entre tecnologías, aunque parezca que son de diferentes ramas de la computación, procesamiento de datos y electrónica, tuvo un buen acoplamiento y rendimiento, por lo tanto, se considera como un éxito el lograr que estas tres áreas, se combinen en un proyecto que incursiona en un campo no tan explorado, que son los análisis gráficos en RV.

RECOMENDACIONES

1. Existen diferentes tecnologías para la implementación de microservicios, desde grandes clústeres provistos por empresas como *Google* o *Amazon Web Services* que generan costos por utilización y tamaño, hasta los administrados por la empresa, dentro de estos o en servidores in-situ. Para este experimento se utilizó la segunda opción y es viable si se desea incursionar o iniciar en el ambiente de microservicios en contenedores *Docker*.

2. Para ámbitos como *Big Data*, también *AWS* provee una gran variedad de servicios, de los cuales no hay preocupación de escalar ni mantener infraestructura, los servicios son completamente integrados. Es útil únicamente si tenemos un volumen de datos muy grande y una de sus ventajas es que permite *data streaming*, en otras palabras, transmisión de datos en tiempo real. Si se está analizando grandes cantidades de datos cambiantes es posible utilizarlo. Por ser un pequeño experimento, se recomienda, como se menciona en los resultados, utilizar alguna versión *community de Hadoop, Apache Spark* o algún otro software. Uno con buen desempeño y con las herramientas adecuadas para iniciar es la máquina virtual de o *Sandbox* Cloudera, disponible para *Virtual Box*, *Hyper-V* o bien *Docker*.

3. Relativo a la creación de RV, el software más recomendado es *Unity*, ya que combina el lenguaje C# para el uso de *scripts* para darle animación o acciones a cada escena. Tiene una curva de aprendizaje mucho menor que *Unreal Engine*, siendo este otro motor para la creación de prototipos,

puede ser un poco difícil de utilizar ya que utiliza C++ como lenguaje base para los scripts de programación. Si ya tiene una base para programar sería más recomendable este último motor. Por su lado *Unity* ofrece flexibilidad para manipular todos los objetos nativos, incluso sobrescribir botones de mandos, ajustar cámaras e iluminación, entre otros.

4. El experimento realizado para este proyecto fue aplicado a 25 personas de las cuales el 50 % son profesionales de la informática y otro 50 % son personas que no utilizan tecnología. Siendo el análisis de datos una de las partes más importantes de la toma de decisiones se recomienda enfocar futuros experimentos hacia un público objetivo definido, realizando también un estudio estadístico de mercado para asegurar que se llegue al grupo objetivo.

REFERENCIAS

1. Apache Foundation. (2018). *Welcome to Apache Pig!* California, EU.: Apache Foundation. Recuperado de http://pig.apache.org/

2. Bellgardt, M., Pick, S., Zielasko, D., Vierjahn, T., Weyers, B. y Kuhlen, T. (2017). *Utilizing Immersive Virtual Reality in Everyday Work.* Recuperado de https://www.vci.rwth-aachen.de/publications/0000/02193/

3. Capriolo, E., Wampler, D. y Rutherglen, J. (2012). *Programming Hive: Data warehouse and query language for Hadoop*. California, EU.: O'Reilly Media, Inc.

4. Cheng, X. (2018). *3D Mixed Reality vs 2D Visualizations: Decision Support for Comparing Financial Data.* (Tesis de Maestría). University of Utrecht, Utrecht, Holanda.

5. Lowood, H. (2000, 2020), Virtual Reality*, Encyclopaedia Britannica* [versión electrónica]. Chicago, EU.: Encyclopaedia Britannica Inc., Recuperado de https://www.britannica.com/

6. Franck, G. (Octubre de 1994). Viewing a Graph in a Virtual Reality Display is Three Times as Good as a 2D Diagram. En C. Ware (Presidente), *Visual Languages.* Simposio llevado a cabo en el congreso del Institute of Electrical and Electronics Engineers. Missouri, EU.

7. Garzía, R. (1996). *Vision and Reading.* California, EU.: Mosby Inc.

8. Groffman, S. (2006). *The Relationships between Visual Perceptual Problems and Learning.* Philadelphia, EU.: Evolve Publishing House.

9. Hurwitz, J., Nugent, A., Halper, F. y Kaufman, M. (2013). *Big Data for Dummies, Nueva* Jersey, EU.: John Wiley & Sons, Inc.

10. Hwan, W.-Y., Su, J.-H. Huang, Y.-M. y Dong, J.-J. (2009). *A Study of Multi-Representation of Geometry Problem Solving with Virtual Manipulatives and Whiteboard System.* Educational Tecnology & Society, 12 (3), 229-247.

11. Mayer-Schönberger, V. y Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work and think.* Londres, Reino Unido: Jhon Murray editores.

12. Parker, G., Franck, G. y Ware, C. (2000). Visualization of Large Nested Graphs in 3D: Navigation and Interaction*. The Center for Coastal and Ocean Mapping*, *32*(22), 6-28. Recuperado de http://ccom.unh.edu/publications

APÉNDICES

Apéndice 1. **Inicio de la experimentación, movimiento a través de movimiento de la cabeza y control a través de panel táctil lateral**

Fuente: elaboración propia.

Apéndice 2. **Fase final, usuario ajustando la configuración del dispositivo de RV antes de iniciar la prueba**

Fuente: elaboración propia.

Apéndice 3. **Fase final, usuario utilizando dispositivo de RV con controles ajustados y visualizando los diagramas gráficos previamente renderizados**

Fuente: elaboración propia.

Apéndice 4. **Fase final, usuario utilizando dispositivo de RV con controles ajustados y visualizando los diagramas gráficos previamente renderizados**

Fuente: elaboración propia.