



Universidad
Tecnológica
del Perú

Facultad de Ingeniería
Ingeniería de Sistemas e Informática

Tesis:

**“IMPLEMENTACIÓN DE UN CUBO MULTIDIMENSIONAL DE EVENTOS SAFE
PARA MEJORAR EL ANÁLISIS DE INFORMACIÓN QUE GENERAN LOS
VEHÍCULOS EN LA PLATAFORMA WEB CLOCATOR EN COMSATEL PERÚ S.A.C”**

Gary Manuel Ygnacio Pinedo

para optar el Título Profesional de
Ingeniero de Sistemas e Informática

Asesor: Jaime Oliver Velarde Araujo

Lima – Perú
Mayo del 2022

DEDICATORIA

Dedico este proyecto, con todo mi amor, a mis padres y mi hermano, quienes me han apoyado en todo tiempo y me impulsan a desarrollarme profesionalmente. A mi padre, por enseñarme que la vida es una lucha constante y que con esfuerzo puedo lograr mis sueños; a mi madre, por enseñarme a enfrentar las dificultades y nunca darme por vencido; a mi hermano, quien siempre ha sido mi referente como hombre íntegro, lleno valores y con su ejemplo me inspira a ser una mejor persona; Y sobre todo me lo dedico a mí, que aun en tiempos difíciles pude desarrollar esta tesis con éxito.

AGRADECIMIENTOS

En primer lugar, agradezco a Dios por darme la vida, la oportunidad de estudiar una carrera universitaria y ayudarme durante todo los años dedicados a terminar los estudios sin desviarme de su camino y siendo luz donde él me ponía.

A mis padres Manuel Walther Ygnacio y Judith Pinedo, a mi hermano Jean Pierre Ygnacio, por apoyarme siempre, aun cuando están lejos siempre se hicieron sentir cerca de mí.

A mi mejor amigo Fernando Legua SC, por animarme a terminar la tesis y siempre brindarme su apoyo incondicional.

A mis compañeros gatunos Normen y Love que me acompañaron durante todo el proceso de tesis, por las amanecidas que se dieron a mi lado y por enseñarme que el amor también se expresa sin palabras.

A mi buen amigo Emilio Carmelo Navarrini quien aún desde lejos siempre se mostró amigo y me apoyo durante el proceso de tesis y mi carrera universitaria

A la familia Ottaviano, Sara, Andrea, Patrizia y Ernestino quienes significan mucho para mi familia y que fueron parte fundamental en mi vida.

A mis abuelos Manuel Ygnacio Correa y Jesús Hernández Serrano, quienes trajeron al mundo un gran hombre quien es mi padre.

A mi abuela María Bautista Domínguez quien trajo a la vida una gran mujer quien es mi madre.

Tabla de Contenido

Resumen.....	15
Abstract.....	16
Introducción	17
Capítulo I. Planteamiento del Problema	18
1.1 Descripción de la Realidad Problemática	18
1.2 Formulación del Problema.....	19
1.2.1 Problema General.....	19
1.2.2 Problemas Específicos	19
1.3 Determinación de Objetivos.....	20
1.3.1 Objetivo General.....	20
1.3.2 Objetivos Específicos.....	20
1.4 Hipótesis	20
1.4.1 Hipótesis General.....	20
1.4.2 Hipótesis Específicas	20
1.5 Justificación de la Investigación	21
1.5.1 Teórica	21
1.5.2 Práctica.....	21
1.5.3 Metodológica	22
1.5.4 Delimitación del Estudio.....	22

1.5.4.1	Espacial.....	22
1.5.4.2	Temporal.....	23
1.5.4.3	Conceptual.....	23
Capítulo II. Marco Teórico		23
2.1	Antecedentes de la Investigación.....	23
2.2	Bases Teóricas	29
2.2.1	Cubo Multidimensional	29
2.2.1.1	Tipos de Almacenamiento de Información en los Cubos Multidimensionales (OLAP).....	30
2.2.1.1.1	MOLAP.....	30
2.2.1.1.2	ROLAP.....	31
2.2.1.1.3	HOLAP.....	31
2.2.1.2	Datos Precalculados.....	32
2.2.1.3	Velocidad de Respuesta.....	32
2.2.2	Análisis de Información.....	33
2.2.2.1	Interpretación de la Información	33
2.2.2.2	Investigación Cuantitativa	34
2.2.2.3	Establecer Categorías	35
2.2.3	Metodología de Ralph Kimball.....	35
2.2.3.1	Planificación.....	37
2.2.3.2	Análisis de Requerimientos.....	38

2.2.3.3	Modelado Dimensional.....	38
2.2.3.4	Diseño Físico.	39
2.2.3.5	Diseño de Sistema ETL.	40
2.2.3.6	Diseño de Arquitectura Técnica.	40
2.2.3.7	Selección de Productos e Instalación.....	41
2.2.3.8	Desarrollo de Aplicaciones para Usuarios Finales.	41
2.2.3.9	Implementación.	41
2.2.3.10	Mantenimiento y Crecimiento.	41
2.2.4	Metodología de Bill Inmon.....	42
2.2.5	MySQL	43
2.2.6	Apache Superset.....	43
2.2.7	SSIS (SQL Server Integration Services).....	44
2.3	Contexto de la Investigación.....	44
Capítulo III. Metodología de Investigación		46
3.1	Diseño de Investigación	46
3.1.1	Diseño	46
3.1.2	Tipo.....	47
3.1.3	Enfoque.....	47
3.1.4	Población.....	48
3.1.5	Muestra	48
3.1.6	Operacionalización de Variables	49
3.1.6.1	Cubo Multidimensional.....	49

3.1.6.2	Análisis de Información.....	49
3.2	Instrumentos de Investigación / Herramientas.....	52
3.2.1	Técnicas	52
3.2.2	Instrumentos.....	52
3.2.2.1	Validación del Instrumento.	53
3.2.3	Confiabilidad.....	53
3.3	Metodología de la Implementación de la Solución.....	54
3.3.1	Selección de la Metodología.....	54
3.3.2	Desarrollo de la Metodología – Ralph Kimball.....	56
3.3.2.1	Fase 01. Planificación.....	56
3.3.2.2	Fase 02. Definición de los Requerimientos del Negocio.....	56
3.3.2.3	Fase 03. Modelo Dimensional.	57
3.3.2.4	Fase 04. Diseño Físico.....	58
3.3.2.5	Fase 05. Diseño del Sistema de ETL.....	58
3.3.2.6	Fase 06. Implementación.....	59
3.3.2.7	Fase 07. Especificación y Desarrollo de Aplicaciones BI.....	59
3.4	Metodología para la Medición de Resultados de la Implementación	59
3.4.1	Metodología de Medición.....	59
3.4.2	Medición de los Indicadores	60
3.5	Cronograma de Actividades.....	61
3.6	Costos Directos	62

3.6.1	Costos de Personal y Materiales	62
3.7	Costos Indirectos.....	62
3.8	Costos Fijos.....	63
3.9	Costos Variables	63
3.10	Presupuesto	64
Capítulo IV. Desarrollo de la Solución.....		73
4.1	Propuesta de Solución.....	73
4.1.1	Fase 01. Planificación	73
4.1.1.1	Caso del Negocio.....	73
4.1.1.2	Definición de los Requerimientos del Negocio.....	74
4.1.2	Fase 02. Diseño del Modelo Dimensional	75
4.1.3	Fase 03. Diseño Físico	75
4.1.4	Fase 04. Diseño ETL.....	84
4.1.5	Fase 05. Implementación	85
4.1.5.1	Validar Información.	85
4.1.6	Fase 06. Diseño de Aplicación BI.....	86
4.2	Prototipos de la Aplicación.....	91
4.2.1	Sistema del cubo	91
4.2.2	Sistema ETL.....	92
4.2.3	Sistema BI.....	94
4.3	Medición de la Solución	98

4.3.1	Contrastación de Hipótesis	64
4.3.1.1	Hipótesis General.	64
4.3.1.2	Hipótesis Específica 01.	65
4.3.1.3	Hipótesis Específica 02.	66
4.3.1.4	Hipótesis Específica 03.	68
4.3.1.5	Resultados de Método de Reprocesamiento.	69
4.3.2	Análisis de Indicadores Cuantitativos.....	98
4.3.2.1	Comparación pretest-postest.....	99
4.3.2.2	Variable Dependiente Análisis de Información.....	104
4.3.2.3	Comparación Pretest-Postest.	105
Capítulo V. Conclusiones y Recomendaciones		110
5.1	Discusión y Conclusiones	110
5.1.1	Discusión.....	110
5.1.2	Conclusiones.....	112
5.2	Recomendaciones.....	113
6	Referencias	114
7	Anexos.....	119

Lista de Tablas

Tabla 1 Validación del instrumento.....	53
Tabla 2 Resultados de confiabilidad.....	54
Tabla 3 Contraste de hipótesis de estudio.....	64
Tabla 4 Prueba de Wilcoxon.....	69
Tabla 5 Prueba de Wilcoxon. Efecto de la propuesta de cubo multidimensional de eventos Safe en la investigación cuantitativa que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.	70
Tabla 6 Prueba de Wilcoxon. Efecto de la propuesta de cubo multidimensional de eventos Safe en la interpretación de la información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.	71
Tabla 7 Prueba de Wilcoxon. Efecto de la propuesta de cubo multidimensional de eventos Safe en establecer categorías que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.....	72
Tabla 8 Pretest - análisis descriptivo de la variable cubo dimensional.....	98
Tabla 9 Postest - análisis descriptivo de la variable cubo dimensional	98
Tabla 10 Comparación de la media de la variable “cubo dimensional” y sus dimensiones	99
Tabla 11 Pretest - análisis descriptivo de la variable análisis de información.....	104
Tabla 12 Postest - análisis descriptivo de la variable análisis de información	105
Tabla 13 Comparación de la media de la variable análisis de información y sus dimensiones .	105

Lista de Figuras

Figura 1 Tareas del ciclo de vida en la metodología de Ralph Kimball	36
Figura 2 Arquitectura de un data Warehouse según Bill Inmon.....	42
Figura 3 Matriz de operacional de variables.....	51
Figura 4 Características de las metodologías.....	55
Figura 5 Cronograma de actividades	61
Figura 6 Detalle sobre costos directos.	62
Figura 7 Detalle sobre costos indirectos.	63
Figura 8 Detalle sobre costos fijos.....	63
Figura 9 Detalle sobre costos variables.	63
Figura 10 Detalle sobre presupuesto.....	72
Figura 11 Índices de dimensión conductor.	76
Figura 12 Índices de dimensión flota.....	76
Figura 13 Índices de dimensión compañía.....	77
Figura 14 Índices de dimensión vehículo.	77
Figura 15 Índices de dimensión tipo vehículo	78
Figura 16 Índices de dimensión tiempo.	78
Figura 17 Índices de tabla hechos fact_esafe.....	78
Figura 18 Estructura de tabla dimensional conductor.....	79
Figura 19 Estructura de tabla dimensional flota	79
Figura 20 Estructura de tabla dimensional compañía	80
Figura 21 Estructura de tabla dimensional vehículo.....	80

Figura 22 Estructura de tabla dimensional tipo vehículo.....	81
Figura 23 Estructura de tabla dimensional tipo vehículo.....	81
Figura 24 Estructura de tabla dimensional tipo vehículo.....	82
Figura 25 Estructura del sistema ELT	84
Figura 26 Resultados de carga en la tabla hechos.....	85
Figura 27 Conexión a base de datos desde Apache Superset	87
Figura 28 Configuración y creación dataset	88
Figura 29 Resultado final del filtro para las compañías	89
Figura 30 Dataset para llevar el conteo de las unidades analizadas	89
Figura 31 Resultado del dataset: un cuadro con el conteo total de las unidades analizadas en todo el cubo multidimensional.....	90
Figura 32 Estructura multidimensional tipo estrella para el análisis de información de los eventos Safe generados en la plataforma web Clocator.....	91
Figura 33 Flujo de carga ETL.....	92
Figura 34 Query para realizar a dimensión vehículo	93
Figura 35 Relaciones internas, base de datos origen y cubo multidimensional.....	93
Figura 36 Filtros para detallar la información del cubo.....	94
Figura 37 Métricas	94
Figura 38 Grupos de evento principales.	95
Figura 39 Tipos de evento.....	95
Figura 40 Tipos de vehículo	96
Figura 41 Dashboard cubo multidimensional Esafe	97
Figura 42 Prueba Chi-cuadrado para comprobar la hipótesis general	65

Figura 43 Prueba Chi-cuadrado para comprobar la hipótesis específica 01	66
Figura 44 Prueba Chi-cuadrado para comprobar la hipótesis específica 02	67
Figura 45 Prueba Chi-cuadrado para comprobar la hipótesis específica 03	69
Figura 46 Comparación pretest – postest de la variable “cubo multidimensional”	100
Figura 47 Comparación pretest – postest de la dimensión “almacén de información”	101
Figura 48 Comparación pretest – postest de la dimensión velocidad de respuesta	102
Figura 49 Comparación pretest – postest de la dimensión datos precalculados	103
Figura 50 Comparación pretest – postest de la variable análisis de información	106
Figura 51 Comparación pretest – postest de la dimensión investigación cuantitativa	107
Figura 52 Comparación pretest – postest de la dimensión interpretación de la información	108
Figura 53 Comparación pretest – postest de la dimensión establecer categorías	109

Lista de Anexos

Anexo 1. Matriz de consistencia.....	119
Anexo 2. Ficha de investigación.....	120
Anexo 3. Declaración de autenticidad y no plagio	120
Anexo 4. Formulario de autorización de publicación.....	122
Anexo 5. Reporte de similitud por Turnitin.....	123

Resumen

El presente proyecto de investigación se propone implementar un cubo multidimensional de eventos Safe para mejorar el análisis de información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C, cuya misión es capturar y procesar los eventos generados por los vehículos y, de esta forma, facilitar el estudio de la información.

Este trabajo tiene un enfoque cuantitativo de tipo aplicado, para el cual se obtuvo una muestra de 24 personas del área de TI. El instrumento usado para recolectar los datos fue la encuesta, que permitió tener datos acerca del estado actual respecto al análisis de información de los eventos Safe y evaluar la eficiencia del proyecto posterior a la implementación del cubo multidimensional. El instrumento estuvo validado por el juicio de expertos, con un nivel de confianza de 0.70 para la variable independiente y 0.78 para la variable dependiente.

En el desarrollo del proyecto y la creación del cubo multidimensional se utilizó la metodología de Ralph Kimball, que posibilitó la elaboración de un plan y unas tareas que ayuden a crear un almacén de datos sólidos y cumplan con los objetivos. Los resultados fueron satisfactorios, pues el cubo elaborado facilita el estudio de los eventos que generan los vehículos en la plataforma web Clocator. Como parte del resultado se obtuvo la evidencia en el análisis realizado a la hipótesis general, donde el valor obtenido de $X^2 = 7.464$ (superior al X^2 crítico = 3.36) y $p = 0.006$, así $p < 0.05$, entonces es posible afirmar que la implementación de un cubo multidimensional de eventos Safe permite mejorar el análisis de información.

Palabras clave: cubo multidimensional, Ralph Kimball, análisis de información, plataforma web, gerencia TI.

Abstract

The thesis project aims to implement a multidimensional cube of Safe events to improve the analysis of information generated by vehicles in the web platform Clocator in Comsatel Peru S.A.C, which has the mission to capture and process the events generated by vehicles and thus facilitate the analysis of information. The project by its nature has a quantitative approach of applied type for which a sample of 24 people from the IT area was obtained; The instrument used to collect information was the survey that allowed to have information about the current state regarding the analysis of information of the Safe events and after the implementation of the multidimensional cube to evaluate the efficiency of the project; The instrument was validated by the judgment of experts with a confidence level of 0.70 for the independent variable and 0.78 for the dependent variable. In the development of the project and creation of the multidimensional cube we used Ralph Kimball's methodology, which allows us to develop a plan and tasks that help to create a solid data warehouse that meets the objectives. At the end of the project the results were satisfactory since the multidimensional cube allows to analyze the events generated by the vehicles in the Clocator web platform, as part of this result the evidence is obtained in the analysis performed to the general hypothesis where the obtained value of $X^2 = 7.464$ (higher than the critical $X^2 = 3.36$) and $p = 0.006$, so $p < 0.05$, then it is possible to affirm that the implementation of a multidimensional cube of Safe events allows to improve the information analysis.

Keywords: Multidimensional cube, Ralph Kimball, information analysis, web platform, IT management.

Introducción

Actualmente, muchas empresas trabajan con el cubo multidimensional para analizar y explotar su información. Esta herramienta, debido a su estructura, proporciona ventajas para dicha labor y cuenta con métricas precalculadas, por lo que los gerentes de TI u otros interesados pueden analizar y visualizar de manera sencilla e intuitiva la información que tiene almacenada, pues hace posible optimizar los tiempos de respuesta ante consultas, pero si estas se hacen directamente en bases de datos tardarían mucho tiempo. Por tal motivo, la estructura de este cubo es más confiable, además de que reduce los riesgos de saturar la base de datos con consultas complejas y proporciona información detallada.

Partiendo de lo anterior, este estudio tiene como objetivo implementar un cubo multidimensional de eventos Safe para mejorar el análisis de información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C bajo la metodología de Ralph Kimball, para capturar y procesar los eventos que ayudarán al análisis y monitoreo de los vehículos, de modo que se obtenga información detallada sobre compañías, flotas, conductores, tipo de vehículo y por fechas.

En cuanto a la estructura, este trabajo comprende cinco capítulos. En el primer capítulo se exponen la problemática, se formula el problema y se determinan los objetivos, la hipótesis, la justificación y la delimitación de la investigación. En el segundo se evidencia el marco teórico, se analizan los antecedentes, así como las bases teóricas de las variables dependiente e independiente. En el tercer capítulo se detalla el diseño de la investigación, el tipo, el enfoque, la población, la muestra, la operacionalización de la variable independiente y las dependientes. También se presentan los instrumentos utilizados, al igual que la técnica, la validez y el grado de

confiabilidad del instrumento basadas en el juicio de expertos. Así mismo, se explica la metodología aplicada para la solución, que es la de Ralph Kimball, la cual ayuda a crear rápidamente almacenes de datos sólidos que se adaptan al crear diseños para el análisis de información. Posteriormente, en el cuarto capítulo se observa el desarrollo de la solución, además del análisis y la interpretación de los resultados. Por último, el quinto capítulo contiene las discusiones, las conclusiones y las recomendaciones en función de lo expuesto a lo largo del estudio.

Capítulo I. Planteamiento del Problema

1.1 Descripción de la Realidad Problemática

Durante los últimos años, la explotación de la información en las organizaciones se ha convertido en una herramienta estratégica, pues analizar los datos correctos resulta útil para la toma de decisiones y ayuda a prepararse ante acontecimientos futuros o a conocer el estado actual de la empresa. Por ello, contar con una estructura adecuada para el almacén de datos y manejar correctamente la información permitirá que la gerencia tenga un panorama claro antes de decidir qué acciones tomar, de manera que impulse el desarrollo de la empresa, en busca de mejorar sus utilidades a largo plazo y fortalecer sus debilidades.

La empresa Comsatel Perú S.A.C tiene una plataforma web llamada Clocator, la cual permite a los usuarios hacer seguimiento de sus flotas, dado que estas unidades cuentan con dispositivos instalados que monitorean el comportamiento mientras están en recorrido. Sin embargo, toda la información que es generada diariamente no está siendo analizada por la plataforma, debido al gran volumen de datos que generan los eventos denominados Safe;

también se tiene que la plataforma no genera reportes específicos para conocer los eventos producidos por empresa, como flota, tipo de vehículo, conductor y fechas, y tampoco para categorizar todos estos datos por los diferentes tipos de eventos Safe que existen. Así pues, se requiere implementar un cubo multidimensional que contribuya a analizar los eventos Safe en la plataforma web Clocator.

1.2 Formulación del Problema

1.2.1 Problema General

PG. ¿De qué manera la implementación de un cubo multidimensional de eventos Safe permite mejorar el análisis de información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C?

1.2.2 Problemas Específicos

01. ¿De qué manera un cubo multidimensional permite interpretar la información de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C?

02. ¿De qué manera un cubo multidimensional permite una investigación cuantitativa de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C?

03. ¿De qué manera un cubo multidimensional permite establecer categorías que ayuden al análisis de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C?

1.3 Determinación de Objetivos

1.3.1 Objetivo General

OG. Implementar un cubo multidimensional de eventos Safe para mejorar el análisis de información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.

1.3.2 Objetivos Específicos

01. Implementar un cubo multidimensional para interpretar la información de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

02. Implementar un cubo multidimensional para realizar una investigación cuantitativa de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

03. Implementar un cubo multidimensional para establecer categorías que ayuden al análisis de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

1.4 Hipótesis

1.4.1 Hipótesis General

HG. La implementación de un cubo multidimensional de eventos Safe permite mejorar el análisis de información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.

1.4.2 Hipótesis Específicas

01. La implementación de un cubo multidimensional permite interpretar la información de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

02. La implementación de un cubo multidimensional permite una investigación cuantitativa de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

03. La implementación de un cubo multidimensional permite establecer categorías que ayuden al análisis de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

1.5 Justificación de la Investigación

1.5.1 Teórica

Este trabajo brinda información relacionada con la implementación de un cubo multidimensional y el análisis de información, el cual se realiza con base en la metodología de Ralph Kimball; adicionalmente, posibilita la implementación del proyecto desde cero. Para todo ello, se toman como base diferentes estudios realizados en torno al tema que ocupa a este trabajo, entre ellos tesis y artículos científicos.

Los resultados obtenidos demuestran que al implementar un cubo multidimensional se mejora el análisis de información sobre los eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C, lo cual hace posible ayudar a la empresa a mejorar el análisis de su información almacenada. Así mismo, con esta implementación las respuestas se obtienen de manera detallada y mucho más rápida que al solicitarlas directamente de la base de datos.

1.5.2 Práctica

La justificación práctica radica principalmente en obtener los requerimientos necesarios para hacer la estructura del nuevo almacén de datos (cubo multidimensional), el cual debe permitir el análisis requerido por los interesados de la gerencia de TI, a fin de ayudarlos a

analizar los eventos generados a través de métricas calculadas y consultas detalladas que sean rápidas, pues actualmente generar este tipo de información de modo preciso no es posible, y de serlo implicaría mucho tiempo para extraerla de la base de datos, debido a la gran cantidad de datos que se almacena en esta. Así pues, con la implementación del cubo multidimensional se crearán nuevas oportunidades para mejorar o anticipar acontecimientos desfavorables, teniendo disponible la información detallada y actualizada.

1.5.3 Metodológica

Para esta investigación, la justificación es de tipo aplicada, puesto que el objetivo es ayudar al área de gerencia de TI a analizar la información capturada mediante la plataforma web Clocator, dado que actualmente no se cuenta con un análisis de los eventos generados. En cuanto a su contribución, este trabajo puede ayudar a los involucrados a tener datos detallados para supervisar, controlar y monitorear la información de eventos que emiten los dispositivos instalados en los vehículos; para ello se eligió la metodología Kimball, con el propósito de elaborar el proyecto a mediano plazo.

1.5.4 Delimitación del Estudio

1.5.4.1 Espacial.

La presente investigación se lleva a cabo en la empresa Comsatel Perú S.A.C, que tiene una sede principal con una gerencia de TI, la cual está conformada por 26 trabajadores del área TI (gerentes, jefes y empleados). En este caso se tomó una muestra delimitada para realizar el estudio, y todos los involucrados fueron la fuente de información que facilitó la medición del contexto actual y del contexto después de la implementación.

1.5.4.2 Temporal.

El trabajo tendrá un alcance de aproximadamente seis meses, y su conclusión está programada a mediados de enero de 2022.

1.5.4.3 Conceptual.

Este proyecto aplica la metodología de Ralph Kimball como parte de la mejora por medio de la implementación del cubo multidimensional. El objeto de estudio es la plataforma web Clocator, la cual captura los eventos Safe. Cabe destacar que la presente investigación abarca la planificación del proyecto y define las necesidades para tomar los requerimientos adecuados que aporten al cumplimiento del objetivo. Por esa razón, se presentan fundamentos nacional e internacionales, y no hay delimitación para el desarrollo del estudio.

Capítulo II. Marco Teórico

2.1 Antecedentes de la Investigación

En los últimos años han surgido muchas investigaciones que abordan el tema de los cubos multidimensionales, también llamados cubos OLAP, los cuales se han convertido en parte de la estrategia de las organizaciones para provechar toda la información que poseen sus almacenes de datos. Llegados a este punto cabe preguntarse ¿en qué radica el éxito de los cubos multidimensionales? y ¿de qué manera ayudan a las organizaciones?

Los buenos resultados de estos cubos radican en el levantamiento de la información y el diseño estructural del que está conformado. Tal como explicó Poveda (2018), el primer paso para elaborar el diseño de un cubo OLAP es establecer qué dimensiones y qué métricas harán parte de

él, considerando algunos estándares basados en datos administrados por la organización.

Además, el autor enfatizó en la realización de una buena estructura, y parte de ello es definir bien las dimensiones, así que antes de hacer cualquier proceso de ETL (extracción, transformación y carga) para cargar el cubo de información se debe tener claro cuál es la necesidad de la empresa y el objetivo que este tendrá, a fin de cumplir los objetivos trazados.

En esta línea, Sacolick (2020) expuso que, ante todo, los datos deben ser depurados; luego, dependiendo de su volumen, variedad y velocidad, es probable que se encuentren múltiples bases de datos internas y almacenes de datos en nube. La depuración de datos ayudará a realizar el precálculo, el cual será útil para el análisis antes de poblar el cubo multidimensional mediante el proceso de ETL. Cabe señalar que la depuración a la que se hace referencia tiene que ver con quitar campos innecesarios que no contribuirán al análisis o al cálculo de las métricas, lo que ayuda a mejorar la velocidad de respuesta, teniendo en cuenta que a medida que se vaya insertando información, el volumen de los datos aumentará y la velocidad de respuesta se verá comprometida.

En lo concerniente al filtro de información, la disponibilidad de datos y la importancia del cubo multidimensional dentro de la organización, Violino (2017) indicó que es necesario tener presente que los datos correctos siempre deben estar disponibles y traducidos al lenguaje comercial, de modo que el usuario comprenda el resultado y pueda usarse para generar oportunidades o mejorar procesos. Lo anterior permite un análisis óptimo, el cual ayudará a los interesados en el negocio.

Es importante recalcar que la información, una vez filtrada y procesada, debe ser entendible para todos los involucrados, en busca de que la organización sea más competente respecto a sus rivales, lo cual se logra al anticiparse a hechos u eventos que puedan perjudicarlos,

a través de la mejora y el ofrecimiento de un mejor servicio a sus clientes. Aunado a esto, es preciso definir bien a las personas que proporcionarán esta información para analizarla, pues no todos pueden tomar decisiones importantes dentro de la organización. Por ello, las métricas que serán calculadas deben ir acordes a los requerimientos, con el propósito de satisfacer la necesidad al solucionar o cumplir un objetivo.

Por su parte, Rolón (2021) planteó que cada cubo multidimensional está estructurado por una tabla de hechos y sus dimensiones, por lo que cuenta con una tabla central denominada hechos, que contiene las llaves principales de las diferentes dimensiones. Dichas llaves sirven para establecer una relación entre la tabla central y las dimensiones que están a su alrededor, las cuales deben de contener información que permita analizar el cubo correctamente, dado que ayudarán a describir los datos calculados en la tabla hechos. Las dimensiones, además, facilitarán la agrupación y la descripción del cubo, al igual que la creación de una estructura correcta y bien definida para este, lo que hará más eficiente el tiempo de consulta y, por ende, se obtendrán respuestas más rápidas.

Desde el punto de vista de Orlandi (2020), es bastante popular representar un modelo dimensional a modo de tabla de hechos, junto a sus tablas dimensionales, o presentar un modelo en estrella o esquema de copo de nieve, donde la rama proviene de la tabla de dimensiones, no solo de la de hechos. En concordancia, este autor señaló que existen dos tipos de modelos para cubos multidimensionales, a saber: el modelo estrella, el cual comprende a la tabla hechos como centro y alrededor sus dimensiones; y la estructura copo de nieve, cuyas dimensiones se relacionan tanto con las de la tabla hechos (principal) como con otras. En la actualidad, se considera que el modelo estrella tiene mejor implementación, debido a que únicamente depende de la tabla hechos, por lo que se obtiene una respuesta más rápida.

Por otro lado, las organizaciones, para asegurar un manejo responsable de la información que poseen y obtener todos los resultados que desean, deben contar con personal calificado y capaz de proporcionar la facilidad para que esa información sea explotada, porque si no es gestionada por el personal capacitado para la implementación de cubos multidimensionales, todos los recursos se pueden desperdiciar.

De acuerdo con Pratt (2020), las organizaciones no pueden tener buenas prácticas de limpieza de datos sin un programa de gobierno de datos completamente implementado, tampoco les es acceder o integrar los datos que poseen, puesto que permanecen encerrados en silos departamentales ni sabrán qué datos son necesarios. Tal como señaló el autor, las organizaciones que no tienen control y gobierno de sus datos no podrán gestionar eficazmente la información y la explotación de esta será limitada.

Por lo anterior, es sumamente importante que, antes de implementar los almacenes de datos, las organizaciones definan de modo oportuno los recursos esenciales para trabajar con los datos, de lo contrario, su competencia en el mercado será limitada frente sus rivales directos, situación que conllevará al fracaso, y sin conocer sus debilidades y fortalezas quedarán en una posición de desventaja. De ahí la relevancia no solo de invertir en tecnología de infraestructura, sino también en profesionales que tengan habilidades para gestionar, ordenar y analizar la información almacenada en las bases de datos. Esto, sin duda, impulsará la mejora continua de la organización para que sea más eficaz.

Según Olavsrud (2020), el análisis de datos es un campo que continúa creciendo a medida que las organizaciones siguen invirtiendo en análisis para respaldar la transformación digital. En este contexto, mantenerse actualizado con las últimas tendencias es fundamental para garantizar que la organización adopte las estrategias de análisis necesarias para obtener el éxito

durante los próximos años. Como apuntó el autor, existen empresas que utilizan la tecnología de forma correcta y explotan bien su información, debido a que han identificado la importancia de saber manejar las estrategias, por consiguiente, invierten en nuevas tecnologías y en profesionales con la capacidad de extraer los datos vitales que ayuden a mejorar sus procesos, acoplar su información y alinearla a decisiones eficaces.

Sumado a esto, algunas compañías sacan el máximo provecho de la información y la llevan a otro nivel, por lo que pueden anticiparse y predecir ciertos comportamientos. Actualmente, el análisis de datos lleva todo a un nivel de competencia entre las organizaciones, es decir, la que más datos posee tendrá ventaja sobre el otro, y así se reduce el índice de fracaso y se aumentan las probabilidades de éxito.

Teniendo en cuenta a Wayner (2020) se tiene que, en cierto momento, las computadoras pasaron de recolectores datos a “bolas de cristal” que predicen el futuro (unos pocos segundos, pocos días e incluso años) por medio de la evaluación de datos. Así pues, estructurar un cubo multidimensional o base de datos de manera sólida hará que soporte un gran volumen de información y esta podrá analizarse y usarse favorablemente para anticiparse a ciertos eventos, considerando que a más información, más conocimiento. Esto, además, mejorará y fortalecerá los procesos dentro de la organización, al alinear sus objetivos, corregir errores y, por tanto, ser más competentes.

Acerca de las ventajas de analizar los datos se puede decir que los más interesados en manejarlos y analizarlos correctamente son los dueños de las organizaciones, negocios, etc. En la opinión de Peña et al. (2019), en los cubos multidimensionales y los reportes gerenciales hoy en día se utilizan en cada uno de los sectores empresariales, pues pueden resumir información para tomar decisiones.

Según estos autores, para la toma de decisiones a nivel gerencial es fundamental generar reportes a partir de un buen análisis, lo cual les permitirá decidir qué hacer, los guiará a buscar soluciones, proponer mejoras, fortalecer debilidades y mejorar constantemente a través de los resultados, que irán acordes a las métricas y los requerimientos establecidos para cumplir con el objetivo. Por ello, también son relevantes las herramientas que faciliten la visualización de estos datos mediante *dashboard's*. También mencionaron que analizar y explotar la información siempre va de la mano con una herramienta que haga posible modelar el estudio en *dashboard* (Peña et al., 2019).

Finalmente, Micha (2021) argumentó que el análisis de datos, en apariencia, es simple si se usan aplicaciones hechas o un ERP sofisticado; empero, se trata de algo complejo, dado que estas aplicaciones disponen de informes estándar que exponen información estáticamente, de modo que no ayudan profundizar en el análisis de datos, adentrarse en ellos o manejarlos a través de diversas perspectivas.

Como apuntó el autor, el desarrollo de la estructura que almacenará la información no es fácil de realizar, pues todo parte de los requerimientos para el desarrollo, luego se estructura un buen almacén de datos, se definen sus dimensiones y se realiza la carga de información. Entonces, hacerlo con herramientas estándar no garantizará un buen análisis de los datos y, por ende, no será posible profundizar en los detalles, los cuales son provechosos si se diseña un cubo multidimensional eficiente que se ajuste a los requerimientos (Micha, 2021).

2.2 Bases Teóricas

2.2.1 *Cubo Multidimensional*

La tecnología de procesamiento analítico en línea – OLAP, también llamada cubo multidimensional, “posibilita la utilización óptima de los almacenes de datos, a fin de mejorar el análisis de la información y proporcionar respuestas inmediatas para las consultas analíticas repetitivas y complejas” (Auribox Training, 2017, párr. 1). Los cubos están conformados por una tabla como elemento principal (llamada hechos), en la cual se hallan datos precalculados y las claves primarias de sus dimensiones, las cuales se ubican a su alrededor como parte esencial de su estructura, por lo que contribuyen a agrupar, identificar y detallar los datos que están en la tabla principal.

El objetivo de los cubos OLAP es proporcionar reportes que permitan una lectura rápida de la información almacenada, de modo que faciliten la toma de decisiones. De acuerdo con Fuentes y Valdivia (2010), el uso de gráficos y cubos multidimensionales ocasiona una mayor dinámica, por lo que es posible relacionarse con los datos que se manejan en el Data Mart; asimismo, las posibilidades de exportar esto a archivos tipo Excel o PDF brindan más portabilidad del reporte para los usuarios.

Según la conceptualización del cubo OLAP, sus componentes se pueden mostrar, por ejemplo, cada dimensión tiene uno o varios niveles, que son agrupaciones ordenadas de manera jerárquica al interior de una dimensión. Para ilustrar esta explicación se propone la “ubicación” como dimensión, que puede componerse por estos niveles: continente, región, país y ciudad; los cuales tienen dentro de sí al anterior, por ejemplo, los continentes contienen regiones y así sucesivamente. Estos niveles posibilitan la obtención rápida de resultados totales de información;

en el caso del ejemplo, con los cuatro niveles se pueden obtener las ventas totales por continente, región, país y ciudad. Así pues, los cubos están compuestos, básicamente, por dimensiones, miembros (estos dos últimos resultan de las tablas dimensionales) y celdas (que surgen de la tabla de hechos).

2.2.1.1 Tipos de Almacenamiento de Información en los Cubos

Multidimensionales (OLAP).

2.2.1.1.1 MOLAP.

Este almacenamiento “ofrece un mayor rendimiento, pero requiere más espacio porque los datos del cubo son actualizados cada vez que se procesa, mientras que los cambios desde el origen de los datos se actualizan periódicamente” (Pesquera, 2020, párr. 10). El tipo de cubo MOLAP es un sistema que utiliza una base de datos multidimensional para almacenar información multidimensionalmente, debido a que emplea dos niveles en su arquitectura (BDMD - base de datos multidimensional y motor analítico), encargada de manejar, acceder y obtener los datos; por otro lado, como aplicación, ejecuta las consultas OLAP. De esta manera se puede indicar que MOLAP está diseñado para guardar pocos datos, lo que limita en cuanto al tamaño de la base de datos. Por ello, los usuarios que necesiten datos agregados y sumariados requerirán una arquitectura MOLAP (García, 2012).

2.2.1.1.2 ROLAP.

En el caso de ROLAP no usa almacenamiento multidimensional, por lo que es más lento al momento de procesar las consultas, pero posibilita acceder a los datos en tiempo real y usa menos espacio (Pesquera, 2020).

Por lo anterior, es posible señalar que no usa tres niveles en su arquitectura. Además, dirige el almacenamiento de datos, OLAP brinda la función analítica y, para la presentación, una herramienta especial. En cuanto a la aplicación, en este nivel se ejecutan las consultas o las búsquedas del usuario (García, 2012).

2.2.1.1.3 HOLAP.

Este tipo de almacenamiento es híbrido entre MOLAP y ROLAP, por lo cual realiza doble función: ordena la información en un MOLAP y lo restante como ROLAP. Para esto hace uso de dos clases de bases de datos, donde guarda la información: relacional y multidimensional, con base en los requerimientos de los datos que deben ser procesados.

2.2.1.2 Almacén de información

En el crecimiento constante de las bases de datos y sus usos, integran los datos que provienen de diferentes fuentes con fines de intereses específicos para la organización, cuando los datos se almacenan ya no pueden eliminarse, esto sirve para mantener un historial de aquello guardado (Salvador et al., 2017, pp. 11-14). En ese sentido, una vez que los datos son almacenados podrán ser utilizados para analizarlos.

2.2.1.3 Datos Precalculados.

El cubo tiene información precalculada en la tabla hechos, por lo que el tiempo de respuesta resulta ser reducido. En caso de estar bien diseñado el cubo, será igual de rápido hacer la consulta sobre las ventas de cualquier ciudad, un país entero o todas las ventas de la compañía (Business Intelligence, 2008).

2.2.1.4 Velocidad de Respuesta.

La flexibilidad y la velocidad requerida para ayudar a los analistas del negocio es proporcionada por los sistemas OLAP, dado que brindan ciertas ventajas; de acuerdo con ello, se puede decir que es un modelo multidimensional y de carácter intuitivo que hace más rápido escoger, recorrer y explorar datos al usar lenguaje analítico para consultas, de modo que esto ofrece la oportunidad de explorar las relaciones más intrínsecas entre los datos de la compañía y la realización de cálculos sobre la información más consultada, lo que posibilitará una respuesta óptima a las consultas hechas correctamente (Auribox Training, 2017).

Cabe señalar que el procesamiento OLAP facilita la extracción, de manera sencilla y selectiva, de los datos que necesita el usuario entre un gran volumen de información; su rapidez radica en su estructura agrupada jerárquicamente.

2.2.2 *Análisis de Información*

En la actualidad, la tecnología ha ido avanzando exponencialmente, de tal forma que el análisis de información se ha vuelto clave en muchas organizaciones para asegurar el éxito, pues proporciona ventajas respecto a los rivales directos con los que se compete en el mercado, por este motivo, muchas empresas invierten en analistas que ayuden a explotar la información que tienen almacenada en la base de datos. Tal como indicó Violino (2017), todos los proyectos importantes de análisis de datos tienen que garantizar una óptima gestión de la información y la calidad, de lo contrario, las posibilidades de fracaso serán mayores. En tal sentido, el análisis de la información es de suma importancia en los proyectos de la organización, pues ayuda a definir estrategias de negocio, fortalecer debilidades y anticiparse a acontecimientos de todo ámbito.

2.2.2.1 **Interpretación de la Información**

Es un proceso mediante el cual se “inspeccionan, limpian y transforman los datos, a fin de obtener información útil para derivar en conclusiones concretas que permitan esclarecer y tomar decisiones” (Euroinnova Business School, s.f., párr. 1). También se puede definir como parte del análisis de datos a través del que se recopila información y se obtienen resultados. Este proceso se divide en dos tipos:

- **Interpretación de datos cualitativos:** es obtener conclusiones y conceptos de datos heterogéneos, no estructurados ni expresados cuantificable o numéricamente.

- **Interpretación de datos cuantitativos:** es una muestra numérica; consiste en obtener respuesta de acuerdo con los datos que se tengan (Euroinnova Business School, s.f.)

El fin del análisis y la interpretación de datos es conseguir información provechosa y utilizable para tomar decisiones, sin importar sus características. La interpretación puede resumir y detallar los datos, identificar las relaciones existentes, comparar variables, métricas y ajuste de datos o realizar pronósticos para anticiparse a los hechos.

(Tecnologías Información, 2018, párr. 5-10)

Según lo encontrado en Tecnologías Información (2018), el análisis de los datos debe empezar en los objetivos del trabajo, lo que significa, a partir del motivo de la realización de la investigación; luego, la información sobre las dimensiones tiene que tomarse a detalle para obtener resultados coherentes con los objetivos planteados en la creación del cubo multidimensional.

2.2.2.2 Investigación Cuantitativa

Se caracteriza por usar técnicas cuantitativas, así que está relacionada con mediciones, usar magnitudes, observar y analizar unidades, muestreo, tratamiento de estadísticas. En *Metodología de la investigación*, Ñaupás et al. (2018) indicaron que el enfoque cuantitativo hace uso de la recogida y el análisis de los datos para responder preguntas investigativas y comprobar hipótesis; igualmente, a través de la estadística inferencial y descriptiva, de formular y probar hipótesis, tratamiento de datos estadísticos, diseño formal del tipo de investigación, el muestreo, entre otros, tiene confianza en la medición de las variables y los instrumentos para investigar.

2.2.2.3 Establecer Categorías

Categorizar consiste en dar un título o nombre y conceptualizar las unidades analíticas; cabe aclarar que en cada categorías es posible determinar subcategorías. La categorización puede llevarse a cabo de dos formas diferentes, pero complementarias, a saber: inductiva o deductivamente. En la forma inductiva, las categorías surgen de los datos, con base en patrones y recurrencias existentes en estos; mientras tanto, en la forma deductiva, las categorías surgen de los marcos teóricos (Marín et al., 2016).

2.2.3 Metodología de Ralph Kimball

En este trabajo, la metodología usada abarca lo que Ralph Kimball llamó “ciclo de vida dimensional del negocio”, que se enfoca en el diseño y el desarrollo de almacenes de datos, los cuales buscan elaborar estructuras sólidas que permitan soportar grandes cantidades de datos y, en cuanto a estructura, responder rápidamente ante consultas. Kimball planteó cuatro principios, los cuales fueron organizados por Rivadera (2010) de la siguiente manera.

- Enfocarse en el negocio: se deben identificar los requerimientos que la organización o el negocio necesita, usarlos para establecer relaciones sólidas y luego afinar su análisis y los conocimientos de quienes las implementan.
- Construir infraestructura óptima de información: hacer una base de datos única, sencilla, completa y eficaz, en la cual se evidencie el gran número de requerimientos determinados en la empresa.
- Hacer entregas en incrementos significativos: diseñar el almacén de datos (DW) en aumentos entregables en plazos de 6-12 meses. Para determinar el orden de

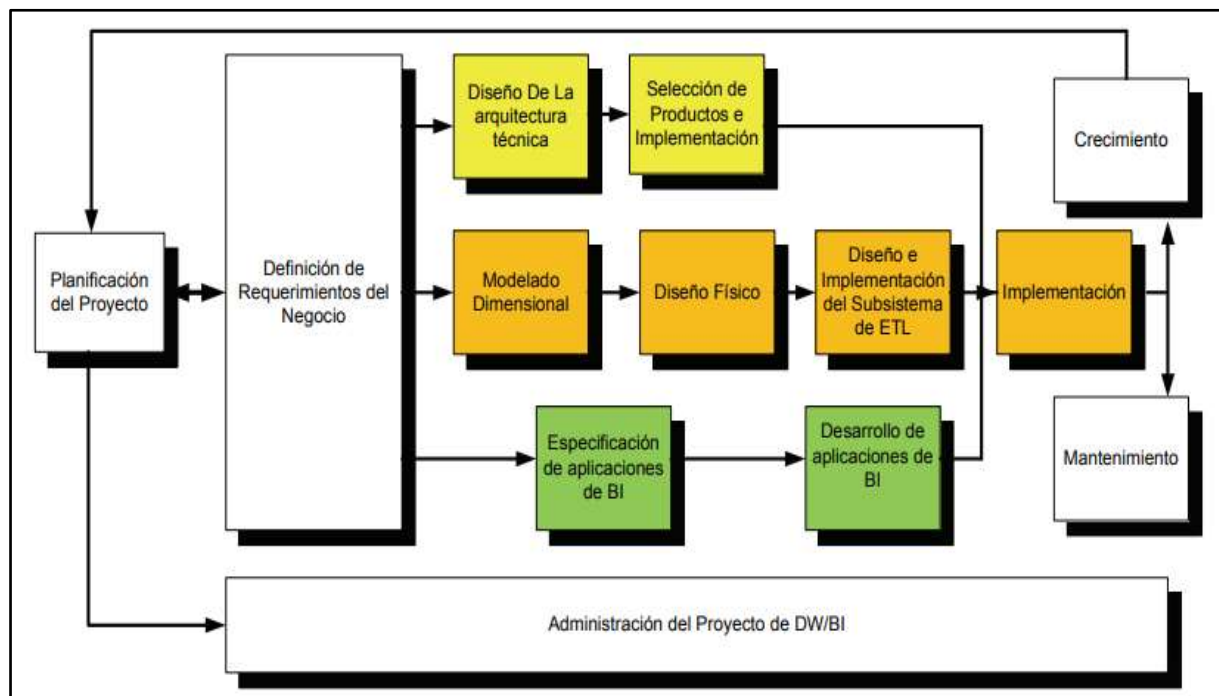
aplicación de los incrementos hay que usar el valor de negocio de cada elemento identificado. En este aspecto, es parecida a otras metodologías ágiles para construir software.

- Ofrecer una solución completa: brindar todo lo requerido para dar valor a los usuarios de la compañía. Para iniciar, es necesario crear un almacenamiento para datos sólido, con un buen diseño, de calidad y buen acceso. Además, se deberán dar aplicaciones para informes y análisis avanzados, capacitación, soporte, etc. (Rivadera, 2010).

Finalmente, para la construcción de un esquema que brinde solución a DW/BI (Datawarehouse/Business Intelligence) es bastante complejo, pero Kimball propuso una metodología simple para ello.

Figura 1

Tareas del ciclo de vida en la metodología de Ralph Kimball



Nota. Tomado de La metodología de Kimball para el diseño de almacenes de

datos (Data warehouses), por G. R. Rivera, 2010, p. 59, en *Cuadernos de la Facultad*, (5).

Como se observa en la Figura 1 se resaltan dos cuestiones: el papel central de la necesidad de definir los requerimientos de la empresa, al ser la base de las tareas siguientes e influyen en el plan de proyecto; y los tres caminos que hacen énfasis en áreas diferentes:

- Tecnología (camino superior): se relaciona con tareas asociadas a un software en especial, como Microsoft SQL Analysis Services.
- Datos (camino del medio): el modelo dimensional se diseña e implementa, después se desarrolla el subsistema de ETL para cargar el DW.
- Aplicaciones de Inteligencia de Negocios (camino inferior): aquí se hallan tareas de diseño y desarrollo de aplicaciones de negocios para usuarios finales.

(Rivadera, 2010, p. 59)

Partiendo de lo anterior, a continuación se mencionan, como parte fundamental de la estructura del plan propuesto por Kimball en su metodología, las diferentes fases que comprende y que ayudan a que la implementación del proyecto sea exitosa. De acuerdo con Rivadera (2010), desde la planificación y los requerimientos, todas las rutas de combinan cuando el sistema se instala.

2.2.3.1 Planificación.

En esta fase se determina el propósito del proyecto de DW/BI (Data Warehouse / Business intelligence) y sus tareas, que son definir el alcance, identificar y programar las tareas, planificar el uso de los recursos y elaborar la documentación final que representará el plan del proyecto. Siguiendo a Rivadera (2010), en esta parte se define la manera en que se realizará la

gestión de las subfases que comprenden todo el proyecto, por lo que se divide en monitoreo del proceso, rastreo de problemas y desarrollo de un plan de comunicación comprensiva que direcciona tanto la empresa como las áreas de TI.

2.2.3.2 Análisis de Requerimientos.

En la definición de los requerimientos se entrevista a los interesados del negocio, con el objetivo de evaluar los requerimientos necesarios para estructurar el proyecto. En esta etapa se debe aprender todo lo que se pueda acerca de la organización, los competidores, las fortalezas, las debilidades, la industria y los clientes. Tal como lo explicó Rivadera (2010), todos los informes posibles de la organización se deben leer, así como rastrear documentos de estrategia interna, analizar la competencia y lo que dice la prensa. De esta manera, el análisis de la información será eficaz y posibilitará la construcción de un proyecto sólido que contribuirá a la consecución de los objetivos.

2.2.3.3 Modelado Dimensional.

De acuerdo con Rivadera (2010), crear un modelo dimensional es un proceso dinámico e iterativo; es posible decir que elaborar la estructura multidimensional debe soportar cargas de datos que serán insertadas reiterativamente para elaborar cálculos de forma reiterativa. Acerca del proceso iterativo se tiene que consiste en cuatro pasos, que son elegir el proceso de negocio, establecer el nivel de granularidad, escoger las dimensiones apropiadas e identificar las medidas que contendrán datos calculados en la tabla de hechos.

2.2.3.4 Diseño Físico.

Para elaborar un almacén de datos robusto y sólido que soporte todo lo que se va desarrollar en el proyecto se deben tener en cuenta factores relevantes, por ello es importante definir bien qué equipos serán usados. En atención a esto, Rivadera (2010) elaboró un diseño físico donde se deben de responder las siguientes preguntas:

- ¿Cómo puede determinar cuán grande será el sistema de DW/BI?
- ¿Cuáles son los factores de uso que llevarán a una configuración más grande y más compleja?
- ¿Cómo se debe configurar el sistema?
- ¿Cuánta memoria y servidores se necesitan?
- ¿Qué tipo de almacenamiento y procesadores?
- ¿Cómo instalar el software en los servidores de desarrollo, prueba y producción?
- ¿Qué necesitan instalar los diferentes miembros del equipo de DW/BI en sus estaciones de trabajo?
- ¿Cómo convertir el modelo de datos lógico en uno de datos físicos en la base de datos relacional?
- ¿Cómo conseguir un plan de indexación inicial?
- ¿La partición debe usarse en las tablas relacionales? (Rivadera, 2010, pp. 67-68)

Citando a Rivera (2010), como parte de la implementación física de la infraestructura, la cual será pieza fundamental para el funcionamiento del cubo multidimensional, antes de invertir en tecnología se debe conocer qué tipo de equipos de infraestructura se necesitarán, con el propósito de contar con el equipo adecuado para un funcionamiento eficaz.

2.2.3.5 Diseño de Sistema ETL.

Una vez definida la parte de infraestructura, y considerando tanto el requerimiento como el análisis de la información que serán necesario para la evaluación y el cumplimiento de los objetivos por los que será creado, se procede a preparar la información para trasladarla al cubo multidimensional. En esta fase se llevan a cabo los procesos de ETL, que son:

- Extracción: se toma la información de un *data warehouse*, la cual será útil para el análisis.
- Transformación: los datos son transformados según los requerimientos establecidos, por ejemplo, elaborar cálculos, agrupar, unificar, distinguir, eliminar, etc.
- Carga: teniendo toda la información transformada en conformidad con lo que se necesita, se poblará el cubo multidimensional (dimensiones y tabla hechos).

Como lo expuso Rivadera (2010), si el sistema ETL se diseña adecuadamente puede extraer datos de los sistemas de origen de datos, aplicar distintas reglas para optimizar la calidad de estos, consolidar la información de distintos sistemas y, finalmente, cargar los datos al DW.

2.2.3.6 Diseño de Arquitectura Técnica.

Los ámbitos de *data warehousing* necesitan que se integren diversas tecnologías. Para este aspecto, se deben considerar estos factores, a saber: requerimientos de la compañía, ambientes técnicos de la actualidad y directrices técnicas-estratégicas, que se planean hacia el futuro, a fin de determinar la arquitectura de dicho ambiente, en cuanto a su diseño.

Antes de comenzar debe contar con un plan, dado que no es solo volver a ordenar y explotar la información. Tal como en una construcción, los planos se usan para expresar lo que

quieren el cliente y el arquitecto, además de medir los esfuerzos y los materiales requeridos para realizar la obra (documentación, planificación, comunicación, mantenimiento, etc.).

2.2.3.7 Selección de Productos e Instalación.

Al hacer uso del diseño de la arquitectura técnica como base es preciso evaluar y escoger componentes de esta, como el hardware, el motor de base de datos, las herramientas de acceso, entre otros. Seguidamente, se instala y se prueban los componentes en un ambiente integrado de *data warehousing*.

2.2.3.8 Desarrollo de Aplicaciones para Usuarios Finales.

En conformidad con distinción de aplicaciones para los usuarios últimos, el progreso de estas implica configurar el metadata y construir reportes determinados.

2.2.3.9 Implementación.

Esta es la parte final de las etapas que comprenden la metodología de Ralph Kimball, en la cual se implementa el almacén de datos. En cuanto se cargue la información quedará lista para ser objeto de análisis y, así, permitirá validar si todo el proceso fue correcto; y cuando se evalúe el funcionamiento de la implementación será útil para apoyar a la organización en la toma de decisiones.

2.2.3.10 Mantenimiento y Crecimiento.

Data Warehousing es un proceso de etapas bien definidas, que tiene comienzo y fin, pero de naturaleza espiral, pues acompaña a la evolución de la organización durante toda su

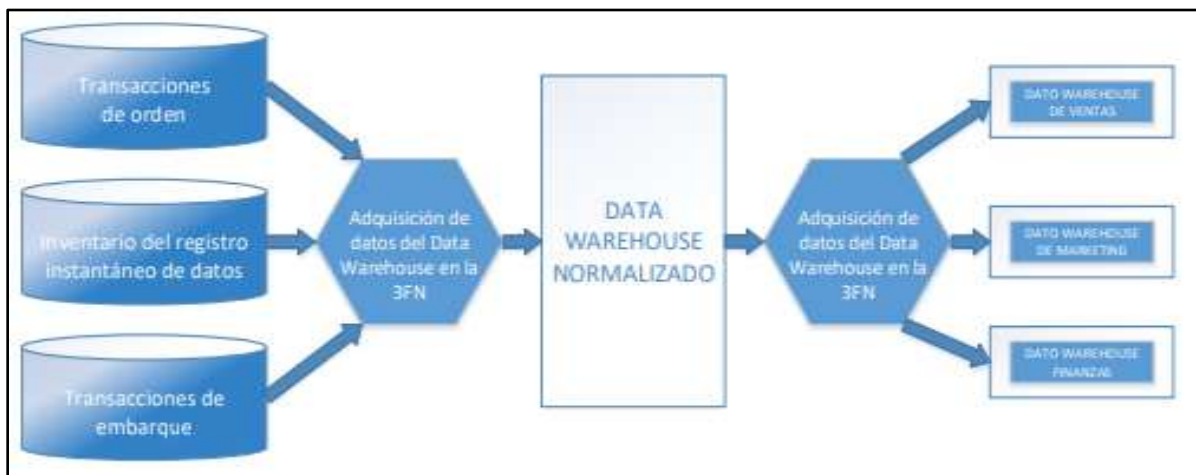
historia. Para poder seguir la evolución de las metas trazadas es necesario continuar con los relevamientos de forma constante. (Eguila y Parco, 2007, p. 136)

2.2.4 Metodología de Bill Inmon

Uno de los primeros autores que habló sobre los almacenes de datos fue Bill Inmon, quien planteó la metodología descendente (*top-down*) a la hora de implementar un almacén de este tipo, porque así todos los datos corporativos se consideran de mejor manera. Según Silva et al. (2019), la transferencia de la información OLTP de la empresa a un solo destino es fundamental, por lo que todos los datos se centran en un análisis que debería cumplir con características de un almacén de datos.

Figura 2

Arquitectura de un data Warehouse según Bill Inmon



Nota. Tomado de Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones, por G. E. Silva, 2019, p. 403, en *Ciencia Digital*, 3(3.4).

En este tipo de metodología, los *data marts* se implementan después de terminar el desarrollo del *data warehouse* completo de la organización. El enfoque que diseñó Inmon utiliza la forma para normalizar y construir la estructura de la entidad, de modo que evita la redundancia de datos tanto como sea posible. Esto es sumamente importante porque hace posible identificar los requisitos comerciales y evita irregularidades en la actualización de datos. Una de las ventajas de este enfoque es que las bases de datos diseñadas son robustas y sólidas ante los cambios comerciales, además de que contiene perspectivas de tipo dimensionales de los datos en el mercado.

2.2.5 MySQL

Es una base de datos con código abierto que actualmente se encuentra entre las más populares. Desde la perspectiva de Yegulalp (2017), el código abierto es muy útil y usado por muchos desarrolladores de aplicaciones web; en este sentido, es una base de datos de la cual se puede tener información de todo tipo para solucionar problemas o seguir una línea de aprendizaje, debido a su amplia utilización por parte de diferentes profesionales. Aunado a esto, se califica como un gestor de datos con buen rendimiento y rapidez para responder.

2.2.6 Apache Superset

Esta es una aplicación web para inteligencia de negocios, que sirve para modelar y visualizar la información en *dashboard's*, permitiendo la exploración y la visualización de sus datos a través de gráficos circulares, geospaciales, barras, filtros de información, etc. Entre sus ventajas está que es rápido, liviano y muy intuitivo para su desarrollo.

2.2.7 SSIS (*SQL Server Integration Services*)

Microsoft SQL Server Integration Services (SSIS) facilita la realización de soluciones de tipo integración con alto rendimiento. Entre sus integraciones se encuentran los paquetes de ETL para el almacenamiento de datos; también permite ejecutar consultas de tipo SQL y envío de mensajes de correo electrónico, tener orígenes y destinos de datos para obtenerlos y almacenarlos, y hacer transformaciones para depurar, añadir, mezclar y copiar datos.

2.3 Contexto de la Investigación

La presente investigación tiene como objetivo implementar un almacén de datos denominado cubo multidimensional OLAP, el cual contendrá información resumida y calculada acerca de los eventos que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C., que servirá para analizar, categorizar y monitorear los eventos que se generan diariamente.

El proceso de elaboración del proyecto consistió, en primer lugar, en seguir los pasos planteados en la metodología de Ralph Kimball, con el propósito de elaborar un sistema sólido que satisfaga las necesidades planteadas en los requerimientos; para ello se tomaron como referencia las siete fases propuestas en la metodología. Sumado a esto, en busca de lograr el objetivo propuesto, se llevaron a cabo las siguientes actividades:

1. Planificar el proyecto: se elaboró el plan de duración y se eligió el equipo que lo conformó, así como los requerimientos, entre ellos los accesos a los sistemas.
2. Se hizo una reunión con los interesados en el proyecto para tomar los requerimientos que ayudarían a elaborar el diseño del nuevo almacén de datos multidimensional.

3. El modelo del almacén de datos se diseñó siguiendo los requerimientos establecidos previamente; además, se escogieron las tablas dimensionales que tendrá el cubo OLAP y las métricas que harán parte de su contenido. Aparte de la tabla principal (hechos), para este desarrollo y análisis de eventos Safe se consideraron las siguientes tablas:
- Tabla vehículo: contiene la información característica de las unidades.
 - Tabla compañía: tiene la información de las compañías a las que pertenecen las unidades.
 - Tabla flota: a donde pertenecen los vehículos, los cuales están asociados a la compañía.
 - Tabla tipo vehículo: ayuda a categorizar según el tipo de eventos que se están generando.
 - Tabla conductor: a fin de obtener detalles específicos.
 - Tabla tiempo: como dato, tiene categorizadas las fechas en las que se obtuvieron los eventos.
 - Tabla de hechos o principal: contiene las llaves de las demás tablas para relacionar sus respectivas categorías. Al interior de esta tabla se encontrarán las medias calculadas respecto a los eventos generados para cada vehículo, por ejemplo, la cantidad de eventos por cada tipo, los kilómetros recorridos, el tipo evento, etc.

Toda esta data será trasladada diariamente al cubo, mediante un proceso ETL que se efectuará con la herramienta SSIS, la cual será elaborada con un orden establecido, primero se cargarán las tablas secundarias y por último la tabla principal de hechos, de modo que se depure la información del origen para obtener datos precisos en el cubo multidimensional. Previo al

traslado de información se amplió el tamaño del almacenamiento a 1TB en el servidor que estará alojado el cubo OLAP, a fin de que físicamente pueda soportar la cantidad de datos y no se presenten problemas al procesar la información.

Una vez cumplidos estos pasos se implementará el nuevo almacén de datos, a través del cual se finalizará con el modelado de todos los datos cargados, que serán visualizados y modelados por medio de la herramienta *BI Apache superset*, donde el usuario podrá ver y analizar la información. Este *dashboard* será visible desde cualquier lado, pues la herramienta está diseñada en una plataforma web, lo que será de mucha ayuda para los interesados del área de gerencia de TI, quienes actualmente no cuentan con esta información y, por ello, no pueden aprovechar los datos que se almacenan en su base de datos central, así que el proyecto les permitirá tener detalles por empresa, vehículo, conductor, flota y por tipo de evento.

El proyecto iniciará desde la toma de requerimientos por parte de los involucrados para hacer el levantamiento de información, después se diseñará la estructura que tendrá el cubo, sus métricas en la tabla central (hechos) y sus dimensiones. Luego de esto se realizará la carga ETL, poblando las dimensiones y finalmente su tabla central. En cuanto la información sea cargada en el cubo se realizará un *dashboard* para la visualización y el análisis de los datos.

Capítulo III. Metodología de Investigación

3.1 Diseño de Investigación

3.1.1 Diseño

En esta investigación se aplicó un diseño metodológico no experimental, debido a que la información registrada en la base de datos no puede ser manipulada ni alterada, pues será tomada

para el análisis. Esta metodología permite analizar la información recolectada de las flotas de vehículos activos que actualmente tienen un contrato en la empresa Comsatel Perú S.A.C., lo cual se debe a la falta de análisis de los eventos generados, y de esta manera se podrán evaluar para la toma de decisiones.

Acerca del diseño de investigación no experimental, Hernández (1998) indicó que con este resulta imposible la manipulación de las variables, “de hecho, no hay condiciones o estímulos a los que se expongan los sujetos del estudio, sino que estos son observados en su ambiente natural, en su realidad” (como se citó en Jiménez et al., 2021, p. 158).

3.1.2 Tipo

Por su naturaleza, el presente estudio es de tipo aplicado y tiene un enfoque cuantitativo, así que se propone una herramienta que le permita a la empresa Comsatel Perú S.A.C explotar la información que actualmente tiene en su almacén de datos, el cual, a través de un cubo OLAP, facilitará la obtención de métricas que ayuden al análisis de eventos Safe generados por los vehículos en la plataforma web Clocator.

3.1.3 Enfoque

El enfoque utilizado es cuantitativo, que, como explicó Romero (1997), tiene una innegable potencia para el tratamiento de los datos más concretos y más ecológicos, por lo que su objetivo es recolectarlos y analizarlos para contestar las preguntas planteadas en la investigación y probar las hipótesis establecidas.

3.1.4 Población

De acuerdo con Arias et al. (2016), la población es un conjunto de casos definido, limitado y accesible, que servirá como referente para seleccionar la muestra, la cual debe cumplir con requisitos predeterminados que ayuden a obtener los resultados para la investigación. Cabe señalar que con población no solo se hace referencia a seres humanos, sino también a muestras biológicas, animales, hospitales, autos, etc. En este caso se consideraron 26 personas (gerentes, jefes de TI y colaboradores a su cargo) que se encargan de gestionar, analizar y revisar la información sobre los vehículos, y, además, pertenecen a las áreas de operaciones y de procesos, gerencia y área de sistemas.

3.1.5 Muestra

Tal como indicó López (2004), una muestra es un subconjunto de la población objeto de investigación y constituye un elemento representativo para obtener resultados con base en fórmulas, lógica u otros. Teniendo en cuenta el tamaño poblacional se aplicó un muestreo probabilístico intencional simple, en el que todos los participantes del estudio tienen la misma probabilidad de ser seleccionados aleatoriamente (Hernández et al., 2014). El tamaño de la muestra se calculó con la fórmula probabilística para poblaciones finitas inferiores a 100 000 individuos o elementos (Palomino et al., 2015).

$$n = \frac{N * Z^2 * p * (1 - p)}{(N - 1) * E^2 + Z^2 * p * (1 - p)}$$

Donde n es el tamaño de la muestra, N es la cantidad de elementos que conforman la población, Z es la desviación estándar que posibilita alcanzar el nivel de confianza deseado, E es el error máximo permitido y p es la probabilidad de éxito. Por tanto, dado que $N = 26$; $Z = 1.96$; $E = 0.05$ y $p = 0.5$, se tiene que:

$$n = \frac{26 * 1.96^2 * 0.5 * (1 - 0.5)}{(26 - 1) * 0.05^2 + 1.96^2 * 0.5 * (1 - 0.5)} = 24.41 \approx 24$$

Esto indica que la muestra se constituye por 24 empleados.

3.1.6 Operacionalización de Variables

3.1.6.1 Cubo Multidimensional.

Desde el punto de vista de Siesquen (2019), el cubo es un vector con múltiples dimensiones, pero su nombre puede indicar, erróneamente, que solo posee tres dimensiones. En el cubo se almacena la información, dentro de cada dimensión, ordenadamente y eso se logra con estructuras multidimensionales que se llaman cubos OLAP, los cuales poseen datos precalculados y añadidos.

Así mismo, por medio de un cubo multidimensional es posible tener diferentes perspectivas y resultados calculados, al igual que una respuesta rápida, la cual podrá almacenar y procesar grandes volúmenes de datos para ser analizados.

3.1.6.2 Análisis de Información.

Según Simão (2010), tanto el análisis como el trabajo realizado sobre los datos obtenidos son piezas claves en una investigación cuantitativa, con los cuales se puede explicar el proceso cíclico insertado en cada una de las etapas de la investigación y cuyo objetivo es contestar,

triangular y validar todo el estudio obtenido. Analizar la información implica organizar las formas de establecer categorías, modelos, unidades descriptivas, además de interpretar la información, de modo que el análisis tenga sentido y significado, y que las categorías se expliquen, en busca de identificar relaciones entre las dimensiones descriptivas.

Así pues, siguiendo los argumentos del autor, el análisis de la información servirá para establecer grupos que faciliten la comprensión del objeto en estudio, por lo que los resultados de este agrupamiento se reforzarán con métricas que permitan definir con claridad la problemática. Por tal motivo, en la actualidad, analizar la información favorece mucho a las organizaciones, porque posibilita el conocimiento detallado de lo que está sucediendo y la anticipación de acontecimientos.

Figura 3

Matriz de operacional de variables

Variables	Dimensión	Indicadores	Fórmulas	Items	Escala de Medición de	Instrumento
Cubo Multidimensional	Almacén de información	Eventos		1. ¿Considera usted que los eventos generados por los vehículos describen su comportamiento durante su recorrido?	[5] Siempre [4] Casi siempre [3] A veces [2] Casi nunca [1] Nunca	Encuesta
	Velocidad de respuesta	Tiempo		2. ¿Considera usted que mediante el cubo multidimensional se ha optimizado el tiempo para obtener resultados acerca de los eventos Safe generados?		Encuesta
		Eficiencia		3. ¿Considera usted que es eficaz el procesamiento de datos en el cubo multidimensional?		Encuesta
	Datos pre calculados	Cálculo		4. ¿Considera usted que los cálculos realizados en el cubo multidimensional de eventos Safe permiten analizar y responder las necesidad para lo cual fue implementado?		Encuesta
Análisis de Información	Investigación cuantitativa	Recolección de datos	$\% \text{ Eventos para analizar} = \frac{\text{total de eventos registrados en el cubo multidimensional}}{\text{Total de eventos registrados en base de datos}} \times 100$	5. ¿Considera usted que se recolecta la mayor cantidad de eventos Safe de los vehículos para realizar un análisis?		Encuesta
	Interpretación de la información	Filtro de información	$\% \text{ Vehículos para analizar} = \frac{(\text{total de vehículos} - \text{total de vehículos inactivos})}{\text{Total de vehículos registrados}} \times 100$	6. ¿Considera usted que se utiliza el filtro necesario para obtener la información detallada de los eventos Safe generados?		Encuesta
				7. ¿Considera usted que los eventos recolectados para cada vehículo se ajustan a la información real que generan los dispositivos instalados?		Encuesta
	Establecer categorías	Categorías	$\% \text{ Eventos categoría a} = \frac{\text{total de eventos categoría A}}{\text{Total de eventos}} \times 100$ $\% \text{ Eventos categoría b} = \frac{\text{total de eventos categoría B}}{\text{Total de eventos}} \times 100$	8. ¿Considera usted que se tiene una cantidad considerable de resultados que permita el análisis de eventos y ayude a la toma de decisiones?		Encuesta
				9. ¿Considera usted que los resultados analizados permiten tener un panorama claro acerca de los eventos Safe generados?		Encuesta
				10. ¿Considera usted que los resultados obtenidos permiten mejora del servicio que brindan los dispositivos instalados en los vehículos?		Encuesta

3.2 Instrumentos de Investigación / Herramientas

3.2.1 Técnicas

En esta investigación se empleará la encuesta para recoger los datos acerca del proceso que será implementado, a fin de analizar la información. Esto se llevará a cabo con las personas involucradas e interesados en la implementación, las cual les será útil para su labor. Como apuntó Montes (2000), la encuesta es un sistema de preguntas cuyo propósito es obtener datos para una investigación. También resulta ser un auxiliar eficaz en la observación científica.

Las encuestas se realizarán en modalidad virtual, debido a la situación sanitaria generada por el covid-19, acatando el distanciamiento social y guardando la cuarentena, dado que la interacción entre personas está prohibida, pues se busca evitar la propagación del virus.

3.2.2 Instrumentos

El instrumento de recolección de datos es un formulario con 10 preguntas que se crearon en la nube a través de la herramienta de Google Forms; su finalidad es medir la eficiencia de la implementación del cubo dimensional, así como las ventajas y las respuestas que se obtienen a través de ella, midiendo la conformidad de los interesados en el proyecto para obtener resultados que permitan determinar si con la implementación del proyecto se satisfacen las necesidades para las que fue creado. También se empleó la escala de Likert, que ayudará a conocer la perspectiva de la muestra con respecto a la implementación de un cubo multidimensional para analizar los eventos Safe que los vehículos generan en la plataforma Clocator.

3.2.2.1 Validación del Instrumento.

Para validar el instrumento se sometió a la revisión de dos expertos, tal como se detalla en la Tabla 1.

Tabla 1

Validación del instrumento

N.º	Nombre y apellidos	Especialidad	Pertinencia	Relevancia	Claridad	Criterio
1	Experto 1	Magíster en Telecomunicaciones – Redes Móviles.	100 %	100 %	100 %	Aplicable
2	Experto 2	Doctor en Ingeniería de Sistemas.	100 %	100 %	100 %	Aplicable

3.2.3 Confiabilidad

La evaluación para la prueba de confiabilidad del instrumento indicó su cualidad para mediciones que serán realizadas consecutivamente al mismo sujeto o grupo, de modo que se puedan obtener resultados más parecidos. Para esto, en el presente proyecto se empleó el coeficiente alfa de Cronbach, cuyos valores deben ser superiores a 0.70 para obtener resultados confiables, lo cual indicará que el coeficiente es confiable. Teniendo en cuenta a Oviedo y Campo (2005), el mínimo valor que se acepta para el coeficiente es de 0.70, y los valores que estén por debajo no se consideran confiables.

Tabla 2*Resultados de confiabilidad*

Variable	N.º elementos	Alfa de Cronbach	Criterio
Cubo multidimensional	4	0.704	Confiable
Análisis de información	6	0.784	Confiable

3.3 Metodología de la Implementación de la Solución**3.3.1 Selección de la Metodología**

Para aplicar una metodología que se ajuste a las necesidades de la implementación de este proyecto se hizo una comparación entre la metodología de Ralph Kimball y el método de Bill Inmon (Figura 4), que poseen características relevantes para crear, diseñar e implementar, por lo cual se presentan las más destacables.

Figura 4

Características de las metodologías

Características	Metodología de Bill Inmon	Metodología de Ralph Kimball
Informar necesidades informes centrados en el proceso comercial o el equipo		X
Plazo del proyecto Tiempos de entrega rápidos	X	X
Plan de reclutamiento prospectivo Elegir un equipo más grande de profesionales para la gestión del almacén de datos.	X	
Cambios Frecuentes Sujeto a cambios en el informe con estabilidad en los almacenes de origen		X
Principios de la organización Costo de implementación y solución rápida a las necesidades del negocio		X

Como se ha dicho, en esta investigación se aplica la metodología de Ralph Kimball, debido a las características y las ventajas que posee en la implementación de un almacén de información, pues logra que esta sea mucho más rápida, además de que no requiere tantos profesionales, en comparación con la metodología de Bill Inmon, y su costo de diseño es menor. Para reforzar la justificación de esta elección se indican las ventajas que obtendrán al utilizar esta metodología:

1. El modelado dimensional es más rápido de construir.
2. Posee un esquema de tipo estrella, donde los operadores de datos pueden comprenderlo fácilmente, lo que simplifica las consultas y el análisis.
3. Permite la recuperación rápida de datos, puesto que los datos se liberan en la tabla de hechos y dimensiones.
4. Es controlable, o sea, no requiere de muchas personas para su gestión.

3.3.2 Desarrollo de la Metodología – Ralph Kimball

3.3.2.1 Fase 01. Planificación.

En primer lugar, se hará una reunión con los interesados, a fin de conocer y visionar la funcionalidad del proyecto, con la intención de identificar el tiempo, el costo y el personal idóneo que estará a cargo del desarrollo y la implementación desde el inicio. También se abordarán las tecnologías hardware y software que se utilizarán, y se planificarán reuniones para ir mostrando el avance en el proyecto. Por otro lado, se podrá proyectar el alcance que tendrán y desde el principio se solicitarán los permisos necesarios para tener acceso completo a la información, sin limitación o atraso en los avances.

3.3.2.2 Fase 02. Definición de los Requerimientos del Negocio.

En este punto, con los interesados en la implementación del proyecto se deben definir los requerimientos. Esta tarea es una de las primeras y será la base fundamental para identificar los recursos necesarios en la elaboración del almacén de datos multidimensional. Aunado a esto, se definirá lo que se desea solucionar, mejorar y analizar.

Adicionalmente, se identificará la base datos que contiene la información que será procesada, las tablas intervienen en el proyecto y la seguridad de los datos. Luego se definirá aquello que se espera analizar para obtener las métricas y los resultados. Posteriormente, se tomarán los objetivos principales para la implementación y, una vez resueltos, responderán las dudas de los interesados.

Uno de los primeros pasos en esta fase es elegir el área a modelar, que es decidido por los interesados en la implementación del proyecto, la cual depende fundamentalmente de los requerimientos recalentados en la etapa anterior. Por tal razón se deben tener claras las necesidades, debido a que son la base para definir bien el área que abarcará la solución, lo que hará posible tener un mejor panorama de lo que se desea alcanzar.

3.3.2.3 Fase 03. Modelo Dimensional.

Este proceso de diseño inicia con un modelo dimensional de alto nivel, a partir de los procesos priorizados a través de los requerimientos identificados en el punto anterior. El proceso de modelado consiste en cuatro pasos:

1. Elegir la área del negocio donde se desea hacer el proyecto.
2. Establecer el nivel de detalle.
3. Elegir las dimensiones involucradas en la implementación.
4. Identificas las medidas y los cálculos que se emplearán en la tabla de hechos (tabla central del esquema estrella).

Siguiendo la línea de tareas, en esta etapa se identifican las tablas involucradas para crear el almacén de datos, las cuales surgen de manera natural durante la determinación de los requerimientos. Estas tablas de dimensiones contienen una serie de atributos que brindan un

detalle sobre una medida en la tabla de hechos, la cual permitirá describir e identificar los datos; es decir, la información almacenada en la tabla hechos podrá ser mejor descrita por medio de sus dimensiones y atributos; en el caso de este proyecto se cuenta con dimensiones que servirán de apoyo para describir la información, teniendo en cuenta que el eje central es la tabla hechos, pues contiene las medidas precalculadas que resumen la información.

3.3.2.4 Fase 04. Diseño Físico.

En esta etapa se diseñará físicamente el modelo, pero antes de esto es necesario establecer la capacidad de almacenamiento necesaria y cuánta memoria y servidores se necesitarán. Esta información ayudará a definir el tipo de datos y la longitud de los caracteres donde se ingresará la información, lo que contribuirá a optimizar y reducir espacios en memoria, los cuales podrían aprovecharse de mejor manera durante el almacenamiento. Una vez que se determinen los hechos, las métricas, las dimensiones y los atributos existentes se establecerán las claves foráneas que se deben incluir en el modelo físico, además de las respectivas claves primarias en cada dimensión y tabla de hechos.

3.3.2.5 Fase 05. Diseño del Sistema de ETL.

Para esta fase es fundamental que se hayan definido bien los requerimientos e identificado las tablas de las que se extraerá la información de la base datos principal. El sistema ETL se diseñará adecuadamente para aplicar diferentes procedimientos que ayuden a incrementar la calidad y la consistencia de estos, así como a consolidar la información proveniente del origen de los datos, los cuales se irán cargando de manera progresiva, según el

orden establecido en la estructura del ETL, a saber: primero las dimensiones y por último la tabla de hechos (es la principal y contiene las llaves foráneas y métricas de las dimensiones).

3.3.2.6 Fase 06. Implementación.

La implementación culmina en cuanto se valida el éxito de los pasos previos, por lo que está lista para ser procesada con herramientas de inteligencia de negocios que posibiliten la realización de un análisis, y así se habrá logrado implementar el cubo OLAP satisfactoriamente.

3.3.2.7 Fase 07. Especificación y Desarrollo de Aplicaciones BI.

Al terminar la implementación del cubo multidimensional se desarrollará un modelo de *dashboard* que permita visualizar y analizar la información almacenada en el cubo; para este punto, las aplicaciones BI proporcionan herramientas e informes de análisis que van desde simples formatos fijos hasta aplicaciones analíticas elaboradas con algoritmos complejos, las cuales, en la metodología de Kimball, se dividen en dos categorías fundamentadas en el nivel de sofisticaciones, que son llamadas informes estándar y aplicaciones analíticas. Los primeros son elaboraciones simples que suelen ser estáticas, mientras que las segundas facilitan el análisis profundo e incluyen algoritmos complejos para su desarrollo. A efectos de esta investigación se utilizarán las aplicaciones analíticas.

3.4 Metodología para la Medición de Resultados de la Implementación

3.4.1 Metodología de Medición

Para la medición de datos se empleó la estadística descriptiva e inferencial, que posibilitó el análisis descriptivo de la variable dependiente e independiente con la contratación de

hipótesis. En este proyecto, la población solo estuvo conformada por trabajadores del área de TI, quienes fueron encuestados en su totalidad.

La estadística descriptiva, como su nombre lo indica, intenta describir algo cuantitativamente. De acuerdo con Álvarez y Barreda (2020), esta herramienta es útil para la recolección de datos, así como la interpretación y la elaboración de gráficas que precisan de conocimientos básicos del tema, con base en una valoración de la evidencia objetiva. Mientras, para la prueba de las hipótesis se empleó la estadística inferencial Chi-cuadrado, que permitió obtener resultados que responden a las hipótesis planteadas para el proyecto.

3.4.2 *Medición de los Indicadores*

En cuanto a la medición de la variable dependiente y de los indicadores se usan las siguientes fórmulas:

- **Medición de recolección de datos**

$$\% \text{ Eventos para analizar} = \frac{\text{total de eventos registrados en el cubo multidimensional}}{\text{Total de eventos registrados en base de datos}} \times 100$$

- **Medición para filtro de información**

$$\% \text{ Vehiculos para analizar} = \frac{(\text{total de vehículos} - \text{total de vehículos inactivos})}{\text{Total de vehículos registrados}} \times 100$$

- **Medición para categorías**

$$\% \text{ Eventos categoría } a = \frac{\text{total de eventos categoría } A}{\text{Total de eventos}} \times 100$$

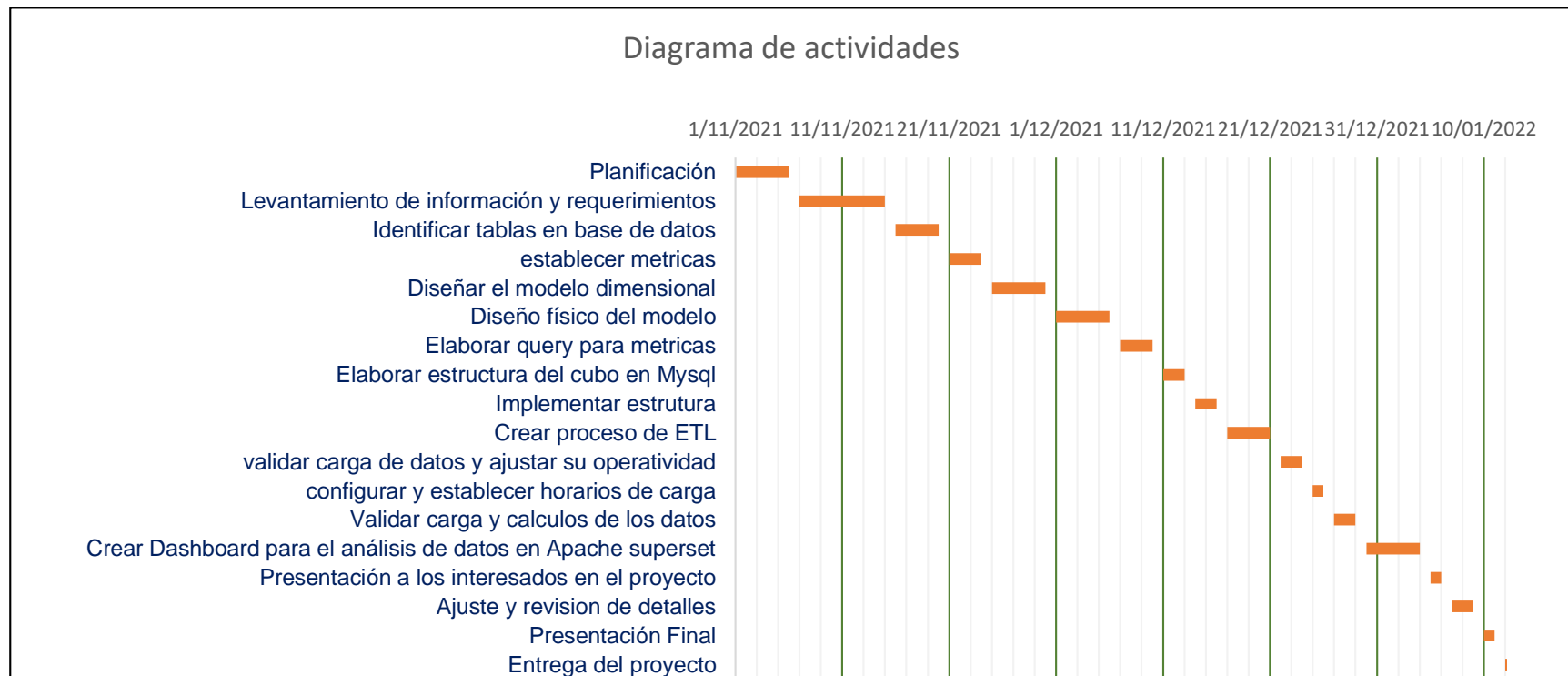
$$\% \text{ Eventos categoría } b = \frac{\text{total de eventos categoría } B}{\text{Total de eventos}} \times 100$$

3.5 Cronograma de Actividades

Contiene las fechas previstas en las que se realizará el proyecto, desde el inicio hasta el final.

Figura 5

Cronograma de actividades



3.6 Costos Directos

3.6.1 Costos de Personal y Materiales

En este apartado se describe el costo directo necesario para la implementación, que, en este caso, es el pago por los casi tres meses de trabajo para el gestor, la laptop que se usa (debido a la coyuntura del covid-19 el trabajo se realiza de manera remota y la empresa asigna un ordenador a cada empleado) y la ampliación de la capacidad del servidor a 1TB para que no se presenten problemas de almacenamiento.

Figura 6

Detalle sobre costos directos.

	Descripcion	Cantidad	Unitario	Tiempo de uso en días	TOTAL
Personal	Gestor de base de datos	1	S/2,500.00	75	
Hardware	Laptop	1	S/3,200.00	75	
	Discos duros 1T	1	S/419.00	365	
					S/11,119.00

3.7 Costos Indirectos

Para los costos indirectos se toman en cuenta el internet, el plan de celular y la aplicación Zoom para las reuniones que se asignarán durante el proyecto. Estos materiales influyen en el avance del proyecto, por lo cual se consideran costos indirectos.

Figura 7

Detalle sobre costos indirectos.

	Descripción	Cantidad	Unidad de medida	Costo por unidad	Total
Costo de Materiales	Internet	3	Mensual	S/ 99.00	S/ 297.00
	Plan Post pago bitel	3	Unidad	S/ 29.90	S/ 89.70
	zoom	3	Mensual	S/ 60.00	S/ 180.00
	Total				S/ 566.70

3.8 Costos Fijos

Como se mencionó, el trabajo se viene realizando remotamente desde casa, por ello se consideran costos fijos los servicios que se presentan en la Figura 8.

Figura 8

Detalle sobre costos fijos.

	Descripción	Cantidad	Unidad de medida	Costo por unidad	Total
Costo de Servicios	Luz	6	Mensual	S/ 130.00	S/ 780.00
	Agua	6	millar	S/ 35.00	S/ 210.00
	Total de Servicio				S/ 990.00

3.9 Costos Variables

Para estos costos se tiene en cuenta la asesoría que brindará un gestor de base de datos sénior.

Figura 9

Detalle sobre costos variables.

Descripción	Cantidad	Unidad de medida	Costo por unidad	Total
Gestor Senior en Base de datos	1	Mes	S/ 3,800.00	S/ 3,800.00

3.9.1 Contratación de Hipótesis

En la Tabla 3 aparecen los resultados de las pruebas Chi-cuadrado para el contraste de las hipótesis de la investigación.

Tabla 3

Contraste de hipótesis de estudio

Variable / dimensión	Prueba Chi-cuadrado	
	Valor de Chi-cuadrado de Pearson (X^2)	Sig. (p)
Análisis de información	7.464	0.006
Interpretación de la información	7.464	0.006
Investigación cuantitativa	4.052	0.044
Establecer categorías	4.531	0.033

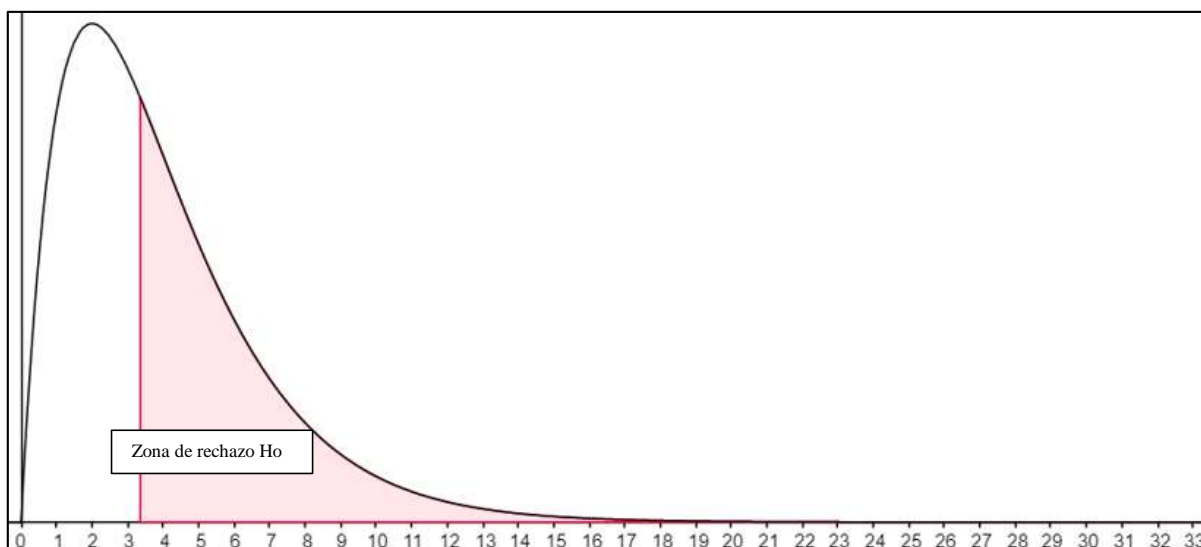
3.9.1.1 Hipótesis General.

- **H₀** = la implementación de un cubo multidimensional de eventos Safe no permite mejorar el análisis de la información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.
- **H_a** = la implementación de un cubo multidimensional de eventos Safe permite mejorar el análisis de la información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.
- **Nivel de significancia** = 5 % (0.05).
- **Decisión** = como el valor obtenido de $X^2 = 7.464$ (superior al X^2 crítico= 3.36) y $p = 0.006$, así $p < 0.05$, entonces se rechaza la hipótesis nula (H₀), como se aprecia en la Figura 42, por lo que es posible afirmar que la implementación de un cubo

OLAP de eventos Safe hace posible mejorar el análisis de información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.

Figura 10

Prueba Chi-cuadrado para comprobar la hipótesis general



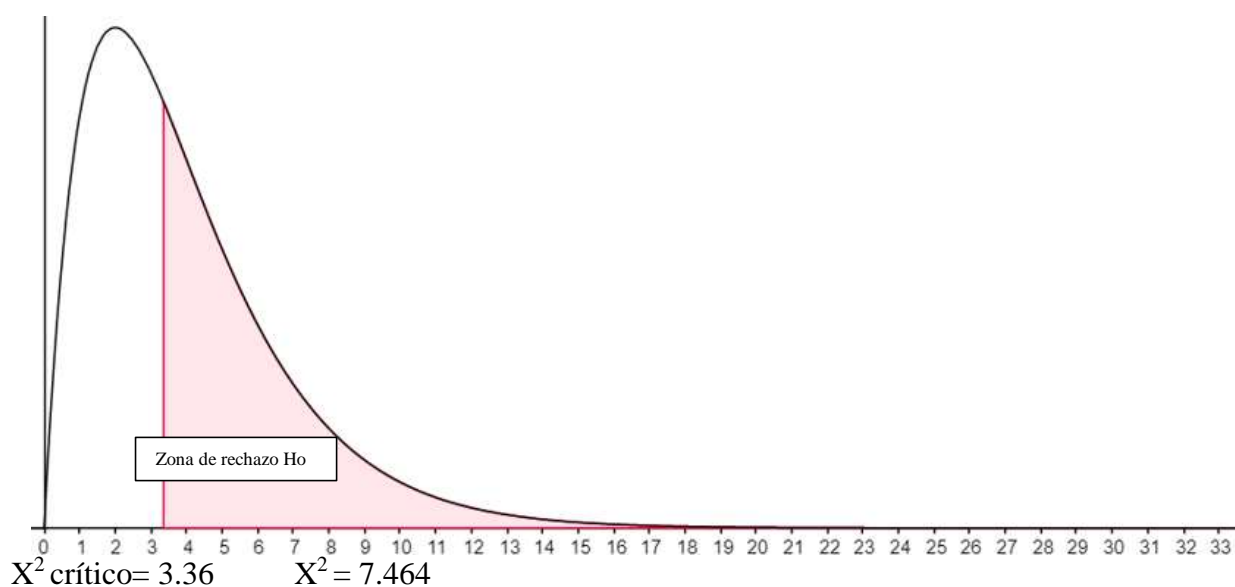
3.9.1.2 Hipótesis Específica 01.

- **H₀** = la implementación de un cubo multidimensional no permite una investigación cuantitativa de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.
- **H_a** = la implementación de un cubo multidimensional permite una investigación cuantitativa de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.
- **Nivel de significancia** = 5 % (0.05).

- **Decisión** = como el valor obtenido de $X^2 = 7.464$ (superior al X^2 crítico= 3.36) y $p = 0.006$, así $p < 0.05$, entonces se rechaza la hipótesis nula (H_0), como se observa en la Figura 43, de modo que la implementación de un cubo multidimensional facilita una investigación cuantitativa de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

Figura 11

Prueba Chi-cuadrado para comprobar la hipótesis específica 01



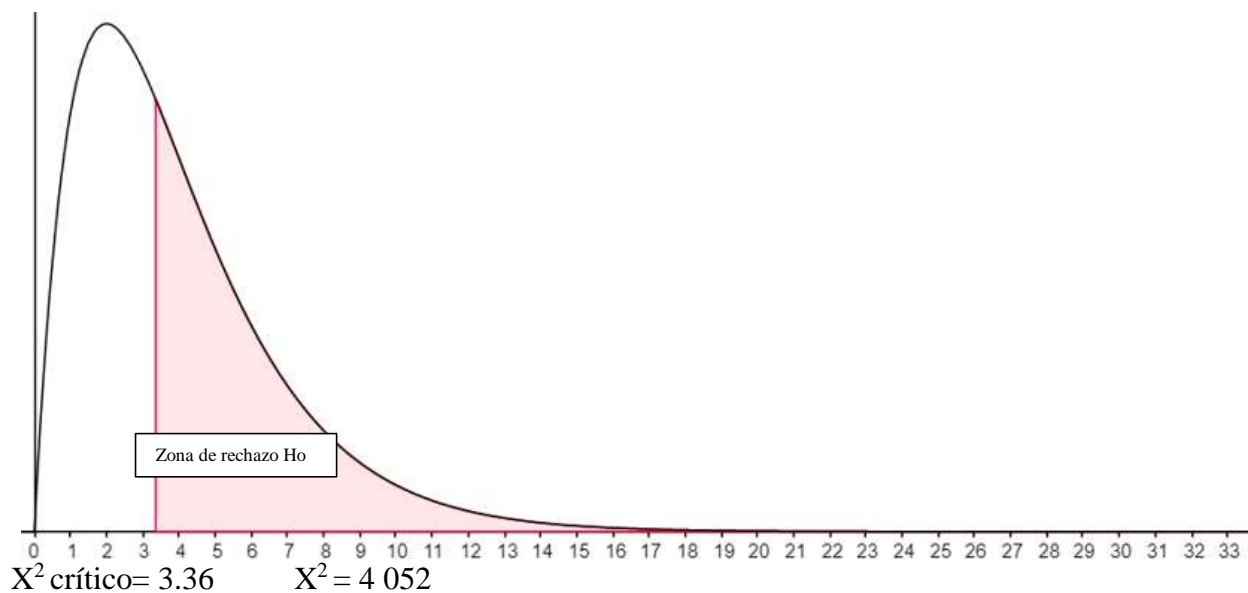
3.9.1.3 Hipótesis Específica 02.

- **H0** = la implementación de un cubo multidimensional no permite interpretar la información de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

- **Ha** = la implementación de un cubo multidimensional permite interpretar la información de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.
- **Nivel de significancia** = 5 % (0.05).
- **Decisión** = como el valor obtenido de $X^2 = 4.052$ (superior al X^2 crítico= 3.36) y $p = 0.006$, así $p < 0.05$, entonces se rechaza la hipótesis nula (H_0), como se puede ver en la Figura 44, por lo que es posible afirmar que la implementación de un cubo multidimensional permite interpretar la información de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

Figura 12

Prueba Chi-cuadrado para comprobar la hipótesis específica 02

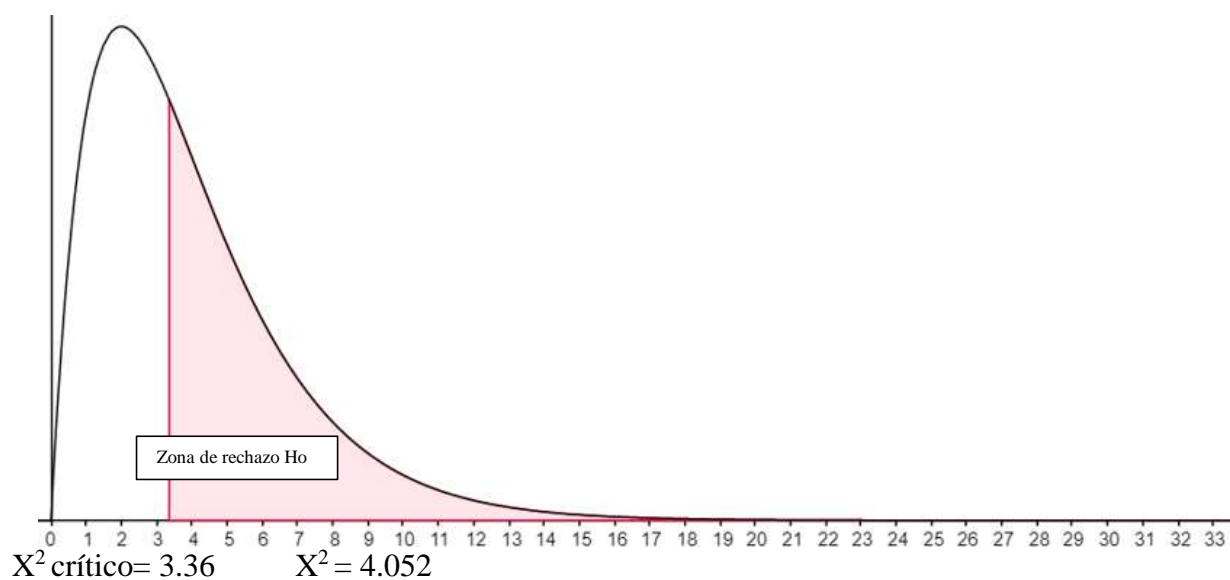


3.9.1.4 Hipótesis Específica 03.

- **H₀** = la implementación de un cubo multidimensional no permite establecer categorías que ayuden al análisis de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.
- **H_a** = la implementación de un cubo multidimensional permite establecer categorías que ayuden al análisis de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.
- **Nivel de significancia** = 5 % (0.05).
- **Decisión** = como el valor obtenido de $X^2 = 4.531$ (superior al X^2 crítico= 3.36) y $p = 0.006$, así $p < 0.05$, entonces se rechaza la hipótesis nula (H₀), como se muestra en la Figura 45, por ello se tiene que la implementación de un cubo multidimensional permite establecer categorías que ayuden al análisis de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C.

Figura 13

Prueba Chi-cuadrado para comprobar la hipótesis específica 03



3.9.1.5 Resultados de Método de Reprocesamiento.

En esta sección se evalúa el efecto de la propuesta de cubo multidimensional de eventos Safe en el análisis de información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C. (Tabla 4).

Tabla 4

Prueba de Wilcoxon

Rangos	n	Rango promedio	Suma de rangos	Z	Sig. (p)
Negativos	0	0.00	0.00	-4.33	0.000
Positivos	24	12.50	300.00		
Empates	0				

- **H0** = no hay cambio en la variable análisis de información

- **Ha** = la variable “análisis de información” mejora, luego de la implementación de la propuesta.
- **Nivel** de significancia = 5 % (0.05).
- **Decisión** = como el valor obtenido de $Z = -4.33$ y $p = 0.000$, así $p < 0.05$ (Tabla 5), entonces se rechaza la hipótesis nula (H_0), de modo que el análisis de la información que generan los vehículos en la plataforma Web Clocator en Comsatel Perú S.A.C. mejora con la propuesta, así como la percepción de los 24 colaboradores acerca de esta variable.

Tabla 5

Prueba de Wilcoxon. Efecto de la propuesta de cubo multidimensional de eventos Safe en la investigación cuantitativa que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.

Rangos	n	Rango promedio	Suma de rangos	Z	Sig. (p)
Negativos	0	0.00	0.00	-4.56	0.000
Positivos	23	12.50	300.00		
Empates	1				

- **H0** = no hay cambio en la dimensión investigación cuantitativa.
- **Ha** = la dimensión “investigación cuantitativa” mejora, luego de la implementación de la propuesta.
- **Nivel de significancia** = 5 % (0.05).
- **Decisión** = como el valor obtenido de $Z = -4.56$ y $p = 0.000$, así $p < 0.05$ (Tabla 6), entonces se rechaza la hipótesis nula (H_0), así que la investigación cuantitativa que generan los vehículos en la plataforma Web Clocator en Comsatel Perú S.A.C. mejora

con la propuesta, al igual que la percepción de 23 de los 24 colaboradores sobre esta variable.

Tabla 6

Prueba de Wilcoxon. Efecto de la propuesta de cubo multidimensional de eventos Safe en la interpretación de la información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.

Rangos	n	Rango promedio	Suma de rangos	Z	Sig. (p)
Negativos	0	0.00	0.00	-4.90	0.000
Positivos	24	12.50	300.00		
Empates	0				

- **H₀** = no hay cambio en la dimensión interpretación de la información.
- **H_a** = la dimensión “interpretación de la información” mejora, luego de la implementación de la propuesta.
- **Nivel de significancia** = 5 % (0.05).
- **Decisión** = como el valor obtenido de $Z = -4.90$ y $p = 0.000$, así $p < 0.05$ (Tabla 7), entonces se rechaza la hipótesis nula (H_0), por ello se tiene que la interpretación de la información que generan los vehículos en la plataforma Web Clocator en Comsatel Perú S.A.C. mejora con la propuesta y también la percepción de los 24 colaboradores en cuanto a esta dimensión.

Tabla 7

Prueba de Wilcoxon. Efecto de la propuesta de cubo multidimensional de eventos Safe en establecer categorías que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C.

Rangos	n	Rango promedio	Suma de rangos	Z	Sig. (p)
Negativos	0	0.00	0.00	-4.34	0.000
Positivos	24	12.50	300.00		
Empates	0				

- **H₀** = no hay cambio en la dimensión establecer categorías.
- **H_a** = la dimensión “establecer categorías” mejora, luego de la implementación de la propuesta.
- **Nivel de significancia** = 5 % (0.05).
- **Decisión** = como el valor obtenido de $Z = -4.34$ y $p = 0.000$, así $p < 0.05$ (Tabla 7), entonces se rechaza la hipótesis nula (H_0), por lo que es posible afirmar que el establecimiento de categorías que generan los vehículos en la plataforma Web Clocator en Comsatel Perú S.A.C. mejora con la propuesta, así como la percepción de los 24 colaboradores acerca de esta variable.

3.10 Presupuesto

Figura 14

Detalle sobre presupuesto.

Descripcion	Cantidad	TOTAL
Costos Directos	S/ 11,119.00	
Costos Indirectos	S/ 566.70	
Costos Fijos	S/ 990.00	
Costos Variable	S/ 3,800.00	
		S/16,475.70

Capítulo IV. Desarrollo de la Solución

4.1 Propuesta de Solución

4.1.1 Fase 01. Planificación

Se llevó a cabo una reunión con los interesados en la elaboración del proyecto, quienes habían identificado la problemática y la necesidad de implementar un almacén de datos para analizar la información que generan los dispositivos instalados en los vehículos. En relación con esto detallaron lo siguiente:

4.1.1.1 Caso del Negocio.

Actualmente, Comsatel Perú S.A.C tiene una plataforma que permite monitorear y supervisar los vehículos que cuentan con contrato vigente en la empresa; además, se determinó que los eventos emitidos por los dispositivos instalados en las unidades no son vistos en la plataforma ni se pueden detallar, por lo que tampoco es posible analizarlos. Esta situación podría ser perjudicial, pues los dispositivos reciben actualizaciones dependiendo el tipo de vehículos, entonces, no hay forma de saber si la calibración de los dispositivos está bien dada y si los eventos Safe que generan se ajustan a la realidad; por ejemplo, si se trazó una ruta para una flota en ciudad, pero registra continuamente el evento Safe denominado “inicio trocha”, se podría

detectar que la calibración se encuentra sensible y el evento que está generando no se ajusta a la realidad, pero no se sabe con qué frecuencia los conductores producen mayor cantidad de eventos, si cuando manejan ocho horas o menos etc.

Toda esta información es útil porque parte del servicio que brinda Comsatel es calificar a los conductores por su comportamiento durante su recorrido, y ese récord de puntaje es enviado a la compañía a la cual pertenece, junto con una calificación. Así, se entiende que no hay un reporte en la plataforma web para Comsatel que les permita a los interesados visualizar ese detalle de manera general para todas las empresas y de forma detalla por compañía, por eso hay que idear un modo de capturar toda esa información y analizarla óptimamente.

4.1.1.2 Definición de los Requerimientos del Negocio.

En esta esta etapa ya se había comprendido la problemática y la necesidad, así que el siguiente paso fue identificar estos mismos aspectos para la implementación. Los requerimientos fueron:

1. Filtrar la información por fechas de manera total y por rango de días.
2. Filtrar por compañía, conductor, vehículo, tipo de vehículo, placa, conductor y flota.
3. Mostrar la distancia recorrida total.
4. Mostrar el total de vehículos.
5. Mostrar el kilometraje.
6. Mostrar las horas de conducción.
7. Agrupar los eventos en dos categorías ADAS y Driving Behavior, y mostrar el total generado.
8. Mostrar todos los eventos existentes y el total generado.

9. Mostrar el tipo de vehículo y la cantidad total existente.

4.1.2 Fase 02. Diseño del Modelo Dimensional

En esta etapa se procedió a crear el modelo dimensional en función de los requerimientos mencionados; en el proceso se identificaron las tablas de la base de datos general y se eligieron los atributos que serían de utilidad para describir las métricas en el cubo, y, posteriormente, se estructuraron y relacionaron con la tabla principal de hechos denominada fact_esafe.

- a. **dim_flota:** contiene la información sobre las flotas pertenecientes a cada compañía, por lo que es un factor de indicación al momento de analizar, por ejemplo, cuál es la flota con más eventos Safe registrados para la compañía A.
- b. **dim_compania:** tiene el detalle acerca de las compañías que tienen contrato en Comsatel.
- c. **dim_tipovehiculo:** aquí se encuentra información referente al tipo de vehículo que registra eventos en Clocator (camión, auto, ambulancia, etc.).
- d. **dim_conductor:** se registra el detalle acerca del conductor de los vehículos.
- e. **dim_vehiculo:** almacena información relacionada con los vehículos, como placa, motor, marca, etc.
- f. **dim_tiempo:** contiene datos sobre las fechas en las que se cargaron los eventos, los cuales se resumen en diferentes tipos, como mes, día, bimestre, trimestre, semana, entre otros.

4.1.3 Fase 03. Diseño Físico

Con la finalidad de optimizar recursos de almacenamiento y guardar los datos definidos que permitan ser rápidos al momento de realizar una consulta utilizando filtros, se seleccionó el tipo de dato idóneo e índice que favorezcan estas operaciones.

- **Índices seleccionados:** para este punto se consideró el detalle nombre para cada atributo como parte de un índice, pues el filtro que se hará en el diseño de BI se usará con frecuencia en la descripción para obtener resultados.
- **Dimensión conductor:** en busca de obtener resultados particulares por nombre de conductor se elaboró el índice llamado de esta forma y el índice para ConductorId, así, el filtro de búsqueda en las uniones de las consultas será más rápido.

Figura 15

Índices de dimensión conductor.

Table	Key_name
dim_conductor	PRIMARY
dim_conductor	ConductorId_UNIQUE
dim_conductor	ConductorId_UNIQUE
dim_conductor	INDX_ConductorNombre

- **Dimensión flota:** para obtener resultados acerca de flotas específicas, filtrando únicamente por estas, se creó el índice para nombre debido al filtro que se tendrá.

Figura 16

Índices de dimensión flota.

Table	Key_name
dim_flota	PRIMARY
dim_flota	FlotaId_UNIQUE
dim_flota	FlotaId_UNIQUE
dim_flota	INDX_FlotaNombre

- **Dimensión compañía:** utilizada con el objetivo de tener resultados sobre compañías, las cuales servirán como dato para hacer el filtro.

Figura 17

Índices de dimensión compañía

Table	Key_name
dim_compania	PRIMARY
dim_compania	CompaniaId_UNIQUE
dim_compania	INDX_CompaniaNombre

- **Dimensión vehículo:** en busca de obtener los resultados que describirán a los vehículos y sus diferentes características. Dada la cantidad de datos, estos índices ayudan a tener una respuesta más rápida ante cualquier consulta.

Figura 18

Índices de dimensión vehículo

Table	Key_name
dim_vehiculo	PRIMARY
dim_vehiculo	VehiculoId_UNIQUE
dim_vehiculo	VehiculoId_UNIQUE
dim_vehiculo	INDX_VehiculoPlaca

- **Dimensión tipo vehículo:** se creó un índice para la descripción del tipo vehículo.

Figura 19*Índices de dimensión tipo vehículo*

Table	Key_name
dim_tipovehiculo	PRIMARY
dim_tipovehiculo	TipoVehiculoId_UNIQUE
dim_tipovehiculo	TipoVehiculoId_UNIQUE
dim_tipovehiculo	INDX_TipoVehiculoNombre

- **Dimensión tiempo:** se elaboró el índice para la fecha, el cual permitirá acelerar el proceso mientras se hace un filtro.

Figura 20*Índices de dimensión tiempo*

Table	Key_name
dim_tiempo	PRIMARY
dim_tiempo	PRIMARY
dim_tiempo	id_fecha_UNIQUE
dim_tiempo	id_fecha_UNIQUE
dim_tiempo	fecha_UNIQUE
dim_tiempo	INDX_TiempoFecha

- **Tabla hechos:** para esta tabla se creó el índice para todas las claves de las dimensiones.

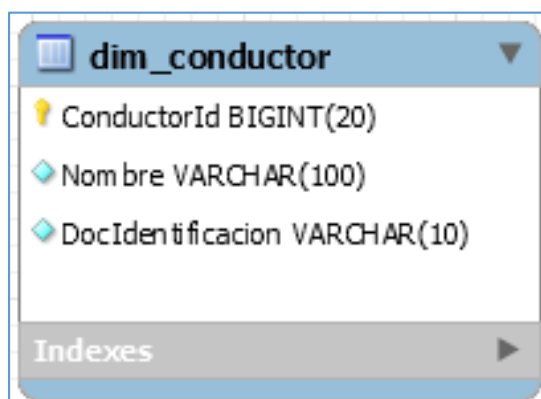
Figura 21*Índices de tabla hechos fact_esafe*

Table	Key_name
fact_esafe	id_fecha
fact_esafe	ConductorId
fact_esafe	VehiculoId
fact_esafe	TipoVehiculoId
fact_esafe	FlotaId
fact_esafe	CompaniaId

- **Tipos de datos seleccionados para cada tabla:** en este punto se definió el tipo de dato en cada tabla (dimensión y hechos), lo cual permitirá almacenar información de manera óptima y esto favorecerá el uso de los recursos físicos de almacenamiento.
- **Dimensión conductor:** se consideraron tres atributos principales que ayudarán a identificar al conductor.

Figura 22

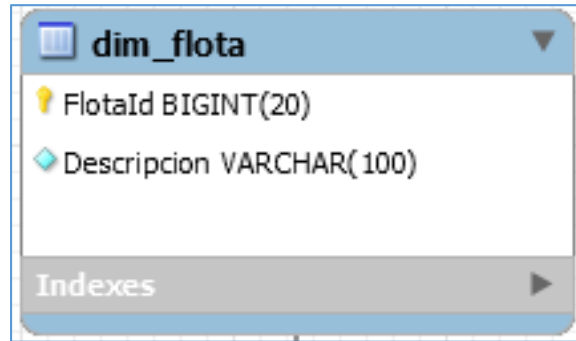
Estructura de tabla dimensional conductor



- **Dimensión flota:** describe a los vehículos pertenecientes a cada empresa.

Figura 23

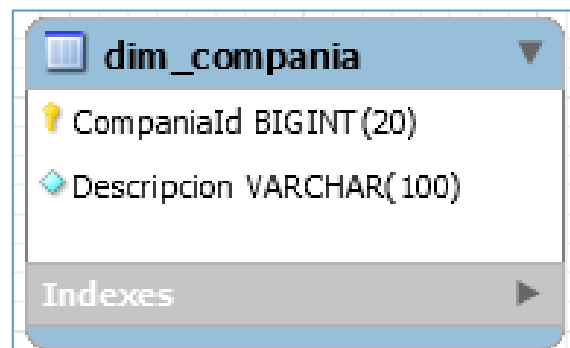
Estructura de tabla dimensional flota



- **Dimensión compañía:** contiene información sobre la compañía a la que está asociada cada vehículo.

Figura 24

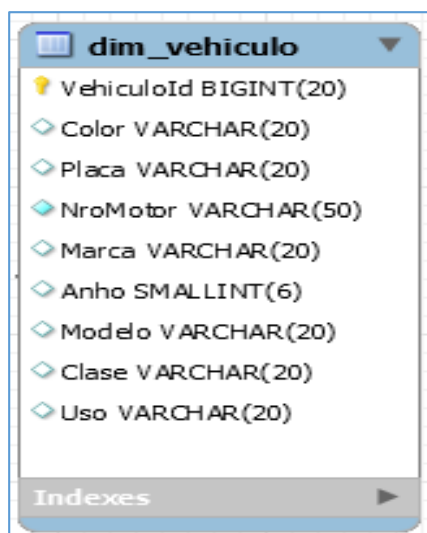
Estructura de tabla dimensional compañía



- **Dimensión vehículo:** contiene información relacionada con los vehículos.

Figura 25

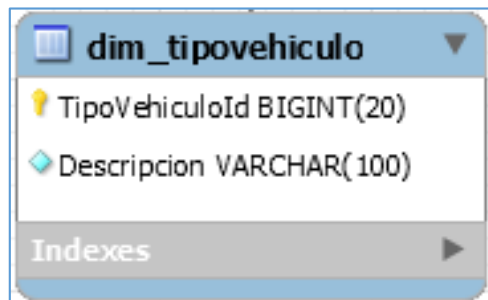
Estructura de tabla dimensional vehículo



- **Dimensión tipo vehículo:** contiene información referente a los diversos tipos de vehículos registrados en la base de datos.

Figura 26

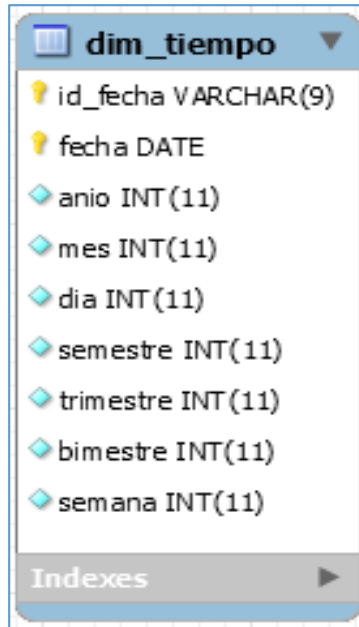
Estructura de tabla dimensional tipo vehículo



- **Dimensión tiempo:** en esta tabla se detallan la fecha y la categoría actuales, por ejemplo, año, mes, día, etc.

Figura 27

Estructura de tabla dimensional tipo vehículo



The image shows a screenshot of a database table structure for a table named 'dim_tiempo'. The table has the following columns:

Column Name	Data Type
id_fecha	VARCHAR(9)
fecha	DATE
anio	INT(11)
mes	INT(11)
dia	INT(11)
semestre	INT(11)
trimestre	INT(11)
bimestre	INT(11)
semana	INT(11)

At the bottom of the screenshot, there is a tab labeled 'Indexes' with a right-pointing arrow.

- **Tabla hechos:** es la tabla principal y contiene las llaves de las diferentes dimensiones que están a su alrededor, así como las métricas.

Figura 28

Estructura de tabla dimensional tipo vehículo

Column Name	Data Type
id_fecha	VARCHAR(9)
ConductorId	BIGINT(20)
VehiculoId	BIGINT(20)
TipoVehiculoId	BIGINT(20)
FlotaId	BIGINT(20)
CompaniaId	BIGINT(20)
Kilometraje	INT(11)
SegundosConduccion	BIGINT(20)
Distancia	DOUBLE
VelocidadMaxima	INT(11)
Eventos	INT(11)
Accidente	INT(11)
Aceleracion	INT(11)
Frenada	INT(11)
GiroIzquierda	INT(11)
GiroDerecha	INT(11)
GiroIzquierdaFrenando	INT(11)
GiroDerechaFrenando	INT(11)
GiroIzquierdaAcelerando	INT(11)
GiroDerechaAcelerando	INT(11)
IngresoOvalo	INT(11)
CambioCarril	INT(11)
Sobrepasar	INT(11)
RompeMuelle	INT(11)
InicioTrocha	INT(11)
FinTrocha	INT(11)
ConductaAgresiva	INT(11)
ConductaPeligrosa	INT(11)
EventosMobileye	SMALLINT(6)
CambioCarrilID	SMALLINT(6)

22 more...

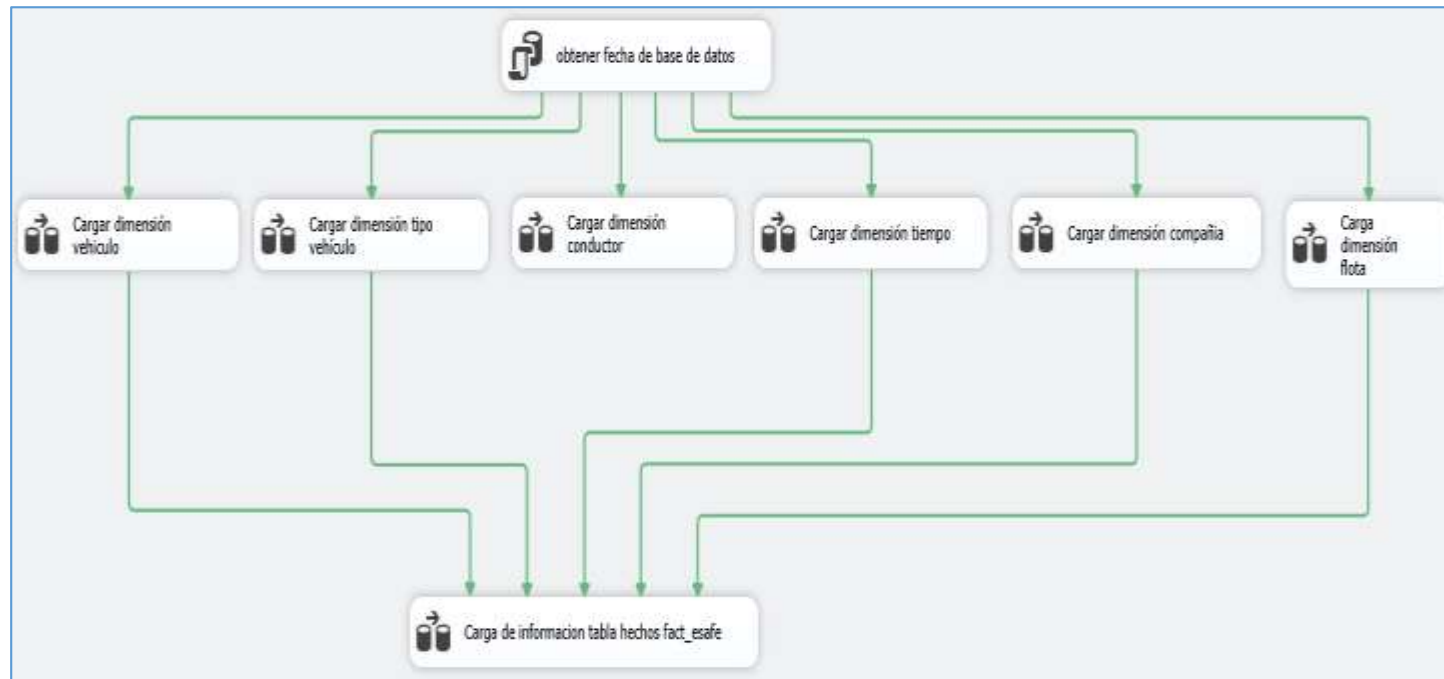
Indexes ▶

4.1.4 Fase 04. Diseño ETL

Se creó el diseño de la carga ETL, la cual inicia con el uso de un *script* para seleccionar el almacén de datos cubo_esafe y obtener la fecha actual. Por otro lado, la carga de datos lleva un orden: se comienza con la carga de todas las dimensiones y al finalizar con éxito se empieza a llenar la tabla de hechos fact_esafe.

Figura 29

Estructura del sistema ELT



4.1.5 Fase 05. Implementación

En este punto, el flujo de carga fue exitoso, por lo que se tiene la información cargada en el cubo multidimensional creado en MySQL server, la cual debería contar con datos listos para ser analizados.

4.1.5.1 Validar Información.

Como se mencionó, la carga de información primero se da en las dimensiones, y al tener éxito en todas se procede a cargar la tabla principal junto con las métricas. Finalmente, para validar la carga correcta se revisará la tabla principal fact_esafe.

Figura 30

Resultados de carga en la tabla hechos

Id_fecha	ConductorId	VehiculoId	TipoVehiculoId	PlacaId	CompaniaId	Kilometraje	SegundoConductor	Distancia	VelocidadMaxima	Eventos	Accidente	Asistencia	Prensado	GiroIzquierda	GiroDerecha	GiroIzquierdaPrensado	GiroDerechaPrensado
20210824	0	1318	1	13031	2017	6	157	0.250	7	0	0	0	0	0	0	0	0
20210824	0	1318	1	13031	2017	6	203	0.758	36	1	0	0	1	0	0	0	0
20210824	0	1318	1	13031	2017	6	480	3.281	38	0	0	0	0	0	0	0	0
20210824	0	2791	1	14489	2017	20	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2791	1	14489	2017	20	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2791	1	14489	2017	20	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	2568	12.939899999999997	40	5	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	24	0.034	0	0	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	2210	7.71	49	5	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	12	0.025	1	0	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	686	2.0138999999999996	44	1	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	1424	4.120000000000001	38	2	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	141	0.032	1	0	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	3849	14.812	53	6	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	772	2.8848999999999997	30	3	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	118	0.082	1	0	0	0	0	0	0	0	0
20210824	0	2949	1	11530	2017	3255	60	0.017	2	0	0	0	0	0	0	0	0
20210824	0	2978	1	11530	2017	1424	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2978	1	11530	2017	1424	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2978	1	11530	2017	1424	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2978	1	11530	2017	1424	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2978	1	11530	2017	1424	0	0	0	0	0	0	0	0	0	0	0
20210824	0	2978	1	11530	2017	1424	0	0	0	0	0	0	0	0	0	0	0
20210824	0	3883	1	11811	2017	9	112	0.6129000000000001	20	0	0	0	0	0	0	0	0
20210824	0	3883	1	11811	2017	9	42967	6.762899999999998	41	4	0	1	0	0	0	0	0

Una vez realizada la consulta a la base de datos se obtuvieron resultados, lo cual indica que la implementación del cubo fue exitosa y tanto las dimensiones como la tabla central se encuentran con información lista para ser modelada en un *dashboard* que permitirá su análisis.

4.1.6 Fase 06. Diseño de Aplicación BI

Respecto a la elaboración del diseño BI, este se llevó a cabo con la herramienta Apache Superset, está desplegado en el servidor de Comsatel Perú S.A.C y es una herramienta que suele utilizarse. Primeramente, para desarrollar cada modelo que facilitará los filtros o la visualización de resultados, deberá hacerse la conexión a la base de datos que contiene la información a analizar, luego se realizará la consulta directamente al cubo a través de *query* y se almacenará en pequeñas tablas denominadas *dataset's*.

Figura 31

Conexión a base de datos desde Apache Superset



The screenshot shows the 'Edit database' dialog in Apache Superset. The dialog is titled 'Edit database' and has a close button (X) in the top right corner. Below the title, it indicates the database type is 'MYSQL' and the name is 'cubo_esafe_prod'. There are two tabs: 'BASIC' and 'ADVANCED'. The 'BASIC' tab is selected. The form contains the following fields:

- HOST**: 127.0.0.0
- PORT**: 5432
- DATABASE NAME**: cubo_esafe
- USERNAME**: gygnacio
- PASSWORD**: [masked]
- DISPLAY NAME**: cubo_esafe_prod
- ADDITIONAL PARAMETERS**: [empty]
- SSL**: [toggle off]

At the bottom, there are two buttons: 'CLOSE' and 'FINISH'.

En cuanto se estableció la conexión al cubo multidimensional se elaboraron los *dataset's* con sus respectivas *queries*. En la Figura 27 se observa la configuración y la creación de un *dataset*, el cual tiene la *query* general, donde se obtiene la información de compañías que, en este caso, ayudará al filtro

Figura 32*Configuración y creación dataset*

Be careful. Changing these settings will affect all charts using this dataset, including charts owned by other people.

SOURCE METRICS 1 COLUMNS 6 CALCULATED COLUMNS 0 SETTINGS

Physical (table or view) Virtual (SQL)

VIRTUAL

DATABASE
mysql

SCHEMA
cubo_esafe

DATASET NAME
FILTRO_COMPANIA

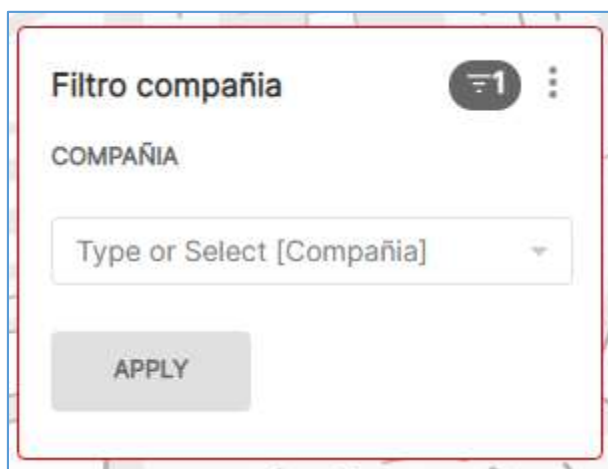
SQL ⓘ

```
1 -- Note: Unless you save your query, these tabs will NOT persist if you clear your cookies or change browsers.
2
3
4 SELECT distinct(c.Descripcion) as Compania,f.Descripcion as Flota,co.Nombre,v.Placa ,tv.Descripcion as TipoVehiculo,t.fe
5 from fact_esafe e inner join dim_tiempo t on t.id_fecha =e.id_fecha
6 inner join dim_compania c on c.CompaniaId =e.CompaniaId
7 inner join dim_flota f on f.FlotaId =e.FlotaId
8 inner join dim_conductor co on co.ConductorId=e.ConductorId
9 inner join dim_vehiculo v on v.VehiculoId =e.VehiculoId
10 inner join dim_tipovehiculo tv on tv.TipoVehiculoId =e.TipoVehiculoId
11
12
```

Click the lock to make changes.

Figura 33

Resultado final del filtro para las compañías



Filtro compañía

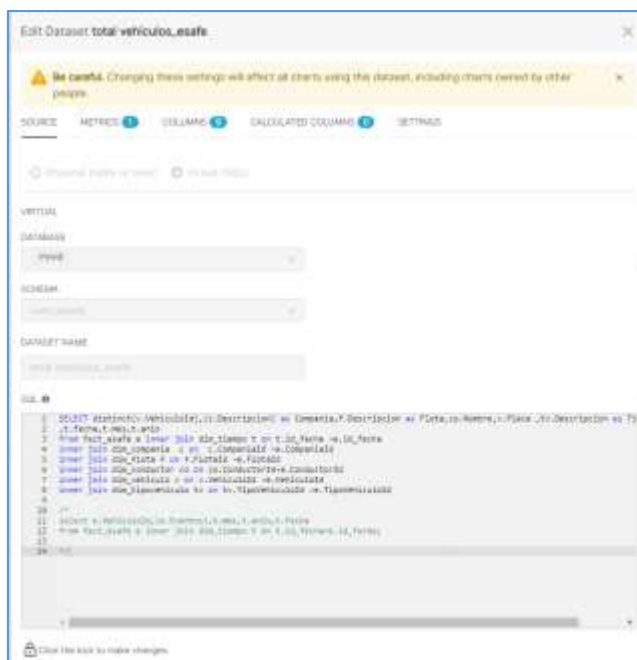
COMPañIA

Type or Select [Compañía]

APPLY

Figura 34

Dataset para llevar el conteo de las unidades analizadas



Edit Dataset: total_vehiculos_esafe

WARNING: Changing these settings will affect all charts using this dataset, including charts owned by other people.

SOURCE METRICS COLUMNS CALCULATED COLUMNS SETTINGS

VIRTUAL

DATABASE: [dropdown]

SCHEMA: [dropdown]

TABLENAME: [dropdown]

```
SQL
SELECT count(*) AS total_vehiculos, count(*) AS total_vehiculos
FROM total_vehiculos AS total_vehiculos
WHERE total_vehiculos.compañía LIKE 'A%'
GROUP BY total_vehiculos.compañía
```

Click the lock to make changes.

Figura 35

Resultado del dataset: un cuadro con el conteo total de las unidades analizadas en todo el cubo multidimensional



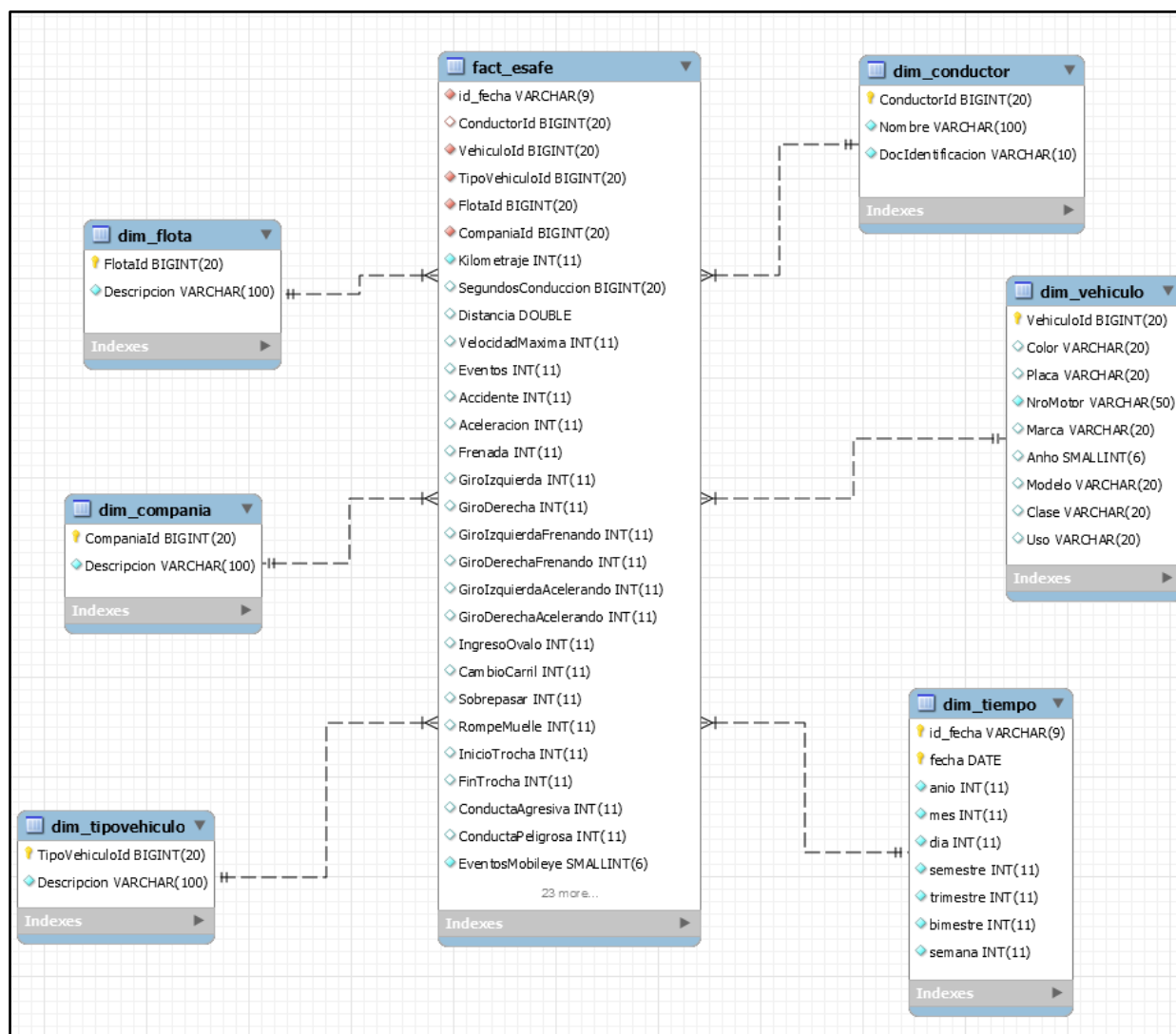
Este será el procedimiento que se hará para todos los cuadros y *dashboard's* que se mostrarán en el análisis; cada diseño tiene una estructura de *query* similar, como se observa en la Figura 30 y en el resultado obtenido de ello, que aparece en la Figura 31, la única variante son los campos a analizar, de modo que con los filtros se tendrá un resultado más rápido y detallado.

4.2 Prototipos de la Aplicación

4.2.1 Sistema del cubo

Figura 36

Estructura multidimensional tipo estrella para el análisis de información de los eventos Safe generados en la plataforma web Clocator



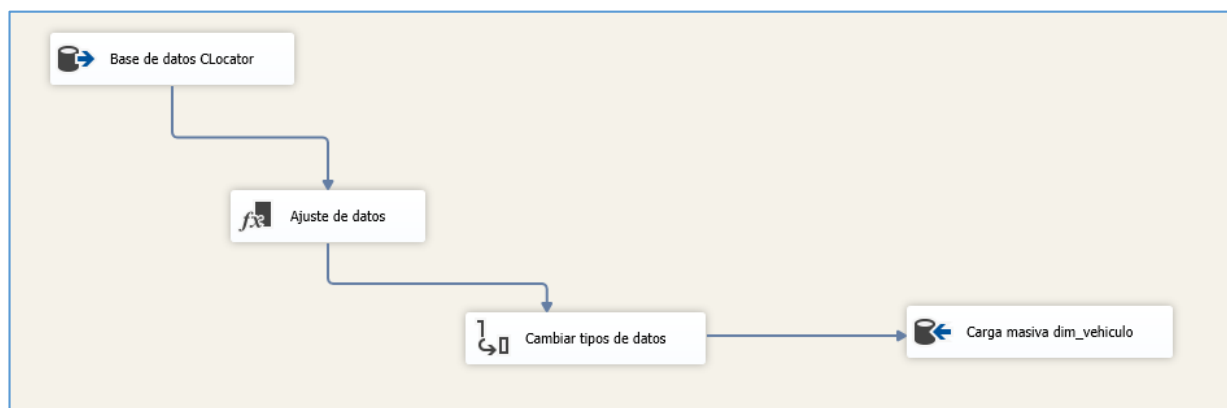
En la elaboración del sistema del cubo (ver Figura 32) se tuvieron en cuenta las relaciones entre la tabla principal (hechos) y las dimensiones; este esquema se denomina estrella, y las dimensiones no están vinculadas entre sí, sino con la tabla principal. Esto facilitará la extracción de la información y la realización de consultas más rápidas.

4.2.2 Sistema ETL

Luego de elaborar el esquema del diseño ETL se creó la estructura interna que posibilitará la obtención de los datos contenidos en la base de datos principal, la cual permitirá transformarlos, ajustarlos y cargarlos al nuevo almacén de datos (cubo multidimensional).

Figura 37

Flujo de carga ETL



El interno del origen de datos tendrá la *query* que aparece en la Figura 33 para realizar el proceso de extracción de información.

Figura 38

Query para realizar a dimensión vehículo

```

/*dim_vehiculo*/

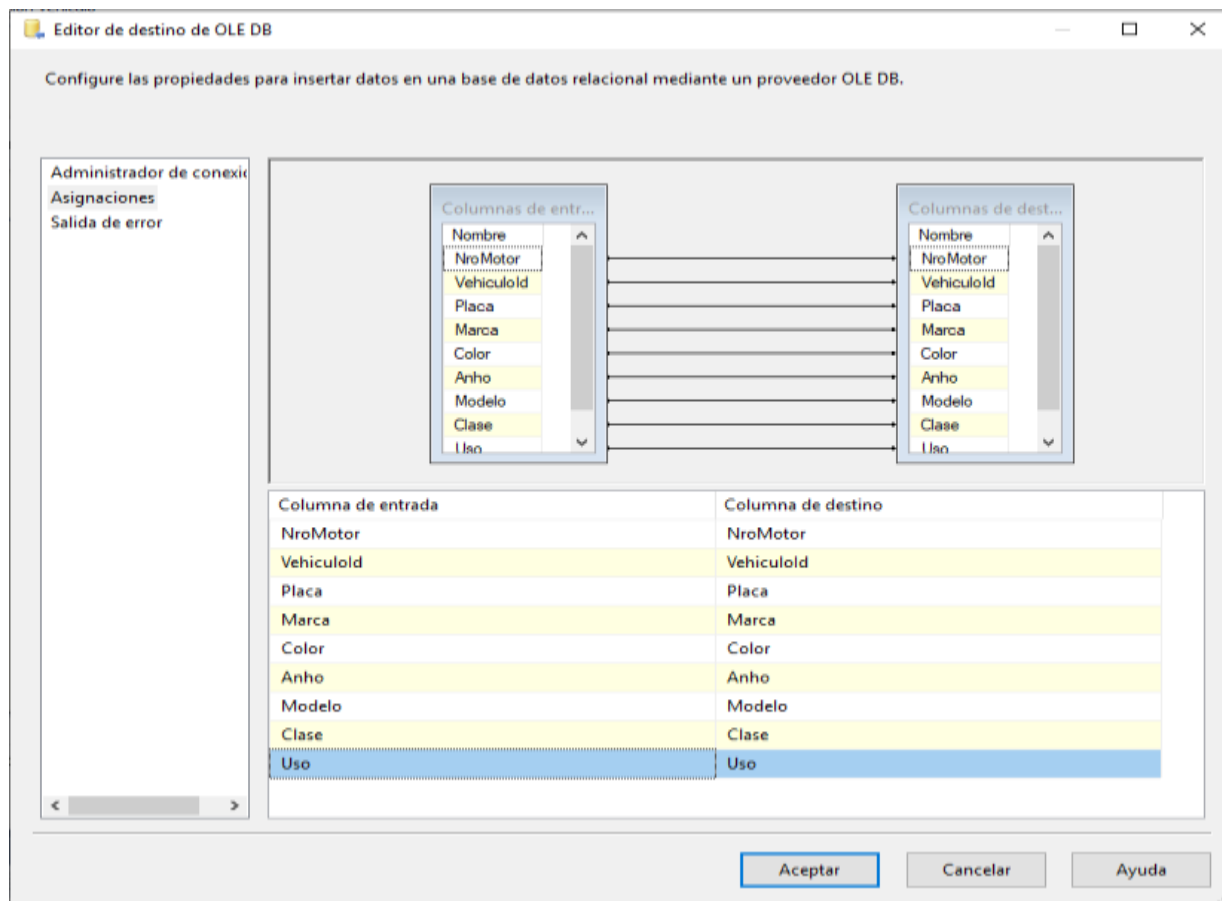
select distinct(v.VehiculoId),v.Color,v.Placa,v.NroMotor,v.Marca,v.Anho,v.Modelo,v.Clase,v.Us
from ComsatelToMaster.TB_ESAFESUM e inner join ComsatelToMaster.TB_VEHICULO v on v.VehiculoId=e.VehiculoId
inner join ComsatelToMaster.TB_FLOTA f on f.FlotaId=v.FlotaId
inner join ComsatelToMaster.TB_COMPANIA c on c.CompaniaId=f.CompaniaId
inner join ComsatelToMaster.TB_TIPOVEHICULO tv on tv.TipoVehiculoId=v.TipoVehiculoId
where e.Fecha='2021-12-02' and e.VehiculoId in (select VehiculoId from ComsatelToMaster.TB_VEHICULO);

```

Posteriormente, se establecerán las relaciones desde la base de datos origen hacia el cubo multidimensional, por lo que todos los datos se tendrán que relacionar.

Figura 39

Relaciones internas, base de datos origen y cubo multidimensional



4.2.3 Sistema BI

Cuando la carga de información al cubo multidimensional se realizó con éxito, al igual que la creación de los *datasets* en la herramienta BI *apache superset*, se elaboró el modelo BI para el análisis de la información.

En lo concerniente a los filtros para este análisis, son los que ayudarán a obtener resultados detallados por compañía, flota, conductor, tipo de vehículo, placa o únicamente por fecha, como se observa en la Figura 36.

Figura 40

Filtros para detallar la información del cubo



Luego se colocará el detalle de las métricas para los vehículos y las horas de conducción, donde se mostrará el acumulado de toda la información almacenada en el cubo, porque aún no se realiza ningún filtro previo.

Figura 41

Métricas



Actualmente existen dos categorías para todos los eventos Safe generados en la plataforma web Clocator, a saber: ADAS, que hace referencia a los eventos emitidos por los

vehículos, y *Driving Behavior*, que describe el comportamiento del conductor durante el recorrido

Figura 42

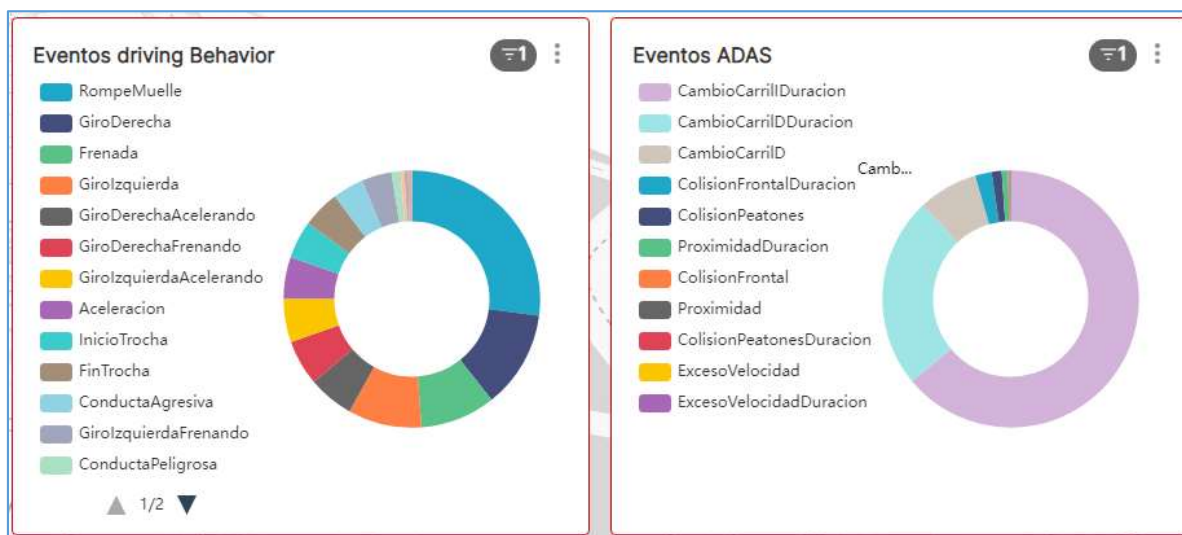
Grupos de evento principales.



También se mostrará el análisis detallado de los eventos generados para cada categoría, desde los más hasta los menos generados.

Figura 43

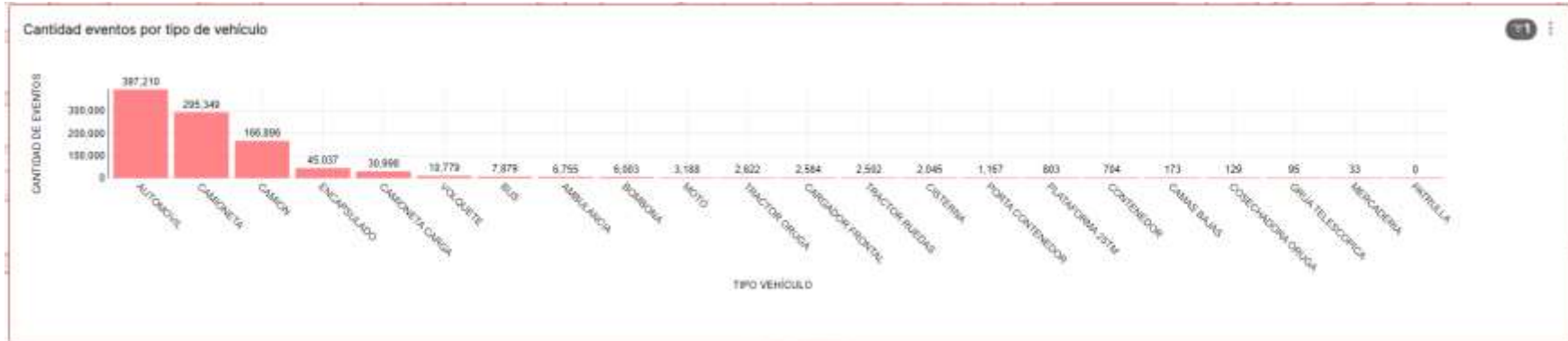
Tipos de evento



Finalmente, se presentará el detalle de los eventos generados para cada tipo de vehículo, lo cual ayudará a identificar qué tipo de unidades producen estos eventos con mayor frecuencia.

Figura 44

Tipos de vehículo

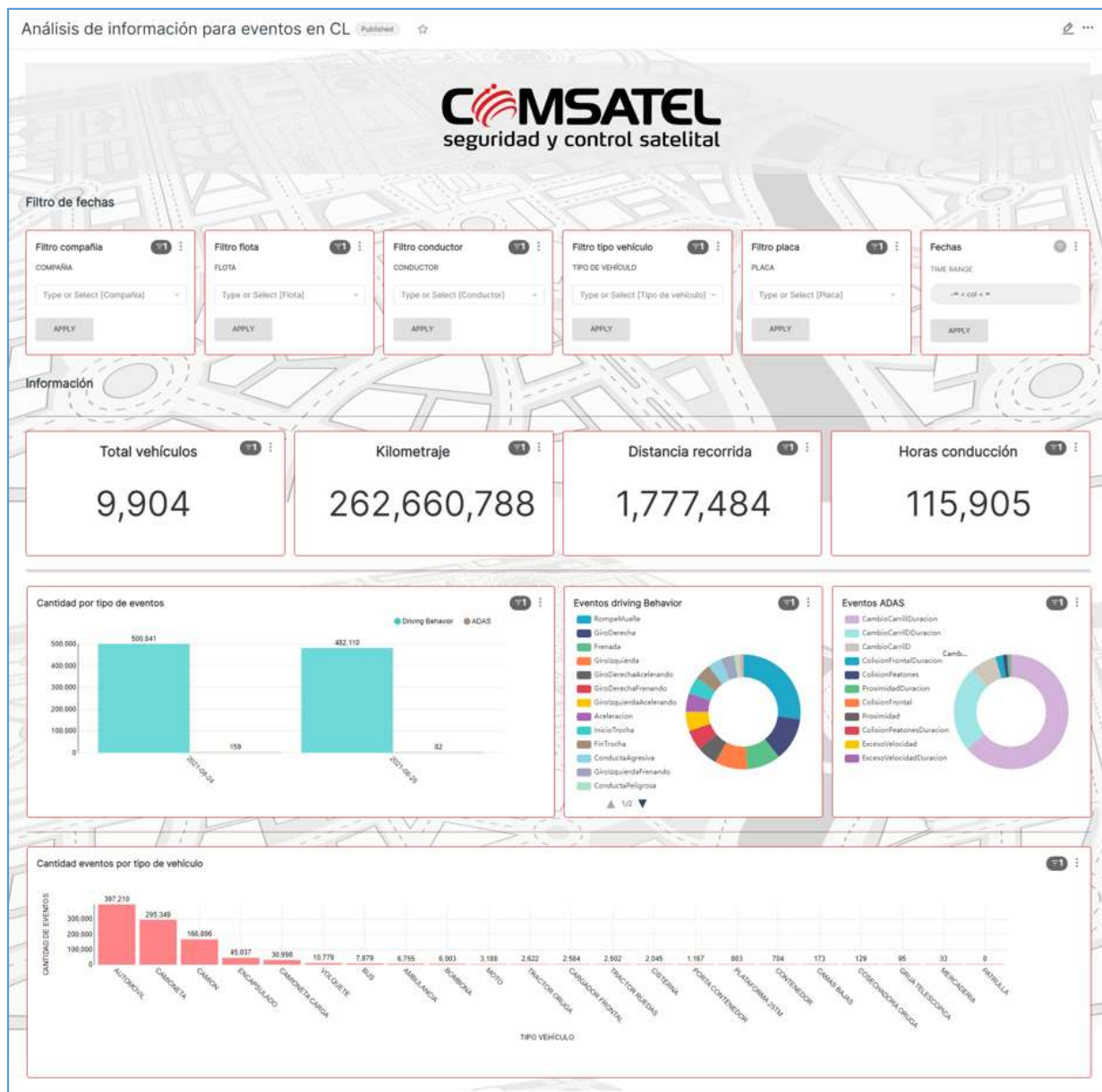


Debido a que la herramienta BI se realiza en la nube, la información estará disponible para los interesados en el proyecto, cada vez que ellos deseen visualizarlo. Para eso se tendrá:

- Prototipo de *Dashboard* para el análisis de información.
- Prototipo para el análisis de los eventos generados en la plataforma web Clocator, donde se encontrará un análisis detallado acerca de estos eventos.

Figura 45

Dashboard cubo multidimensional Esafe



4.3 Medición de la Solución

4.3.1 Análisis de Indicadores Cuantitativos

La variable independiente es el cubo multidimensional. Los resultados descriptivos del pretest para esta variable se presentan en la Tabla 8.

Tabla 8

Pretest - análisis descriptivo de la variable cubo dimensional

Variable / dimensión	Nivel						Total	
	Bajo		Moderado		Alto		n	%
	n	%	n	%	n	%		
Cubo multidimensional	13	54.2	11	45.8	0	0.0	24	100.0
Almacén de información	17	70.8	7	29.2	0	0.0	24	100.0
Velocidad de respuesta	16	66.7	8	33.3	0	0.0	24	100.0
Datos precalculados	19	79.2	5	20.8	0	0.0	24	100.0

Con base en los resultados descritos en la Tabla 8, el 54.2 % (n = 13) de la valoración de la variable “cubo dimensional” se clasificó en el nivel bajo; el 70.8 % (n = 17) también evaluó en ese nivel a la dimensión “almacén de información”; el 66.7 % (n = 16) calificó a la dimensión “velocidad de respuesta” y el 79.2 % (n = 19) a la dimensión “datos precalculados”.

Luego de la implementación, partiendo de los resultados contenidos en la Tabla 9, el 87.5 % (n = 21) calificó a la variable “cubo dimensional” en el nivel bajo; el 37.5 % (n = 9) evaluó en ese nivel a la dimensión “almacén de información”; el 87.5 % (n = 21) a la dimensión “velocidad de respuesta” y el 41.3 % (n = 10) a la dimensión “datos precalculados”.

Tabla 9

Postest - análisis descriptivo de la variable cubo dimensional

Variable / dimensión	Nivel						Total	
	Bajo		Moderado		Alto		n	%
	n	%	n	%	n	%		
Cubo multidimensional	0	0.0	21	87.5	3	12.5	24	100.0
Almacén de información	8	33.3	9	37.5	7	29.2	24	100.0
Velocidad de respuesta	0	0.0	21	87.5	3	12.5	24	100.0
Datos precalculados	7	29.2	10	41.3	7	29.2	24	100.0

4.3.1.1 Comparación pretest-postest.

Tal como se aprecia en la Tabla 10, la valoración media de la variable “cubo multidimensional”, según los participantes del estudio, aumentó de 8.17 a 12.67, lo que representa un incremento porcentual de 55.08 %. Respecto a las dimensiones almacén de información, velocidad de respuesta y datos precalculados se aprecia incrementos del 65.82 %, 69.93 % y 63.02 %, respectivamente.

Tabla 10

Comparación de la media de la variable “cubo dimensional” y sus dimensiones

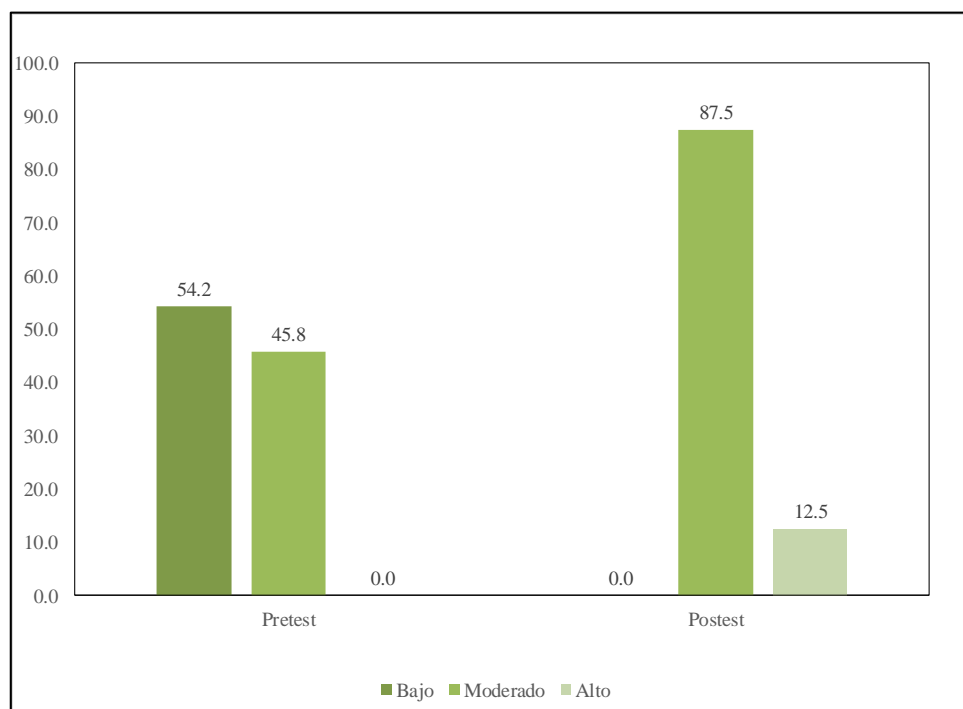
Variable / dimensión	Pretest	Postest	Var. absoluta	Var. Porcentual (%)
Cubo multidimensional	8.17	12.67	4.50	55.08 %
Almacén de información	1.96	3.25	1.29	65.82 %
Velocidad de respuesta	4.29	6.29	3.00	69.93 %
Datos precalculados	1.92	3.13	1.21	63.02 %

A continuación, en la Figura 46, se aprecia que el 54.2 % de los encuestados calificaron la variable “cubo multidimensional” en un nivel bajo antes de la propuesta, pero después este

porcentaje se redujo a 0.0 %, y quienes la evaluaron en el nivel moderado ahora se ubican en 87.5 %.

Figura 46

Comparación pretest – postest de la variable “cubo multidimensional”

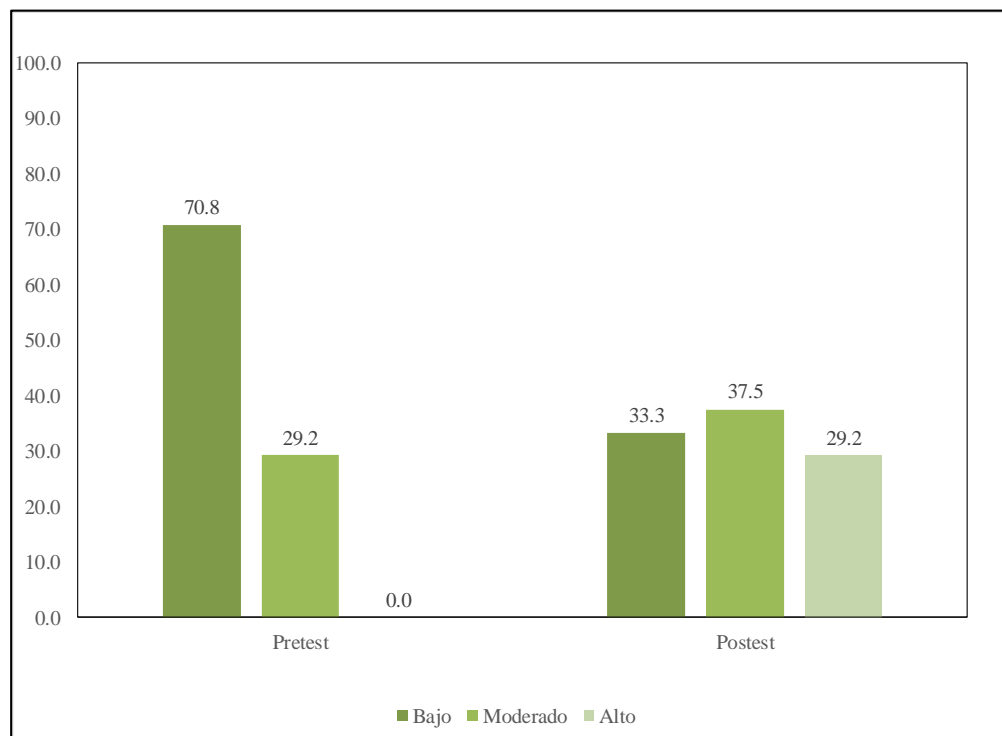


Teniendo en cuenta los datos que aparecen en la Figura 47, el 70.8 % de los encuestados calificaron la dimensión “almacén de información” en un nivel bajo previo a la propuesta, pero

luego de esta, el valor se redujo a 33.3 %, y quienes la evaluaron en el nivel moderado ahora se ubican en 37.5 %.

Figura 47

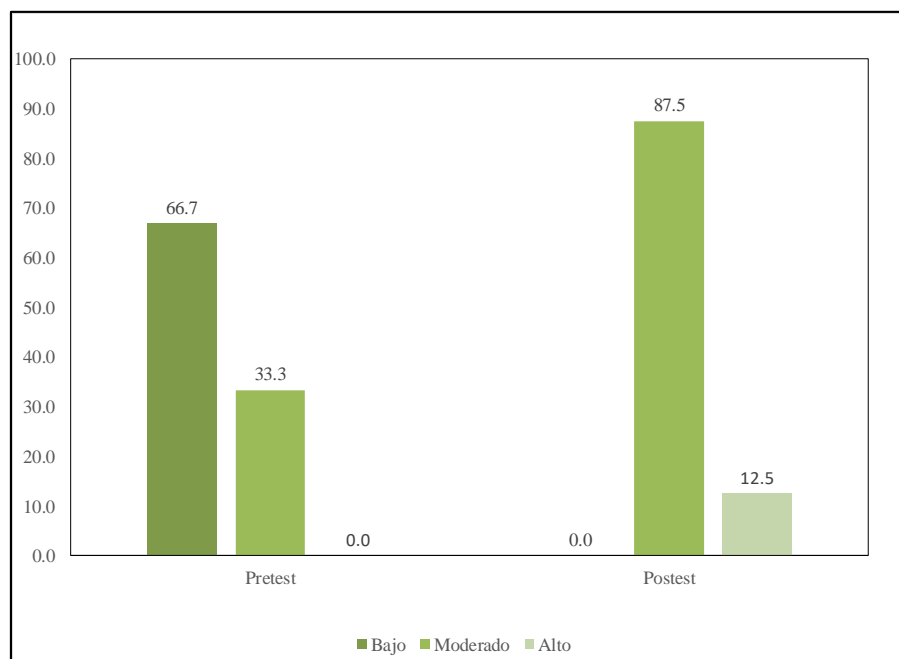
Comparación pretest – postest de la dimensión “almacén de información”



En la Figura 48 se puede ver que el 66.7 % de los encuestados calificaron la dimensión “velocidad de respuesta” en un nivel bajo antes la propuesta, pero después este valor se redujo a 0.0 %, y quienes la evaluaron en el nivel moderado ahora se ubican en 87.5 %.

Figura 48

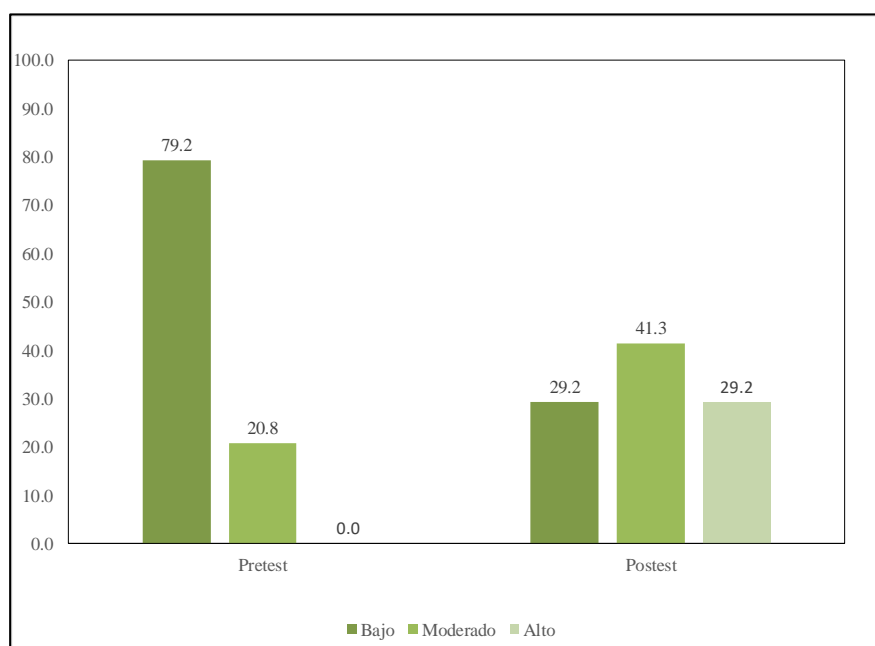
Comparación pretest – postest de la dimensión velocidad de respuesta



Tal como se ilustra en la Figura 49, el 79.2 % de los encuestados calificaron la dimensión “datos precalculados” en un nivel bajo previo a la propuesta, pero luego este valor se redujo a 29.2 %, y quienes la evaluaron en el nivel moderado ahora se ubican en 41.3 %.

Figura 49

Comparación pretest – postest de la dimensión datos precalculados



4.3.1.2 Variable Dependiente Análisis de Información.

- **Pretest:** los resultados descriptivos del pretest para la variable “análisis de información” se presentan en la Tabla 11.

Tabla 11

Pretest - análisis descriptivo de la variable análisis de información

Variable / dimensión	Nivel						Total	
	Poco satisfactorio		Regular		Satisfactorio		n	%
	n	%	n	%	n	%		
Análisis de información	19	79.2	5	20.8	0	0.0	24	0.0
Investigación cuantitativa	21	87.5	3	12.5	0	0.0	24	0.0
Interpretación de la información	19	79.2	5	20.8	0	0.0	24	0.0
Establecer categorías	18	75.0	6	25.0	0	0.0	24	0.0

Según los resultados descritos en la Tabla 11, el 79.2 % (n = 19) de la valoración de la variable “análisis de información” se clasifica en el nivel poco satisfactorio; el 87.5 % (n = 21) de la dimensión “investigación cuantitativa” también la ubicó en ese nivel; al igual que el 79.2 % (n = 19) de la dimensión “interpretación de la información” y el 75.0 % (n = 18) de la dimensión “establecer categorías”.

- **Posttest:** luego de la implementación se obtuvieron los resultados presentados en la Tabla 12, donde el 87.5 % (n = 21) de la valoración de la variable “análisis de información” se clasificó en el nivel regular, al igual que el 79.2 % (n = 19) de la dimensión “interpretación de la información” y el 54.2 % (n = 19) de la dimensión “establecer categorías”. Acerca de la dimensión “investigación cuantitativa”, el 58.3 % (n = 14) la valoró en el nivel satisfactorio.

Tabla 12*Postest - análisis descriptivo de la variable análisis de información*

Variable / Dimensión	Nivel						Total	
	Poco satisfactorio		Regular		Satisfactorio		n	%
	n	%	n	%	n	%		
Análisis de información	0	0.0	21	87.5	3	12.5	24	100.0
Investigación cuantitativa	0	0.0	10	41.7	14	58.3	24	100.0
Interpretación de la información	0	0.0	19	79.2	5	20.8	24	100.0
Establecer categorías	1	4.2	13	54.2	10	41.7	24	100.0

4.3.1.3 Comparación Pretest-Postest.

Como se muestra en la Tabla 13, de acuerdo con los participantes del estudio, la valoración media de la variable “análisis de información” aumentó de 10.63 a 21.25, lo cual representa un incremento porcentual de 99.91 %. En cuanto a las dimensiones de investigación cuantitativa, interpretación de la información y establecer categorías se aprecian incrementos del 68.71 %, 123.08 % y 95.65 %, respectivamente.

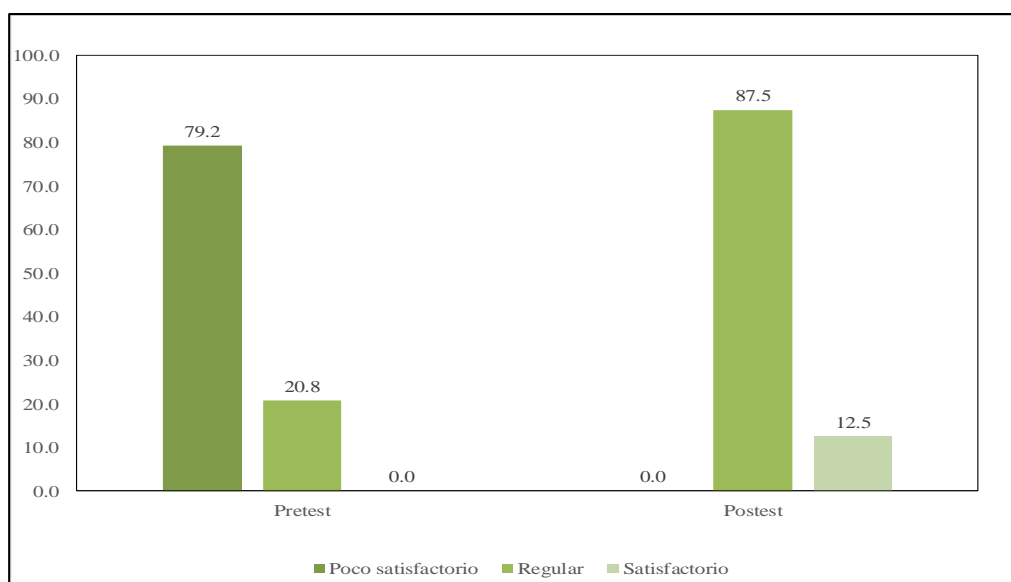
Tabla 13*Comparación de la media de la variable análisis de información y sus dimensiones*

Variable / dimensión	Pretest	Postest	Var. absoluta	Var. Porcentual (%)
Análisis de información	10.63	21.25	10.62	99.91 %
Investigación cuantitativa	1.63	2.75	1.12	68.71 %
Interpretación de la información	3.25	7.25	4.00	123.08 %
Establecer categorías	5.75	11.25	5.50	95.65 %

Teniendo en cuenta la información que aparece en la Figura 50, el 79.2 % de los encuestados calificaron la variable “análisis de información” en un nivel poco satisfactorio antes de la propuesta, pero luego este valor se redujo a 0.0 %, y quienes la evaluaron en el nivel regular ahora se ubican en 87.5 %.

Figura 50

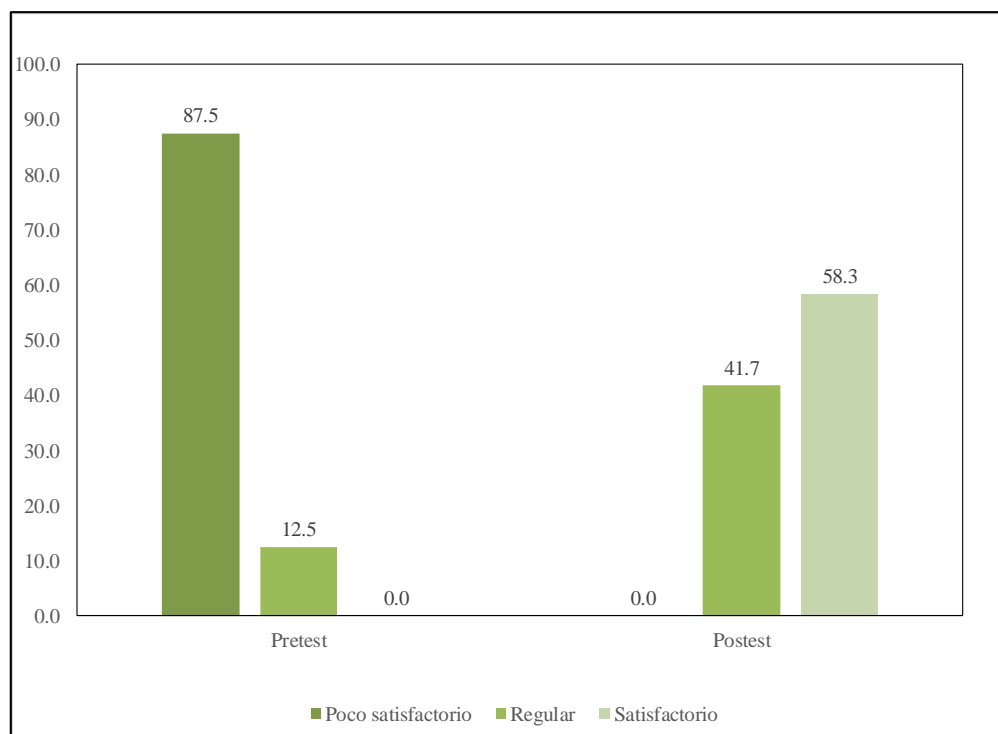
Comparación pretest – postest de la variable análisis de información



En la Figura 51 se observa que el 87.5 % de los encuestados calificaron la dimensión “investigación cuantitativa” en un nivel poco satisfactorio previo a la propuesta, pero después este valor se redujo a 0.0 %, y quienes la evaluaron en el nivel satisfactorio ahora se ubican en 58.3 %.

Figura 51

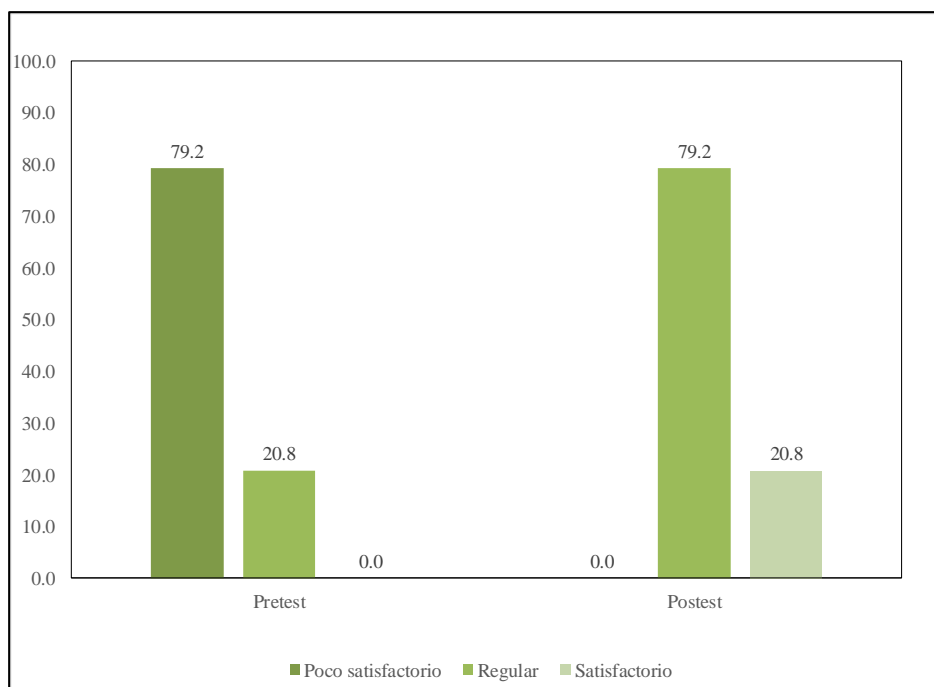
Comparación pretest – posttest de la dimensión investigación cuantitativa



Según los datos contenidos en la Figura 52, el 79.2 % de los encuestados calificaron la dimensión “interpretación de la información” en un nivel poco satisfactorio previo a la propuesta, pero después este valor se redujo a 0.0 %, y quienes la evaluaron en el nivel regular ahora se ubican en 79.2 %.

Figura 52

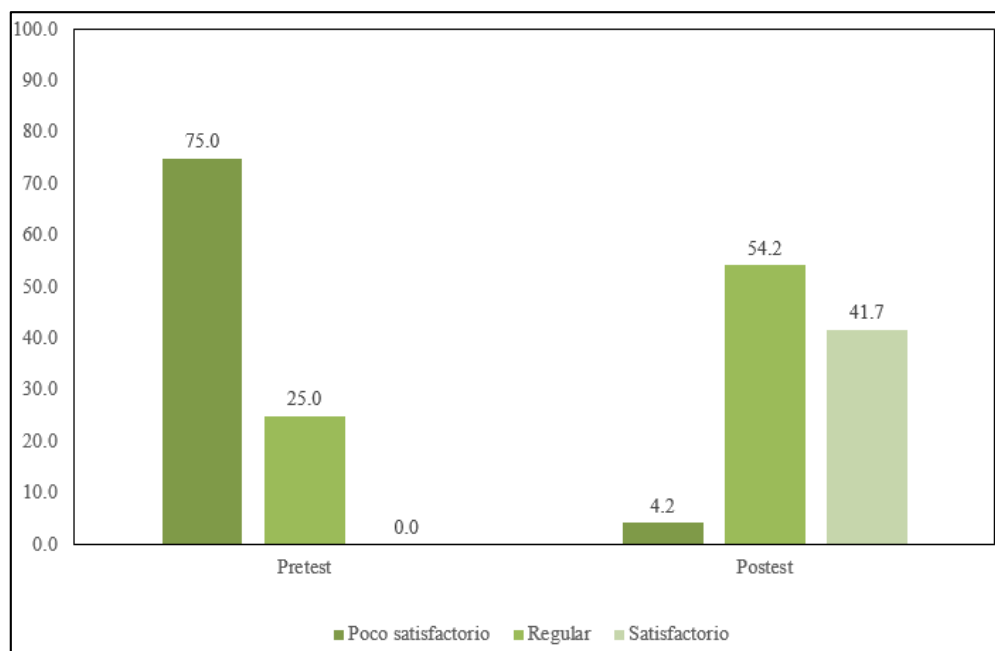
Comparación pretest – posttest de la dimensión interpretación de la información



Con base en la información de la Figura 53, el 75.0 % de los encuestados calificó la dimensión “establecer categorías” en un nivel poco satisfactorio antes de la propuesta, pero luego este valor se redujo a 4.2 %, y quienes la evaluaron en el nivel regular ahora la ubican en 41.7 %.

Figura 53

Comparación pretest – posttest de la dimensión establecer categorías



Capítulo V. Conclusiones y Recomendaciones

5.1 Discusión y Conclusiones

5.1.1 Discusión

En esta investigación se diseñó una propuesta de un cubo multidimensional que permite el análisis de información de eventos Safe generados en la plataforma web Clocator en Comsatel Perú S.A.C., utilizando la metodología propuesta por Ralph Kimball. Así, se determinó que esta implementación, según los colaboradores de la empresa, hace posible mejorar el análisis de esta información y, por ende, la toma de decisiones en un 99.91 %, de modo que se logró una mejora en la valoración de los 24 encuestados ($Z = -4.90$, $p < 0.05$). Esto se sustenta en los argumentos de Violino (2017), quien señaló que este diseño se sustenta en el almacenamiento de datos

correctos y traducidos al lenguaje comercial para facilitar su comprensión, por lo cual es posible anticipar eventos que puedan poner en riesgo a una organización.

De acuerdo con Rolón (2021), el diseño del cubo multidimensional propuesto dispone de una tabla central denominada hechos, que contiene las llaves principales de las diferentes dimensiones, lo cual facilitó la descripción de los datos calculados para garantizar una disponibilidad efectiva de la información. Esto, como indicó Orlandi (2020), se ha vuelto muy común, pues la dependencia únicamente es con la tabla hechos, por lo que se tiene una respuesta más rápida.

Por otro lado, desde el punto de vista de Pratt (2020), para asegurar la efectividad de la propuesta se requiere de personal calificado para ayudar a que esa información sea explotada, puesto que con estas habilidades podrán gestionar, ordenar y analizar la información almacenada en las bases de datos, lo que desembocará en mejoras en la organización. En concordancia, Olavsrud (2020) sostuvo que no solo se requieren inversiones tecnológicas, sino también profesionales que contribuyan a mejorar sus procesos, acoplar su información y alinear lo favorable a decisiones eficaces.

Adicionalmente, en la propuesta se establecieron los estándares que deben basarse en los datos administrados por la organización (Poveda, 2018), para lo cual se determinaron, en primer lugar, las necesidades de la empresa y después se definió el objetivo que tendrá el cubo multidimensional. Otro resultado importante de este trabajo es la mejora en la investigación cuantitativa ($Z = -4.56$, $p < 0.05$), con un crecimiento de 68.71 %. Esto, en palabras de Peña et al. (2019), refleja la oportunidad de analizar y explotar los datos para buscar soluciones, proponer mejoras, fortalecer debilidades y potenciar las organizaciones. Así, con la implementación de la propuesta, la información ha dejado de ser estática (Micha, 2021) y se pueden generar

requerimientos en función de las necesidades actuales, aunque también se maneja la visión de adaptarse a los cambios.

Sumado a lo anterior se observó una mejora en la interpretación de la información ($Z = -4.90$, $p < 0.05$), con un crecimiento de 123.08 %, de manera que se alcanza una mayor velocidad de respuesta y el procesamiento del volumen de datos aumenta de un tiempo más corto (Sacolick, 2020). Por último, se identificó una mejora de 95.65 % ($Z = -4.34$, $p < 0.05$) en el establecimiento de categorías, lo que, en términos de Wayner (2020), se evidencia en brindar soporte a un gran volumen de información que podrá ayuda a establecer dimensiones, así como controlar y gobernar los datos.

5.1.2 Conclusiones

Partiendo de lo explicado en esta investigación, con base en distintas perspectivas teóricas y en los resultados de los diferentes análisis detallados a lo largo del estudio, a continuación se muestra las conclusiones a las que se llegó en el desarrollo de este trabajo.

- Con la propuesta de un cubo multidimensional de eventos Safe se logró mejorar en un 99.91 % el análisis de información que generan los vehículos en la plataforma web Clocator en Comsatel Perú S.A.C., por lo que se llegó al nivel regular.
- Gracias a la propuesta se mejoró la investigación cuantitativa, pues se evidencia una mejora del 68.71 % en la apreciación de los colaboradores, lo que permitió pasar al nivel satisfactorio. Además, la propuesta también incrementó un 123.08 % la valoración de la interpretación de los datos, pasando del nivel poco satisfactorio al regular.

- A través de la propuesta se logró el establecimiento de categorías en un 95.65 %, así que se pasó del nivel poco satisfactorio al regular.

5.2 Recomendaciones

A fin de optimizar la toma de decisiones de la empresa, considerando también la mejora en la disponibilidad, el control y el gobierno de los datos, se recomienda:

- Proponer un programa de capacitación de los empleados para analizar los datos y diseñar requerimientos para atender las necesidades internas y externas de información.
- Implementar mecanismos de seguridad que aseguren la disponibilidad, la integridad, y el resguardo de la información de la aplicación propuesta.
- Implementar diferentes cubos multidimensionales con base en el método de Ralph Kimball, con el objetivo de analizar los datos sobre diferentes temas que ayuden a mejorar el servicio que ofrece Comsatel Perú S.A.C

6 Referencias

- Álvarez, E. D., y Barreda, L. (2020). La estadística descriptiva en la formación investigativa del instructor de arte. *Conrado*, 16(73), 100-107.
<https://conrado.ucf.edu.cu/index.php/conrado/article/view/1277>
- Arias, J., Villasís, M. Á., y Miranda, M. G. (2016). El protocolo de investigación III: la población de estudio. *Revista Alergia México*, 63(2), 201-206.
- Auribox Training. (2017). *¿Qué son los cubos OLAP?* <https://blog.auriboxtraining.com/business-intelligence/los-cubos-olap/>
- Business Intelligence. (2008). *Cubos*. <https://www.businessintelligence.info/definiciones/cubos-olap.html>
- Eguila, A., y Parco, A. A. (2007). *Implementación de una herramienta de inteligencia de negocios para la administración de justicia sobre una metodología ad-hoc. [Tesis de grado]*. Universidad Nacional Mayor de San Marcos.
<https://hdl.handle.net/20.500.12672/2645>
- Euroinnova Business School. (s.f.). *¿Qué es la interpretación de datos?*
<https://www.euroinnova.co/blog/que-es-la-interpretacion-de-datos#:~:text=La%20interpretaci%C3%B3n%20de%20datos%20es,esclarecer%20la%20toma%20de%20decisiones.>
- Fuentes, L., y Valdivia, R. (2010). Incorporación de Elementos de Inteligencia de Negocios en el Proceso de Admisión y Matrícula de una Universidad Chilena. *Ingeniare. Revista Chilena de Ingeniería*, 18(3), 383-394. <http://dx.doi.org/10.4067/S0718-33052010000300012>

García, J. M. (2012). *Conceptos de tecnología OLAP*.

https://groups.google.com/g/itecnodgo_ninteligentes_ad_12/c/Ns8JHCK6BWk

Hernández, R., Fernández, C., y Baptista, P. (2014). *Metodología de la Investigación. 6a edición*.

McGraw Hill.

Jiménez, M. D., Zavala, M. E., y Téllez, M. I. (2021). Estrategias de Mercadotecnia Social. Caso

Psicología y Desarrollo, S.A de C.V. *Revista de Investigaciones Universidad del*

Quindío, 22(S1), 152-162. <https://doi.org/10.33975/riuuq.vol33nS1.489>

López, P. L. (2004). Población, Muestra Y Muestreo. *Punto Cero*, 9(8), 69-74.

López A., y López, Salvador. (2017). Las tecnologías de bases de datos al servicio de la metría de la información y del conocimiento científico. *Investigación bibliotecológica*, 31(spe),

11-14. <https://doi.org/10.22201/iibi.24488321xe.2017.nesp1.57883>

Marín, A., Hernández, E., y Flores, J. (2016). Metodología para el análisis de datos cualitativos en investigaciones orientadas al aprovechamiento de fuentes renovables de energía.

Revista Arbitrada Interdisciplinaria Koinonía, 1(1), 60-75.

Micha, J. C. (2021). *Modelo de Inteligencia de Negocios como Soporte a la Toma de Decisiones Estratégicas en el Centro Pre Universitario CEPUNC de la Universidad Nacional de Cajamarca*. Universidad Nacional de Piura .

<http://repositorio.unp.edu.pe/handle/20.500.12676/2879>

Montes, G. (2000). Metodología y técnicas de diseño y realización de encuestas en el área rural.

Temas Sociales(21), 39-50.

Ñaupas, H., Valdivia, M. R., Palcios, J. J., y Romero, H. E. (2018). *Metodología de la*

investigación. Cuantitativa - Cualitativa y Redacción de la Tesis. Ediciones de la U.

- Olavsrud, T. (2020). *4 tendencias en análisis de datos que dominarán en el 2020*. Recuperado el 6 de octubre de 2021, de <https://cioperu.pe/articulo/29825/4-tendencias-en-analisis-de-datos-que-dominaran-en-el-2020/?p=3>
- Orlandi, A. M. (2020). *Tecnología big data y analítica aplicada a la gestión de recursos humanos en una entidad financiera. Un caso de estudio. [Tesis de maestría]*. Universidad Tecnológica Nacional. <http://hdl.handle.net/20.500.12272/5228>
- Oviedo, H. C., y Campo, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572-580.
- Palomino, J., Peña, J., Zevallos, G., y Orizano, L. (2015). *Metodología de la investigación*. Editorial San Marcos.
- Peña, C., Risco, G., Cardoza, L., Ubillus, E., y Olaya, C. (2019). Modelo bioeconómico que pronostica las consecuencias técnicas, biológicas y económicas en el cultivo de langostino. *Manglar. Revista de Investigación Científica*, 16(1), 3-10.
<http://dx.doi.org/10.17268/manglar.2019.002>
- Pesquera, C. (2020). *Tipos de modelos OLAP. ¿Tabular o Multidimensional?* Recuperado el 11 de octubre de 2021, de <https://carlospesquera.com/tipos-olap-tabular-multidimensional/>
- Poveda, C. A. (2018). *Diseño e implementación de un cubo multidimensional OLAP usando software libre: estudio de caso sector comercial repuesto automotriz. [Tesis de maestría]*. Universidad Autónoma de Bucaramanga.
<http://hdl.handle.net/20.500.12749/3437>
- Pratt, M. K. (2020). *Por qué siguen fallando las iniciativas de análisis de datos*. Recuperado el 11 de octubre de 2021, de <https://cioperu.pe/articulo/30603/por-que-siguen-fallando-las-iniciativas-de-analisis-de-datos/>

- Rivadera, G. R. (2010). La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). *Cuadernos de la Facultad*(5), 56-71.
- Rolón, M. D. (2021). *Metodología para el desarrollo de Datamarts en Universidades Nacionales del Consorcio SIU. Caso de Estudio: Área Graduados de la FCEQyN. [Tesis de maestría]* . Universidad Nacional de Misiones.
- Romero, A. (1997). Reciprocidad y convergencia de los enfoques cuantitativo y cualitativo. Una experiencia de investigación en salud. *Revista Costarricense de Salud Pública*, 16(11).
- Sacolick, I. (2020). *Cómo elegir una plataforma de análisis de datos*. Recuperado el 10 de octubre de 2021, de <https://cioperu.pe/articulo/30575/como-elegir-una-plataforma-de-analisis-de-datos/>
- Siesquen, C. A. (2019). *Solución de inteligencia de negocios para la gestión del gasto, en una entidad financiera de Piura; 2017. [Tesis de maestría]*. Universidad Nacional de Piura. <http://repositorio.unp.edu.pe/handle/20.500.12676/2233>
- Silva, G. E., Zapata, V. M., Morales, K. P., y Toaquiza, L. M. (2019). Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones. *Ciencia Digital*, 3(3.4), 397-418. <https://doi.org/10.33262/cienciadigital.v3i3.4..922>
- Simão, V. L. (2010). *Formación Continuada y varias voces del profesorado de educación infantil de Blumenau: una propuesta desde dentro*. Universitat de Barcelona.
- Tecnologías Información. (2018). *Interpretación de los Datos: Objetivos, Tipos y Beneficios*. <https://www.tecnologias-informacion.com/interpretacion-datos.html>
- Violino, B. (2017). *Cómo evitar los grandes errores de análisis de datos*. Recuperado el 10 de octubre de 2021, de <https://cioperu.pe/articulo/24189/como-evitar-los-grandes-errores-de-analisis-de-datos/>

Wayner, P. (2020). *15 herramientas de análisis predictivo*. Recuperado el 5 de octubre de 2021,
de <https://cioperu.pe/articulo/30593/15-herramientas-de-analisis-predictivo/>

Yegualp, S. (2017). *Se presentala primera versión candidata de MySQL 8.0*.

<https://cioperu.pe/articulo/24424/se-presentala-primera-version-candidata-de-mysql-80/>

7 Anexos

Anexo 1. Matriz de consistencia

Anexo 2. Ficha de investigación

Anexo 3. Declaración de autenticidad y no plagio

Anexo 4. Formulario de autorización de publicación

Anexo 5. Reporte de similitud por Turnitin