



**UNIVERSIDAD DE LA FRONTERA
FACULTAD DE INGENIERÍA Y CIENCIAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN E INFORMÁTICA**

**“IMPLEMENTACIÓN DE UNA PLATAFORMA DE BUSINESS INTELLIGENCE PARA
EL CONTROL DE MÉTRICAS DE NEGOCIO MEDIANTE UN CUADRO DE MANDOS
INTEGRAL”**

**PLAN DE TRABAJO ACTIVIDAD DE TITULACIÓN,
MODALIDAD PRÁCTICA PROFESIONAL CONTROLADA
PARA OPTAR AL TÍTULO DE INGENIERO
INFORMÁTICO**

**PROFESOR GUÍA: MG. JORGE ALBERTO HOCHSTETTER
DIEZ**

**DIEGO NICOLÁS VILLAGRÁN CASTRO
2020**

**“IMPLEMENTACIÓN DE UNA PLATAFORMA DE BUSINESS INTELLIGENCE PARA
EL CONTROL DE MÉTRICAS DE NEGOCIO MEDIANTE UN CUADRO DE MANDOS
INTEGRAL”**

**DIEGO NICOLAS VILLAGRAN CASTRO
N° DE MATRÍCULA: 18195986K14
CARRERA: INGENIERÍA INFORMÁTICA
NIVEL CURSADO: IX NIVEL**

**JORGE ALBERTO HOCHSTETTER DIEZ
Profesor Guía**

**ANIA LORENA CRAVERO LEAL
Profesor evaluador**

**RICARDO ANTONIO GACITÚA TORRES
Profesor evaluador**

**MAURICIO DIEGUÉZ REBOLLEDO
Director departamento de ciencias de la computación e informática**

Agradecimiento a mi Madre, Hermana y Abuela

Índice de contenidos

Capítulo 1 Introducción.....	9
1.1 Descripción del problema.....	10
1.1.1 Exposición general del problema.....	10
1.1.2 Nivel actual del problema.....	11
1.1.3 Solución del problema.....	11
1.2 Objetivos.....	12
1.2.1 Objetivo general.....	12
1.2.2 Objetivos específicos.....	12
Capítulo 2. Antecedentes Generales.....	13
2.1 Marco conceptual.....	13
2.1.1 Gestión de los datos.....	13
2.1.2 Inteligencia de negocios.....	14
2.1.3 Software como un servicio.....	14
2.1.4 Métricas de negocio en software como un servicio.....	14
2.1.4.1 Métrica de ingresos recurrentes.....	14
2.1.4.2 Métrica de la tasa de abandono.....	14
2.1.4.3 Métrica para costos adquisición de clientes.....	15
2.1.4.4 Métrica sobre el ciclo de vida del cliente.....	15
2.2 Contexto del mercado.....	15
2.2.1 Soluciones para disminuir la deserción.....	15
2.2.2 Soluciones de visualización de métricas.....	15
2.3 Tecnologías asociadas.....	16
2.2.1 Uso de Elasticsearch.....	16
2.2.1 Uso de Logstash.....	16
2.2.1 Uso de Kibana.....	16
Capítulo 3 Metodología.....	17
3.1 Metodología Moss.....	17
3.2 Metodología Moss para proyectos Business Intelligence.....	17
3.3.- Plan de trabajo.....	18
Capítulo 4 Actividades realizadas.....	21
4.1 Justificación.....	21
4.2 Planeación.....	21
4.3 Análisis de negocio.....	22
4.3.1 Toma de requerimientos.....	22
4.3.1.1 Documento de requerimientos de negocio (BRD).....	22
4.3.1.2 Documento de especificaciones de requisitos de software (SRS).....	22
4.3.1.2.1 Requerimientos funcionales.....	22
4.3.1.2.2 Requerimientos no funcionales.....	22
4.3.2 Casos de uso.....	22
4.4 Diseño.....	23
4.4.1 Datos disponibles.....	23
4.4.2 Diseño de la base de datos.....	24
3.4.5 Gráficos de visualización.....	24
3.4.5.1 Gráficas para RR.....	24
3.4.5.2 Gráficas para CAC.....	24
3.4.5.3 Gráficas para CHURN.....	25

4.5	Tecnología.....	25
4.5.1	Solucion a implementar.....	26
4.5.1.1	ELK Stack.....	26
4.5.1.1.1	Elasticsearch.....	26
4.5.1.1.2	Logstash.....	26
4.5.1.1.3	Kibana.....	26
4.5.1.1.4	Beats.....	26
4.6	Despliegue e implementacion.....	27
4.6.1	Entorno de instalación.....	27
4.6.1.1	Datos técnicos entorno de instalación.....	27
4.6.1.2	Prerrequisitos de software en el entorno de instalación.....	27
4.6.1.3	Prerrequisitos de seguridad en el entorno de instalación.....	27
4.6.2	Instalar Elasticsearch.....	27
4.6.3	Instalar Kibana.....	27
4.6.4	Instalar Logstash.....	28
4.6.5	Instalar Beat.....	28
4.6.6	Coneccion con JDBC.....	29
Capítulo 5	Resultados.....	31
5.1	Resultado objetivo 1.....	31
5.2	Resultado objetivo 2.....	32
5.3	Resultado objetivo 3.....	35
5.4	Resultado objetivo 4.....	36
5.5	Comentarios.....	39
Capítulo 6	Conclusión.....	40
6.1	Discusión.....	40
6.2	Conclusión.....	40
6.3	Trabajo futuro.....	41
	Nomenclatura.....	42
	Referencias.....	43
	Bibliografia.....	44
	Anexos.....	45

Indice de tablas

Tabla 1: Representacion plan de trabajo objetivo especifico 1.....	18
Tabla 2: Representacion plan de trabajo objetivo especifico 2.....	19
Tabla 3: Representacion plan de trabajo objetivo especifico 3.....	19
Tabla 4: Representacion plan de trabajo objetivo especifico 4.....	20
Tabla 5: Datos de decompañias registradas.....	23
Tabla 6: Datos para los usuarios registrados.....	23
Tabla 7: Datos para las cuentas pagadas.....	23
Tabla 8: Comando en promp para instalar ElasticSearch.....	27
Tabla 9: Comando en promp para instalar Kibana.....	28
Tabla 10: Comando en promp para instalar Logstash.....	28
Tabla 11: Posibilidades de extenciones para Beats.....	28
Tabla 12: Comando en promp para instalar Beats.....	29
Tabla 13: Coneccion JDBC mediante edicion de filebeat.yml.....	29
Tabla 14: Ejemplo de consulta para 12000 usuarios.....	29
Tabla 15: Configuracion para alimentacion de ElastiSearch.....	30
Tabla 16: Formulario de capturacion de requerimientos de negocio.....	32
Tabla 17: Formulario que captura requerimientos funcionales.....	33
Tabla 18: Formulario que captura requerimientos no funcionales.....	34

Indice de imagenes

Imagen 1: Capitalizacion de mercado de 16.528 Mil millones para Slack inc.....	9
Imagen 2: Patallazo al software EasyCargo para la cubicacion de carga en camiones y contenedores.....	10
Imagen 3: Cuadro de mandos de EasyCargo para desarrollo de negocios. Visitas, registros vs costos de publicidad.....	11
Imagen 4: Propuesta de esquema para la valorizacion del dato como elemento para la toma de decisiones basadas en conocimiento.....	13
Imagen 5: Grafico para representar metrica RR.....	24
Imagen 6: Grafico para representar metrica CAC.....	25
Imagen 7: Grafico para representar metrica CHURN.....	25
Imagen 8: Dominios identificados para la verificacion de informacion.....	31
Imagen 9: Resultado diseño para los datos disponibles.....	35
Imagen 10: Resultado despliegue y testeo de Kibana.....	36
Imagen 11: Verificacion y pruebas de los datos para visualizacion.....	36
Imagen 12: Consulta a una fila o elemento en la tabla.....	37
Imagen 13: Graficacion de contador para distribucion de ip para usuarios.....	37
Imagen 14: Selector de espacio para trabajar en Kibana.....	38
Imagen 15: Resultado cuadro de mandos primeras iteraciones.....	38
Imagen 16: Metricas seleccionadas para el rendimiento.....	39

Capítulo 1 Introducción

En el año 2011 Joel Spolsky lanzó Trello, el cual para el año 2016 tuvo un valor de 400 millones de dólares. El 2013, Stewart Butterfield y asociados lanzaron al mercado Slack, el que actualmente tiene un capitalización de mercado de 15 mil millones de dólares en el 2020 [1].

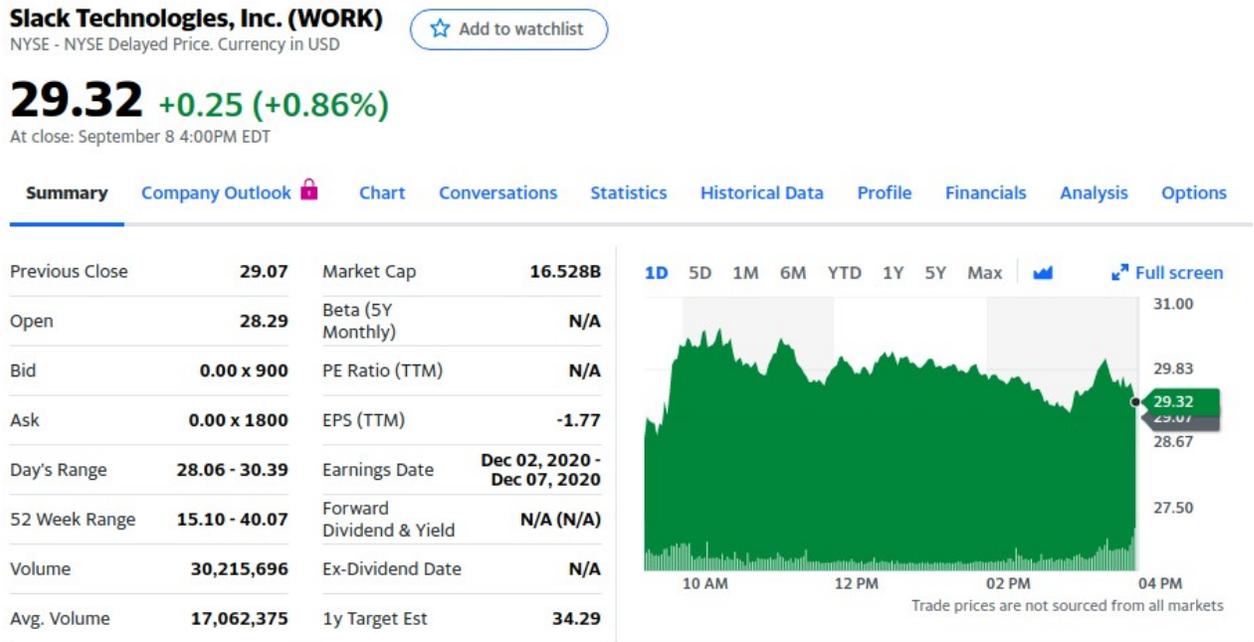


Imagen 1: Capitalización de mercado de 16.528 Mil millones para Slack inc

Ambas son empresas que ofrecen software como un servicio, y que tuvieron grandes retornos por inversión. Esto hace parecer que no hay mejor momento para ser una empresa de software como servicio que ahora. Si bien esto es cierto en muchos aspectos, la verdad es que la mayoría de las empresas de software como servicio no llegan ni al millón de dólares.

Sólo el cuatro por ciento de las empresas de software como servicio alcanzan un millón de dólares en ingresos y solo el 0.04 por ciento llega a los 10 millones de dólares, y a medida que aumentan las cifras de ingresos, esa cifra se reduce aún más [2].

Por lo que en el presente trabajo se presenta al Business Intelligence como una herramienta para la toma de decisiones en procesos de negocio para alcanzar mejores cifras de negocio.

Y parafraseando al científico Lord Kelvin [3] :

“...cuando se puede medir lo que se dice y expresarlo en números, se sabe algo al respecto...”

Por lo que el Business Intelligence será la herramienta para medir, expresar en números y tomar acción.

1.1 Descripción del problema

1.1.1 Exposición general del problema

En el año 2016 se fundó en República Checa la empresa Bee Interactive [4] con el software como servicio EasyCargo el cual está destinado a automatizar la cubicación en espacios de carga tales como contenedores, carrocerías y camiones, para empresa de logística, transporte y suministros entre muchas otras que suman aproximadamente sesenta mil las empresas registradas con ochenta mil usuarios inscritos en total .



Imagen 2: Patallazo al software EasyCargo para la cubicacion de carga en camiones y contenedores

EasyCargo al ser un software como servicio genera ingresos por suscripción por lo que este está sujeto a métricas particulares que producen un gran impacto en los resultados futuros y en los objetivos de crecimiento. Si un cliente está contento con el servicio se quedará mucho tiempo y el beneficio que se puede obtener con ese cliente se incrementará considerablemente. Por otro lado, si un cliente no está contento abandonará rápidamente y el negocio muy probablemente perderá la inversión que hizo para conseguir a ese cliente. Esto crea una dinámica sobre la “captación” del cliente y la “retención” del cliente. Y debido a la importancia de estos parámetros nos centraremos en una gran cantidad en métricas que ayuden a entender la retención y la pérdida de clientes. Y son las que finalmente ayudan a responder tanto las preguntas más básicas de salud del negocio, así como planificar la toma de acciones para escalar la valorización de la empresa [5].

Para EasyCargo el problema radica en que solo el 4.7% de sus usuarios totales han sido suscripciones de pago y que de los cuales solo el 33% son activos al día de hoy. Y que si bien permite mantener la empresa, esto no alienta al equipo a pisar el acelerador.

1.1.2 Nivel actual del problema

La empresa Bee Interactive implementó el uso de Google Analytics para el seguimiento de las campañas de posicionamiento web, publicidades dirigidas, datos demográficos entre otros de gran valor, pero de bajo impacto en la dirección económica. Además, en la empresa no existe una solución al estudio de los datos generados por los logs del servidor. Y por otra parte tampoco se utilizan los datos planteados en esta descripción del problema para el análisis de la adquisición y retención de clientes. Así como tampoco se utilizan herramientas para analizar, administrar y visualizar los datos.

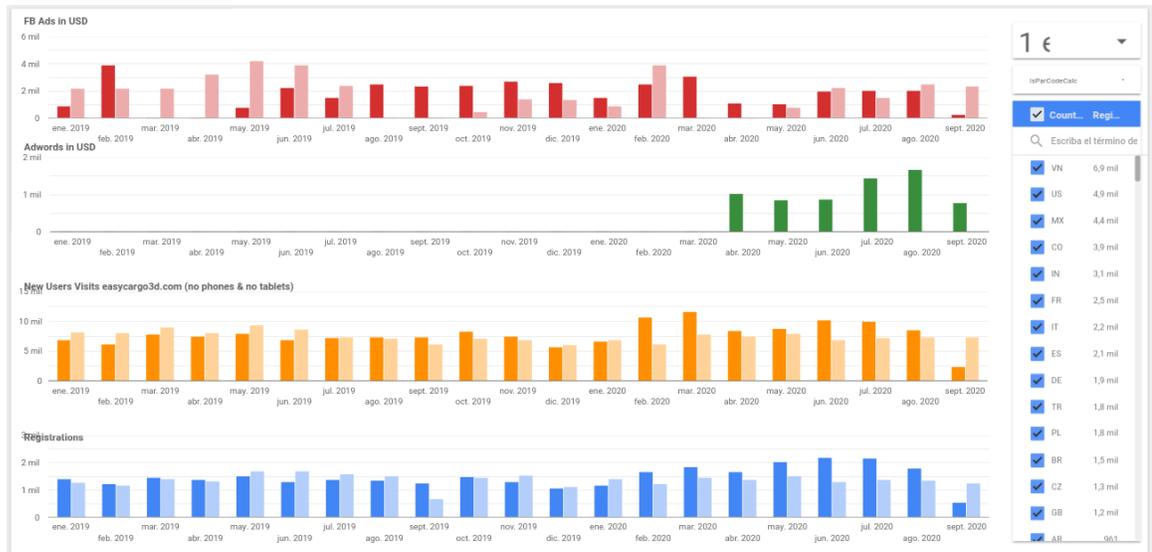


Imagen 3: Cuadro de mandos de EazyCargo para desarrollo de negocios. Visitas, registros vs costos de publicidad

1.1.3 Solución del problema

Se propone iniciar prácticas de análisis de datos mediante un conjunto de procesos para la detección de métricas e indicadores clave del negocio mediante un conjunto de herramientas de software de código abierto para el manejo de datos, conocido como ELK Stack, el cual es en primera instancia la unión de tres plataformas divididas en: ElasticSearch, utilizado para almacenar registros. LogStash, utilizada tanto para el envío como para el procesamiento y almacenamiento de registros. Y Kibana, que es una herramienta de visualización (una interfaz web) que se aloja a través de Nginx o Apache y permite a los usuarios acceder a datos de cualquier fuente, en cualquier formato, y buscar, analizar y visualizar esos datos en tiempo real.

1.2 Objetivos

1.2.1 Objetivo general

- Implementar una plataforma de Business Intelligence para el análisis de datos del área de desarrollo de negocios en una empresa que ofrece software como un servicio.

1.2.2 Objetivos específicos

- Verificar los principales conceptos, técnicas, tecnologías utilizadas y enfoques metodológicos existentes en los procesos de proyectos de Business Intelligence.
- Analizar los requerimientos para la implementación del sistema.
- Elaborar base de datos para facturaciones, logs, cuentas de usuario y procesos de negocio de la empresa.
- Desplegar herramienta de Business Intelligence con base de datos integrada en servidor de prueba.

Capítulo 2. Antecedentes Generales.

A continuación, se presentan los fundamentos teóricos relacionados a la problemática afrontada, donde se explican algunos conceptos relevantes para el entendimiento del documento, y se entrega un marco teórico conceptual, una exposición de contexto del mercado de estas soluciones y las tecnologías de la información disponibles para abordar las necesidades en la inteligencia de negocios.

2.1 Marco conceptual

Como referencia al dominio del problema, se identifican “los datos” como el concepto director del desarrollo de la teoría, explicando en el punto 2.1.1, dando inicio al área de business intelligence en el 2.1.2, refiriéndose al conjunto de estrategias y herramientas disponibles para los fines del proyecto. A continuación, encuentra la sección 2.1.3 comentando el origen de la modalidad de funcionamiento de la empresa en la que se desarrolla este proceso para finalizar en una introducción a las métricas para la empresas de software como servicio, en el punto 2.1.4.

2.1.1 Gestión de los datos

La gestión de los datos se ha convertido en el eje del desarrollo laboral de varias organizaciones en el mundo que busca acceder a niveles altos de competitividad, implementado sistemas con el objetivo de identificar, capturar y compartir conocimiento dentro de una corporación.

La aplicación de la gestión de datos, al contexto de las empresas ha supuesto una importante fuente de ventajas competitivas que puede garantizar el correcto funcionamiento y la supervivencia de las organizaciones, en el actual escenario económico que se caracteriza por la dura competencia y globalización de mercados.

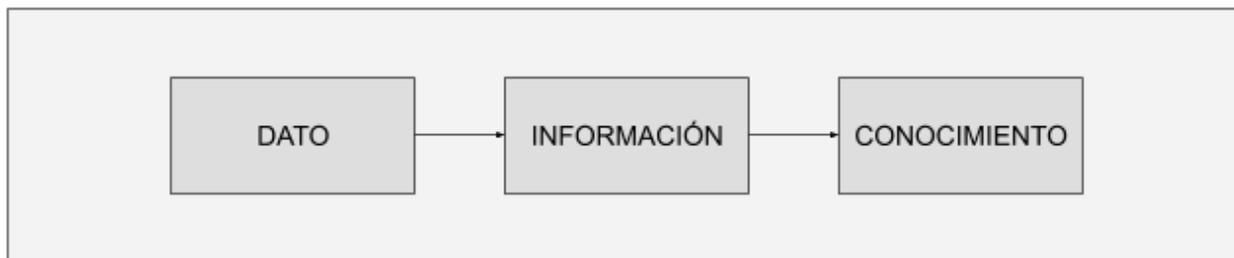


Imagen 4: Propuesta de esquema para la valorización del dato como elemento para la toma de decisiones basadas en conocimiento.

Teniendo en cuenta esto, la gestión se encarga de diseñar e implementar sistemas con el objetivo de identificar, capturar y compartir el conocimiento de una organización, de forma tal que pueda ser convertido en valor para la organización, logrando objetivos perseguidos por la propia compañía.

2.1.2 Inteligencia de negocios

Se denomina inteligencia de negocios, inteligencia comercial o BI (del inglés business intelligence), al conjunto de estrategias, aplicaciones, datos, productos, tecnologías y técnicas, los cuales están enfocadas a la administración y creación de conocimiento mediante el análisis de los datos existentes en una organización [6].

En la actualidad está estrechamente relacionada con el área de ciencia de datos y la ingeniería informática dado los procesos y métodos de recopilación, almacenamiento y análisis de datos de operaciones o actividades comerciales en empresas ya sean tecnológicas o tradicionales con el fin final de optimizar el rendimiento de procesos de negocio apoyando la toma de decisiones generando informes, alertas, cuadros de mando y sugerencias pre definidas o diseñadas mediante IA.

2.1.3 Software como un servicio

La capacidad proporcionada al consumidor de utilizar las aplicaciones del proveedor que se ejecutan en una infraestructura en la nube. Se puede acceder a las aplicaciones desde varios dispositivos cliente a través de una interfaz de cliente ligero, como un navegador web (por ejemplo, correo electrónico basado en web) o una interfaz de programa. El consumidor no administra ni controla la infraestructura de la nube subyacente, incluida la red, los servidores, los sistemas operativos, el almacenamiento o incluso las capacidades de las aplicaciones individuales, con la posible excepción de la configuración limitada de la aplicación específica del usuario. [7].

2.1.4 Métricas de negocio en software como un servicio

Las métricas de software como un servicio son herramientas que se diseñan a partir de la información financiera de la empresa. Son necesarios para medir la estabilidad, la capacidad de endeudamiento, la capacidad de generar liquidez, los rendimientos y los beneficios de la entidad. Las métricas presentadas en este trabajo están centradas en el tiempo de estadía del cliente dentro del software y los ingresos recurrentes [8].

2.1.4.1 Métrica de ingresos recurrentes

Los ingresos recurrentes o RR (por sus siglas en inglés) son una métrica de los ingresos generados por las cuentas de clientes suscritas al software como servicio en un determinado espacio de tiempo, describe una proporción que puede ser en espacios de tiempo mensuales (MRR) o anuales (ARR).

2.1.4.2 Métrica de la tasa de abandono

La tasa de abandono o CHURN (por el sustantivo en inglés) es una métrica que representa el porcentaje de clientes que abandonan las suscripciones del software como un servicio en un determinado espacio de tiempo, habitualmente en meses (MONTHLY CHURN RATE) y años (ANNUAL CHURN RATE).

2.1.4.3 Métrica para costos adquisición de clientes

La métrica de costo de adquisición de un cliente o CAC (por sus siglas en inglés) expresa el valor bruto promedio pagado para obtener un nuevo suscriptor. Se obtiene por la sumatoria de pasivos en un determinado tiempo .

2.1.4.4 Métrica sobre el ciclo de vida del cliente

La métrica del ciclo de vida del cliente o LTV (por sus siglas en inglés) es la medición de los ingresos brutos promedio que generan un clientes participando con la suscripción a la empresa de software como un servicio. Utilizado normalmente para identificar el crecimiento futuro, ya que es construido a partir ARR y CHURN.

2.2 Contexto del mercado

Cuando se trata del mercado, impulsar el crecimiento es un desafío para todas las empresas. A corto plazo, el crecimiento del software como un servicio aumenta con la adquisición de clientes, pero a largo plazo la rotación se activa y domina incluso la estrategia de crecimiento más agresiva crea un techo de crecimiento que puede ser increíblemente difícil de superar.

Superar la rotación y romper el techo de crecimiento requiere un enfoque implacable en el crecimiento. Para EasyCargo el problema radica desde la base, entendiendo que rara vez se analizaron los datos de log o nunca se estimó el costo por cliente. Por lo que no se ha desarrollado alguna estrategia basada en los datos. Pero en el mercado, hay una amplia variedad de servicios basados en la experiencia de los expertos acompañados de plataformas que presentan cuadros de mando para apoyar la toma de decisiones con el objetivo de cerrar los puntos de fuga. Expondremos los principales.

2.2.1 Soluciones para disminuir la deserción.

Amplitude por ejemplo, enfocado en conocer las prácticas de los clientes con un producto y extraer la información para mitigar la fuga de clientes, así mismo mixpanel, identificando tendencias, de por qué hay clientes fieles y otros volátiles. Por otra parte, bajo el perfil de no limitar a solo generar con los clientes pagos recurrentes, sino que a hacer ingresos por suscripción, existe Chargebee ofreciendo portales de pago de distintas fuentes.

Observamos que las principales soluciones enfocadas a disminuir la deserción están enfocadas como primera meta el disminuir los usuarios que abandonan en etapas tempranas del embudo de adquisición, siendo estos de los más caros al momento de perder.

2.2.2 Soluciones de visualización de métricas

Es preciso visualizar periódicamente la salud del negocio y constantemente verificar mediante número e indicadores y en tiempo real el resultado de las acciones y decisiones. Por lo que en el mercado encontramos soluciones basadas en software como servicio para el monitoreo de las principales métricas de la empresa. Un caso es Revenue Story, completa en métricas y de sencilla visualización, muy similar y del mismo modo y con la intención de empoderar a las empresas al proporcionar una ventana en tiempo real a su empresa existe Klipfolio, robusta y basada en el manifiesto SMART para la selección de métricas. Finalmente, para un nivel avanzado Geckoboard, la solución más robusta [9].Cada una de estas soluciones están basadas en las ideas David Skok de Matrix Partners, Christoph Janz de Point Nine Capital, Joel York de Kaotic Flow.

2.3 Tecnologías asociadas

De la misma forma que las empresas ofrecen la posibilidad de generar soluciones para disminuir las pérdidas de clientes, visualizar el comportamiento de los usuarios y mejorar los indicadores para la empresa. Hay disponibles tecnologías que apoyan los procesos de inteligencia de negocio para desarrollar en base al conocimiento disponible soluciones para el almacenamiento, distribución, análisis y visualización de datos. Así como recursos para despliegue de la información y toma de decisión.

Como marco de referencia a las tecnologías de código abierto para el uso mencionado existe el ELK Stack, que son proyectos desarrollados en Java: Elasticsearch, Logstash y Kibana. Elasticsearch es un motor de búsqueda y analítica. Logstash es un canal de procesamiento de datos del lado del servidor y Kibana, para visualizar los datos en un cuadro de mandos [10].

2.2.1 Uso de Elasticsearch

Nos permite indexar grandes volúmenes de datos, para poder consultarlos posteriormente. Elasticsearch se basa en los documentos JSON para poder realizar esta indexación [11]. Es una base de datos distribuida que escala de manera dinámica. Por lo que a mayor demanda podemos ir creciendo en nodos. Llegando a poder almacenar petabytes de información (no siendo el caso para EasyCargo). Entre sus casos de uso para Elasticsearch está buscador de texto, gestor de Logs, análisis de seguridad y análisis de métricas.

2.2.1 Uso de Logstash

Permite centralizar la recogida de información, normalizar y redistribuirla. Es posible definir canales de proceso permitiendo recoger los eventos de una multitud de fuentes de datos como input (por ejemplo logs, csv, nosql, sql entre otras muy usadas por EasyCargo), los cuales son para su posterior transformación, enriquecimiento y normalización (filtros) y su distribución a diferentes destinos outputs [12].

2.2.1 Uso de Kibana

Kibana ofrece una amplia gama de funciones para presentar gráficamente los datos obtenidos mediante Logstash (y como complemento Beats de elastic) [13]. Permitiendo además centralizar la gestión de los canales de logstash y filtrar y visualizar la información en el dashboard.

Capítulo 3 Metodología

3.1 Metodología Moss

Para la implementación de Business Intelligence en EasyCargo, se utiliza la metodología Moss, la que fue publicada por Larissa Moss en el libro “Business intelligence roadmap” [14] el año 2003 y que tiene como objetivo la puesta en marcha de Business Intelligence en el entornos empresariales. Pese a la antigüedad sigue siendo una metodología vigente en temas como data integration y gestión de datos ágil. expuestos así en su libro Extreme Scoping: An Agile Approach to Enterprise Data Warehousing and Business del año 2013 [15]. La metodología Moss se compone de seis etapas, las cuales se señalan en la tabla. <Anexo 05>.

3.2 Metodología Moss para proyectos Business Intelligence

A continuación se describe brevemente las actividades de la metodología elegida, la cual se divide en 6 etapas:

1. **Justificación:** Introducción al contexto y caso de negocio, justificación de la necesidad del proyecto.
2. **Planeación:** Inspección de la infraestructura de de datos de la empresa y de los aspectos a implementar.
3. **Análisis de Negocio:** definición de los requerimientos, análisis de datos y creación del prototipo.
4. **Diseño:** Diseño de la base de datos y de procesos ETL que sustentan la implementación de Business Intelligence.
5. **Construcción:** Desarrollo o implementación de la aplicación de Business Intelligence, del ETL y la minería de datos.
6. **Despliegue:** implementación y evaluación del entorno implementado.

3.3.- Plan de trabajo

El plan de trabajo que contempló esta implementación con sus objetivos se señala a continuación:

Tabla 1: Representacion plan de trabajo objetivo especifico 1

Objetivo	Actividad	Periodo
Verificar los principales conceptos, técnicas, tecnologías utilizadas y enfoques metodológicos existentes en los procesos de proyectos de Business Intelligence	Busqueda de literatura repositorio PU y PUCV	1ra semana Mayo
	Selección de información relacionada al Business Intelligence	1ra semana Mayo
	Exploración de metodologías para la implementación de Business Intelligence en una empresa	2da semana Mayo
	Estudiar y seleccionar una metodología para seleccionar	3ra semana Mayo
	Buscar en repositorios de universidades casos de implementaciones similares	4ta semana Mayo y 1ra semana de Junio

Tabla 2: Representacion plan de trabajo objetivo especifico 2

Objetivo	Actividad	Periodo
Analizar los requerimientos para la implementación del sistema	Reunion para conocer en profundidad la necesidad de la empresa	1ra semana de Junio
	Elaboracion de un documento de requerimientos de negocios segun la metodologia seleccionada	2da semana de Junio y 3ra semana de Junio
	Elaborar documento de requerimiento de software	3ra semana de Junio
	Analizar los requerimientos generados con la contraparte del equipo y posibles metricas a usar	4ta semana de Junio y 1ra Semana de Julio
	Generar un documento de casos de uso	4ta semana de Junio y 1ra Semana de Julio
	Bosquejar graficos de metricas necesarias y datos necesarios para ello	3ra semana de Junio y 4ta semana de Junio y 1ra Semana de Julio
	Gestionar con el equipo una muestra de datos	4ta semana de Junio y 1ra Semana de Juli

Tabla 3: Representacion plan de trabajo objetivo especifico 3

Objetivo	Actividad	Periodo
Elaborar base de datos para facturaciones, logs, cuentas de usuario y procesos de negocio de la empresa.	Verificar integridad de los datos	4ta semana de Junio y 1ra Semana de Julio
	Diseñar una base de datos	1ra Semana de Julio
	Poblar base de datos, testear coneccion con JDBC	4ta semana de Junio y 1ra Semana de Julio
	Solicitar datos economicos de rendimiento de la empresa, crear tabla para futuras consultas	4ta semana de Junio

Tabla 4: Representacion plan de trabajo objetivo especifico 4

Objetivo	Actividad	Periodo
Desplegar herramienta de Business Intelligence con base de datos integrada en servidor de prueba.	Puesta en marcha servidor EC2 Amazon	1ra, 2da, 3ra y 4ta semana de Junio
	Instalacion prerequisites de software para el entorno de instalacion	2da, 3ra y 4ta semana de Junio
	Incorporación de base de datos al servidor	4ta semana de junio, 1ra y 2da de Julio
	Instalacion ElastiCsearch, pruebas	2da, 3ra y 4ta semana de Junio y 1ra de Julio
	Instalacion Kibana, Pruebas dashboard ejemplos, y logs disponible	2da semana de Junio
	Instalacion Logstash, pruebas JDBC	2da semana de Julio
	Instalacion Beats, pruebas y lecturas	2da, 3ra y 4ta semana de Julio
	Coneccion con JDBC a postgresql	4ta semana de Julio
	Diseño de cuadro de mandos	2da, 3ra y 4ta semana de Julio y 1ra semana de Agosto
	Presentacion cuadro de mando equipo	4ta semana de Julio y 1ra semana de Agosto
	Iteracion cuadro de mando	1ra semana de Agosto
Mofificaciones finales al cuadro de mando	1ra y 2da semana de Agosto	

A continuación, en el capítulo 4, se describen las actividades realizadas del plan de trabajo:

Capítulo 4 Actividades realizadas

4.1 Justificación

Actualmente la empresa para sus análisis de mercado y de madurez del negocio se respalda de las métricas en la plataforma de Google Analytics, comparando los gastos utilizados en publicidad web y posicionamiento web, versus la cantidad de registros nuevos de este año contra años anteriores, entre otros.

Debido a la disponibilidad de más datos en la empresa, nace la necesidad de implementar alguna herramienta que asista el análisis y la presentación de los datos.

Para la justificación se implementó una plataforma de Business Analytics, que mediante un cuadro de mando entrega visualizaciones sobre el estado del negocio y ciclos de vida de usuarios.

4.2 Planeación

Para implementar el cuadro de mandos basado en la herramienta de Business Analytics se utilizó métricas de costos de adquisición de un cliente, tasa de abandono de clientes, ingresos recurrentes por suscripción y valor de por vida de un cliente. Por lo que para su uso, se elabora una tabla descriptiva para conocer los elementos necesarios y sus ecuaciones.

4.3 Análisis de negocio

4.3.1 Toma de requerimientos

Para la toma de requerimientos en la implementación del Business Intelligence se inicia por el documento de requerimientos de negocio (BRD por sus siglas en inglés) y se complementa con la especificación de requerimientos de software (SRS por sus siglas en inglés).

4.3.1.1 Documento de requerimientos de negocio (BRD)

Para la toma de requerimientos de negocio se realizó una reunión dividida en dos periodos donde se dieron a conocer las necesidades que tenía la empresa en relación a cómo espera que una solución tecnológica pueda apoyar la toma de decisiones y el desarrollo de estrategias para el negocio. Por lo que durante la reunión se establece un objetivo y se identifican cuales son las necesidades de la empresa y de donde se originan estas. Se establecieron los actores involucrados en la solución y se comentaron los supuestos sobre los posibles fallos a encontrar en la implementación. Finalmente se ha diseñado un formulario que permitió la captura de requerimientos.

4.3.1.2 Documento de especificaciones de requisitos de software (SRS)

Una vez que se ha identificado "por qué" que se ha llevado a cabo mediante la creación del BRD. Se inicia el documentar "Qué" requisitos deben cumplirse para satisfacer las necesidades de la empresa .

4.3.1.2.1 Requerimientos funcionales

Para la elaboración de los requerimientos de software se procedió a fijar una fecha con los involucrados en la empresa. Se revisaron nuevamente los requerimientos de negocio y se hizo una lluvia de ideas a lo que esperamos que se pueda hacer una plataforma de business intelligence. Esto dio como resultado un formulario que especifica los requerimientos funcionales de la implementación, captados en el formulario de requerimientos funcionales que contiene una descripción y un código para el diagrama UML de caso de uso.

4.3.1.2.2 Requerimientos no funcionales

Durante la reunión y basado en los requerimientos de negocio, el equipo nuevamente establece que uno de los peores casos es la pérdida de datos. Por lo que se identificaron a raíz de las conversaciones requerimientos no funcionales en torno a la seguridad, usabilidad. Lo capturaron en un formulario de toma de requerimientos.

4.3.2 Casos de uso

Tras la revisión de los documentos de requerimientos se procedió con el diseño de un diagrama UML para los casos de uso con la implementación. Para esto se transformaron los requerimientos en actividades realizables mediante interacciones usuario plataforma. Esta actividad dio como resultado un diagrama UML.

4.4 Diseño

Como se mencionó el uso de la herramienta Logstash sera para indezar los datos de los archivos. Esta herramienta del lado del servidor permite recoger datos de múltiples fuentes simultáneamente, transformarlos y almacenarlos. EasyCargo utilizará las entradas de los archivos de registro generados por los diferentes componentes de la plataforma de servicios. Los códecs se utilizarán para convertir los diferentes formatos de las entradas en otro aceptado por Logstash, para unificar el formato de los registros y tener la posibilidad de leer formatos distintos del texto simple. Además, los filtros nos permiten procesar diferentes eventos.

En este proceso se verificaron la integridad de los datos, se crearon tablas para ser pobladas en Posgrestsq, se poblaron las tablas y se testeó la conexión con JDBC, además para el trabajo futuro se solicitaron datos económicos de rendimiento de la empresa, crear tabla para futuras consultas.

4.4.1 Datos disponibles

A continuación se presentan extractos de los datos conseguidos para las pruebas de conexión de datos y poblamiento de tablas.

Tabla 5: Datos de compañías registradas

DateRegistered	CompanyNumber	IPLocCountry	IPLocCode	Company	IsParCode
2016-01-02 04:26:20	446615	United States	US	1	False

Tabla 6: Datos para los usuarios registrados

CompanyNumber	First Name	Last Name	Email	Roles Text	Created	Date Last Login	Language	Measure System	Partner Code	Status
602247	Rodrigo	martinez	Romartinez@hdao.com	CompanyAdmin	19/03/2019:55	13/08/2023:25	Spanish	Metric	NULL	Active

Tabla 7: Datos para las cuentas pagadas

DateCreated	Company Number	IPLocCountry	IPLocCode	LicenseType	Count	PaymentMethod	PartnerCode
2016-01-15 10:33:21	602247	Slovenia	SI	2	1	2	NULL

4.4.2 Diseño de la base de datos

Basado en los datos, disponibles y la información con respecto al uso que se le puede dar a las variables para formar las métricas se establece una base de datos relacional de tres tablas con la llave de CompanyNumber. Como resultado de este procedimiento se realiza un diagrama ERD de las tablas.

3.4.5 Gráficos de visualización

Para las conclusiones de la toma de requerimientos de negocio y funcionales se establece que la graficación de métricas como un cuadro de mandos es de rigor para el éxito del proyecto, ya que además en base a estos se pueden seguir construyendo los demás análisis que darán pie para la creación de nuevas métricas y planes de mejoramiento del negocio. Por lo que para EasyCargo se presenta posibles formas de visualización de las métricas para el equipo.

3.4.5.1 Gráficas para RR

La presente imagen describe visualmente lo que se espera ver en el cuadro de mando integral para la métrica de ingresos recurrentes.

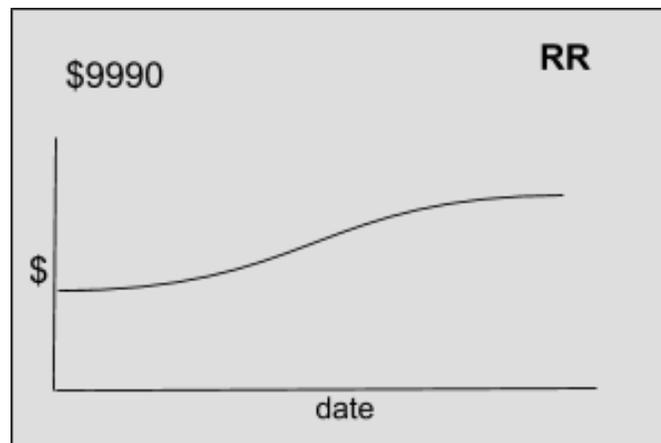


Imagen 5: Gráfico para representar métrica RR

3.4.5.2 Gráficas para CAC

La presente imagen describe visualmente lo que se espera ver en el cuadro de mando integral para la métrica de costos de adquisición de un cliente.

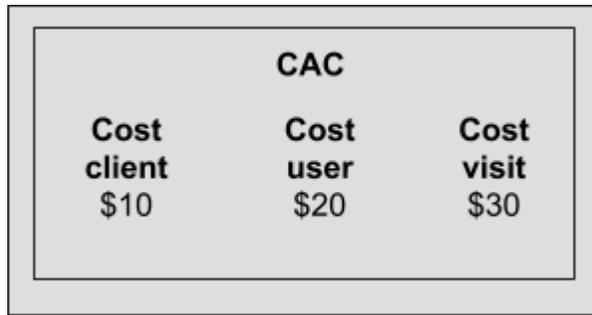


Imagen 6: Grafico para representar metrica CAC

3.4.5.3 Gráficas para CHURN

La presente imagen describe visualmente lo que se espera ver en el cuadro de mando integral para la métrica del porcentaje de deserción de los clientes.

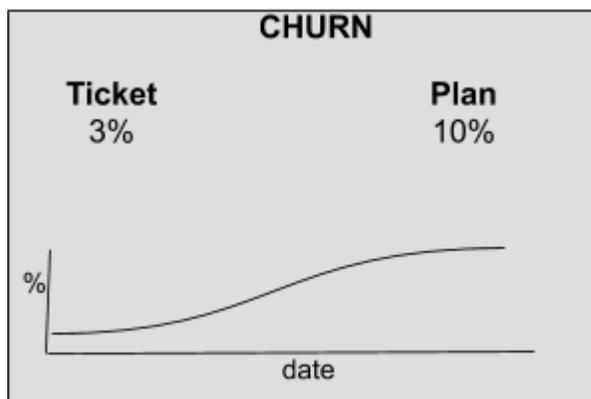


Imagen 7: Grafico para representar metrica CHURN

4.5 Tecnologia

Para EasyCargo darse la posibilidad de construir un software a medida es imposible en este momento, ya que demanda tiempo y mantenerlo cuesta mucho dinero, y los esfuerzos de desarrollo están abocados a el software como servicio en sí. Por otro lado, si se hiciese, el ingeniero que desarrolló el software podría migrar de empresa y sería más costoso aún su mantenimiento independiente de la documentación y las convenciones de código.

Por lo que para EasyCargo se presentan soluciones de mercado robustas en su usabilidad y mantenidas por comunidades y con versiones de pago para uso empresarial. Por lo que para desarrollar la sección 3.5 se presentó la solución a implementar en la compañía.

4.5.1 Solucion a implementar

4.5.1.1 ELK Stack

ELK es la combinación de Elasticsearch, Logstash y Kibana. Se utiliza para proporcionar un enfoque integral en la consolidación, gestión y análisis de registros de los datos. Simplifica la búsqueda e indexación de datos y su análisis y visualización. Proporcionar información en tiempo real a partir de los datos de registro para crear cuadros de mando robustos e intuitivos y editables.

4.5.1.1.1 Elasticsearch

Es el núcleo para almacenar, buscar, procesar e indexar datos. Su característica distintiva es que está distribuida, lo que permite su crecimiento horizontal al almacenar grandes cantidades de datos no estructurados.

4.5.1.1.2 Logstash

Logstash permite llevar datos desde distintos orígenes a Elasticsearch. Ya sea SAP, Oracle, SQL Server, MySQL, archivos de logs, XML, CSV, API-REST, entre otros. Logstash funciona como un ETL para extraer, transformar y editar, pudiendo buscar, analizar y visualizar en tiempo real datos.

4.5.1.1.3 Kibana

Tiene dos funciones concretas, por un lado, es la interfaz gráfica que permite administrar todas las herramientas del stack y por otro, es la herramienta de visualización para la creación de tableros de control, mapas, informes, entre otros para visualizar información.

Mantiene también la gestión de herramientas adicionales al Stack como los módulos de APM, Machine Learning, Canvas, Logging, Infraestructure y muchos otros que permiten ver la información de forma más atractiva para casos de uso específicos.

4.5.1.1.4 Beats

Beats es un complemento para la ingesta de datos, este permite ingresar disponibilidad y métricas de desempeño de servidores, datos de servicios cloud como Lambda, información del tráfico de red de un Sistema Operativo, logs, datos del AWS o Azure, desempeño de Apache, SQL.

4.6 Despliegue e implementacion

El prototipo para EasyCargo con el Stack de herramientas de Elastic es puesto en marcha en una instancia EC2 de Amazon AWS.

4.6.1 Entorno de instalación

Se instalará Elastic Stack en un servidor Ubuntu 18.04. Además, dado que Kibana normalmente solo está disponible en el localhost, se utilizará Nginx como proxy para que sea accesible a través de un navegador.

4.6.1.1 Datos técnicos entorno de instalación

Servidor Ubuntu 18.04 configurado para un usuario no root con privilegios sudo y un firewall configurado con ufw. El VPS cuenta con Ubuntu 18.04, 8GB RAM, 2 CPU.

4.6.1.2 Prerrequisitos de software en el entorno de instalación

Para Elasticsearch y Logstash, es necesario Java 8. Además un servidor web como Nginx.

4.6.1.3 Prerrequisitos de seguridad en el entorno de instalación

Instalar un certificado TLS / SSL.

4.6.2 Instalar Elasticsearch

Para comenzar mediante SSH se hizo la conexión a AWS. Luego en 1| se agrega la KEY de Elasticsearch en APT. Después en 2|, se agrega la lista de estables para el paquete 7 de Elastic. Para finalizar, en 3| se actualiza la lista de paquetes, para que en 4| instalar Elasticsearch. Y para agregar seguridad al entorno se restringe el acceso externo a los puertos para Elasticsearch en 5| . Por lo que modificando el archivo .yml en el atributo “network.host”, se agrega el valor localhost, quitando los comentarios del documento.

Tabla 8: Comando en promp para instalar ElasticSearch

```
1| wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
2| echo "deb https://artifacts.elastic.co/packages/7.x/apt stable main" | sudo tee -a
/etc/apt/sources.list.d/elastic-7.x.list
3| sudo apt update
4| sudo apt install elasticsearch
5| nano /etc/elasticsearch/elasticsearch.yml
```

En este punto ya es posible acceder a Elasticsearch, por lo que es posible hacer las consultas mediante consola o navegador. Por lo que podemos iniciar elasticsearch mediante el administrador systemctl, como se puede ver en |6. Por otra parte y para agregar usabilidad se activa la autocarga para el servidor en |7. Luego en 8| una consulta para el servicio de Elasticsearch.

4.6.3 Instalar Kibana

Para instalar Kibana es necesario previamente la instalación de Elasticsearch, ya que componentes de configuración de puertos son compartidos. Del mismo modo, y mediante el Bash se debe ejecutar lo siguiente (recordando que ya fueron agregados los repositorios en el

punto 3.8.1): 1| para instalar kibana. Luego en 2| mediante el administrador systemctl autocargar kibana en el servidor y iniciar en 3|.

Tabla 9: Comando en promp para instalar Kibana

```
1| sudo apt install logstash
2| sudo systemctl start logstash
3| sudo systemctl enable logstash
```

Como se comentó al inicio, se debe previamente instalar Nginx (aunque para este caso de Ubuntu 18 esta como default) con el propósito de utilizar kibana fuera de la escucha para el localhost. Esto se debe realizar mediante openssl, creando un usuario y su contraseña para el acceso a la interfaz web. Luego es imperativo para su usabilidad crear un archivo bloque para la dirección donde este coincida con la IP del servidor que podrá dirigir el tráfico HTTP del servidor a Kibana que está abierto para localhost:5601. Luego crear un enlace al documento de bloque. Ahora al reiniciar el servicio de Nginx se podrá acceder a la dirección HTTP.

4.6.4 Instalar Logstash

En el comienzo del documento se habla habitualmente del Stack ELK, siendo realmente aquellas tres aplicaciones las que la componen, pero el equipo de elastic dividió funcionalidades, creando Beats además de Logstash siendo el Elastic Stack desde ahora. Entonces, aunque es posible que Beats envíe datos directamente a la base de datos de Elasticsearch, usar Logstash para procesar los datos es parte de los requisitos ya que permite recopilar datos de diferentes fuentes, transformarlos a un formato común y exportarlos a otra base de datos.

Tabla 10: Comando en promp para instalar Logstash

```
1| sudo apt install logstash
2| sudo systemctl start logstash
3| sudo systemctl enable logstash
```

Logstash es una canalización que toma datos en un extremo, los procesa de una forma y los envía a su destino, esta usa dos elementos obligatorios input y output, y un elemento opcional, filter. Los inputs consumen datos de una fuente, los complementos de filtro procesan los datos y los complementos de salida escriben los datos en un destino.

4.6.5 Instalar Beat

El Elastic Stack utiliza varios remitentes de datos livianos llamados Beats para recopilar datos de diversas fuentes y transportarlos a Logstash o Elasticsearch. Estos son los Beats más útiles para necesidades como las de EasyCargo:

Tabla 11: Posibilidades de extensiones para Beats

Beat	Funcion
Filebeat	Recopila y envía archivos de registro.
Metricbeat	Recopilar métricas de sistemas y servicios.
Packetbeat	Recopila y analiza datos de red.

Para instalar Filebeat, al igual que el resto del stack, se puede usar el apt en Bash usando 1|. Para la configuración de la misma forma que en Elasticsearch en base al archivo .yaml usando el editor default de Ubuntu nano, en 2| para activar o desactivar los puertos para Logstash y Elasticsearch.

Tabla 12: Comando en promp para instalar Beats

```
1| sudo apt install filebeat
2| sudo nano /etc/filebeat/filebeat.yml
```

4.6.6. Conexión con JDBC

Primero, comencemos por configurar el complemento de entrada JDBC y enviarlo a un archivo local para probarlo. Aquí está el archivo de configuración inicial de Logstash.

Tabla 13: Conexión JDBC mediante edición de filebeat.yml

```
input {
  jdbc {
    jdbc_driver_library => "/opt/jdbc/mysql-connector-java-5.1.46-bin.jar"
    jdbc_driver_class => "com.mysql.jdbc.Driver"
    jdbc_connection_string => "jdbc: mysql://DireccionAmazon:3306/EayCargoTest"
    jdbc_user => "EasyCargoTest"
    jdbc_password => "EasyCargo2020."
    statement => "SELECT * FROM users LIMIT 12000"
    lowercase_column_names => true
  }
}

output {
  file {
    path => "/tmp/test.log"
    file_mode => 0644
    codec => "json_lines"
  }
}
```

Para el archivo de salida en /tmp/test.log con 12000 consultas a user:

Tabla 14: Ejemplo de consulta para 12000 usuarios

```
{
  "_index": "dbEasyCargo",
  "_type": "_doc",
  "_id": "UwQ07nMBB_IPrhPiuG2Y",
  "_version": 1,
  "_score": null,
  "_source": {
    "IPLocCountry": "Spain",
    "DateCreated": "2020-05-01 14:26:36",
    "Count": 1,
    "PartnerCode": "NULL",
    "LicenseType": 1,
    "@timestamp": "2020-05-01T14:26:36.000-04:00",
    "CompanyNumber": 775514,
    "IsFirstOrder": "No",
    "PaymentMethod": 8,
    "IPLocCode": "ES"
  },
}
```

```
"fields": {
  "@timestamp": [
    "2020-05-01T18:26:36.000Z"
  ],
  "DateCreated": [
    "2020-05-01T14:26:36.000Z"
  ]
},
"sort": [
  1588357596000
]
}
```

Configuración para la salida de los datos para Elasticsearch:

Tabla 15: Configuración para alimentación de ElastiSearch

```
output {
  elasticsearch {
    id => "testout"
    document_id => "%{user_no}"
    hosts => ["testhost", "respaldotesthost"]
    user => "EasyCargoTest"
    password => "EasyCargo2020."
    ssl => "true"
    index => "dbEasyCargo"
  }
}
```

Capítulo 5 Resultados

El trabajo realizado consistió en implementar una plataforma de Business Intelligence para el análisis de datos del área de desarrollo de negocios en una empresa que ofrece software como un servicio, la cual pretende ayudar al equipo con información para desarrollar estrategias de desarrollo de negocio.

A continuación, se discutirá sobre el cumplimiento de los objetivos específicos planteados previamente en este documento y el nivel logrado en la implementación de dichos objetivos. Además, se hará un análisis en cuanto al nivel de usabilidad alcanzado en la implementación, exponiendo sobre la fuente de datos actuales.

5.1 Resultado objetivo 1

“Verificar los principales conceptos, técnicas, tecnologías utilizadas y enfoques metodológicos existentes en los procesos de proyectos de Business Intelligence”

Para comenzar a implementar la plataforma de business intelligence se comenzó por estudiar los dominios de conocimiento relacionados. Logrando una mirada en macro al Business Intelligence mediante el libro de Larissa Terpeluk Moss Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications [17], rescatando conceptos, técnicas, tecnologías y un plan metodológico para la implementación del Business Intelligence en organizaciones expresado en el capítulo 4.

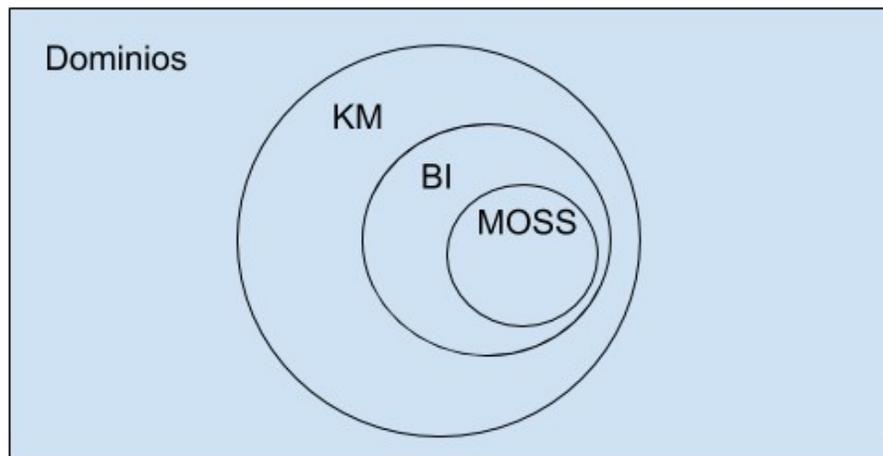


Imagen 8: Dominios identificados para la verificación de información

5.2 Resultado objetivo 2

“Analizar los requerimientos para la implementación del sistema” Desde el punto de vista de requerimientos de negocio se establece que la necesidad principal es de obtener una herramienta que permita tomar decisiones para apoyar el desarrollo de estrategias que permitan maximizar la conversión de usuarios a clientes por suscripción. Lo que da como resultado el documento de requerimientos de negocio BRD.

A continuación se describe la toma de requerimientos de negocio.

Tabla 16: Formulario de capturacion de requerimientos de negocio

Objetivos del proyecto: Implementar una plataforma de business intelligence para el análisis de datos del área de desarrollo de negocios.	
Declaración de necesidades: El fin de la implementación de una plataforma para business intelligence es el análisis de datos para poder comprender cuales son las falencias del negocio en la administración de los usuarios y clientes. La necesidad es crear estrategias que permitan maximizar la transformación de clientes a cuentas de pago, disminuir el radio de usuarios con licencias gratuitas y aumentar el número de clientes con cuentas de suscripción. Poder estimar la fecha de abandono de un cliente y crear estrategias para evitar la deserción de la primera compra y crear un sistema de recompensas para los usuarios fidelizados.	
Alcance del proyecto:	<p>Áreas de la organización que se verán involucradas:</p> <p>Desarrollo de negocios: Equipo que incluye a Jan Sovak, Lukas Broz, Lukas Polame, Diego Villagran.</p> <p>Administrador base de datos: Robert Sojak</p> <p>Procesos del nuevo sistema:</p> <p>Almacenar datos: Ingreso de datos de logs y jdbc.</p> <p>Filtrar información: Filtrar, Analizar texto y estructurarlo, mutar texto, geoposicionar direcciones ip.</p> <p>Desplegar gráficos: Representar datos en gráficos, representar datos en tiempo real, exportación de gráficos como cuadro de mandos.</p>
Supuestos:	Pérdida de datos, pérdida de información, desviación de las métricas, interpretación errónea, prolongación del proyecto.

Por otra parte, para los requisitos no funcionales de software la empresa confía en tener un sistema en el que operar sus datos sea seguro y lo menos invasivo posible con la interacción de otras plataformas. Y como requerimientos funcionales, una plataforma de manejo masiva de datos que permita por una parte la interacción con los datos para análisis y por otro lado la visualización en un cuadro de mandos integral. Requerimientos que dan como resultado los documentos de toma de requerimientos funcionales y no funcionales en el capítulo 4.

A continuación se describen los requerimientos funcionales para la implementación de la herramienta de Business Intelligence: A continuación se

Tabla 17: Formulario que captura requerimientos funcionales

COD	Requerimiento funcional	Descripción	Caso de uso
RF 1.1	El sistema deberá posibilitar la selección arbitraria de información.	Que el usuario pueda elegir información que desea visualizar en el cuadro de mandos.	CU - 01
RF 2.1	El sistema deberá posibilitar el cambio de parámetros.	El usuario pueda cambiar los parámetros de intervalo de fechas.	CU - 02
RF 3.1	El sistema deberá posibilitar la elección de varios parámetros a la vez.	El usuario puede agregar distintos gráficos para el cuadro de mandos.	CU - 03
RF 4.1	El sistema deberá posibilitar usar gráficos para visualización de información.	Las visualizaciones deben poseer gráficos de de disco, torta, barras, etc para graficar ARR, MRR, CHURN.	CU - 04
RF 5.1	El sistema deberá posibilitar la extracción datos	Independiente la fuente de entrada estos deben poder ser agregados	CU - 05
RF 6.1	El sistema deberá posibilitar el formatear datos	Los datos ingresados pueden ser cambiados de formato y editados.	CU - 06
RF 7.1	El sistema deberá posibilitar la distribuir y almacenar datos	Los datos deben ser almacenados de forma independiente y organizada en nodos.	CU - 07
RF 8.1	El sistema deberá posibilitar la distribución de datos para análisis	Posibilidad de conexión de los nodos para hacer análisis de datos mediante ML o DL	CU - 08

En la siguiente tabla se expresa los requerimientos no funcionales obtenidos mediante las reuniones con los involucrados.

Tabla 18: Formulario que captura requerimientos no funcionales

Requerimiento no funcional	Descripción
RNF 1.1	Usabilidad: Se requiere de una plataforma de Business Intelligence de fácil uso, con gráficos comprensibles para un usuario principiantes en desarrollo de negocios.
RNF 2.1	Compatibilidad: Se requiere que la plataforma a implementar sea compatible con los sistemas ya implementados en la empresa y que requiera el menor esfuerzo de integración.
RNF 3.1	Seguridad: Se requiere que la plataforma a implementar sea segura, en relación a los datos que maneja, ya que, son datos sensibles para el usuario y empresa. Estos poseen gran valor comercial por lo que se deben mantener fuera del alcance de terceros.

Como resultado además se incluye el diagrama para los casos de uso en el anexo Anexo 02.

5.3 Resultado objetivo 3

“Elaborar base de datos para facturaciones, logs, cuentas de usuario y procesos de negocio de la empresa” Considerando las dos etapas anteriores y ya teniendo la herramienta seleccionada mediante la metodología Moss para el despliegue de Business Intelligence se solicitan a la empresa muestras de datos relativos a los requerimientos de métricas de visualización expuestas en el capítulo cuatro. Esto da como resultado a partir de los datos de las tablas UserReg, OrdersPaid, RegCompaniesCountry y el diseño de base de datos el capítulo 4.

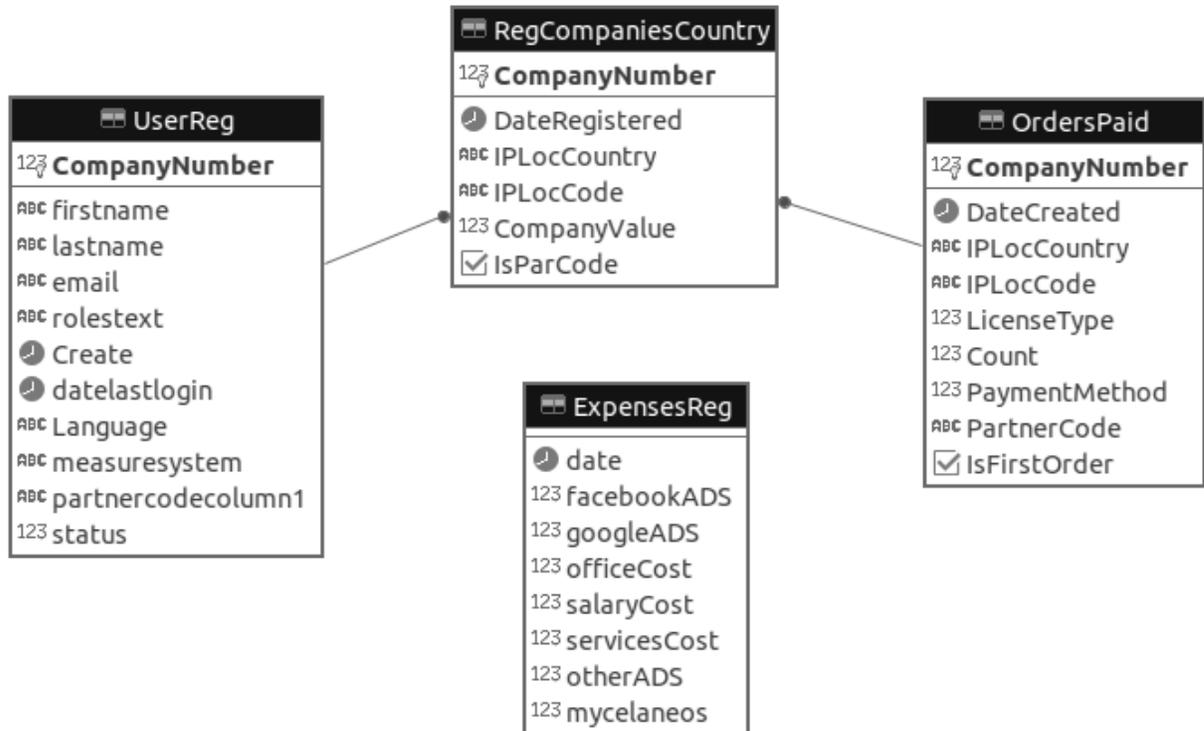


Imagen 9: Resultado diseño para los datos disponibles

5.4 Resultado objetivo 4

“Desplegar herramienta de Business Intelligence con base de datos integrada en servidor de prueba” Finalmente con los datos disponibles, se procedió a la implementación, la que se realizó en un servidor de pruebas como se expone en el despliegue del capítulo 4, el que mediante Logstash y JDBC se agregan los datos a ElasticSearch para poder ser indexados y consultados por Kibana. Como resultado se obtiene un cuadro de mandos integral que refleja las métricas solicitadas. Esto dio origen al lanzamiento en el servidor de la interfaz de kibana que se presenta a continuacion:

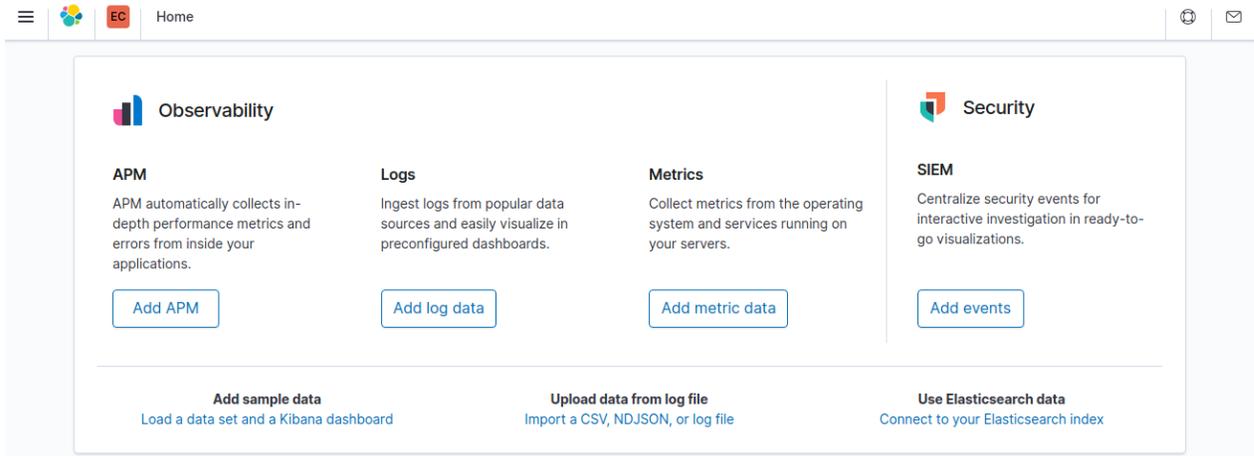


Imagen 10: Resultado despliegue y testeo de Kibana

Tras la conexión de Logstash con ElasticSearch, es posible verificar los datos ingresados, en este caso la totalidad de ellos en el rango año 2016

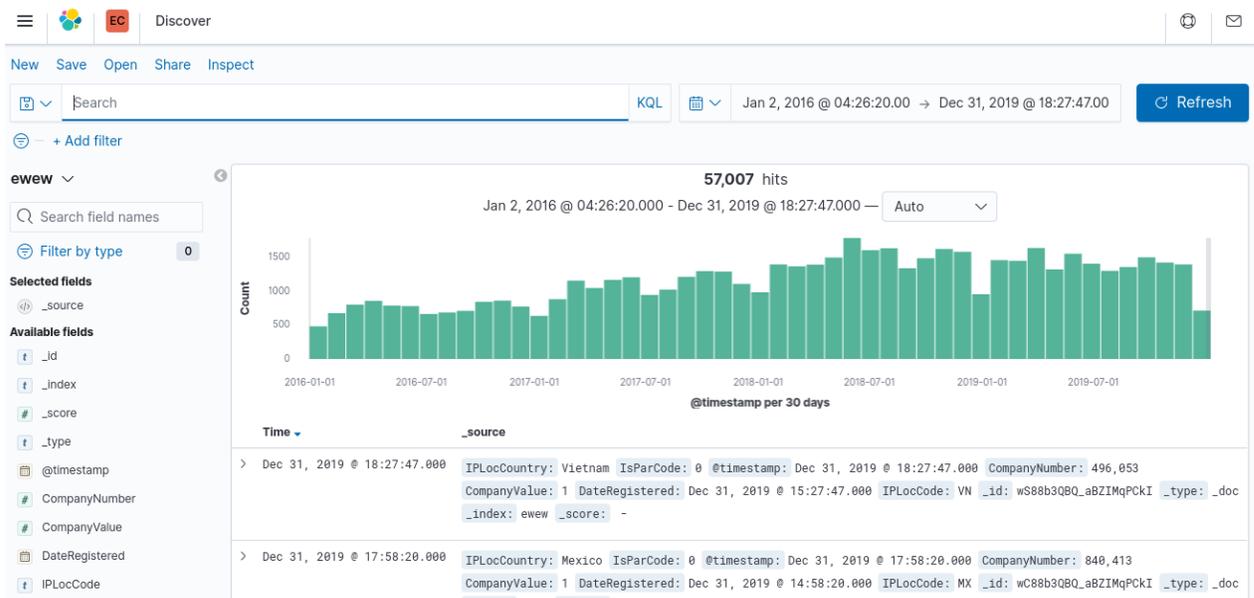


Imagen 11: Verificación y pruebas de los datos para visualización

Ademas, la posibilidad de verificar la disposicion de los datos y descargar muestras de ellos. Para verificar una fila como se ve a continuacion:

Field	Value
@timestamp	Dec 31, 2019 @ 13:53:46.000
# CompanyNumber	809,828
# CompanyValue	1
DateRegistered	Dec 31, 2019 @ 10:53:46.000
f IPLocCode	ES
f IPLocCountry	Spain
f IsParCode	0
f _id	v188b3QBQ_aBZIMqPckI
f _index	ewew
# _score	-
f _type	_doc

Imagen 12: Consulta a una fila o elemento en la tabla

Finalmente, verificacion de los datos creado visualizaciones rapida para con un contador simple. Mostrando la predominancia de usuarios de Vietnam y Estados Unidos de America como se ve a continuacion.

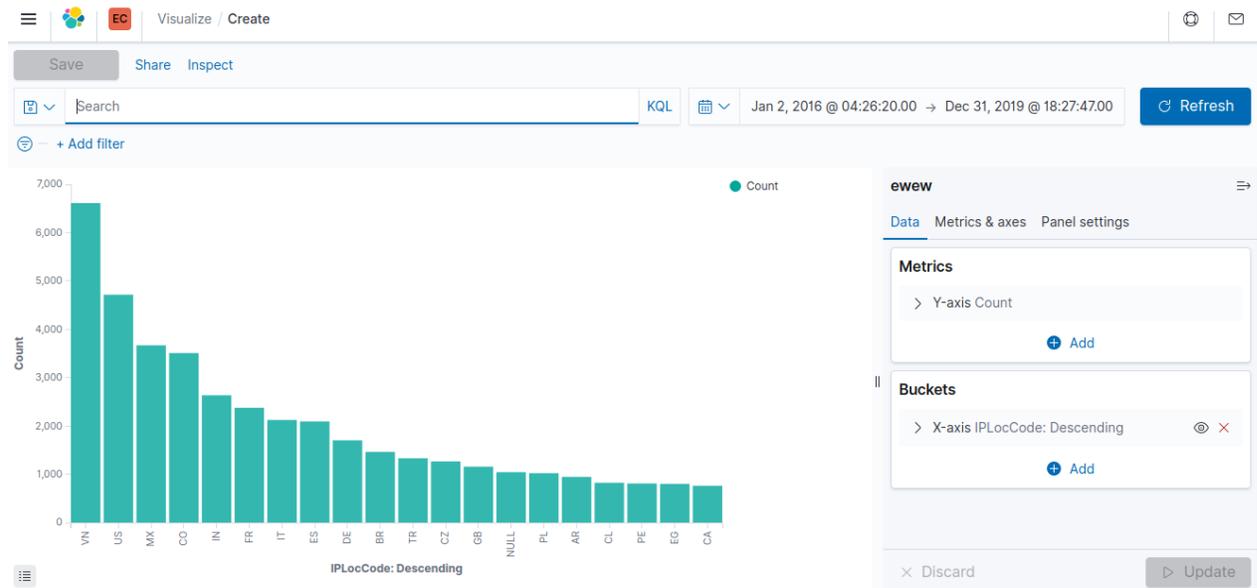


Imagen 13: Graficacion de contador para distribucion de ip para usuarios

En este punto se inicio el proceso de pruebas para desarrollar un cuadro de mandos que mediante los datos mostrar las metricas seleccionadas como relevantes. Las imagenes a continuacion exponen parte del proceso. Donde a continuacion presentamos el espacio de trabajo para EasyCargo donde se encuentran los datos consultados en las anteriores imagenes.

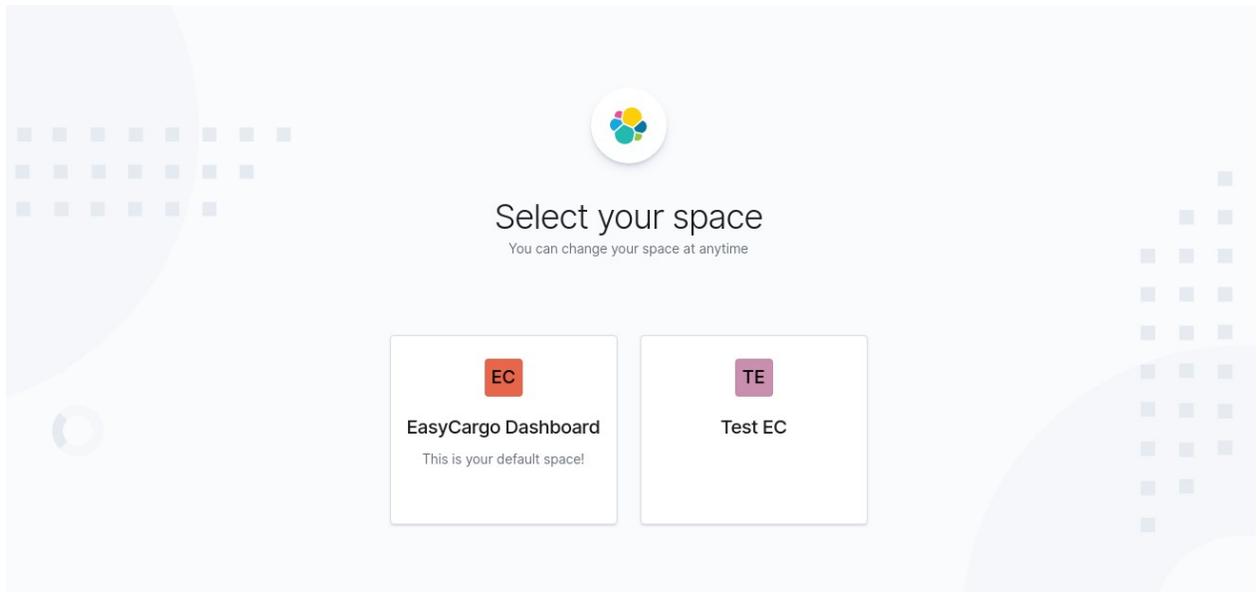


Imagen 14: Selector de espacio para trabajar en Kibana

A continuación se presenta la iteración que fue presentada al equipo de EasyCargo, este cuadro de mandos representa a un costado las variables medidas por la empresa y las nuevas propuestas en el trabajo.

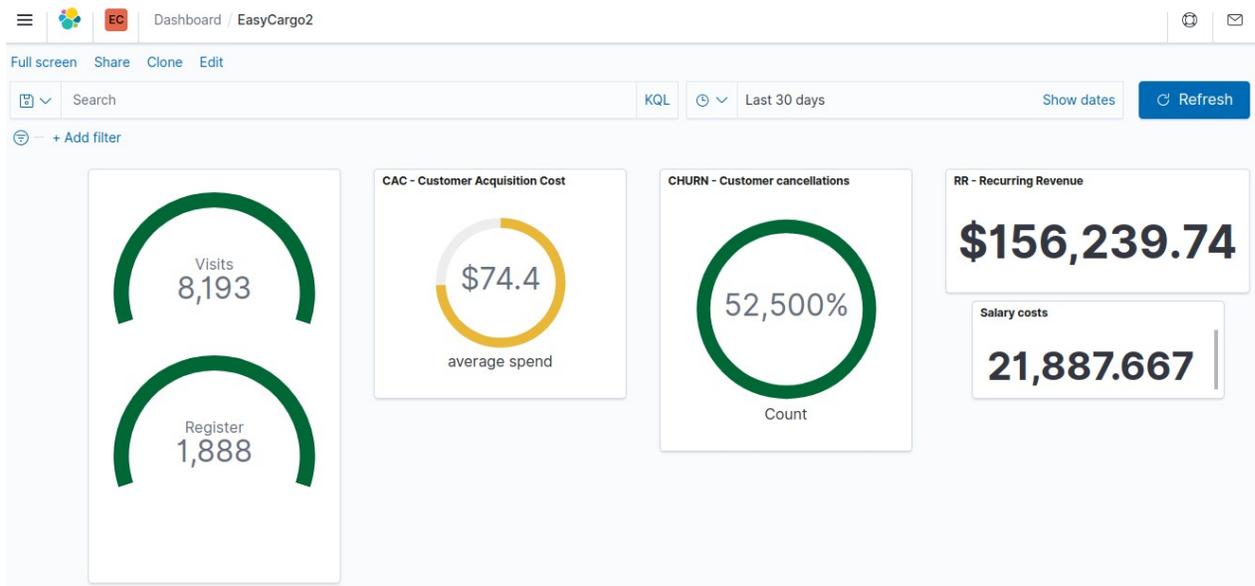


Imagen 15: Resultado cuadro de mandos primeras iteraciones

Si bien, los graficos no cumplen con el diseño del capítulo cuatro esto se debe a que esta representación es acorde a un cuadro de mandos. Y justifica las anteriores diseñadas para un posible reporte.



Imagen 16: Métricas seleccionadas para el rendimiento

5.5 Comentarios

Finalizado el proceso de despliegue, la ronda de pruebas y la exploración de las funcionalidades de la implementación del stack de elastic, se comparte con el equipo los resultados y se expuso el funcionamiento junto a una prueba de las posibilidades. Esta exposición al equipo finalizó con una lluvia de ideas espontánea sobre las cosas que no se están haciendo en la empresa y sobre lo que no se está tomando en cuenta, esto genera la siguiente tabla de reacciones.

- Se debe medir el costo por visitante, usuario y cliente.
- Se debe medir el costo por cliente de tipo ticket, suscripción menor a un año y mayor a un año.
- Se debe medir la tasa de abandono de los usuarios de ticket, mensual y anual.
- Se deben almacenar los costos fijos y variables de la empresa.
- Se debe medir la tasa de renovación de mensual a anual, anual a mensual, mensual a mensual, anual a anual y la conversión de ticket a mensual o anual.
- Se deben medir los procesos de tiempo de conversión de usuario a clientes.
- Se debe estimar el costo de uso de un usuario
- Se debe medir la cantidad de clientes en fase de prueba, contacto y reprobación
- Se debe implementar una estrategia para cada métrica y deben descubrir para ella indicador de medición.

Capítulo 6 Conclusión

En primera instancia, el inicio de la investigación de los términos relacionados con BI permitió visualizar todo el contexto de lo que significa hoy en día y que su uso no sólo está ligado a la empresas tecnológicas sino que también el mercado de minería, retail, construcción y empresas sin fines de lucro, por lo que su uso representa muchos beneficios, y no sólo competitivos, sino que de rendimiento, alcance e incluso de sororidad.

6.1 Discusión

Por otra parte, la implementación de esta plataforma para la empresa establece un llamado al equipo a desarrollar estrategias que permitan mantener a los clientes suscritos, guiarlos a la renovación de las suscripciones y a recuperar aquellos en los que ya se invirtió.

Estos pequeños reportes generados a partir de los datos disponibles expusieron ante todo la fragilidad de la empresa y la dependencia de las suscripción anual para su funcionamiento, destacando que el 48% de los clientes nunca vuelven a renovar y los ratios entre los costos de los distintos estados del cliente que son de 4:4:4 para los estados de visitante, usuario y cliente.

Con respecto a la masividad de los datos, para este caso en la empresa, la cuota era muy baja, sobretodo considerando las tablas ocupadas para los cálculos expuestos en el desarrollo. Esto sin considerar las posibilidades a futuro con la empresa ya que en primera instancia para la implementación del Elasti Stack se planea la implementación de lecturas de Logs y la extracción de datos mediante ETL.

Además considerar para esta etapa la implementación de un perfil de machine learning para el estudio de los datos es necesario, asimismo que por ahora es hacer cluster con los datos. Pero dada las cualidades del entorno implementado, para la empresa esta puede ser una alternativa alcanzable y fácilmente implementable.

Para cerrar el ciclo de esta implementación en base a las métricas medidas nace la necesidad de identificar cotas para cada indicador y en base a esto planificar soluciones que permitan probar cambios y verificar su funcionamiento.

6.2 Conclusión

Con respecto a los objetivos, se evidencia y concluye lo siguiente:

- El objetivo 1 “Verificar los principales conceptos, técnicas, tecnologías utilizadas y enfoques metodológicos existentes en los procesos de proyectos de Business Intelligence” se considera logrado, mediante el estudio de los conceptos relacionados al Business Intelligence, así como su objetivo y técnicas relacionadas, abriendo paso a conocer las especialidades y posibilidades en el mercado laboral. Se evidencia este resultado mediante la implementación de la metodología Moss que guía la puesta en marcha de Business Intelligence en organizaciones que fue estructurado en este trabajo desde el capítulo 4. Por otra parte, sobre este objetivo se estudió también sobre el software como un servicio, sobre cuáles son los driver que dirigen este tipo de negocio y específicamente cuales son los aspectos con los que se puede vincular con el Business Intelligence, proceso que se ve reflejado en el proceso de diseño en el capítulo cuatro donde se exponen según el criterio de Joel York y David Skok métricas que apoyan la toma de decisiones para entornos de software como un servicio.

- Sobre el objetivo 2 de “Analizar los requerimientos para la implementación del sistema”, se consideran completados. Ya que a raíz de las conversaciones con el equipo se inicia el “documento de requerimientos de negocio” con el objetivo de poder conocer los objetivos empresariales y cuales son las necesidades identificadas por el equipo. Ya que independiente de que no existían mediciones anteriores, se sospechaba sobre algunos rendimientos negativos. Basado en esto, se inicia el proceso de toma de requerimientos de software. Dando como resultado los documentos mostrados en el capítulo 4. La formulación de estos requerimientos dio pie para entender cuál era la necesidad en específico para la empresa demostrando que lo que se requería era una herramienta amplia en su usabilidad y robusta en la seguridad de datos. Y que finalmente permitirá estar visualizando constantemente en un cuadro de mando algunas métricas relevantes para el proyecto.
- La base de datos fue diseñada para respaldar los datos suministrados por la empresa, ser conectada mediante el JDBC de Logstash e indexada en el Elasticsearch para desplegarla visualmente con Kibana. Del proceso se obtuvieron aprendizajes con respecto al trabajo en Amazon C2. Basado en esto se considera que el Objetivo 3: “Elaborar base de datos para facturaciones, logs, cuentas de usuario y procesos de negocio de la empresa” está completado.
- Sobre el objetivo 4 de “Desplegar herramienta de Business Intelligence con base de datos integrada en servidor de prueba” se considera completada. El proceso comenzó desde la puesta en marcha de los servicios, para luego explorar el entorno de los datos indexados y probar las visualizaciones mediante los datos, como resultado se generó en primera instancia los graficos segundo las métricas planteadas en este trabajo. Con cada iteración y prueba, finalmente se logran mayores resultados entre el uso de los datos, la generación de cuadros de mando y la consulta de datos a fuentes externas. Como evidencia de esto en los anexos se adjuntan las visualizaciones implementadas para las iteraciones con el equipo.

6.3 Trabajo futuro

Próximamente se realizarán pruebas con la conexión a servidores para la extracción de Logs, se planea identificar los beneficios de esta información y como esta puede respaldar algunas estrategias de precios de venta de la suscripciones por regiones.

Se planean desarrollar procesos ETL con hojas de cálculo para extraer información de costos sobre facturaciones de la empresa, mediante los Beats de Elastic, perfeccionar la cantidad de usuarios que pasan por el embudo de clientes (anexo 06).

Estudiar los casos de uso de Metricbeans de Elastic, para la monitorización. Además se debe estudiar el lenguaje de consulta de datos de Kibana.

Desarrollar e identificar métricas para distintos dominios de la empresa y desarrollar reportes. Medir el beneficio y los costos de usabilidad (anexo 03 y 04).

Estudiar sobre la composición y heurística de un cuadro de mandos debido a la cantidad de información y la presión mental que ejercen. Y sobre la posibilidad de mantener un cuadro de mandos 24/7 para el equipo en oficina y cuáles serían las posibles ventajas y desventajas de estar expuesto a esta información.

Implementar una versión en INTERNET CZ de la red Forpsi en República Checa que sea el punto de generación de control de mando, reportes e indexación de datos.

Nomenclatura

- BI: siglas en inglés de Business Intelligence, comprende las estrategias y tecnologías utilizadas por las empresas para el análisis de datos de información comercial.
- SAAS: siglas en inglés de software as a service. Es un modelo de licencia de software y de entrega en el que el software se licencia por suscripción y está alojado de forma centralizada en cloud computing.
- CLOUD: como referencia a cloud computing en inglés. Es la disponibilidad de brindar recursos de sistemas informáticos, especialmente de almacenamiento de datos, computación e infraestructura sin una gestión activa directa por parte del usuario.
- KM: siglas en inglés de knowledge management. Es el dominio del conocimiento que tiene un enfoque de crear, compartir, utilizar y gestionar el conocimiento y la información de una organización. Se refiere a un enfoque multidisciplinario para lograr los objetivos de la organización haciendo el mejor uso del conocimiento.
- KPI: siglas en inglés de key performance indicators. Se refiere a un indicador de rendimiento para un tipo de medición que evalúan el éxito de una organización o de una actividad concreta.
- ELK STACK: Grupo de aplicaciones presentadas por la firma Elastic que permiten entre muchas cosas la gestión de los datos para el Business Intelligence.
- AWS: siglas en inglés de Amazon Web Services. Son una colección de servicios de computación en CLOUD que en conjunto forman una plataforma de computación ofrecida a través de Internet por Amazon.
- MOSS: Hace referencia a la metodología de Larissa Moss para implementar Business Intelligence en las organizaciones.
- RR: siglas en inglés de recurring revenue. Refiriéndose a los ingresos obtenidos en un periodo de tiempo. Siendo los ingresos recurrentes.
- MRR: siglas en inglés de monthly recurring revenue. Refiriéndose a los ingresos obtenidos en un mes. Siendo los ingresos mensuales recurrentes.
- ARR: siglas en inglés de anual recurring revenue. Refiriéndose a los ingresos obtenidos en un año. Siendo los ingresos anuales recurrentes.
- CHURN: referencia a CHURN RATE con significado en inglés de tasa de abandono. Refiriéndose a la cantidad de suscripciones, usuarios o clientes perdidos en un determinado tiempo.

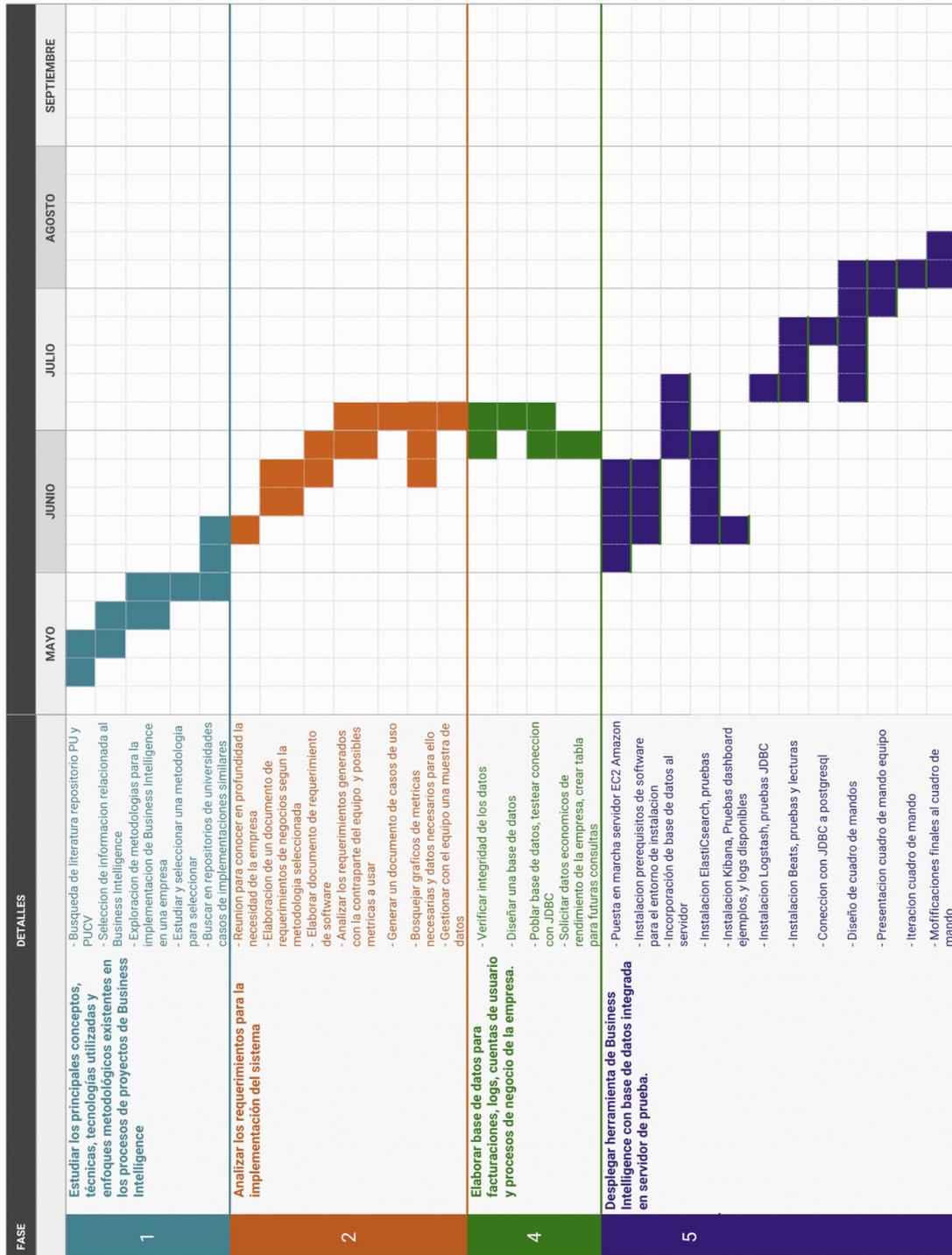
Referencias

- [1] Finance, Y. (2020, September 10). Slack Technologies, Inc. (WORK) Stock Price, News, Quote & History. Retrieved September 10, 2020, from <https://finance.yahoo.com/quote/WORK?p=WORK> . Market Cap (intraday) 4/30/2020 15.00B
- [2] Poyar, K., Fanning, S., Dietz, B., & O'Leary, N. (n.d.). 2019 Expansion SaaS Benchmarks Report. Retrieved September 10, 2020, from <https://openviewpartners.com/expansion-saas-benchmarks/> . OpenView's 2019 report PDF
- [3] Thomson, W. (1889). *Popular Lectures and Addresses*. Place of publication not identified: Publisher not identified. <https://archive.org/stream/popularlecturesa01kelvuoft#page/72/mode/2up>
- [4] Profile, L. (Ed.). (2014). Profile of Bee Informatics. Retrieved June 7, 2020, from <https://ar.linkedin.com/company/bee-interactive-s-r-o->
- [5] SaaS Metrics 2.0 – Detailed Definitions. (2017, June 14). Retrieved July 07, 2020, from <https://www.forentrepreneurs.com/saas-metrics-2-definitions-2/>
- [6] Dedić N. & Stanier C. (2016). Measuring the Success of Changes to Existing Business Intelligence Solutions to Improve Business Intelligence Reporting. *Lecture Notes in Business Information Processing*. Springer International Publishing. Volume 268, pp. 225-236.
- [7] Peter Mell; Timothy Grance (September 2011). The NIST Definition of Cloud Computing (Technical report). National Institute of Standards and Technology: U.S. Department of Commerce. doi:10.6028/NIST.SP.800-145
- [8] York, Joel. (2014). *The Metrics-driven SaaS Business*. 517 York Street San Francisco, CA 94110: Bluenose Analytics.
- [9] Neale, A. (2012). Enhanced Developer Platform Launched By @Geckoboard [Web log review]. Retrieved June 12, 2020, from <http://www.dailydooh.com/archives/73985>
- [10] Sweetman, M. (2020, May 20). Using the ELK stack for log review and visualizations [Web log review]. Retrieved June 7, 2020, from <https://luneba.com/blog/using-the-elk-stack-for-log-review-and-visualizations>
- [11] Co., E. (2020). motor de analítica y análisis distribuido open source Retrieved August 3, 2020, from <https://www.elastic.co/es/what-is/elasticsearch>
- [12]Co., E. (2020). Centraliza, transforma y almacena tus datos Retrieved August 3, 2020, from <https://www.elastic.co/es/logstash>
- [13]Co., E. (2020). interfaz de usuario gratuita y abierta que te permite visualizar los datos de Elasticsearch Retrieved August 3, 2020, from <https://www.elastic.co/es/kibana>
- [14] Moss, L. T., & Atre, S. (2011). *Business intelligence roadmap: The complete project lifecycle for decision-support applications*. Boston: Addison-Wesley.
- [15] Moss, L. T., & Atre, S. (2011). *Business intelligence roadmap: The complete project lifecycle for decision-support applications*. Boston: Addison-Wesley.

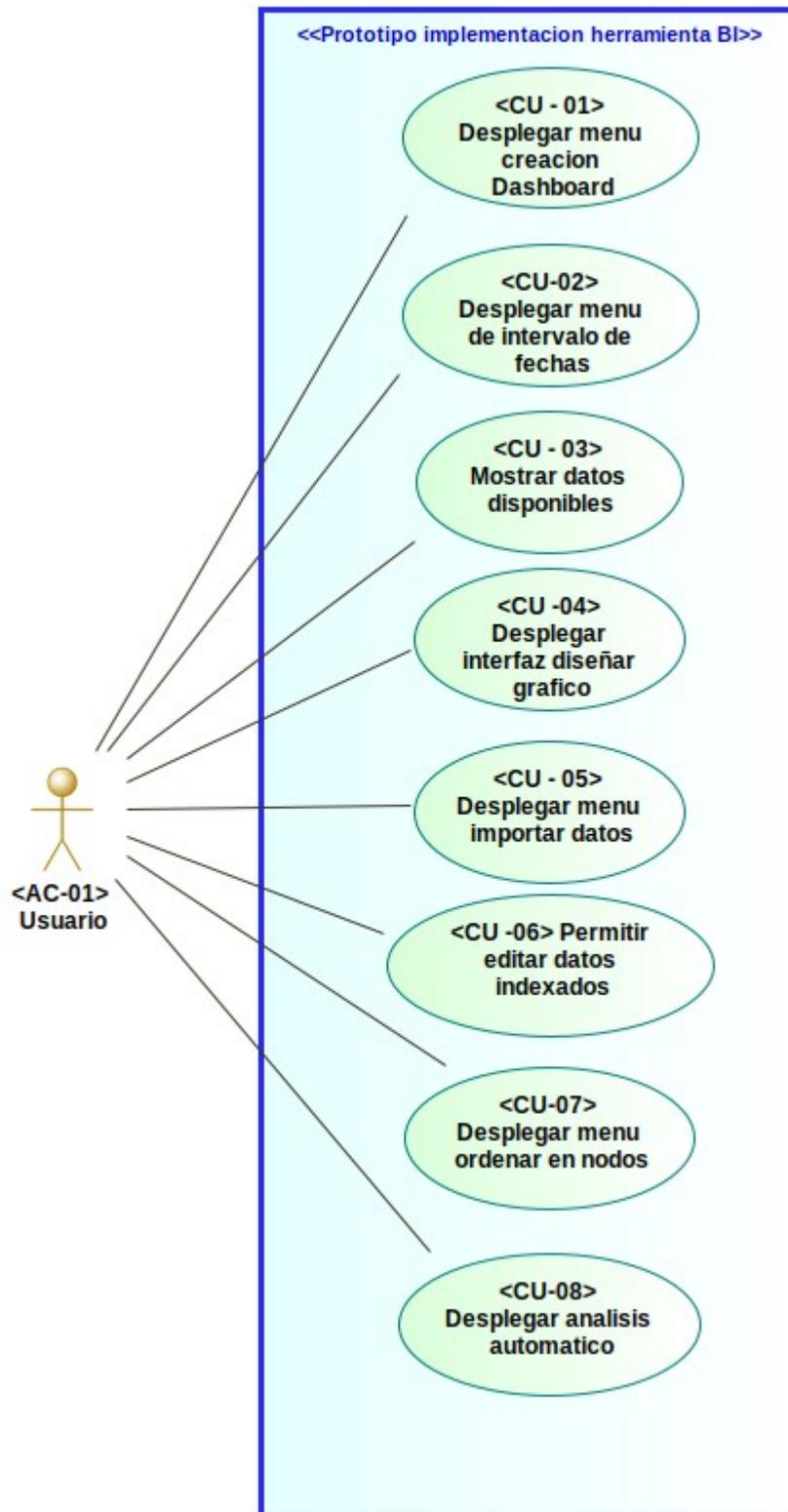
Bibliografia

- Moss, L. T., & Atre, S. (2011). Business intelligence roadmap: The complete project lifecycle for decision-support applications. Boston: Addison-Wesley
- Tiwana, A. (2002). The knowledge management toolkit: practical techniques for building a knowledge management system with cdrom. Prentice Hall PTR.
- Chhajed, S. (2015). Learning ELK stack. Packt Publishing Ltd.
- York, J. (2012). SaaS Sales Models, Strategic and Organizational Choices. Chaotic Flow, 1-14.
- McClure, D. (2007). Startup metrics for pirates: AARRR. Startup Metrics for Product Marketing & Product Management).
- Skok, D. (2016). SaaS metrics 2.0—a guide to measuring and improving what matters. posted at forentrepreneurs. com, Aug, 18.

Anexos



Anexo 01: Carta Gantt para la implementacion



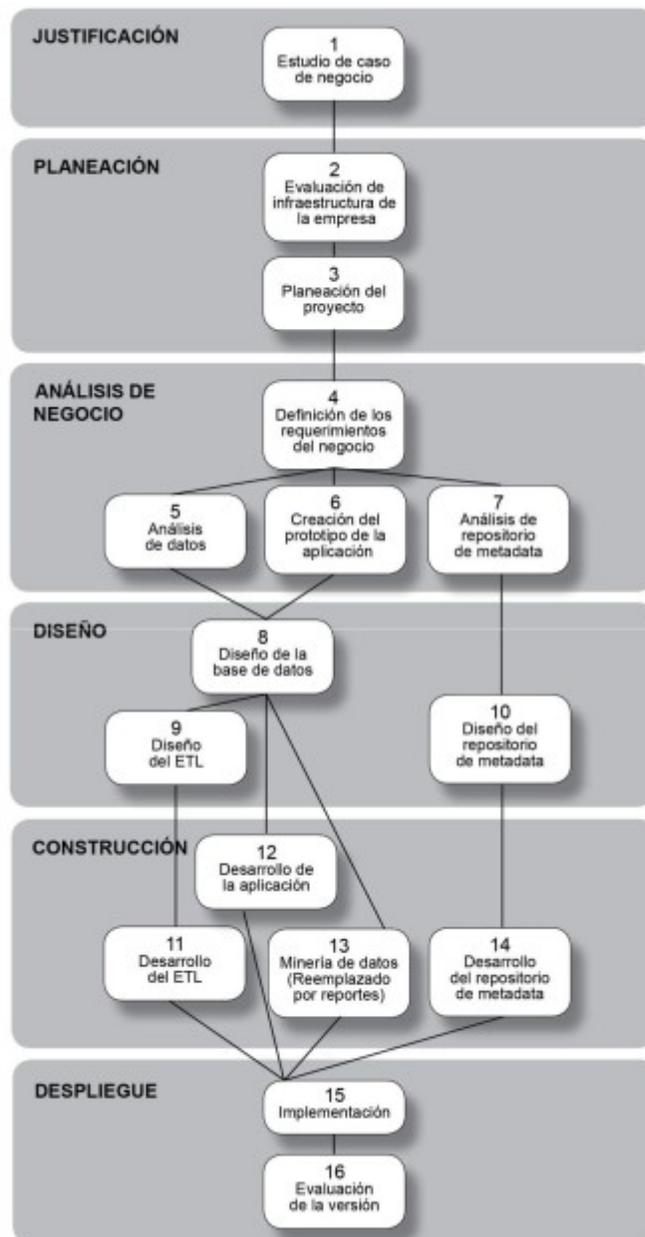
Anexo 02: Diagrama caso de uso para requerimientos

Questions to answer:	<p>¿Qué canal genera más tráfico (#)?</p> <p>¿Qué canal está generando el tráfico más valioso, en otras palabras, funciona mejor (%) en términos de conversión de clientes?</p> <p>¿Qué canal tiene el menor costo de adquisición de clientes (\$), es decir, el costo por cliente convertido?</p> <p>¿Cómo adquirimos usuarios a un costo menor?</p> <p>¿Cuánto cuesta en promedio adquirir un nuevo cliente?</p> <p>¿Dónde están las debilidades en nuestro embudo de conversión?</p> <p>¿Por que medio fidelizamos al cliente?</p>
Business objectives:	<ul style="list-style-type: none"> * Transformar usuarios inscritos a usuarios por prueba * Maximizar cantidad de usuarios por búsquedas * maximizar registros por publicidad rrss *max usuarios por blogs * maximizar usuarios especializado * maximizar renovacion cuentas * Conseguir pruebas de estudiantes/profesores de logistica * minimizar usuarios fantasma * convertir prueba en venta e ticket *convertir ticket en suscripcion
KPIs:	<p>costo por compra de una suscripcion</p> <p>costo tiempo para tomar una suscripcion</p> <ul style="list-style-type: none"> *costo por registro logeo prueba reporte segun canal * costo de usuario por publicidad blog *costo pot usuario a trial * costo por usuaio a suscricop * costo po usuario a renovacion * costo por usuarios de universidades
Metrics:	<p>Registros, logeos, pruebas, reportes, tiempo de estadia</p>
Segmentation:	<p>Demographic</p> <p>Ciudad, Region, Pais, Edad</p> <p>Modelo, OS, buscador web</p> <p>Tipo de empresa</p>
Targets:	<p>10% convertir usuarios de prueba basica a pruebas trial</p> <p>25% trial a ticket</p> <p>5% de ticket a suscripcion</p> <p>20% de renovacion</p>

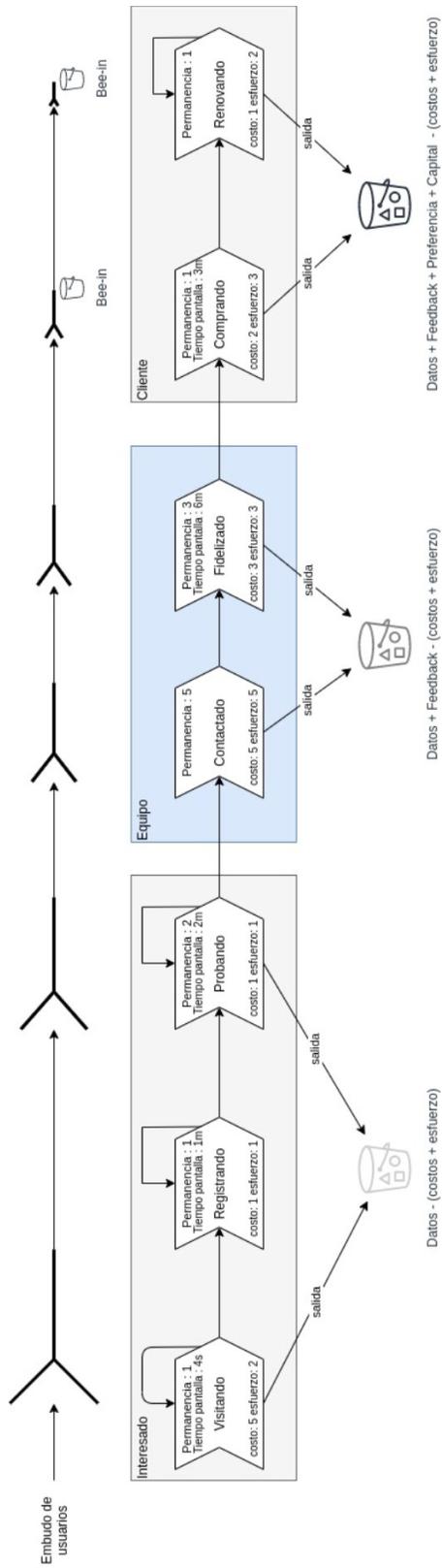
Anexo 03: Exploracion para la adquisicion según modelo AAARR

Questions to answer:	<p>¿cómo puede aumentar la retención de clientes?</p> <p>¿nuestro objetivo de retencion es lograr de que pruebe practicamente si es un producto apto para su produccion?</p> <p>¿de que depende nuestra retencion de clientes ?</p>
Business objectives:	<p>Conseguir un usuario satisfecho en su primera prueba</p> <p>Un usuario que supo si se ajustaba a su empresa o uso</p> <p>Que sea una herramienta que puede usar cuando desea</p>
KPIs:	<p>Mensaje al cerra, ¿esto sirve para tu empresa?</p> <p>Verifico los espacios de carga o vio los precios</p>
Metrics:	<p>tiempo de estadia</p> <p>sesiones distintas</p> <p>navegadores distintos</p> <p>Tasa de adquisición de clientes> Tasa de rotación de clientes = crecimiento</p> <p>Tasa de rotación de clientes> Tasa de adquisición de clientes = Quemar mucho dinero</p>
Segmentation:	<p>cantidad de sesiones distintas</p>

Anexo 04: Exploracion para la retencion según modelo AAARR



Anexo 05: La metodología de Larissa Moss



Anexo 06: Embudo de adquisicion de clientes