



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

1 de 2

Neiva, 03 de marzo de 2020

Señores

CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN

UNIVERSIDAD SURCOLOMBIANA

Ciudad

El (Los) suscrito(s):

WILSON EDUARDO MARTINEZ PERDOMO, con C.C. No. 1075281559, ALEJANDRO POLANCO ESPITIA, con C.C. No.1080186785 Autor(es) de la tesis y/o trabajo de grado titulado SISTEMA DE INFORMACION PUBLICITARIO DE PRODUCTOS EN SUPERMERCADOS Y ALMACENES DE CADENA UTILIZANDO WEB SCRAPING(COMMER) presentado y aprobado en el año 2020 como requisito para optar al título de INGENIERO DE SOFTWARE.

Autorizo (amos) al CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN de la Universidad Surcolombiana para que, con fines académicos, muestre al país y el exterior la producción intelectual de la Universidad Surcolombiana, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios puedan consultar el contenido de este trabajo de grado en los sitios web que administra la Universidad, en bases de datos, repositorio digital, catálogos y en otros sitios web, redes y sistemas de información nacionales e internacionales "open access" y en las redes de información con las cuales tenga convenio la Institución.
- Permita la consulta, la reproducción y préstamo a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato Cd-Rom o digital desde internet, intranet, etc., y en general para cualquier formato conocido o por conocer, dentro de los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia.
- Continúo conservando los correspondientes derechos sin modificación o restricción alguna; puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación del derecho de autor y sus conexos.

Vigilada Mineducación

La versión vigente y controlada de este documento, solo podrá ser consultada a través del sitio web Institucional www.usco.edu.co, link Sistema Gestión de Calidad. La copia o impresión diferente a la publicada, será considerada como documento no controlado y su uso indebido no es de responsabilidad de la Universidad Surcolombiana.



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

2 de 2

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, "Los derechos morales sobre el trabajo son propiedad de los autores" , los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

EL AUTOR/ESTUDIANTE:

Firma: 

EL AUTOR/ESTUDIANTE:

Firma: 

Vigilada Mineducación

La versión vigente y controlada de este documento, solo podrá ser consultada a través del sitio web Institucional www.usco.edu.co, link Sistema Gestión de Calidad. La copia o impresión diferente a la publicada, será considerada como documento no controlado y su uso indebido no es de responsabilidad de la Universidad Surcolombiana.



TÍTULO COMPLETO DEL TRABAJO: SISTEMA DE INFORMACIÓN PUBLICITARIO DE PRODUCTOS EN ALMACENES DE CADENA Y SUPERMERCADOS UTILIZANDO WEB SCRAPING (COMMER).

AUTOR O AUTORES:

Primero y Segundo Apellido	Primero y Segundo Nombre
MARTINEZ PERDOMO	WILSON EDUARDO
POLANCO ESPITIA	ALEJANDRO

DIRECTOR Y CODIRECTOR TESIS:

Primero y Segundo Apellido	Primero y Segundo Nombre

ASESOR (ES):

Primero y Segundo Apellido	Primero y Segundo Nombre
CASTRO SILVA	JUAN ANTONIO

PARA OPTAR AL TÍTULO DE:

FACULTAD: INGNIERIA

PROGRAMA O POSGRADO: INGENIERIA DE SOFTWARE

CIUDAD: NEIVA

AÑO DE PRESENTACIÓN:2020

NÚMERO DE PÁGINAS: 92

TIPO DE ILUSTRACIONES (Marcar con una X):

Vigilada Mineducación

La versión vigente y controlada de este documento, solo podrá ser consultada a través del sitio web Institucional www.usco.edu.co, link Sistema Gestión de Calidad. La copia o impresión diferente a la publicada, será considerada como documento no controlado y su uso indebido no es de responsabilidad de la Universidad Surcolombiana.



Diagramas Fotografías Grabaciones en discos Ilustraciones en general Grabados
Láminas Litografías Mapas Música impresa Planos Retratos Sin ilustraciones
Tablas o Cuadros

SOFTWARE requerido y/o especializado para la lectura del documento:

MATERIAL ANEXO:

PREMIO O DISTINCIÓN (En caso de ser LAUREADAS o Meritoria):

PALABRAS CLAVES EN ESPAÑOL E INGLÉS:

<u>Español</u>	<u>Inglés</u>
1. Supermercados	Supermarket
2. Web Scraping	Web Scraping
3. Sistema	system
4. Información	information
5. software	software

RESUMEN DEL CONTENIDO: (Máximo 250 palabras)

ABSTRACT: (Máximo 250 palabras)

La historia del comercio de importantes supermercados y almacenes de cadena en Colombia tales como Carulla, Almacenes Ley, Olímpica, Éxito entre otros, data desde el siglo pasado en donde los primeros comercios basados en el sector retail (Comercio minorista) implementaron técnicas de publicidad para poder llegar a sus clientes. Empresarios colombianos tuvieron gran influencia en la creación de estos establecimientos y potenciaron el sector retail hasta llegar a ser importantes en el comercio Colombiano. La inversión extranjera ha mostrado interés en este sector del comercio debido al gran crecimiento que representa en la economía colombiana, a las tendencias de consumo y a los modelos de negocio que actúan en el país, ubicando a Colombia



entre los intereses de inversiones extranjeras como WalMart y de grupos importantes como Makro, Carrefour que actualmente operan en el mercado colombiano(Silva, 2011).

La aplicación móvil y web(COMMER) permite capturar, organizar, agrupar, buscar y comparar la información de productos de almacenes de cadena y supermercados en Colombia extrayendo los datos utilizando web scraping (es una práctica que se utiliza como agente para descargar y analizar datos de la web de una manera automática)(Seppe, Baesens, & Seppe, 2018). Además tendrá la opción de registrar los productos de forma manual si la empresa no posee página web, para que cualquier supermercado pequeño o grande pueda ofertar sus productos. De esta forma se puede explotar el sector comercial de los supermercados y almacenes de cadena para poder organizar y unificar en un solo sistema de información todos los datos de sus productos.

APROBACION DE LA TESIS

Nombre Presidente Jurado: FERNANDO ROJAS ROJAS

Firma:

Nombre Jurado: LUIS GREGORIO RAMON CARVAJAL

Firma:

Nombre Jurado: FERLEY MEDINA ROJAS

Firma:

Sistema de información publicitario de productos en almacenes de cadena y supermercados utilizando web scraping (Commer).

Trabajo de grado

Alejandro Polanco Espitia

20132123809

Wilson Eduardo Martínez

20132123608

Asesor: Juan Antonio Castro Silva

Universidad Surcolombiana

Facultad de ingeniería

Programa de Ingeniería de Software

2020

Contenido

Capítulo 1

Inicio y Alcance de proyecto.....	5
1.1 Introducción.....	5
1.2 Problema.....	6
1.3 Antecedentes.....	7
1.3.2 Supermercados que ofrecen productos por internet y sistemas de información que agrupan sus catálogos en Colombia.....	7
1.4 Justificación.....	10
1.5 Área o línea de investigación.....	12
1.6 Objetivos.....	12
1.6.1 Objetivo general:.....	12
1.7 alcance y limitaciones.....	13

Capítulo 2

Planificación.....	13
2.1 Metodología y Materiales.....	13
2.1.1 PMBOK.....	13
2.1.2 Metodologías ágiles.....	14
2.1.2.1 XP.....	14
2.1.2.2 SCRUM.....	15
2.1.2.3 Kanban.....	16
2.1.3 Formas de obtener información de las empresas.....	17
2.1.3.1 Forma Manual.....	17
2.1.3.2 Web Scraping.....	17
2.1.4 Acceder a la mayoría de usuarios.....	18
2.1.4.1 Aplicaciones híbridas.....	18
2.1.4.2 Frameworks de desarrollo híbrido.....	18
2.1.4.2.1 React native.....	18
2.1.4.2.2 IONIC.....	19
2.1.5 Lenguaje de programación.....	20
2.1.5.1 JAVASCRIPT.....	21
2.1.5.2 PYTHON.....	21
2.1.5.3 RUBY.....	21
2.1.5.4 TYPESCRIPT.....	22
2.1.5.5 PHP.....	22
2.1.5.6 JAVA.....	22
2.1.6 Servidores web.....	22

2.1.6.1 WILDFLY	23
2.1.6.2 APACHE	23
2.1.6.3 NGINX	23
2.1.7 Tipos de bases de datos	24
2.1.7.1 Bases de datos No relacionales	25
2.1.7.1.1 Mongo DB	25
2.1.7.2 Bases de datos relacionales	25
2.1.7.2.1 POSTGRES	26
2.1.7.2.2 MYSQL	26
2.1.8 Servicios de virtualización y computación en la nube.....	27
2.1.8.1 Servidor (VPS).....	27
2.1.8.2 Hosting compartido.....	27
2.1.8.3 AMAZON WEB SERVICES (AWS).....	27
2.1.8.4 MICROSOFT AZURE	28
2.1.8.5 Google Cloud Platform.....	28
2.1.9 Secuencia y Actividades que se desarrollaran	29
2.1.10 Cronograma.....	30
2.1.11 Recursos, costos y fuentes de financiación	32
2.1.11.1 Modelo de Negocio	34
2.1.11.2 Mantenimiento del sistema.....	35
2.1.11.3 Proyección anual.	36
2.1.11.3.1 Fuente de ingresos	36
2.1.11.3.2 Punto de equilibrio.....	37
2.1.12.3.3 Retorno de inversión según cada fuente de ingresos.....	38
Capítulo 3	
Ejecución.....	39
3 Análisis y diseño	39
3.1 Recolección de requisitos	39
3.2 Modelado UML	52
3.3 Diseño de base de datos.....	56
3.4 Desarrollo de metodología Kanban	56
3.5 Prototipo del sistema.....	63
3.5 configuración del servidor Amazon EC2	67
3.6 Implementación.....	67
Capítulo 4	
Seguimiento y control	72

4.1 Métricas de calidad de software.....	72
4.1.1 Métrica orientada a la función.....	72
4.1.2 Métrica orientada a la Usabilidad	73
Capítulo 5	
Conclusiones y cierre	76
5.1 Resultados	76
5.2 Conclusiones y recomendaciones	77
5.3 Anexos	78
5.3.1 Plan de proyecto	78
5.3.1.1 Acta de constitución del proyecto o acta de inicio	
Información del proyecto	
Datos	78
5.3.2 Gestión del alcance	81
5.3.2.1 EDT (Estructura de desglose de trabajo).....	81
5.3.2.2 Diccionario de la EDT.....	82
5.3.2 Diccionario de datos	88
Bibliografía	91

ILUSTRACIONES

<u>Ilustración 1 FLUJO DE TRABAJO EN METODOLOGIAS AGILES</u>	<u>15</u>
<u>Ilustración 2 arquitectura web scraping Sistema de información publicitario de productos en almacenes de cadena y supermercados utilizando web scraping (Commer).....</u>	<u>18</u>
<u>Ilustración 3 arquitectura del proyecto Sistema de información publicitario de productos en almacenes de cadena y supermercados utilizando web scraping (Commer).....</u>	<u>24</u>

Capítulo 1

Inicio y Alcance de proyecto

1.1 Introducción

La historia del comercio de importantes supermercados y almacenes de cadena en Colombia tales como Carulla, Almacenes Ley, Olímpica, Exito entre otros, data desde el siglo pasado en donde los primeros comercios basados en el sector retail (Comercio minorista) implementaron técnicas de publicidad para poder llegar a sus clientes. Empresarios colombianos tuvieron gran influencia en la creación de estos establecimientos y potenciaron el sector retail hasta llegar a ser importantes en el comercio colombiano. La inversión extranjera ha mostrado interés en este sector del comercio debido al gran crecimiento que representa en la economía colombiana, a las tendencias de consumo y a los modelos de negocio que actúan en el país, ubicando a Colombia entre los intereses de inversiones extranjeras como WalMart y de grupos importantes como Makro, Carrefour que actualmente operan en el mercado colombiano(Silva, 2011).

Las primeras técnicas de publicidad que optaron estos empresarios fueron los grandes eventos y festividades del país para poder llegar a sus clientes, ocupar espacios de gran cantidad de masas era la mejor forma de hacer publicidad, con la implementación de la imprenta llegaron nuevos métodos para llegar a sus clientes, Panfletos, vallas de publicidad, catálogos en papel. Técnicas que han aportado a la información de sus productos entre empresario y usuario, estas aún se utilizan fortaleciéndose día a día.

Internet ha cambiado la manera de comunicarse en las personas, mediante este actualmente se puede intercambiar información de una forma rápida expandiéndose a miles de usuarios en pocas horas, el auge de las TIC (tecnologías de la información y

comunicación) en Colombia ha llevado a las empresas a hacer uso de las TIC para hacer publicidad y dar un valor agregado a las técnicas de hacer publicidad.

La aplicación móvil y web(COMMER) permite capturar, organizar, agrupar, buscar y comparar la información de productos de almacenes de cadena y supermercados en Colombia extrayendo los datos utilizando web scraping (es una práctica que se utiliza como agente para descargar y analizar datos de la web de una manera automática)(Seppe, Baensens, & Seppe, 2018). Además, tiene la opción de registrar los productos de forma manual si la empresa no posee página web, para que cualquier supermercado pequeño o grande pueda ofertar sus productos. De esta forma se puede explotar el sector comercial de los supermercados y almacenes de cadena para poder organizar y unificar en un solo sistema de información todos los datos de sus productos.

1.2 Problema

El **42,85%** de los principales supermercados y almacenes de cadena en Colombia carecen de una vitrina comercial como aplicación móvil o plataforma web para hacer publicidad, de igual forma los sistemas de información actuales (Ofertia y Tiendeo) que organizan y agrupan la información de productos de supermercados y almacenes de cadena del país utilizan catálogos o folletos digitales que dificultan que el usuario pueda buscar y comparar la información de sus productos.

Por lo tanto, se plantea como problema la carencia de sistemas de información que permitan la búsqueda y comparación de productos en múltiples almacenes de cadena y supermercados.

1.3 Antecedentes

En Colombia el sector minorista(etail) representa una de las principales fuentes de crecimiento del país, entre 2010 y 2014 la participación de esta actividad contribuye al 12 % del crecimiento del PIB y el 44% de los nuevos trabajos generados en el país(Rincon, Niño, & Rodriguez, 2017). La importancia de este sector durante los últimos años en el país ha sido de gran impacto para el desarrollo de la economía en todo el territorio nacional, esto ha llevado a estrategias de crecimiento de las empresas.

Las técnicas de mercadeo y publicidad que utilizan los empresarios como ofertas o promociones de sus productos con fin de llegar a sus clientes y que puedan acceder a sus productos son las formas para poder seguir aumentando sus ventas y seguir posicionándose en el mercado como uno de los sectores influyentes en Colombia.

Actualmente, funcionan dos plataformas de publicidad de supermercados y almacenes de cadena en Colombia Ofertia y Tiendeo, los cuales se encargan de agrupar folletos de los productos de las empresas afiliadas. Estas plataformas muestran solo folletos de las empresas, la búsqueda de información por producto no es eficaz para poder hacer una toma de decisión por parte de los compradores evaluando alternativas.

1.3.2 Supermercados que ofrecen productos por internet y sistemas de información que agrupan sus catálogos en Colombia.

Nombre Empresa	Si ofrecen productos por internet	No ofrecen productos por internet	Empresas que agrupa Tiendeo	Empresas que agrupa Ofertia
Almacenes éxito	X		X	
Olímpica	X		X	X
Alkosto	X		X	
Popular		X		
Confamiliar		X		
Justo y bueno	X			
La 14		X	X	X
Metro		X	X	X
Superior		X		
Tiendas D1		X	X	
Makro	X			

Carrulla	X		X	
Jumbo	X		X	
Surtimax	X		X	
TOTAL	8	6	9	3
%	57.14%	42,85%	64,28%	21,42%

Tabla 1 Supermercados que ofrecen productos por internet y sistemas de información que agrupan sus catálogos en Colombia método de recolección de información Commer 2019.

En la tabla 1 se muestran 14 empresas del sector minorista en Colombia, identificando si ofrecen o no productos por internet de igual forma si hacen parte de los sistemas de información ya existentes que agrupan algunas empresas del sector. De esta manera observamos que el 42.85% de las empresas carecen de una vitrina comercial como aplicación móvil o plataforma web para hacer publicidad.

Ofertia

Desde 2012 está al servicio del consumidor agrupando en su plataforma móvil y web los catálogos de ofertas de las mejores tiendas locales. En su categoría supermercados hasta el momento cuentan con 3 de los principales supermercados (Olímpica, La14 y Metro). Su sistema de búsqueda individual de un producto basado en catalogo se puede hacer por nombre del producto mostrando el catálogo completo de la empresa. Dificulta la búsqueda por producto mostrando catálogos digitales completos de las empresas con información de otros productos que no requiere el cliente.

Tiendeo

Agrupar e indexar el contenido de los catálogos y folletos tanto de pequeños comercios, de supermercados o cadenas de gran distribución, mostrando sus productos a los consumidores y ofreciendo, de una manera sencilla y cómoda(CECARM, 2012).

Es gratuito y sencillo. Muestra todos los catálogos y folletos de las tiendas en las que al usuario pueda interesarle realizar su compra. Junto con las imágenes escaneadas del catálogo, aparecen los datos de contacto de la tienda y además se puede localizar en un mapa el establecimiento cercano a tu domicilio, lo que facilita el trabajo al usuario.

Tiendeo no brinda una búsqueda unificada de cada producto, por lo tanto, el usuario tiene que realizar una búsqueda en cada uno de los folletos de las empresas afiliadas.

1.4 Justificación

El sector digital ha experimentado un fuerte crecimiento, según la primera gran encuesta TIC de 2017 realizada por el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), el 64% de los colombianos usan internet y el 72% de los hogares tiene acceso a Smartphone (MinTIC, 2018).

Las redes sociales, los teléfonos inteligentes y tablets que hoy en día utilizan las personas están cambiando la forma de consumo. No es eficiente tener un catálogo de los productos de una tienda en una página y esperar que los usuarios compren. Cada vez son más utilizados los dispositivos móviles para realizar búsquedas, comparar precios y comprar online, las empresas que quieran sobresalir y vender deberán adaptarse a esta tendencia.

La economía nacional ha tenido profundas transformaciones a partir de los procesos de apertura desde 1990. Por un lado los consumidores han tenido un acceso más amplio a la variada y competitiva oferta proveniente de otros países y, por el otro, las empresas

nacionales han debido soportar ataques publicitarios de la competencia externa, que trata de quedarse con estos mercados estatalmente protegidos(Silva, 2011).

Una de las razones del crecimiento es que los minoristas o cadenas de supermercados se expanden fuera de los grandes centros urbanos y están en búsqueda de ciudades intermedias de más de 100.000 habitantes, mientras Bogotá y Medellín ha experimentado un crecimiento de 4.3 y 6.1 por ciento, respectivamente, las ventas minoristas totales del país crecieron un 7,7 por ciento para el 2014. Entre marzo de 2016 y febrero de 2017 las ventas del comercio minorista en los Grandes Almacenes fueron de \$49,6 billones, lo cual significó una variación de 8,9% con relación al periodo marzo de 2015 – febrero de 2016 cuando las ventas fueron \$45,5 billones, con una variación de 11,9%(Rincon et al., 2017).

El comportamiento del consumidor de este sector ha cambiado, ahora los consumidores quieren llegar de una manera más fácil a la información utilizando las redes o el internet, este proceso es llamado búsqueda de la información dentro de los procesos de toma de decisiones de los compradores, como también evaluando las alternativas donde los consumidores comparan varios productos dentro de una lista de características.

Actualmente funcionan dos plataformas (Ofertia y Tiendeo) que organizan la información en catálogos de almacenes de cadena y supermercados en Colombia, identificamos que el **42,85%** de los principales supermercados y almacenes de cadena del país no cuentan aún con un sistema de información para poder ofertar sus productos. Aun así, estas plataformas que intentan agrupar información de los supermercados y almacenes de cadena en Colombia no alcanzan a la mayoría de supermercados del territorio nacional. De esta manera proponemos desarrollar un sistema de información que mejore los sistemas de información ya existentes utilizando web scraping para poder extraer los

datos de la mayoría los supermercados y almacenes de cadena mostrándolos en un solo sitio web y móvil.

1.5 Área o línea de investigación

Ingeniería de software.

1.6 Objetivos

1.6.1 Objetivo general:

Construir un sistema de información que permita buscar y comparar los precios y características de productos de supermercados y almacenes de cadena en Colombia, utilizando la técnica de web scraping mediante una aplicación móvil y web. Para ayudar a los usuarios en la toma de decisiones en los procesos de compra y a las empresas en la promoción de sus productos.

- Aplicar los procesos de desarrollo de software (análisis, diseño, desarrollo, pruebas, implementación) utilizando las metodologías Pmbok para la gestión del proyecto.
- Implementar un servicio web que permita capturar los datos de los productos de las empresas que tienen un sistema de información o página web, utilizando un algoritmo basado en web scraping en el lenguaje de programación java y la librería Jsoup, con el fin de almacenar una mayor cantidad de productos.

- Desarrollar una aplicación móvil y web utilizando framework ionic, lenguaje php base de datos mysql y metodologías ágiles (Kanban), que permita agregar, buscar y comparar los productos ofrecidos por almacenes de cadena y supermercados en cualquier dispositivo.

1.7 alcance y limitaciones

Desarrollar un sistema de información donde se muestren los productos y ofertas del sector minorista, para que los usuarios hagan una búsqueda eficaz evaluando alternativas en la toma decisiones de sus compras, de esta manera los supermercados y almacenes de cadena puedan subir sus productos y ofertas de manera manual, si no cuentan con una página web de sus productos. Si cuentan con una página web de sus productos se utilizará la técnica de web scraping para poder extraer los datos de los productos, almacenarlos y tratarlos mediante una aplicación móvil y web para ser mostrados al usuario.

Capítulo 2

Planificación

2.1 Metodología y Materiales

2.1.1 PMBOK

La guía Pmbok fue desarrollada por el Project Management Institute(PMI) estableciendo un criterio de buenas prácticas relacionadas con las gestión y administración de proyectos mediante técnicas o herramientas que permiten enfocarse dentro de los 48 procesos según sus 5 áreas del conocimiento.

La guía PMBOK identifica cada uno de los subconjuntos para la dirección de proyectos siendo cada una de estas buenas prácticas aplicables a la mayoría, donde la organización o el equipo son responsables de establecer lo apropiado para cada proyecto (Project Management Institute, 2013) .

El uso de estos procesos PMBOK nos permitirá gestionar nuestro proyecto haciendo uso de lo más necesarios para poder gestionar, administrar y dirigir cada fase de nuestro proyecto.

2.1.2 Metodologías ágiles

Las metodologías ágiles potencian las relaciones interpersonales del equipo de trabajo claves para el éxito del trabajo de desarrollo de software. Las metodologías ágiles han sido pensadas para proyectos en que tiene condiciones como: plazos reducidos, requisitos volátiles o basados en nuevas tecnologías (Escobar & Cobos, 2015).

Las principales metodologías ágiles son SCRUM, XP, RUP, KAMBAN basadas en principios comunes. La entrega temprana y continua de software con valor, también en estas metodologías son bienvenidos los requisitos cambiantes. Siendo esta una ventaja competitiva para el cliente, entre otros principios que dan valor al uso de estas metodologías dentro del trabajo de cada proyecto.

2.1.2.1 XP

XP es una metodología basada en la comunicación la reutilización de código y la realimentación, un proyecto en xp lleva de 10 a 15 ciclos o iteraciones (Pérez Pérez,

2012). Promueve el trabajo en equipo y además propicia una retroalimentación continua entre los involucrados y el equipo de desarrollo, se abarcan 4 fases de desarrollo, Exploración, Planificación, Iteraciones y Puesta en Producción. En la fase de exploración se define el alcance del proyecto mediante historias de usuario los programadores estiman los tiempos de desarrollo en base esa información.

2.1.2.2 SCRUM

En scrum un proyecto se desarrolla en iteraciones (ver ilustración 1) o bloques temporales de un mes donde cada iteración da al usuario un resultado completo según lo requiera. El sprint es un ritmo de los ciclos de scrum se delimita por una planificación y reunión retrospectiva (Pérez Pérez, 2012). Scrum maneja roles, el cliente o product owner, el facilitador o scrum master, el equipo team. También herramientas o técnicas como el product backlog, sprint backlog, burndown. El manejo de roles en esta metodología es el motivo por el cual no fue elegida para la realización de nuestro proyecto por ser un equipo de trabajo pequeño.

Flujo de trabajo en metodologías ágiles

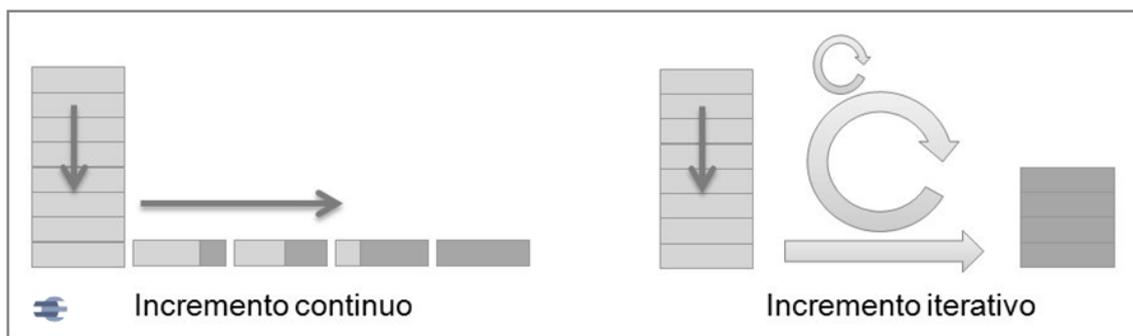


Ilustración 1 FLUJO DE TRABAJO EN METODOLOGIAS AGILES <https://scrummanager.net/blog/2013/09/kanban-las-4-cosas/>

2.1.2.3 Kanban

Es una metodología desarrollada por Toyota para la industria, kanban es una palabra japonesa que significa “tarjetas visuales” y se utiliza para controlar el avance del trabajo. Tiene como objetivo gestionar las actividades de manera general mientras se van completando, kanban no fue creada para la gestión de proyectos de software pero en la actualidad se utiliza en esta área(Manzaba, 2014). Visualizar o mostrar al equipo de trabajo lo que se está haciendo es una herramienta que ayuda a determinar los límites de trabajo, así mismo medir el tiempo que se puede realizar una tarea.

Las tres principales reglas o principios que se basa esta metodología son:

1. Medir el tiempo de cada tarea
2. Límite de trabajo en curso
3. Visualizar el trabajo y sus fases

Se basa en el incremento continuo, una de las aportaciones es que utiliza técnicas visuales para ver la situación de cada tarea representadas en post-it (tarjetas visuales)(Pérez Pérez, 2012). De esta manera Kanban representa de una forma ordenada las actividades que realizamos, identificamos nuestras habilidades medidas en tiempo, limitando a tiempo exacto el trabajo diario en cada actividad teniendo en cuenta lo realizado y lo que falta por hacer.

Kanban no maneja roles es decir un papel asignado y las tareas asociadas a dicho papel crean una identidad en el individuo. De esta manera, pedir que adopten un nuevo papel o un nuevo puesto de trabajo puede ser entendido como un ataque a su identidad. Kanban trata de evitar esa resistencia emocional, entiende que la ausencia de papeles es una ventaja(Pérez Pérez, 2012). De esta manera se escoge esta metodología para el desarrollo

de nuestro proyecto teniendo en cuenta que no se maneja un rol único para cada integrante.

2.1.3 Formas de obtener información de las empresas

2.1.3.1 Forma Manual

Para poder ingresar información de sus productos y almacenarlos en nuestra base de datos, los empresarios deben registrarse en el sistema.

2.1.3.2 Web Scraping

El Web Scraping son un conjunto de técnicas que se utilizan para obtener de forma automática el contenido que hay en páginas web a través de su código HTML. Su principal utilidad es recopilar grandes cantidades de datos de diferentes páginas web (Seppe et al., 2018). Es decir, Web scraping se utiliza para extraer datos de los productos y las empresas (almacenes de cadena y supermercados) de sitios web para poder ser almacenados y procesados.

Entre las distintas opciones para realizar Scraping se escoge por la librería jsoup-1.11.3 de JAVA es una potente herramienta para la extracción y manipulación de los datos además por su facilidad de sintaxis, usando lo mejor de modelo objeto documento (DOM), CSS, y métodos jQuery.

Arquitectura web scraping

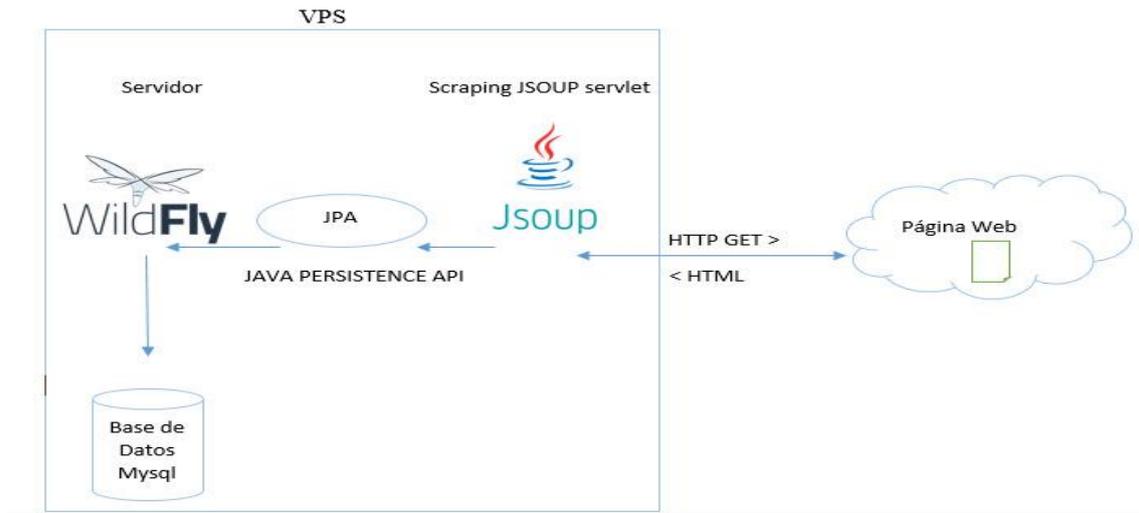


Ilustración 2 arquitectura web scraping Sistema de información publicitario de productos en almacenes de cadena y supermercados utilizando web scraping (Commer).

2.1.4 Acceder a la mayoría de usuarios

2.1.4.1 Aplicaciones híbridas

Las aplicaciones híbridas utilizan tecnologías multiplataforma como HTML, Javascript y CSS, pero se puede acceder a buena parte de las capacidades específicas de los dispositivos (Yaguapaz, 2018). De esta manera son desarrolladas utilizando tecnología web y son ejecutadas dentro de un contenedor web sobre el dispositivo móvil.

2.1.4.2 Frameworks de desarrollo híbrido

Entre los más importantes frameworks de desarrollo híbrido se encuentran ionic y react native, los cuales permiten a los desarrolladores crear aplicaciones en las tecnologías que no conocen, aplicando conocimientos en javascript o en el caso de ionic (angular y TypeScript) esto es lo que estos frameworks tratan de lograr (Kremer, 2019).

2.1.4.2.1 React native

Se introdujo por primera vez durante la conferencia de react.js efectuada por Facebook, react native se utiliza para crear aplicaciones móviles multiplataforma. Desde principios

de 2015 solo apoyaba el desarrollo de aplicaciones IOS sin embargo fue ampliada para poder incluir el soporte de android en septiembre de 2015(Hansson & Vidhall, 2016).

Ventajas

- se ejecuta y organiza los controles de la interfaz de usuario nativos de modo que la aplicación de la interfaz de usuario se ejecuta casi pero no completamente de forma nativa(Kremer, 2019).
- Puede utilizar cada componente nativo de la plataforma

Desventajas

- No comparte componentes a través de plataformas (VextView para Android y UIView para IOS) solo comparte código.

2.1.4.2.2 IONIC

IONIC es un framework de desarrollo de aplicaciones híbridas. Su función principal es aportar al HTML un conjunto de controles para la interfaz gráfica que no están incluidas en este pero que sí son comunes. Estos componentes están contruidos con una combinación de CSS, HTML y JavaScript, se comportan como los controles nativos que se usan normalmente, integrando cordova directamente permitiendo la instalación de los plugins de cordova para utilizarlos según lo requiera(Cano Insa, 2015).

Ventajas

- Animaciones Aceleradas por Hardware
- Mínima manipulación del DOM

- Fácil adaptación del diseño
- Integra cordova y todos sus plugins
- Mayor portabilidad utilizando estándares de web abiertos
- Estabilidad de la plataforma lo que escriba hoy funcionará mañana. Otras herramientas de front-end están creciendo y cambiando todo el tiempo, sin tanto cuidado hacia la compatibilidad hacia atrás(Kremer, 2019).

Desventajas

- Se debe tener conocimientos avanzados en javascript y conocimiento de angular js.
- Por ser Hybrid-Web el rendimiento de la aplicación depende de las mejores prácticas al desarrollar.

De esta manera como equipo de trabajo se escoge ionic framework para el desarrollo de este proyecto por sus ventajas en portabilidad y estabilidad en cada plataforma.

2.1.5 Lenguaje de programación

Existen diferentes tipos de lenguajes los cuales se utilizan para la programación de computadoras, desde el lenguaje máquina que utiliza códigos hexadecimales que representan instrucciones, registros de la cpu, direcciones de memoria o datos, como también el lenguaje ensamblador que utiliza los nemotécnicos para los códigos

hexadecimales, estos lenguajes son llamados de bajo nivel porque se acercan al lenguaje que manejan las computadoras. Los lenguajes de alto nivel son los más cercanos al lenguaje natural y matemático, tienen mayor legibilidad y facilidad de codificación. Tienen una mejor estructura de datos o abstracción procedimental(Yáñez, 2014). A continuación, se nombran de los lenguajes de programación de alto nivel más utilizados en el ambiente web y cuáles de estos son más apropiados para el desarrollo de nuestro proyecto.

2.1.5.1 JAVASCRIPT

Es un lenguaje de programación basado en scripting, los lenguajes scriptings son interpretados en vez de compilados a diferencia de lenguajes tradicionales como Java. Con la aparición de node.js las capacidades de este lenguaje han aumentado hasta desarrollar aplicaciones totalmente desarrolladas en el con módulos reutilizables simplificando la complejidad y facilitando su mantenimiento y evolución(Puerta & Arias, 2014).

2.1.5.2 PYTHON

Es un lenguaje interpretado, multiparadigma de aspectos imperativos, funcionales y orientados a objetos. Cuenta con una gran biblioteca estándar para resolver tareas de utilidad en el desarrollo(Troyano, Cruz, González, Vallejo, & Toro, 2018). Actualmente es utilizado para desarrollo de potentes aplicaciones web y aprendizaje de maquina o maching learning.

2.1.5.3 RUBY

Su creador es Yukihiro Matsumoto fue diseñado para que ayude a los programadores a ser productivos y disfrutar de la programación. Es un lenguaje dinámico orientado a

objetos muy utilizado en el ambiente web, fue base para desarrollo de aplicaciones como twitter, github, entre otras(Vizcay & Zunino, 2017).

2.1.5.4 TYPESCRIPT

Es un lenguaje de programación que permite crear aplicaciones web robustas en JavaScript permite generar código que se ejecuta en cualquier navegador, plataforma o sistema operativo(Yaguapaz, 2018). Su funcionalidad se basa en la transcripción de la lógica de programación o instrucciones de un lenguaje a otro. En nuestro proyecto utilizamos framework IONIC el cual incluye el lenguaje TypeScript.

2.1.5.5 PHP

Es un lenguaje de programación de scripts y server-side creado en 1995, fue pensado y diseñado para desarrollo web pero también usado para desarrollo general. En 2013 estaba instalado en más de 240 millones de sitios web y 2.1 millones de servidores(Juan, Cano, Miguel, & Calafate, 2015). Siendo uno de los lenguajes utilizado en aplicaciones web brinda facilidad de mezclar HTML e integrarse con varios frameworks web. Por la facilidad de integrarse a frameworks se elige este lenguaje para el desarrollo de lado del servidor.

2.1.5.6 JAVA

Java es un lenguaje dinámico y en evolución orientado a objetos, se caracteriza por ser un lenguaje simple, robusto (realiza verificaciones de errores tanto en ejecución como en compilación) y se puede ejecutar en cualquier máquina que tenga el sistema de ejecución runtime y jre(Omar & Flores, 2018). Se elige este lenguaje de programación para la técnica de web scraping por la facilidad de integrar librerías Jsoup y ser un lenguaje simple y robusto.

2.1.6 Servidores web

Los servidores son equipos de buena capacidad de almacenamiento y procesamiento conectados a internet que sirven para alojar información como base de datos, web server, archivos entre otros. Los principales servidores web que existen se pueden utilizar de acuerdo a la necesidad de las empresas.

Servidores según su tecnología:

Java: Apache Tomcat, Glassfish, JBOSS, Oracle WebLogic, IBM WebSphere.

Microsoft: IIS (Internet Information Services)

PHP: Apache HTTP Server

Los servidores más utilizados según datos estadísticos publicados por NETCRAFT son apache, Microsoft, sun, nginx y google(SANGA, 2017).

2.1.6.1 WILDFLY

Es de código abierto y gratis, puede ser usada en diferentes plataformas es una aplicación gestionada flexible y ligera. Representa una nueva versión mejorada del servidor de aplicación JBoss, está escrita en Java e implementa la especificación de Java EE(SANGA, 2017). Este servidor servirá para alojar los servicios de web scraping.

2.1.6.2 APACHE

Es una plataforma de servidores web altamente configurable de código abierto para la creación de páginas y servicios web, considerado como multiplataforma, gratuito, muy robusto(SANGA, 2017).

2.1.6.3 NGINX

Es un servidor web libre de código abierto (open source). Principalmente ha solucionado el problema de manejar una gran cantidad de clientes al mismo tiempo (aproximadamente 10 mil conexiones al mismo tiempo), esta es la diferencia a los demás

servidores(Manzaba, 2014)(MIRANDA, 2016). De esta manera se eligió utilizar este servidor en el proyecto para manejar una gran cantidad de clientes.

La ilustración 3 muestra la arquitectura del proyecto y los lenguajes de programación que se utilizaran en su desarrollo de parte del servidor.

Arquitectura del proyecto

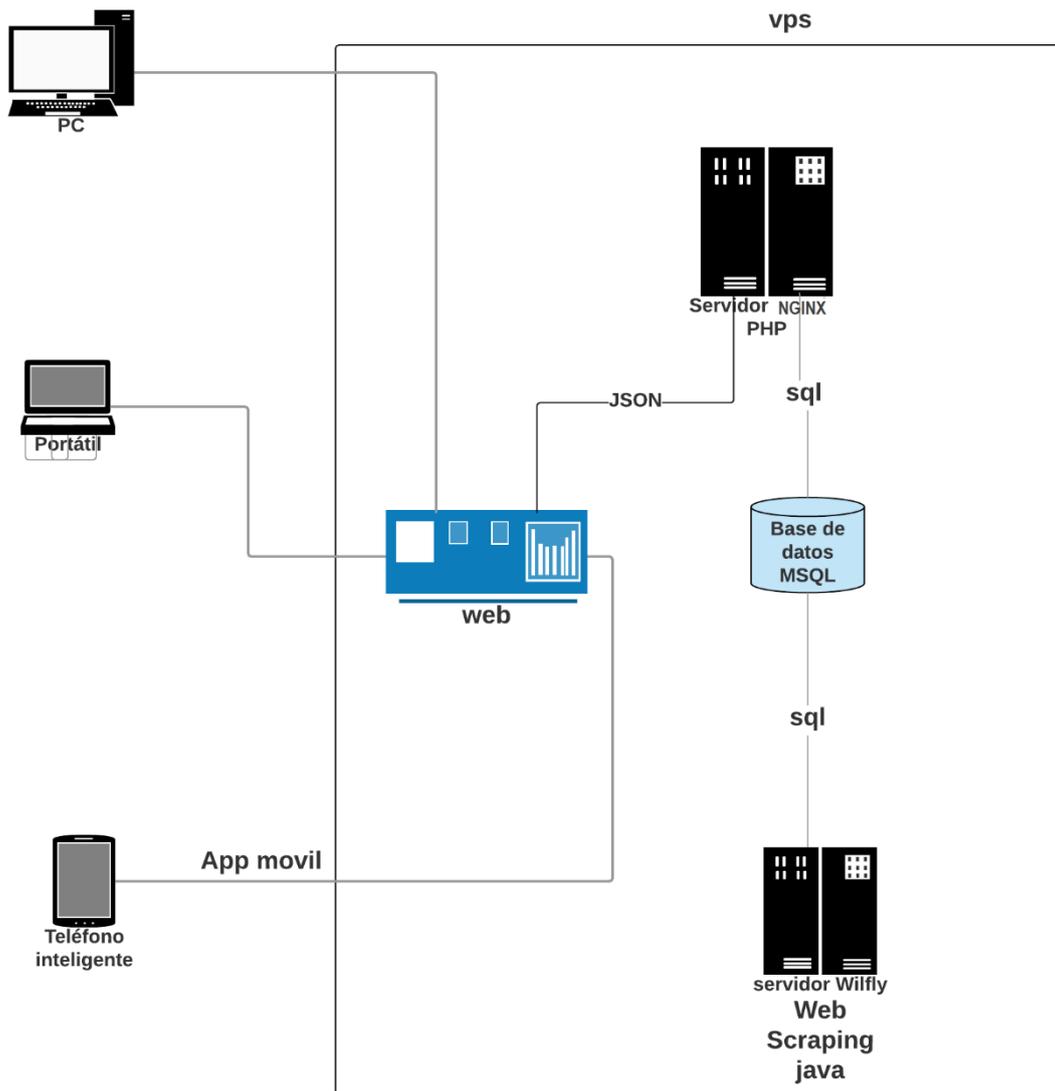


Ilustración 3 arquitectura del proyecto Sistema de información publicitario de productos en almacenes de cadena y supermercados utilizando web scraping (Commer).

2.1.7 Tipos de bases de datos

2.1.7.1 Bases de datos No relacionales

Este término fue usado por primera vez en 1998 para una base de datos relacional que omitía el uso de SQL principalmente se utiliza para almacenar grandes cantidades de datos como Facebook que tiene un gran crecimiento de información diario. Interactúa con sistemas operativos basados en UNIX y poseen alto rendimiento al escalar horizontalmente esto se debe a que sus tablas no están relacionadas(Zhingri, 2015).

2.1.7.1.1 Mongo DB

Está basada en open-source fue desarrollada por 10gen en c++ , lanzada en 2009 perteneciente a la categoría de base de datos documentales. La desventaja de esta base de datos es que tiene una seguridad de transacciones a nivel bajo pero son rápidos en las consulta a sus datos y tienen mayor escalabilidad(Zhingri, 2015). Si es a cuenta velocidad de cada consulta del proyecto, se da importancia a la seguridad de los datos y le dejamos a las buenas prácticas de desarrollo el papel de hacer ágil cada una de las consultas a la base de datos.

2.1.7.2 Bases de datos relacionales

Un modelo de datos es un conjunto de herramientas conceptuales para la descripción de datos su semántica, relaciones y restricciones de consistencia permitiendo representar la información fácilmente manteniendo las características de los metadatos que facilitan las modificaciones disminuyendo los problemas al aplicar una mejora continua, la estructura fundamental del modelo relacional son tablas constituidas por filas y columnas relacionadas entre sí(Zhingri, 2015). Se eligió este modelo relacional en nuestro proyecto por estabilidad en una posterior mejora continua del proyecto, a

continuación nombraremos 2 importantes motores de base de datos relacionales (gratuitos) y cuáles de ellas se elige según el modelo de negocio y sus características.

2.1.7.2.1 POSTGRES

Es un potente motor de base de datos relacional, cuenta con más de 15 años de desarrollo activo y una arquitectura probada ganándose una sólida reputación de fiabilidad e integridad de datos. Se puede ejecutar en sistemas operativos como Linux Unix y Windows(Henríquez, Iglesias, Ramos, & Ropain, 2013).

2.1.7.2.2 MYSQL

Es el segundo motor de base de datos relacional más utilizado en todo el mundo y el primero en open-source(Juan et al., 2015). Desarrollado y distribuido por Oracle Corporation

Características:

- Escrito en C y en C++ es Open source
- Probado con un amplio rango de compiladores diferentes
- Funciona en diferentes plataformas
- Proporciona sistemas de almacenamiento transaccionales y no transaccionales
- Un sistema de reserva de memoria muy rápido basado en threads
- Un sistema de privilegios y contraseñas que es muy flexible y seguro, y que permite verificación basada en el host
- Rápida, fiable, escalable, y fácil de usar.

Mysql puede ejecutarse en computación en la nube como Virtual machine image, MySQL as a service y Managed MySQL cloud hosting(Juan et al., 2015). De esta manera MySQL brinda fascinantes características propias que requiere nuestro proyecto permitiendo aprovechar toda la capacidad del servidor.

2.1.8 Servicios de virtualización y computación en la nube

2.1.8.1 Servidor (VPS)

La virtualización es la manera de implementar recursos de computación que agrupen las diferentes capas de hardware, datos, redes software, y almacenamiento. Una de las ventajas de un vps es la escalabilidad permitiendo ampliar la capacidad del servidor sin necesidad de parar el servicio(PRECIADO, 2013).

2.1.8.2 Hosting compartido

Los servicios de hosting compartidos se limitan a proyectos que no requieran toda la potencia de un servidor, ni de dos o más lenguajes de programación y bases de datos. Se requieren solo para proyectos generales como páginas web. De esta manera no es óptimo el uso de este servicio web para el uso de nuestro proyecto.

2.1.8.3 AMAZON WEB SERVICES (AWS)

Es una de las plataformas que domina el mundo de las TI, considerada la mayor red de centro de datos. AWS cuenta con una gran cantidad de servicios en la nube para empresas que requieran gran flexibilidad, escalabilidad y mantenimiento de datos(Vílchez, 2019). La desventaja de utilizar AWS son los costos de sus servicios, pero es algo que han

logrado reducir para sus usuarios sin afectar la calidad. Amazon Web Services nos brinda todas las capacidades necesarias para la implementación del proyecto de esta manera, se hace uso de EC2 (Amazon Elastic Computing Cloud) que permite computación escalable y controlar los cambios según el proyecto requiera.

2.1.8.4 MICROSOFT AZURE

Es uno de los competidores de AWS brindando servicios en la nube teniendo un crecimiento por la popularidad de sus productos y generando una gran cantidad de ofertas y descuentos para sus usuarios. Es cuanto al uso de otros sistemas operativos que no sean Windows server Azure puede causar restricciones a sus usuarios es una de las desventajas(Vílchez, 2019). En el uso de nuestro proyecto buscamos soluciones complejas al manipular el servidor en cuanto al uso de sistemas operativos Linux de esta manera estos servicios no son los más convenientes para el desarrollo de este.

2.1.8.5 Google Cloud Platform

Es una plataforma de computación en la nube que permite a los usuarios implementar aplicaciones de todo tipo. Su fortaleza se centra en el desarrollo de aplicaciones móviles, servicios como app Engine que crea aplicaciones de forma ágil con un alto nivel de computación, esta plataforma tiene menos servicios (Procesamiento y hosting, Almacenamiento, Herramienta de redes, Macrodatos, aprendizaje automático) con respecto a sus competidores(Vílchez, 2019).

2.1.9 Secuencia y Actividades que se desarrollaran

En la siguiente tabla se definen cada una de las actividades con el número de la actividad y sus predecesoras. Se tiene en cuenta las 5 fases las cuales se identifican como áreas del conocimiento en PMBOK dentro de estas se utilizan 10 procesos PMBOK, en la fase de desarrollo (ejecución) se aplica la metodología Kanban.

Numero	Actividad	Predecesora
1	INICIO	
1.1	Modelo de negocio	
1.2	Acta de inicio	1.1
1.3	Definición de Alcance y Objetivos	1.2
1.4	Plan del Proyecto	1.1,1.2,1.3
2	Planificación	
2.1	Recopilación de información	1.4
2.2	Análisis del sistema	2.1
2.3	Obtener los requisitos	2.2
2.4	Aprobación requisitos	2.3,2.2,2.1
2.5	Análisis de costos	2.4
3	Ejecución	
3.1	Diagramación UML	2.4
3.2	Diseño Base de datos	3.1
3.3	Prototipo del sistema	3.2
3.4	Desarrollo front end	3.3
3.5	Desarrollo back end	3.1,3.2,3.3
3.6	Configuración del servidor	3.4,3.5
3.7	Pruebas	3.6,3.4,3.5

4	Monitoreo y Control	
4.1	Control de requisitos	3.6
4.2	Control de costos	4.1
4.3	Control del cronograma	4.3
5	Cierre	
5.1	Aprobación versión 1.0	4.3
5.2	Instalación	5.1
5.3	Manual	5.2
5.4	Entrega	5.3

2.1.10 Cronograma

	ACTIVIDAD	SEMANAS	DURACION (Días Laborales)
1	INICIO		10
1.1	Modelo de negocio	Semana 1	2
1.2	Acta de inicio	Semana 1	1
1.3	Definición de Alcance y Objetivos	Semana 1	3
1.4	Plan del Proyecto	Semana 2	4
2	Planificación		26
2.1	Recopilación de información	Semana 3, 4	15
2.2	Análisis del sistema		5
2.3	Obtener los requisitos	Semana 5	5
2.4	Aprobación requisitos	Semana 5	0

2.5	Análisis de costos	Semana 6	1
3	EJECUCION		55
3.1	Diagramación UML	Semana 7,8	10
3.2	Diseño Base de datos	Semana 9,10	10
3.3	Prototipo del sistema	Semana 10	2
3.4	Desarrollo front end	Semana 11,12,13,14	28
3.5	Desarrollo back end	Semana 14,15,16,17	28
	Configuración del servidor	Semana 18	5
3.6	Pruebas		5
4	Monitoreo y Control	Semana 19,20	15
4.1	Control de requisitos	Semana 21	5
4.2	Control de costos	Semana 22	5
4.3	Control del cronograma	Semana 23	5
5	Cierre		9
5.1	Aprobación versión 1.0	Semana 23	0
5.2	instalación y capacitación	Semana 24	2
5.3	Manual	Semana 25	3
5.4	Entrega	Semana 25	4
TOTAL		25 semanas	115 días Laborales, (920 horas)

2.1.11 Recursos, costos y fuentes de financiación

Nombre del recurso	Tipo	Tasa estándar	Costo por uso	Total mensual	Responsable
Computador	Material	No aplica	2`000.000	No aplica	Wilson Martínez
Computador	Material	No aplica	2`000.000	No aplica	Wilson Martínez
Vpc amazon web services EC2	Material	No aplica	970.000 anual	112,32 por hora, 88.700,mensual,	Alejandro Polanco
Analista y jefe de proyecto	Trabajo	15.000	0	1`600.000	Wilson Martínez
Diseñador y Desarrollador	Trabajo	15.000	0	1`600.000	Alejandro Polanco

Los costos por cada actividad y recursos asignados

ACTIVIDAD	RECURSOS	COSTO
INICIO		
Modelo de negocio	Analista y jefe de proyecto	240.000
Acta de inicio	Analista y jefe de proyecto	120.000
Definición de Alcance y Objetivos	Analista y jefe de proyecto	360.000
Plan del Proyecto	Analista y jefe de proyecto	480.000
Planificación		
Recopilación de información	Analista y jefe de proyecto	1`845.000
Análisis del sistema	Analista y jefe de proyecto	720.000
Obtener los requisitos	Analista y jefe de proyecto	720.000
Aprobación requisitos	Analista y jefe de proyecto	240.000
Análisis de costos	Analista y jefe de proyecto	120.000
EJECUCION		

Modelado UML	Analista y jefe de proyecto; Diseñador y Desarrollador	1`200.000
Diseño Base de datos	Analista y jefe de proyecto; Diseñador y Desarrollador	1200000
Prototipo del sistema	Analista y jefe de proyecto; Diseñador y Desarrollador	240000
Desarrollo front end	Diseñador y Desarrollador; Computador	4`026.666
Desarrollo back end	Diseñador y Desarrollador; Computador	4`826.666
Configuración del servidor		
Pruebas	VPC y dominio; Diseñador y Desarrollador; Computador	2`066.666
Monitoreo y Control		
Control de requisitos	Analista y jefe de proyecto	720.000
Control de costos	Analista y jefe de proyecto	720.000
Control del cronograma	Analista y jefe de proyecto	720.000
Cierre		
Aprobación versión 1.0	Analista y jefe de proyecto; Diseñador y Desarrollador	0
instalación y capacitación	Diseñador y Desarrollador	240.000
Manual	Diseñador y Desarrollador;	120.000
Entrega	Analista y jefe de proyecto; Diseñador y Desarrollador	240.000
TOTAL		21`174.998

El costo total de la construcción del sistema fue asumido por los integrantes del proyecto.

2.1.11.1 Modelo de Negocio

Commer será totalmente gratuito para sus usuarios teniendo en cuenta que es un sistema de información que ofrece publicidad de supermercados y almacenes de cadena.

Cualquier empresa grande o pequeña podrá hacer uso de la plataforma gratuitamente para agregar sus productos. La propuesta de valor está dada por la publicidad de terceros que quieran pautar anuncios en el sistema haciendo contacto con el administrador para poder concretar el servicio por un valor de 200.000 COP por cada 30 días. El empresario podrá agregar un producto en oferta por un costo de 25.000 COP agregándolo a la página de inicio para poder llegar a más clientes. De igual manera si el empresario desea hacer uso de web scraping tendrá cobro mensual de 80.000 a 200.000 mil pesos dependiendo de la cantidad de productos que maneje su página web.

Plantilla de modelo de negocio

Segmento de clientes:	Sector minorista (retail) en Colombia: Grandes y pequeños comerciantes del sector retail. Grandes almacenes de cadena
Propuesta de valor:	Supermercados o almacenes de cadena en Colombia podrán hacer uso de la plataforma gratuitamente para subir sus productos, si el

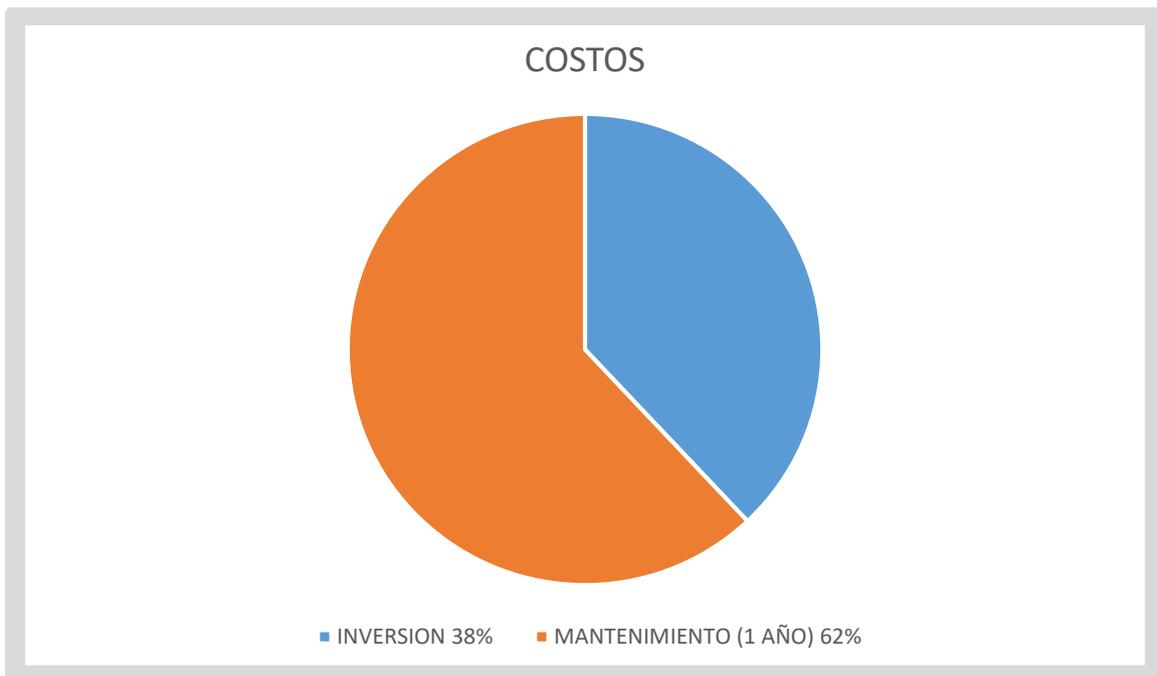
	<p>producto está en oferta aparecerá en la página de inicio y tendrá un costo de 25.000 COP, la publicidad de terceros que quieran pautar anuncios en el sistema haciendo contacto con el administrador tendrá un costo de 200.000 COP por cada 30 días, el servicio de web scraping tendrá un costo de 80.000 COP a 200.000 COP dependiendo de la cantidad de productos.</p>
Canales:	<p>Aplicación móvil y web, donde los empresarios harán uso gratuitamente</p>
Relación con el cliente:	<p>La asistencia a los clientes se hará vía redes sociales formulario de contacto en el sistema, número de teléfono y correo electrónico</p>
Fuente de ingresos:	<p>Publicidad: los clientes pueden pautar en el sistema Web scraping: brindando facilidad a las empresas que ya tienen página web para extraer sus productos ya publicados en sus paginas</p>
Recursos claves	<p>El recurso humano para el desarrollo del proyecto Analista y jefe de proyecto Diseñador y desarrollador</p>
Socios clave	<p>Empresarios del sector retail en Colombia Empresas que ofrecen productos y servicios</p>
Estructura de costes	<p>Los costos están dados en un total de 21'174.998 pesos colombianos en 22 actividades</p>
Actividades clave	<p>Análisis, diseño, desarrollo, implementación, pruebas y mantenimiento del sistema posterior a la aprobación de la versión 1.0</p>

2.1.11.2 Mantenimiento del sistema

Los costos de mantenimiento del sistema los evaluamos de la siguiente manera

Recursos

Recurso	Responsable	Pago mensual
Administrador del sistema	Wilson Martínez	1`400.000
Desarrollador	Alejandro Polanco	1`400.000
Vps mensual	Wilson Martínez	61.588
Dominio mensual	Wilson Martínez	3.333
Total		2`864.921
Total Anual		34`379.052



2.1.11.3 Proyección anual.

2.1.11.3.1 Fuente de ingresos

Publicar producto en oferta	25.000
-----------------------------	---------------

Publicidad incrustada de terceros	200.000
Servicio de Web scraping	200.000

2.1.11.3.2 Punto de equilibrio

El punto de equilibrio es cuanto vamos a vender para poder alcanzar los costos anuales.

COSTOS	TOTAL
fijos anuales	34'368.000
INGRESOS (Publicación en oferta)	25.000
COSTOS FIJOS/INGRESOS	1334 CLIENTES

COSTOS	TOTAL
fijos anuales	34'368.000
INGRESOS (Publicidad incrustada de terceros y web scripting)	200.000
COSTOS FIJOS/INGRESOS	171 clientes

--	--

2.1.12.3.3 Retorno de inversión según cada fuente de ingresos

Publicación en oferta

ROI (Return On Investment) = beneficio obtenido – inversión / inversión

$$66\ 700.000 - 55\ 542.998 / 55\ 542.998 = 0.20087$$

$$0.20087 \times 100 = 20\%$$

Publicación incrustada de terceros y web scraping

$$68\ 800.000 - 55\ 542.998 / 55\ 542.998 = 0.2386$$

$$0.2386 \times 100 = 23,86\%$$

FUENTE DE INGRESOS	\$	CLIENTES ESPERADOS (PUNTO EQUILIBRIOX2)	TOTAL	ROI (Return On Investment) = beneficio obtenido – inversión / inversión
Publicación en oferta	25.000	2668	66`700.000	20%
Publicación incrustada de	200.000	344	68`800.000	23,86%

terceros y web scraping				
----------------------------	--	--	--	--

Capítulo 3

Ejecución

3 Análisis y diseño

3.1 Recolección de requisitos

3.1.1 Requisitos funcionales

Se recolectaron 12 requisitos funcionales del sistema divididos en los 3 roles del sistema a continuación se especifican cada uno de ellos. En el punto 3.4 se especifica la metodología kanban con cada requisito.

Requisitos de usuario

Nombre:	ver productos en oferta
Fecha:	20/02/19
Numero requisito:	1

Contexto:	El usuario Puede ver productos en ofertas en la página de inicio
Autor:	Alejandro Polanco
Entradas:	no aplica
Salidas:	nombre producto, precio producto, precio anterior
Precondiciones:	El usuario no necesita iniciar sesión en el sistema.
Flujo básico de éxito:	inicio, muestra productos en oferta
Estado:	Aprobado

Nombre:	Ver productos por categoría.
Fecha:	24/02/19
Numero requisito:	2

Contexto:	El usuario puede ver las categorías de las empresas que están registradas.
Autor:	Alejandro Polanco
Entradas:	no aplica
Salidas:	Nombre categorías, imagen, precio, nombre, descripción, nombre empresa, dirección, horario, imagen o logo empresa.
Precondiciones:	El empresario ya haya subido sus productos o se haya echo scraping.
Flujo básico de éxito:	Seleccionar categorías, seleccionar tipo de categoría, seleccionar producto, muestra información de producto y empresa.
Estado:	Aprobado

Nombre:	Ver productos por empresa.
Fecha:	30/02/19

Numero requisito:	3
Contexto:	El usuario puede ver las empresas que están registradas y sus productos por empresa.
Autor:	Alejandro Polanco
Entradas:	no aplica
Salidas:	Nombre empresa, Nombre producto, Imagen producto, descripción producto dirección, horario, imagen o logo empresa.
Precondiciones:	El empresario ya haya subido sus productos o se haya echo scraping.
Flujo básico de éxito:	Empresas, seleccionar empresas, seleccionar producto, muestra información de producto y empresa.
Estado:	Aprobado

Nombre:	Buscar productos en el sistema por nombre
Fecha:	3/03/2019
Numero requisito:	4

Contexto:	el usuario puede buscar productos en el sistema por nombre
Autor:	Alejandro Polanco
Entradas:	no aplica
Salidas:	Imagen, precio, nombre, descripción, nombre empresa, dirección, horario, imagen o logo empresa.
Precondiciones:	El empresario ya haya subido sus productos o se haya echo scraping.
Flujo básico de éxito:	Inicio, búsqueda por nombre, seleccionar producto, información de producto y empresa.
Estado:	Aprobado

Nombre:	Seleccionar Ubicación
Fecha:	7/03/19
Numero requisito:	5
Contexto:	El usuario debe seleccionar ubicación para mostrar productos según ubicación

Autor:	Alejandro Polanco
Entradas:	no aplica
Salidas:	Empresas y productos por ubicación
Precondiciones:	No aplica
Flujo básico de éxito:	Seleccionar ubicación , mostrar productos y empresas por ubicación
Estado:	Aprobado

Nombre:	Lista de deseos
Fecha:	10/03/19
Numero requisito:	6
Contexto:	El usuario debe puede agregar productos a la lista de deseos
Autor:	Alejandro Polanco
Entradas:	no aplica
Salidas:	Empresas y productos por ubicación
Precondiciones:	El usuario debe estar registrado y haber iniciado sesión en el sistema
Flujo básico de éxito:	Registro, iniciar sesión , agregar productos a la lista deseos

Estado:	aprobado
---------	----------

Requisitos del empresario

Nombre:	registro de empresario
Fecha:	13/03/19
Numero requisito:	7
Contexto:	El empresario debe hacer registro en el sistema mediante correo y contraseña, identificando número de celular, nombre, correo, matricula mercantil cámara de comercio y contraseña.
Autor:	Alejandro Polanco
Entradas:	nombre empresario, identificación, matricula mercantil, correo, numero de celular, contraseña
Salidas:	registro de empresario exitoso

Precondiciones:	tener matricula mercantil
Flujo básico de éxito:	Registro, agregar campos nombre, identificación, matricula mercantil, correo, numero de celular, contraseña, validar correo de usuario.
Estado:	Aprobado

Nombre:	Agregar productos al sistema
Fecha:	13/03/19
Numero requisito:	8
Contexto:	el empresario puede agregar productos al sistema
Autor:	Alejandro Polanco
Entradas:	nombre empresario, identificación, matricula mercantil, correo, numero de celular, contraseña
Salidas:	Producto agregado exitosamente

Precondiciones:	Haber ingresado al sistema como empresario, y agregar información de empresa.
Flujo básico de éxito:	Ingreso al sistema, agregar producto, agregar campos nombre imagen precio descripción, registrar producto.
Estado:	Aprobado
Nombre:	agregar información de empresa
Fecha:	16/03/19
Numero requisito:	9
Contexto:	el empresario puede agregar información de la empresa
Autor:	Alejandro Polanco
Entradas:	Nombre empresa, logo empresa, dirección empresa, departamento, municipio, horario, teléfono o celular.
Salidas:	Empresa agregada exitosamente
Precondiciones:	haber ingresado al sistema como empresario

Flujo básico de éxito:	Ingreso al sistema, agregar empresa, agregar campos nombre empresa, logo empresa, dirección, departamento, municipio, horario, teléfono o celular, agregar empresa.
Estado:	Aprobado

Nombre:	Ver y editar información empresa
Fecha:	20/03/19
Numero requisito:	10
Contexto:	el empresario puede ver y editar información de empresa
Autor:	Alejandro Polanco
Entradas:	No aplica
Salidas:	Empresa agregada exitosamente
Precondiciones:	Entrar al sistema mediante correo y contraseña, haber registrado información de la empresa.

Flujo básico de éxito:	Ingreso al sistema, ver empresa, editar información empresa.
Estado:	Aprobado

Nombre:	ver catálogo de productos, editar y eliminar productos registrados
Fecha:	25/03/19
Numero requisito:	11
Contexto:	El empresario puede ver su catálogo de productos, editar y eliminar productos registrados.
Autor:	Alejandro Polanco
Entradas:	No aplica
Salidas:	mis productos, nombre, imagen precio y descripción de productos, editar y eliminar

Precondiciones:	Entrar al sistema mediante correo y contraseña, haber agregado información de productos.
Flujo básico de éxito:	Ingreso al sistema, ver mis productos, nombre, precio, descripción de productos, editar o eliminar producto.
Estado:	Aprobado

Requisitos del administrador

Nombre:	Ver, editar y eliminar información de empresas y sus catálogos de productos
Fecha:	30/03/19
Numero requisito:	12
Contexto:	Al administrador puede ver y editar información de empresas y sus productos
Autor:	Alejandro Polanco
Entradas:	Correo y contraseña

Salidas:	Empresas, productos
Precondiciones:	Ingresar al sistema mediante correo y contraseña,
Flujo básico de éxito:	Ingreso al sistema, ver empresas, ver productos
Estado:	Aprobado

Si el empresario especifica una página web de la empresa el administrador debe ingresar a la página ubicar etiquetas de los campos de producto y extraer los datos mediante web scraping el cual requiere un costo de 200000 COP mensual, el pago se hace por consignación a cuenta bancaria.

El scraping web se hará en tiempo de 3 a 5 días después de ingresado el link o dirección de la página.

Validar los anuncios pautados pagos en el sistema

Requerimientos del sistema

Cualquier usuario podrá contactar al administrador para poder pautar cualquier publicación en el sistema por 30 días efectuando su pago de 200.000 COP (medios de pago por consignación a cuenta bancaria).

3.1.2 Requisitos No funcionales.

El sistema debe utilizar VPS para poder instalar y manipular cualquier servidor, lenguaje o base de datos.

El sistema debe tener un Dominio propio

Debe tener aplicación para ios, android y web

3.2 Modelado UML

3.2.1 Diagrama de casos de uso

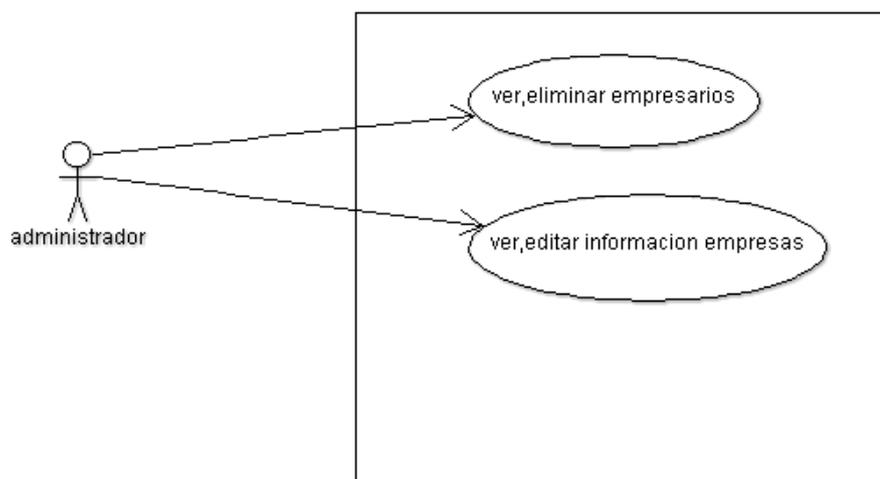
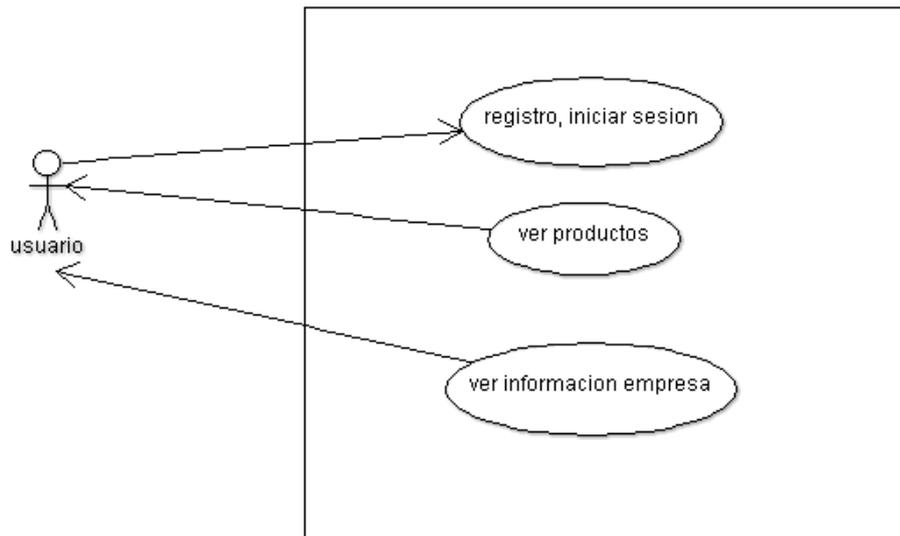
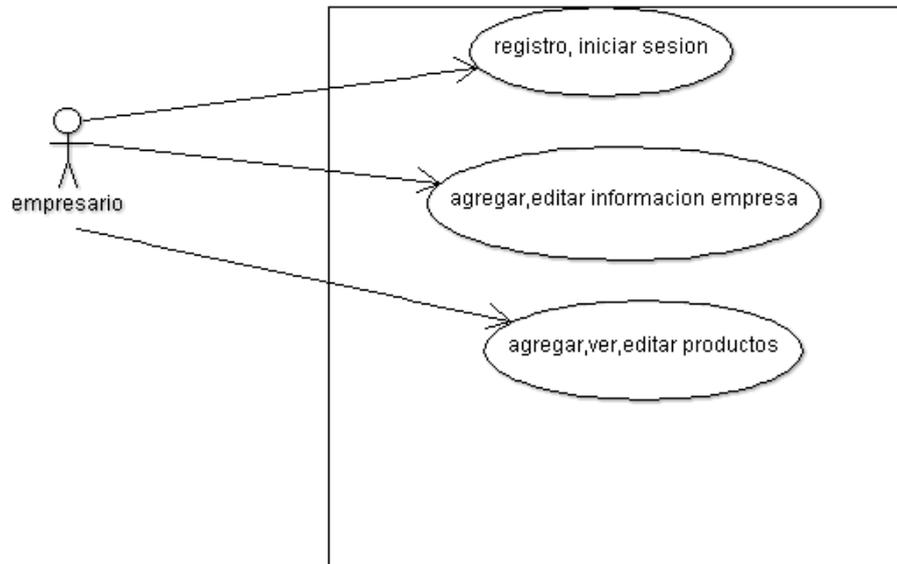
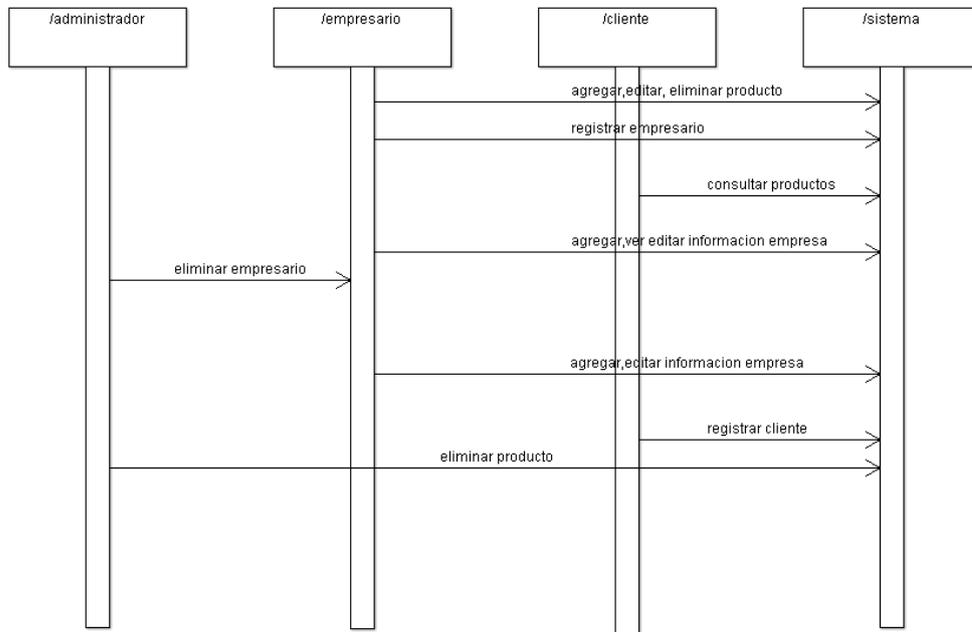
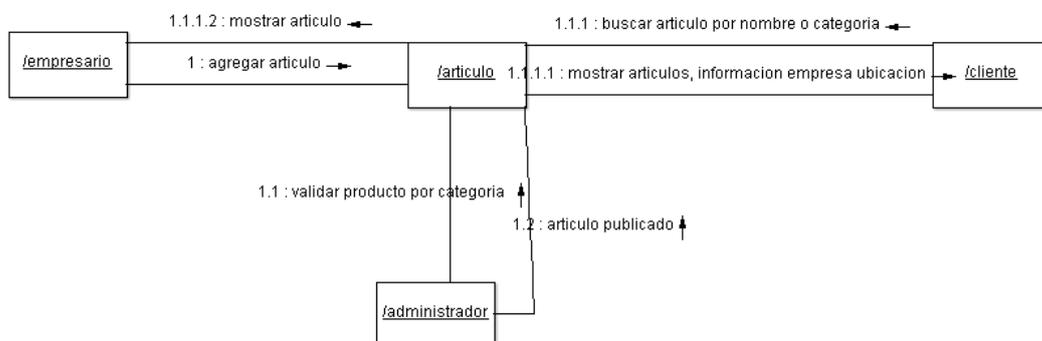


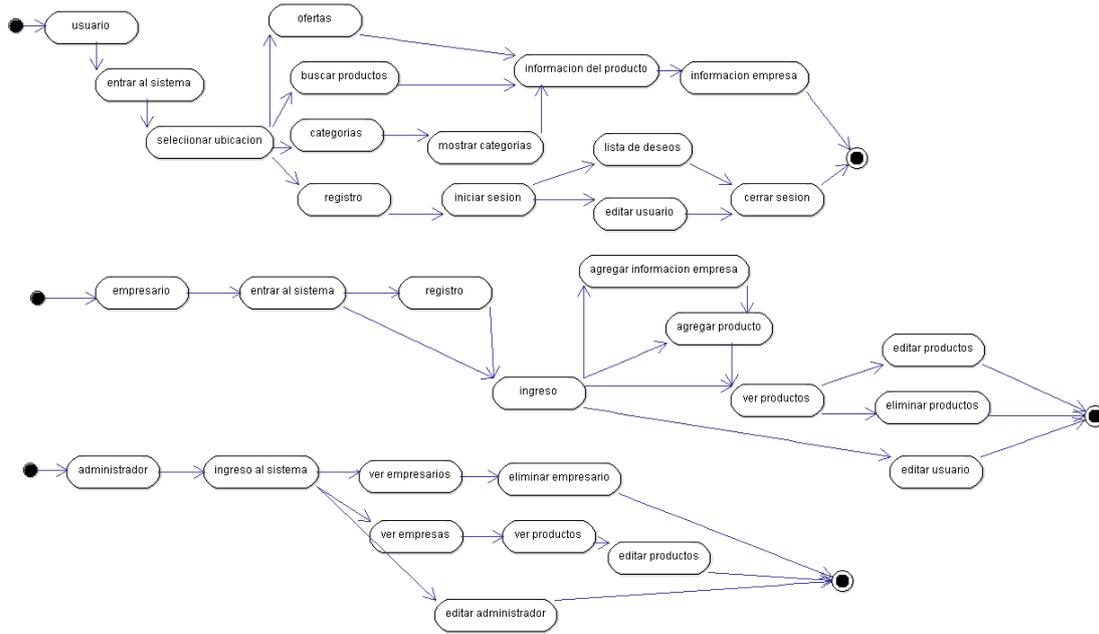
Diagrama de secuencia



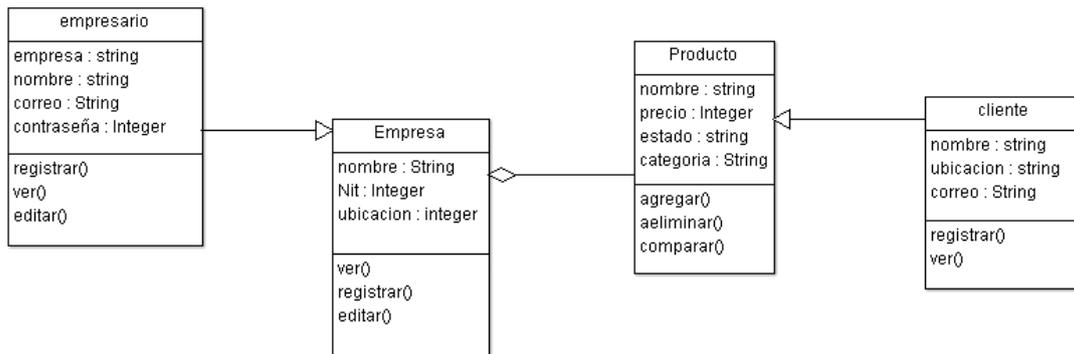
3.2.3 Diagrama de colaboración



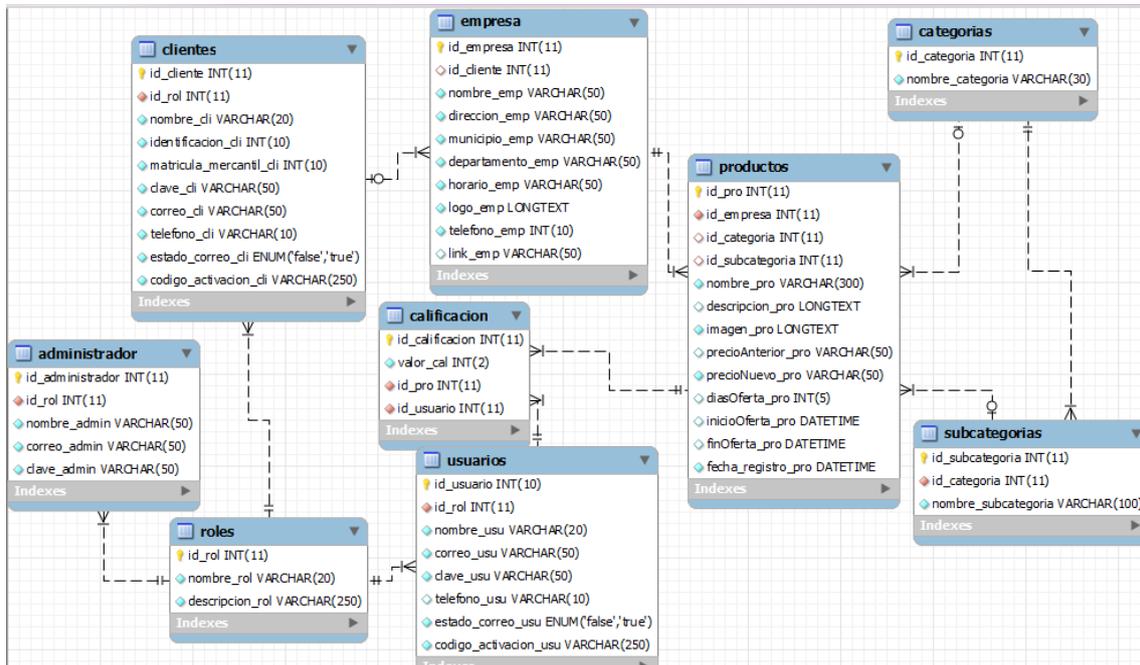
3.2.4 Diagrama de actividades



3.2.5 Diagrama de clases



3.3 Diseño de base de datos



3.4 Desarrollo de metodología Kanban

Tarjetas visuales

Cada tarjeta visual o post-it determina el nombre de la actividad a realizar, los responsables de ella y el estado en que se encuentra.

1) Ver productos en Oferta



Hecho ▾ 25 min ▾

+ lista ☰

Descripción

El usuario Puede ver productos en ofertas en la página de inicio

Adjuntos

Tipo de tarjeta

Prioridad

Fecha de vencimiento ✓

Asignado a

Tarea

▼ normal

▼ 2019-06-12

AP Alejandro Pola... ▼

actualizado hace unos segundos

2) Ver productos por categoría.



Hecho ▾

+ lista ☰

Descripción

El usuario puede ver las categorías de las empresas que están registradas.

Adjuntos

Tipo de tarjeta

Prioridad

Fecha de vencimiento

Asignado a

Tarea

▼ normal

▼

AP Alejandro Pola... ▼

actualizado hace 16 días

3) Ver productos por empresa. ✕

Hecho ▼ + lista ☰

Descripción
El usuario puede ver las empresas que están registradas y sus productos por empresa.

Adjuntos

Tipo de tarjeta	Prioridad	Fecha de vencimiento	Asignado a
Tarea	normal		AP Alejandro Pola... ▼

actualizado hace 16 días

4) Buscar productos en el sistema por nombre ✕

Haciendo ▼ 6 seg ▼ + lista ☰

Descripción
el usuario puede buscar productos en el sistema por nombre

Adjuntos

Tipo de tarjeta	Prioridad	Fecha de vencimiento	Asignado a
Mejora	normal		AP Alejandro Pola... ▼

actualizado hace unos segundos

5) Seleccionar Ubicación



Haciendo ▾

+ lista ☰

Descripción

El usuario debe seleccionar ubicación para mostrar productos según ubicación

Adjuntos

Tipo de tarjeta

Prioridad

Fecha de vencimiento

Asignado a

■ Problema / Error ▾

normal ▾

WM Wilson Eduard... ▾

actualizado hace unos segundos

6) Lista de deseos



Haciendo ▾

+ lista ☰

Descripción

El usuario debe puede agregar productos a la lista de deseos,
El usuario debe estar registrado y haber iniciado sesión en el sistema

Adjuntos

Tipo de tarjeta

Prioridad

Fecha de vencimiento

Asignado a

■ Esperando ▾

normal ▾

AP Alejandro Pola... ▾

actualizado hace 16 días

7) registro de empresario



Haciendo ▾

+ lista ☰

Descripción

El empresario debe hacer registro en el sistema mediante correo y contraseña, identificando número de celular, nombre, correo, matrícula mercantil cámara de comercio y contraseña.

Adjuntos

Tipo de tarjeta

Prioridad

Fecha de vencimiento

Asignado a

■ Esperando

▼ normal



AP Alejandro Pola... ▼

actualizado hace 16 días

8) Agregar productos al sistema



Que hacer ▾

+ lista ☰

Descripción

el empresario puede agregar productos al sistema, debe haber ingresado al sistema como empresario, y agregar información de empresa.

Adjuntos

Tipo de tarjeta

Prioridad

Fecha de vencimiento

Asignado a

■ Tarea

▼ normal



WM Wilson Eduard... ▼

actualizado hace 16 días

9) agregar información de empresa



Que hacer ▾

+ lista ☰

Descripción

el empresario puede agregar información de la empresa, debe haber ingresado al sistema como empresario

Adjuntos

Tipo de tarjeta

Prioridad

Fecha de vencimiento

Asignado a

■ Tarea

▼ normal



AP Alejandro Pola... ▼

actualizado hace 16 días

10) Ver y editar información empresa



Que hacer ▾

+ lista ☰

Descripción

el empresario puede ver y editar información de empresa, debe Entrar al sistema mediante correo y contraseña, haber registrado información de la empresa.

Adjuntos

Tipo de tarjeta

Prioridad

Fecha de vencimiento

Asignado a

■ Tarea

▼ normal



AP Alejandro Pola... ▼

actualizado hace 16 días

✕

11) ver catálogo de productos, editar y eliminar productos registrados

Que hacer ▾ + lista ☰

Descripción

El empresario puede ver su catálogo de productos, editar y eliminar productos registrados, debe haber ingresado al sistema mediante correo y contraseña, haber agregado información de productos.

Adjuntos

Tipo de tarjeta	Prioridad	Fecha de vencimiento	Asignado a
Tarea	▾ normal	▾	AP Alejandro Pola... ▾

actualizado hace 16 días

✕

12) Ver y editar e información de empresas y sus catálogos de productos

Que hacer ▾ + lista ☰

Descripción

Administrador puede ver y editar información de empresas y sus productos, debe Ingresar al sistema mediante correo y contraseña,

Adjuntos

Tipo de tarjeta	Prioridad	Fecha de vencimiento	Asignado a
Tarea	▾ normal	▾	WM Wilson Eduard... ▾

actualizado hace 16 días

Equipo de trabajo

Nombre			
Alejandro Polanco Espitia	xespitia@hotmail.com	✎ editar	🚫 suspender
Wilson Eduardo Martinez ●●●	wilson-t1994@hotmail.com	✎ editar (usted)	

Tablero Kanban

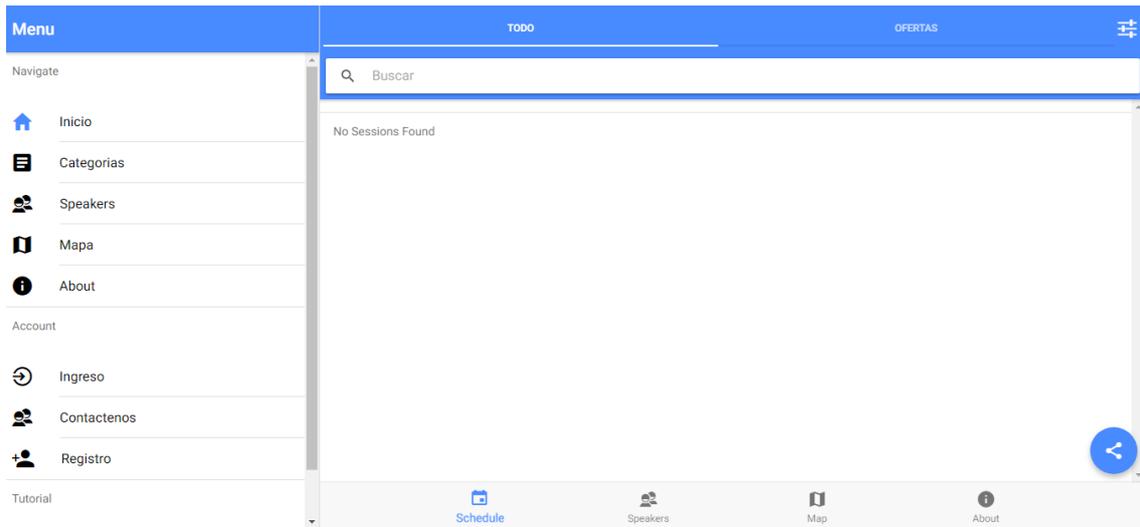
El tablero kanban agrupa todas las actividades a realizar y el estado en que se encuentran para ser visualizadas por todo el equipo de trabajo, así visualizar los límites de trabajo.



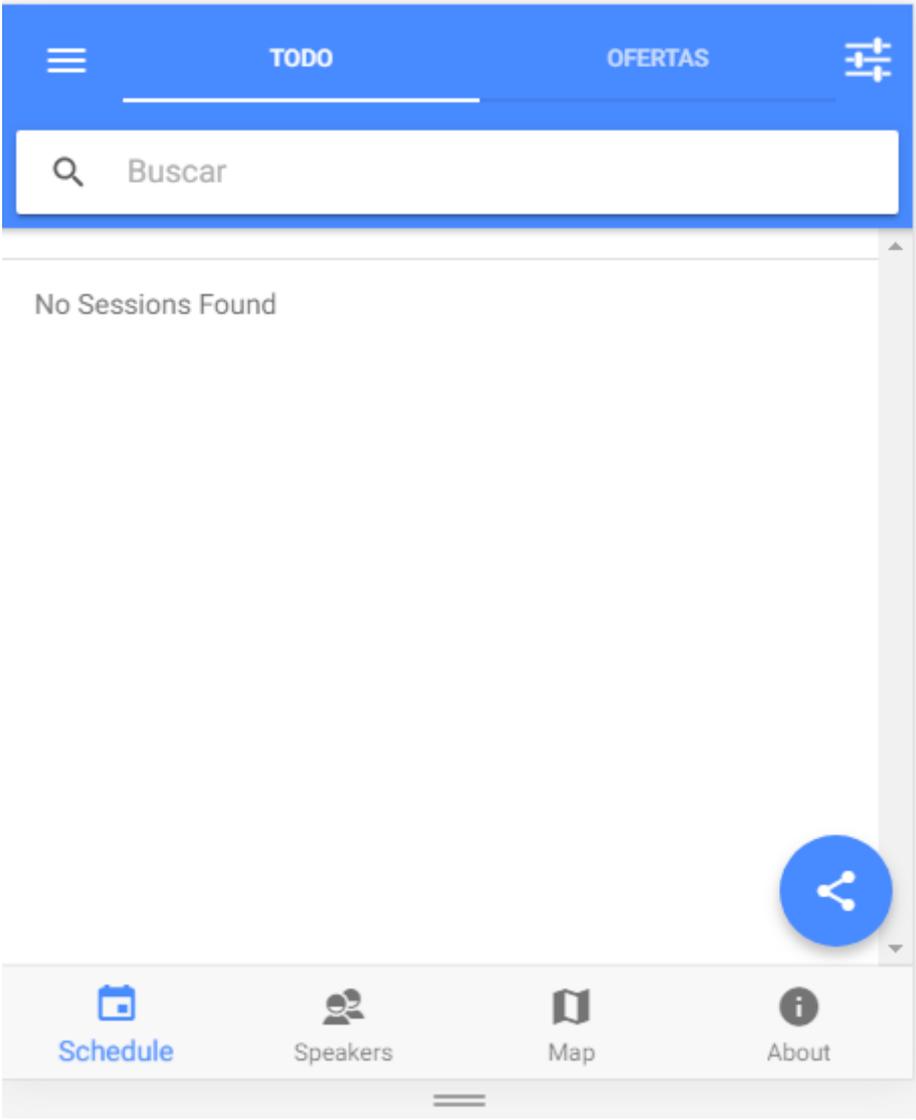
3.5 Prototipo del sistema

La elaboración del prototipo del sistema muestra cómo se visualizará para poder ser desarrollado.

Prototipo web vista principal



Prototipo móvil vista principal



Menu

OFERTAS



Navigate



Inicio



Categorias



Speakers



Mapa



About

Account



Ingreso



Contactenos



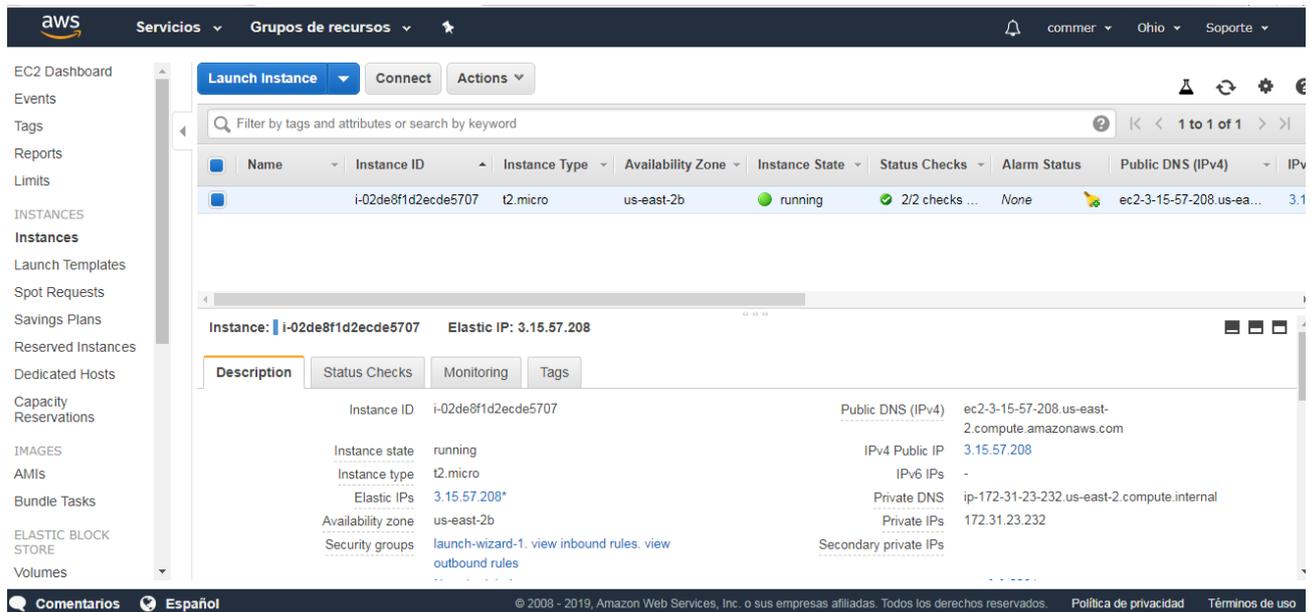
Registrar



About

3.5 configuración del servidor Amazon EC2

Creamos la instancia de ec2 con el sistema operativo UBUNTU 16.4, recursos 2gb de memoria ram 4 cores de procesamiento.



La instalación del servidor nginx

3.6 Implementación

En la implementación utilizamos 4 empresas, dos para la técnica de web scraping y dos para que hicieran uso de la plataforma y subir sus productos manualmente.

Web scraping

```

54
55 public ComerlineServlet() {
56     super();
57     // TODO Auto-generated constructor stub
58 }
59
60 // tiempo
61 // java.util.Date utilDate = new Date();
62 // java.sql.Date Timestamp = new java.sql.Date(utilDate.getTime());
63
64 /**
65  * @see HttpServlet#doGet(HttpServletRequest request, HttpServletResponse
66  *     response)
67  */
68 protected void doGet(HttpServletRequest request, HttpServletResponse response)
69     throws ServletException, IOException {
70
71     Connection conn = null;
72
73     ArrayList<String> list = new ArrayList<String>(); // guardar urls de categoria
74     ArrayList<String> list2 = new ArrayList<String>(); // guardar urls de ofertas
75     ArrayList<String> list3 = new ArrayList<String>(); // guardar urls de nombre de cada producto
76
77     // java.net.UnknownHostException
78     System.setProperty("systemProp.https.proxyHost", "localhost");
79     System.setProperty("systemProp.https.proxyPort", "8080");
80
81     String url = "https://www.alkomprar.com/ofertas";
82     String url2 = "https://www.alkomprar.com";
83
84     Elements elements = Jsoup.connect(url).timeout(60000).get().select(
85         ".content > .container-fluid > .sections > .sectionofertas > .col-12.boton-ofertas-general > a");// etiquetas

```

```

185
186 for (int x = 0; x < list3.size(); x++) {
187     Document doc = Jsoup.connect(list3.get(x)).timeout(120000)
188         .userAgent("Mozilla/5.0 (Windows NT 6.1; Win64; x64; rv:25.0) Gecko/20100101 Firefox/25.0")
189         .timeout(60000).ignoreHttpErrors(true).get();// se conecta a cada uno de los links
190         // nombre del producto
191
192     Elements name = doc.select(".product-name.font-title.font-title--sub-title-2 > h1");
193
194     if (!name.isEmpty()) {
195
196         String namearticulo = name.text();
197         namearticulo = namearticulo.replaceAll("'", ""); // quitar apostrofe
198
199         Elements descripcion = doc.select(".tab-details__content.col-md-6 > ul > li");
200
201         Elements precio = doc.select(
202             ".price-block.col-xs-12.col-sm-12.col-md-12.col-lg-12.d-inline.no-padding > .row > .col-xs-12 > span.col-xs-12.font-title
203
204         Elements precioAnterior = doc.select(
205             ".price-block.col-xs-12.col-sm-12.col-md-12.col-lg-12.d-inline.no-padding > .row > .col-xs-12 > span.col-xs-12.font-title
206
207         Elements imagen = doc.select(".image-modal-zoom-out-cover > img");
208
209         Elements li = doc.select(".breadcrumb-section.hidden-xs.hidden-sm > ol > li > a");
210
211         Element categoria = li.get(1); // obtener el segundo elemento
212         Element subcategoria = li.get(2); // obtener el tercer elemento
213
214         String names = namearticulo;
215         String priceold = precioAnterior.text();
216         String pricenew = precio.text().replaceAll("[^0-9]", "");
217         String image = imagen.attr("src");

```

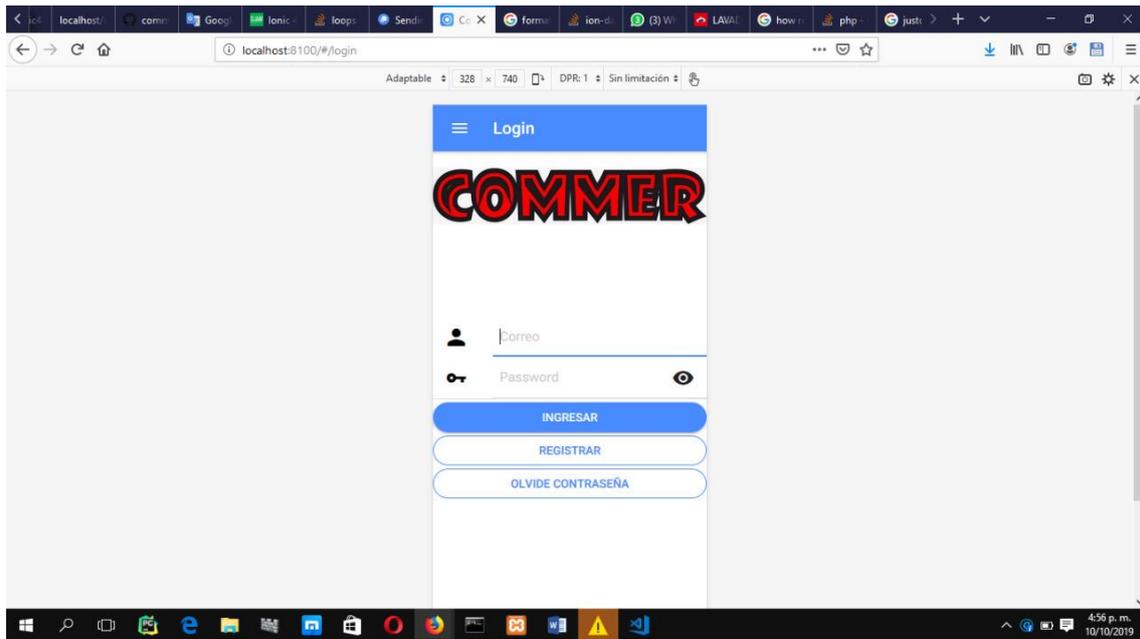
```

1 package comerline;
2
3 import java.io.Serializable;
11
12 /**
13  * Entity implementation class for Entity: ProductosScraping
14  *
15  */
16
17
18 @Entity
19 @Table(uniqueConstraints=
20 @UniqueConstraint(columnNames = {"nombre_pro"}))
21
22 public class Productos implements Serializable {
23     private static final long serialVersionUID = 1L;
24
25
26 @Id
27 @TableGenerator(name = "ProductosScraping_Gen", initialValue = 1)
28 @GeneratedValue(strategy=GenerationType.AUTO)
29 // @GeneratedValue(strategy=GenerationType.AUTO)
30 // @TransactionAttribute(TransactionAttributeType.REQUIRES_NEW)
31
32     private Integer id_pro;
33     private String nombre_pro;
34     private String precioAnterior_pro;
35     private String precioNuevo_pro;
36     private String descripcion_pro;
37     private String imagen_pro;
38     private String categoria_pro;
39     private Integer id_empresa;
40     private Timestamp fecha_registro_pro;

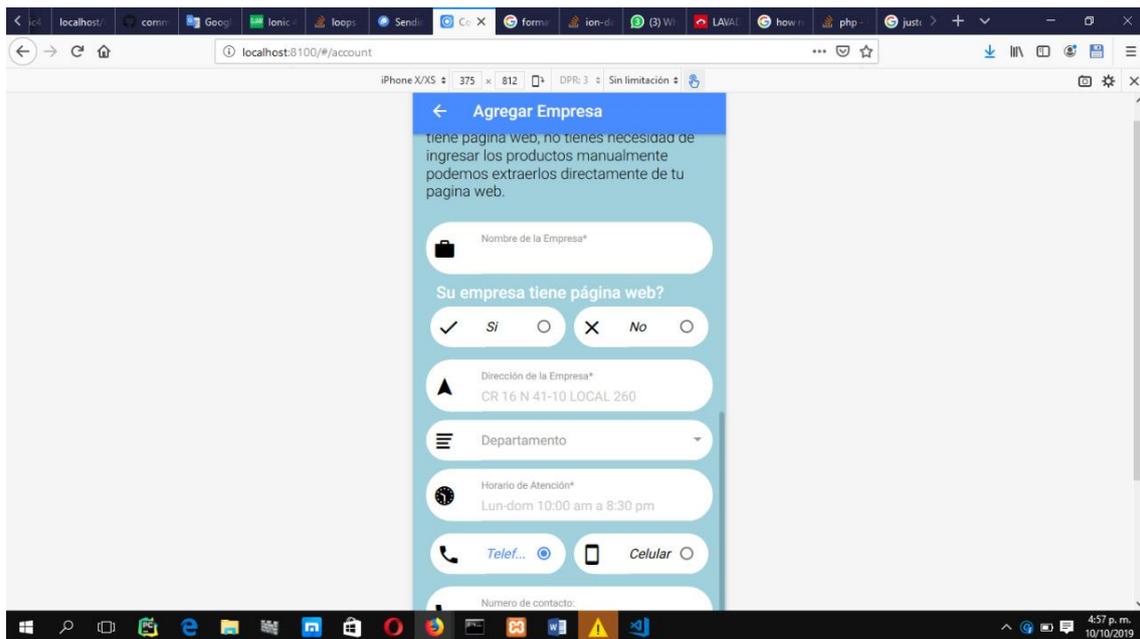
```

Forma Manual

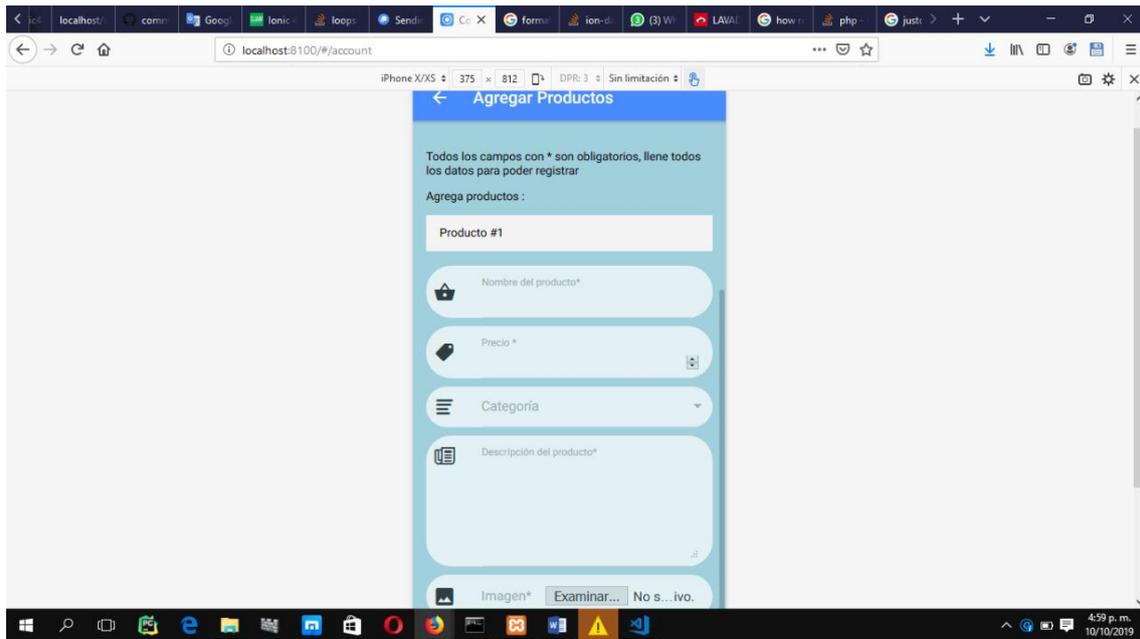
Se hace un registro e inicio de sesión de usuarios.



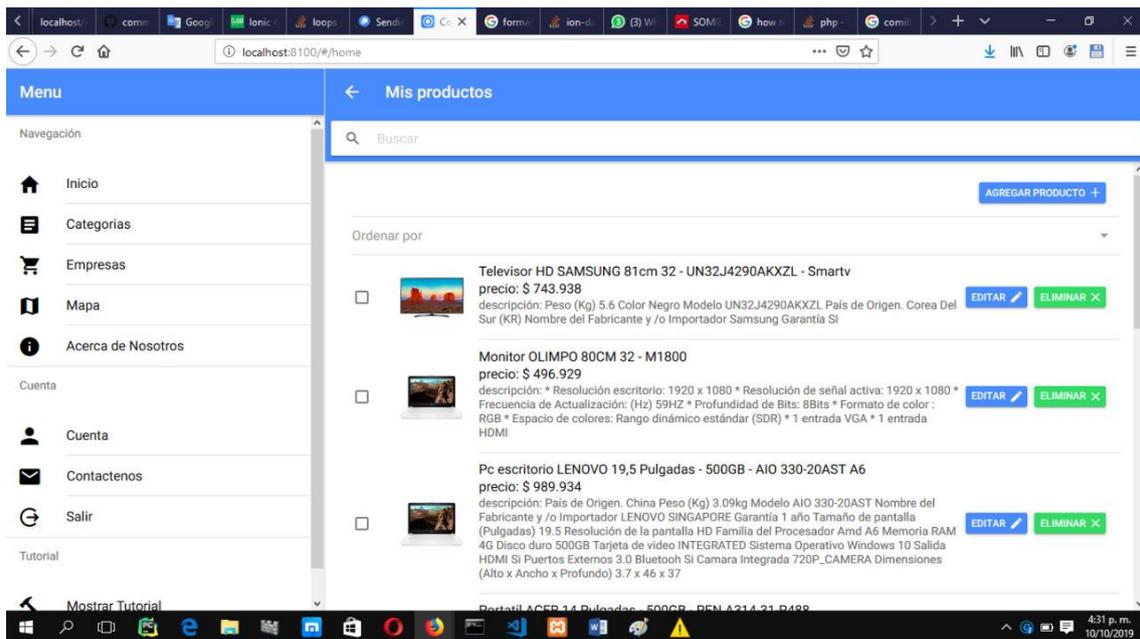
Agrega información de empresa



Agrega información del producto



Ver editar productos



<http://www.commerapp.com/>

Seguimiento y control

4.1 Métricas de calidad de software

Las métricas de calidad de software se han definido en estándares desde la época del 1991 identificando en ellos cada uno de los componentes y características medibles en el proceso de evaluación de productos de software. El estándar ISO/IEC 9126 se compone de 4 partes

Modelo de calidad

Métricas externas

Métricas internas

Métricas para la calidad en uso

Este estándar propone la calidad de productos de software basados en la funcionabilidad, fiabilidad, usabilidad, eficiencia, mantenibilidad y portabilidad que son definidos en el modelo más actual en las normas ISO 25000(Lopez, Sánchez, & Antonio, 2016).

4.1.1 Métrica orientada a la función

La métrica orientada a la función se utiliza para predecir el tamaño de un sistema

El roll de Empresario en el sistema cumple 3 entradas 2 salidas

Parámetro	Simple	Mediana	Compleja	Total
Entradas	3	(3)4	6	12
Salidas	2	(3)5	7	15

Consultas	3	(4)4	6	16
Archivos	2	(1)10	15	10
Interacciones externas	0	(2)7	10	14
				67

Producto Completo

Utilizamos la fórmula del modelo de Pressman

$$PF = \text{cuenta-total} * (0.65 + 0.01 * \sum Fi)$$

Donde cuenta-total es la suma de todas las entradas PF obtenidas en la tabla y Fi (i=1 a 14) son “valores de ajuste de complejidad”. Para el propósito de este ejemplo, asumimos que $\sum Fi$ es 45 (un producto moderadamente complejo) Por lo tanto:

$$Pf = 67 \times (0.65 + 0.01 \times 45) = 74$$

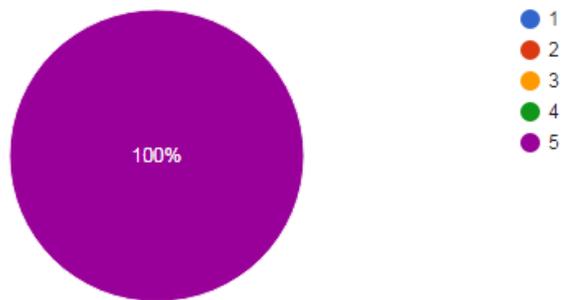
4.1.2 Métrica orientada a la Usabilidad

EFFECTIVIDAD EFICIENCIA Y SATISFACCION DEL USUARIO

Se realizó la siguiente encuesta a 2 usuarios del sistema para medir la usabilidad del software.

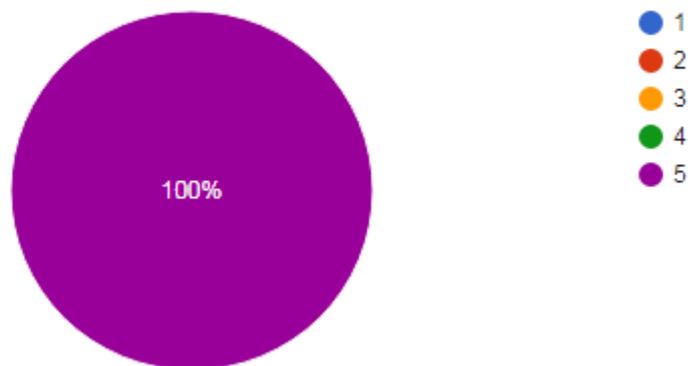
Me imagino que la mayoría de la gente aprendería a usar este sistema muy rápidamente.

2 respuestas



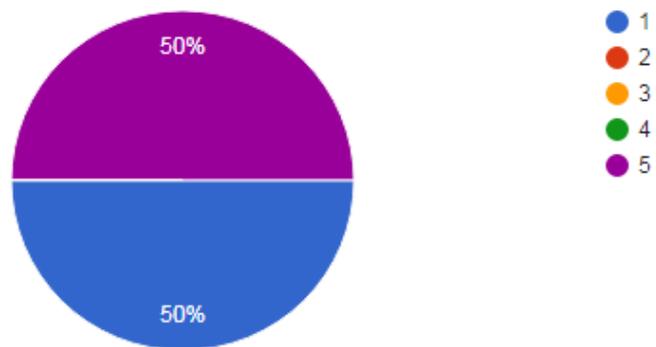
Me sentí muy seguro usando el sistema.

2 respuestas



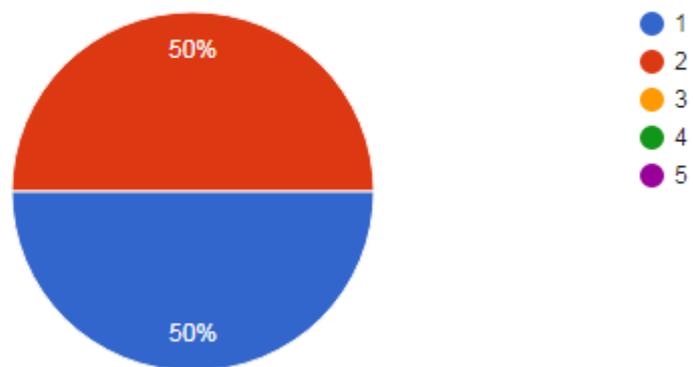
Necesitaba aprender muchas cosas antes de poder comenzar con este sistema.

2 respuestas



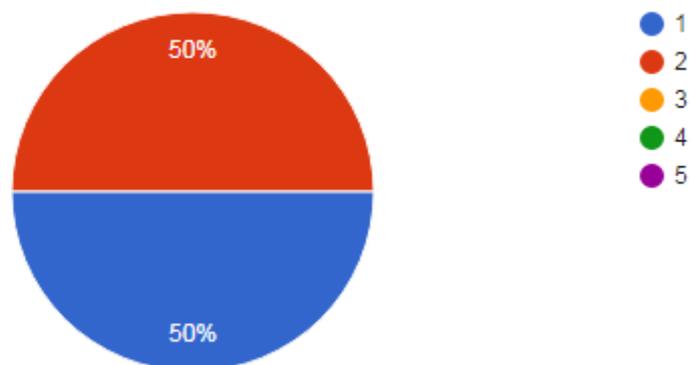
Encontré el sistema innecesariamente complejo.

2 respuestas



¡Pensé que había demasiada inconsistencia en este sistema!

2 respuestas



TOTAL, PUNTUACION OPTENIDA 33 DE 50

LA USABILIDAD DEL SISTEMA ES DE 66%

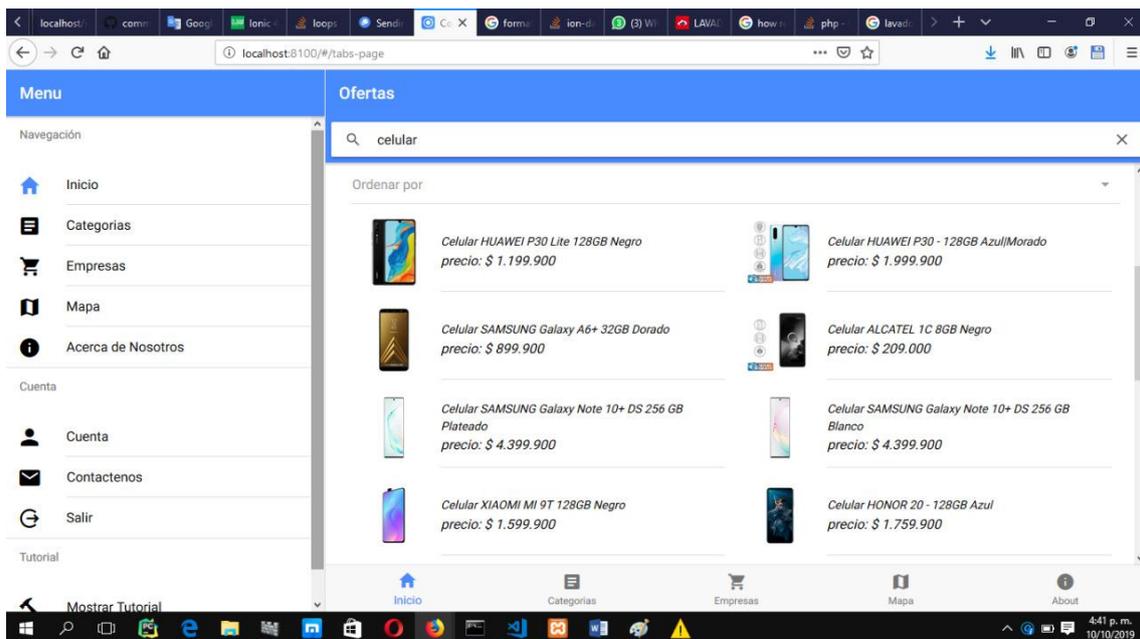
Capítulo

5

Conclusiones y cierre

5.1 Resultados

Se aplicó la técnica de web scraping y se capturaron los datos de dos empresas Alkomprar y Éxito promocionando sus productos. Al ser capturados se organizan, agrupan y almacenan para poder ser mostrados al usuario en la aplicación móvil y web.



Aplicamos la metodología de desarrollo kanban para mejorar el flujo de trabajo, 10 procesos pmbok para la gestión del proyecto teniendo un flujo de trabajo continuo aceptando cambios durante la ejecución del proyecto.

<http://www.commerapp.com/>

5.2 Conclusiones y recomendaciones

Aplicando los procesos de gestión de proyectos PMBOK, 5 áreas del conocimiento y 10 procesos con sus herramientas EDT (Estructura de Desglose del Trabajo) y técnicas juicio de expertos para toma de decisiones teniendo en cuenta la experiencia en el área aplicada, la metodología ágil Kamban y el ciclo de vida del software. Se desarrolló el proyecto de forma eficaz en donde Kanban permitió la visualización completa del flujo de trabajo y cada una de las actividades a realizar en el área de desarrollo. Así mismo se evalúa la calidad del software con métricas de calidad orientada a la función y a la usabilidad del sistema.

Se desarrolló la aplicación móvil y web permitiendo agregar, buscar y comparar productos en el sector minorista ejecutándose en cualquier dispositivo. Esto mediante el desarrollo híbrido desde framework IONIC. De esta manera influir en la toma de decisiones de los usuarios.

El uso del lenguaje php en el servidor nos permitió recibir y hacer peticiones del CRUD (crear, leer, actualizar y borrar) cabe resaltar que el uso de un framework en el servidor permitiría un desarrollo ágil. Sin embargo, el buen uso de un lenguaje nativo como php

y la experiencia de un equipo de desarrollo en este lenguaje influyen en la forma eficaz de realizarlo.

Los servicios de computación en la nube de AWS (AMAZON WEB SERVICES) permitió la instalación de un servidor web NGINX (que principalmente permite el uso de usuarios al mismo tiempo 10 mil aprox) instalado sobre UBUNTU SERVER con 2 gb de memoria ram, 2 cores de procesamiento y 100 gb de almacenamiento totalmente escalable. Así mismo se controlan cualquier tipo de cambios en la configuración del servidor tales como el uso de lenguajes de programación requeridos o en el caso de web scraping que se ejecuta mediante un cron de Linux programado desde el servidor.

Se implementó un servicio web utilizando web scraping con las librerías de java jsoup capturando los datos de 2 empresas del sector minorista. De esta manera se almacenaron y trataron la mayor cantidad de productos de algunas empresas que ya proveen un sistema de información.

5.3 Anexos

5.3.1 Plan de proyecto

5.3.1.1 Acta de constitución del proyecto o acta de inicio

Información del proyecto

Datos

Empresa / Organización	Universidad Surcolombiana
Proyecto	Sistema de información publicitario de productos en almacenes de cadena y supermercados utilizando web scraping (Commer).

Fecha de preparación	30/11/2018
Cliente	Empresarios del sector minorista (supermercados y almacenes de cadena)
Patrocinador principal	Wilson Martinez, Alejandro Polanco
Gerente de proyecto	Wilson Martinez

Patrocinador / Patrocinadores

Nombre	Cargo	Departamento / División	Rama ejecutiva (Vicepresidencia)
Wilson Martinez	Jefe de proyecto	gerencia	
Alejandro Polanco	Analista y desarrollador	desarrollo	

Propósito y justificación del proyecto

Construir un sistema de información que permita capturar, organizar, agrupar, buscar y comparar los precios y características de productos de supermercados y almacenes de cadena en Colombia, utilizando la técnica de web scraping para obtener la información de los productos de manera automática mediante una aplicación móvil y

web para mejorar los sistemas de información existentes, estableciendo comunicación directa entre el usuario y empresa e incentivar el comercio de esta manera que sean de fácil acceso y búsqueda desde la mayoría de dispositivos.

Requerimientos de alto nivel

Requerimientos del producto

El software tiene que ser portable a la mayoría de dispositivos móviles desarrollando página web y aplicaciones Android y IOS

Requerimientos del proyecto

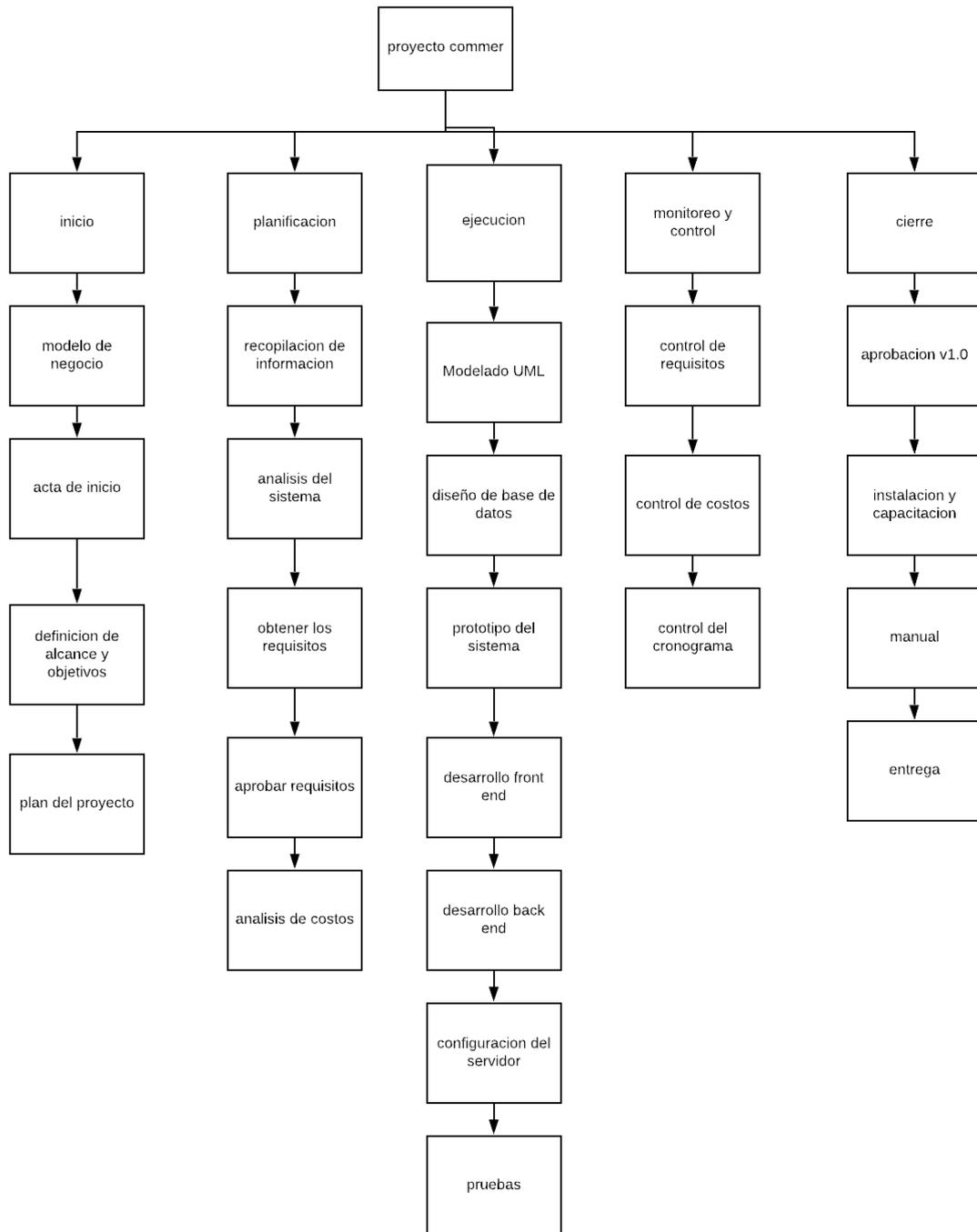
Se tendrá en cuenta la metodología ágil y las buenas prácticas de desarrollo.

Elegir la tecnología más adecuada para el desarrollo del proyecto.

5.3.2 Gestión del alcance

5.3.2.1 EDT (Estructura de desglose de trabajo)

Es la descomposición jerárquica del trabajo a realizar definiendo un alcance del proyecto.



5.3.2.2 Diccionario de la EDT

Código EDT	1	Nombre del paquete de trabajo	Modelo de negocio
Descripción		Se realiza el modelo de negocio con el equipo de trabajo	
Criterio de aceptación		100%	
Requisitos de calidad			
Entregables		Si	
Duración estimada		2 días	Costo 240000
Código EDT	2	Nombre del paquete de trabajo	Acta de inicio
Descripción		Se realiza el acta de inicio según los requisitos del proyecto	
Criterio de aceptación		100%	
Requisitos de calidad			
Entregables		Si	
Duración estimada		1 día	Costo 120000
Código EDT	3	Nombre del paquete de trabajo	Definición de alcance y objetivos
Descripción		Se definen el alcance y los objetivos generales y específicos del proyecto	
Criterio de aceptación			
Requisitos de calidad		100%	
Entregables		Si	
Duración estimada		3 días	Costo 360000
Código EDT	4	Nombre del paquete de trabajo	Plan de proyecto
Descripción		Se estruc	
Criterio de aceptación		100%	
Requisitos de calidad			
Entregables		Si	
Duración estimada		2 días	Costo 240000
Código EDT	5	Nombre del paquete de trabajo	Recopilación de información
Descripción		Se aplica el método de recopilación de información para hacer análisis del sistema	
Criterio de aceptación		100%	

Requisitos de calidad			
Entregables		Si	
Duración estimada		2 días	Costo 1845000
Código EDT	6	Nombre del paquete de trabajo	Análisis del sistema
Descripción		Se realiza el análisis del sistema	
Criterio de aceptación			
Requisitos de calidad		100%	
Entregables		Si	
Duración estimada		6 días	Costo 720000
Código EDT	7	Nombre del paquete de trabajo	Obtener los requisitos
Descripción		Se aplican los métodos para obtener los requisitos del sistema	
Criterio de aceptación		100%	
Requisitos de calidad			
Entregables		Si	
Duración estimada		6 días	Costo 720000
Código EDT	8	Nombre del paquete de trabajo	Aprobación de requisitos
Descripción		Se realiza reunión para aprobar requisitos del sistema	
Criterio de aceptación		100%	
Requisitos de calidad			
Entregables		Si	
Duración estimada		1 día	Costo 120000
Código EDT	9	Nombre del paquete de trabajo	Análisis de costos
Descripción		Se realiza el análisis de costos por cada actividad a realizar	
Criterio de aceptación			
Requisitos de calidad		100%	
Entregables		Si	
Duración estimada		1 días	Costo 120000

Código EDT	10	Nombre del paquete de trabajo	Diseño UML
Descripción		Se realiza el diseño de base de datos teniendo en cuenta el análisis del sistema y los requisitos obtenidos	
Criterio de aceptación			
Requisitos de calidad		100%	
Entregables		Si	
Duración estimada		5 días	Costo 1200000
Código EDT	11	Nombre del paquete de trabajo	Prototipo del sistema
Descripción		Se realiza un prototipo del sistema identificando el diseño UML previamente realizado	
Criterio de aceptación			
Requisitos de calidad		100%	
Entregables		Si	
Duración estimada		1 días	Costo 240000

Código EDT	12	Nombre del paquete de trabajo	Desarrollo front end
Descripción		Se desarrollan las vistas del sistema previamente validando cada uno de los campos de datos y se preparan para ser enviados al servidor	
Criterio de aceptación			
Requisitos de calidad		100%	
Entregables		Si	
Duración estimada		28 días	Costo 4026666
Código EDT	13	Nombre del paquete de trabajo	Desarrollo back end

Descripción	se realiza el desarrollo back end teniendo en cuenta el análisis del sistema y los datos enviados por el front end		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		
Duración estimada	28 días	Costo	4826666
Código EDT	14	Nombre del paquete de trabajo	Configuración del servidor
Descripción	Se realiza la configuración del servidor teniendo en cuenta la arquitectura del proyecto aprobada en el análisis		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		
Duración estimada	5 días	Costo	1400000
Código EDT	15	Nombre del paquete de trabajo	pruebas
Descripción	Se realizan las pruebas de software teniendo en cuenta las métricas de calidad		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		
Duración estimada	5 días	Costo	1266666
Código EDT	16	Nombre del paquete de trabajo	Control de requisitos
Descripción	Se hace control a los requisitos aprobados durante el desarrollo		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		

Duración estimada	6 días	Costo	720000
Código EDT	17	Nombre del paquete de trabajo	Control de costos
Descripción	Se realiza un control a costos por cada actividad		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		
Duración estimada	6 días	Costo	720000
Código EDT	18	Nombre del paquete de trabajo	Control del cronograma
Descripción	Se realiza un control y ajuste del cronograma		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		
Duración estimada	6 días	Costo	720000
Código EDT	19	Nombre del paquete de trabajo	Aprobación primera versión
Descripción	Se realiza la aprobación de la primera versión		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		
Duración estimada	0 días	Costo	0
Código EDT	20	Nombre del paquete de trabajo	Instalación y capacitación
Descripción	Se realiza la instalación del software y se capacita para su manejo		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		

Duración estimada	2 días	Costo	240000
Código EDT	21	Nombre del paquete de trabajo	manual
Descripción	Se definen el alcance y los objetivos generales y específicos del proyecto		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		
Duración estimada	1 días	Costo	240000
Código EDT	22	Nombre del paquete de trabajo	Entrega
Descripción	Se definen el alcance y los objetivos generales y específicos del proyecto		
Criterio de aceptación			
Requisitos de calidad	100%		
Entregables	Si		
Duración estimada	1 días	Costo	240000

5.3.2 Diccionario de datos

administrador

Columna	Tipo de dato	Pk	comentario
id_admin	INT(11)	X	
id_rol_admin	INT(11)		
nombre_admin	VARCHAR(50)		

correo_admin	VARCHAR(50)		
clave_admin	VARCHAR(50)		

calificacion

Columna	Tipo de dato	Pk	comentario
id_calif	INT(11)	X	
valor_calif	INT(11)		
id_pro_calif	INT(2)		
id_usuario_califi	INT(11)		
clave_admin	INT(11)		

clientes

Columna	Tipo de dato	Pk	comentario
id_clien	INT(11)	X	
id_rol_client	INT(11)		
nombre_client	VARCHAR(20)		
identificacion_client	INT(10)		
matricula_mercantil_client	INT(10)		
clave_client	VARCHAR(50)		
correo_cli	VARCHAR(50)		
telefono_cli	VARCHAR(10)		
estado_correo_cli	ENUM('false','true')		
codigo_activacion_cli	VARCHAR(250)		

productos

Columna	Tipo de dato	Pk	comentario
id_produ	INT(11)	X	
id_empresa_produ	INT(11)		
id_categoria_produ	INT(11)		
id_subcategoria_produ	INT(11)		
nombre_produ	VARCHAR(300)		
descripcion_produ	LONGTEXT		
imagen_produ	LONGTEXT		
precioAnterior_produ	VARCHAR(50)		
precioNuevo_produ	VARCHAR(50)		
diasOferta_produ	INT(5)		
inicioOferta_pro	DATETIME		
finOferta_pro	DATETIME		

fecha_registro_pro	DATETIME		
--------------------	----------	--	--

empresa

Columna	Tipo de dato	Pk	comentario
id_empre	INT(11)	X	
id_cliente_empre	INT(11)		
nombre_empre	VARCHAR(50)		
direccion_empre	VARCHAR(50)		
municipio_empre	VARCHAR(50)		
departamento_empre	VARCHAR(50)		
horario_empre	VARCHAR(50)		
logo_empre	LONGTEXT		
telefono_empre	INT(10)		
link_empre	VARCHAR(50)		

Roles

Columna	Tipo de dato	Pk	comentario
id_rol	INT(11)	X	
nombre_rol	VARCHAR(50)		
descripcion_rol	VARCHAR(250)		

subcategorias

Columna	Tipo de dato	Pk	comentario
id_subca	INT(11)	X	
id_categ_subca	INT(11)		
nombre_subca	VARCHAR(100)		

Usuarios

Columna	Tipo de dato	Pk	comentario
id_usuar	INT(11)	X	
id_rol_usuar	INT(11)		
nombre_usuar	VARCHAR(50)		
correo_usuar	VARCHAR(50)		
clave_usuar	VARCHAR(50)		
telefono_usuar	VARCHAR(50)		

estado_correo_usuar	VARCHAR(50)		
codigo_activacion_usuar	LONGTEXT		

Bibliografía

- Cano Insa, E. (2015). BiciPark: Web App con Ionic y acceso a servicios Rest, 37–38. Retrieved from http://rua.ua.es/dspace/bitstream/10045/49988/1/BiciPark_Aplicacion_Web_para_dispositivos_moviles_con_I_CANO_INSA_ESTEBAN.pdf
- CECARM. (2012). Analizamos el primer portal que aglutina geográficamente catálogos y folletos de tiendas y centros comerciales. Retrieved from <https://www.cecarm.com/servlet/s.SI?METHOD=DETALLENOTICIA&sit=c,731,m,2629&id=32715>
- Escobar, F., & Cobos, C. (2015). Business Processes as a Strategy to Improve Requirements Elicitation in Extreme Programming, (December).
- Hansson, N., & Vidhall, T. (2016). Effects on performance and usability for cross-platform application development using React Native.
- Henríquez, N., Iglesias, A., Ramos, L. A., & Ropain, Y. (2013). Postgresql una alternativa efectiva en las empresas, 1–5.
- Juan, T., Cano, C., Miguel, C., & Calafate, T. (2015). Diseño de una aplicación Android para la recogida de datos de tráfico y su comunicación a un servidor.
- Kremer, M. (2019). Ionic vs. React Native: una guía de comparación. Retrieved from <https://ionicframework.com/enterprise/resources/articles/ionic-vs-react-native-a-comparison-guide>
- Lopez, A. V., Sánchez, A., & Antonio, G. (2016). Definición de Métricas de Calidad para Productos de Software, (4700), 483–488.
- Manzaba, J. G. (2014). utilizando una metodología ágil aplicada en una fábrica de software en la ciudad de Guayaquil ., 27(Diciembre), 1–36.
- MIRANDA, V. R. Z. (2016). DESARROLLO DE UNA APLICACIÓN WEB UTILIZANDO EL SERVIDOR NGINX EN LA COMPAÑÍA “GROUP TEKTRON.”
- Omar, D., & Flores, A. (2018). *DISEÑO DE APLICACIONES Y SERVICIOS CON JAVA*.
- Pérez Pérez, M. J. (2012). Guía Comparativa de Metodologías Ágiles. *Universidad de Valladolid*, 3–117. Retrieved from <https://uvadoc.uva.es/bitstream/10324/1495/1/TFG-B.117.pdf>
- PRECIADO, J. A. (2013). VIRTUALIZACIÓN DE SERVIDORES PARA LA NUBE DE LA CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES.
- Project Management Institute. (2013). *Pmbook*.
- Puerta, D. E., & Arias, L. F. B. (2014). Aplicaciones Modulares en JavaScript Modular Applications in JavaScript Revista Digital TIA Tecnología Investigación y Academia, (1), 17–29.
- Rincon, H. M., Niño, F. L., & Rodríguez, D. (2017). Retail en Colombia 2010 - 2015 : Un estudio a

partir del análisis financiero integral como elemento de soporte para la toma de decisiones.

SANGA, L. V. C. (2017). SISTEMA DE AUTOGESTIÓN DE LA SALUD PARA PACIENTES CON DIABETES Y ASMA, DESARROLLADO E IMPLEMENTADO EN UNA PLATAFORMA ANDROID; CON MONITOREO DE UNA APLICACIÓN WEB EN PHP DIRIGIDA A LOS MÉDICOS TRATANTES, ENFOCADO EN LA INFRAESTRUCTURA TECNOLÓGICA PARA EL E.

Seppe, A., Baesens, B., & Seppe. (2018). *Practical Web Scraping for Data Science*. Retrieved from <https://link.springer.com/content/pdf/10.1007%2F978-1-4842-3582-9.pdf>

Silva, H. (2011). Comportamiento de las superficies de retail en Colombia, Harold Silva Guerra, 2011.

Troyano, J. A., Cruz, F., González, M., Vallejo, C. G., & Toro, M. (2018). Introducción a la Programación con Python , Computación Interactiva y Introducción a la Programación con Python , Computación Interactiva y Aprendizaje Significativo, (July).

Vílchez, R. E. B. (2019). ARQUITECTURA DE BACK END CON AMAZON WEB SERVICES (AWS) PARA SISTEMAS ESCOLARES.

Vizcay, P. D., & Zunino, A. (2017). Restful Objects para Ruby , un framework para el desarrollo de Web Services RESTful por Resumen.

Yaguapaz, L. (2018). UNIVERSIDAD TÉCNICA DEL NORTE Facultad de Ingeniería en Ciencias Aplicadas Carrera de Ingeniería en Sistemas Computacionales ESTUDIO DEL FRAMEWORK IONIC 2 PARA EL DESARROLLO DE APLICACIONES MÓVILES HÍBRIDAS Trabajo de grado previo a la obtención del título.

Yáñez, L. H. (2014). Fundamentos de la programación.

Zhingri, C. A. V. (2015). "Análisis de rendimiento entre la base de datos relacional: MySQL y una base de datos no relacional: MongoDB."