



PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito final para obtener el
Título de INGENIERO DE SISTEMAS

www.bdigital.ula.ve

DESARROLLO DE UN MODELO SELECTIVO DE TECNOLOGÍAS BIG DATA PARA CONSULTORÍA EMPRESARIAL

Por

Br: Laura Alejandra Rosales Nencka

Tutor: Prof. Luis Alfonso Dávila

Mayo de 2024

©2024 Universidad de Los Andes Mérida, Venezuela

C.C. Reconocimiento

RESUMEN

DESARROLLO DE UN MODELO SELECTIVO DE TECNOLOGÍAS BIG DATA PARA CONSULTORÍA EMPRESARIAL

Br. Laura Alejandra Rosales Nencka

Proyecto de grado - Investigación de Operaciones

El presente Proyecto de Grado, aborda el desarrollo de un modelo selectivo de tecnologías Big Data para consultoría empresarial. Para ello, se realizó una revisión bibliográfica y webgráfica sobre las tecnologías Big Data más utilizadas en la actualidad, así como también, una selección de las variables más influyentes en la elección de una tecnología Big Data para consultoría empresarial.

Se indagó sobre aportes de la Investigación de Operaciones para el desarrollo o adecuación de un modelo selectivo de tecnologías Big Data, que sirva de soporte a la toma de decisiones gerenciales, en base a las necesidades y requerimientos específicos de una empresa, en cuanto a aspectos como escalabilidad, facilidad de uso, disponibilidad, accesibilidad, costos, soporte técnico y seguridad entre otros factores.

Para el modelado se siguió un enfoque metodológico producto de la combinación del enfoque de Aracil J y el enfoque de Montgomery el cual permitió evolucionar desde un modelo mental hasta un modelo matemático selectivo.

Para validar el modelo selectivo se desarrolló una aplicación informática para dispositivos móviles, la cual fue cargada con datos obtenidos en sondeo realizado sobre 100 ofertas de empleo Big Data, en empresas ubicadas Latinoamérica, USA y parte de Europa y Asia.

Palabras claves: Tecnologías Big Data, Modelo Selectivo, Consultoría Empresarial, Análisis de Datos.

CONTENIDO

INTRODUCCIÓN.....	1
CAPÍTULO I.....	3
IDENTIFICACIÓN DEL PROYECTO.....	3
1.1 . Antecedentes de la investigación.....	3
1.2 . Definición del problema	4
1.3. Objetivos.....	5
1.3.1. Objetivo general.....	5
1.3.2. Objetivos Específicos	5
1.3 . Alcance del estudio.....	5
1.4 . Justificación.....	6
1.5. Metodología utilizada	6
CAPÍTULO II	8
MARCO TEÓRICO	8
2.1. Qué se entiende por Big Data.....	8
2.2. Modelos Big Data para procesamiento de datos	10
2.3. Lenguajes de Programación.....	11
2.4. Mejores proveedores de base de datos en la nube.....	13
2.5. Data warehouse (almacén de datos)	14
2.6. Herramientas de Big Data.....	18
2.7. Modelado del sistema.....	19
CAPÍTULO III.....	28
MODELADO DEL SISTEMA.....	28
3.1.Modelo Mental	29
3.2.Modelo Verbal o descriptivo	29
3.3. Modelo lógico.....	31
3.4. Modelo formal o matemático	34

CAPÍTULO IV	37
APLICACIÓN DEL MODELO SELECTIVO PONDERADO.....	37
4.1. Investigación y selección de las variables para el modelo selectivo	37
4.1.1. Variables tomadas en cuenta para la selección de hardware para Big Data.....	37
4.1.2. Variables tomadas en cuenta para la selección de Sistema Operativa para Big Data.....	38
4.1.3. Variables tomadas en cuenta para la selección de Base de Datos para Big Data.....	40
4.1.4. Variables tomadas en cuenta para la selección de Framework para Big Data	41
4.1.5. Variables tomadas en cuenta para la selección de lenguajes de programación para Big Data	43
4.1.6. Variables tomadas en cuenta para la selección de servicio en la nube para Big Data.....	44
4.2. Asignación del valor y la ponderación para cada variable	45
CAPÍTULO V.....	50
DESARROLLO DE LA APLICACIÓN MÓVIL	50
5.1. Entorno de Desarrollo	50
5.2. Estructura de la Aplicación Móvil.....	50
5.3. Prueba de la Aplicación Informática	51
5.3.1. Pantalla de ingreso.....	51
5.3.2. Pantalla históricos de consultorías	52
5.3.3. Pantalla consultorías de Big Data.....	52
5.3.4. Pantallas de Selección de Hardware	53
5.3.5. Pantallas de Selección de Sistema Operativo	54
5.3.6. Pantallas de Selección de Bases de Datos Relacionales	54
5.3.7. Pantallas de Selección de Bases de Datos No Relacionales.....	55
5.3.8. Pantallas de Selección de Frameworks	55
5.3.9. Pantallas de Selección de Lenguajes de Programación	56
5.3.10. Pantallas de Selección de Servicio en la Nube	56
CONCLUSIONES	58
RECOMENDACIONES	59
REFERENCIAS BIBLIOGRÁFICAS	60

Índice de Tablas

Tabla 4.1. Variables seleccionadas para la tecnología Hardware	37
Tabla 4.2. Variables seleccionadas para la tecnología Sistema Operativo	39
Tabla 4.3. Variables seleccionadas para la tecnología Base de Datos	40
Tabla 4.4. Variables tomadas en cuenta para seleccionar el Framework	41
Tabla 4.5. Variables seleccionadas para la tecnología Lenguaje de Programación.....	43
Tabla 4.6. Variables seleccionadas para la tecnología Servicio en la Nube	44
Tabla 4.7. Valores de las variables para cada producto de la tecnología Hardware.....	49
Tabla 4.8. Valores de las variables para cada producto de la tecnología Sistema Operativo....	49
Tabla 4.9. Valores de las variables para cada producto de la tecnología Base de Datos Relacional.....	50
Tabla 4.10. Valores de las variables para cada producto de la tecnología Base de Datos No Relacional.....	50
Tabla 4.11. Valores de las variables para cada producto de la tecnología Framework.....	51
Tabla 4.12. Valores de las variables para cada producto de la tecnología Lenguaje de Programación.....	51
Tabla 4.13. Valores de las variables para cada producto de la tecnología Servicio en la Nube..	52

Índice de Figuras

Figura 2.1. Enterprise Data Warehouse.....	15
Figura 2.2. Operational Data Store.	16
Figura 2.3. Data Mart.....	17
Figura 2.4. Data Lake.....	17
Figura 2.5. Carácter iterativo de las fases de construcción de un modelo.	23
Figura 2.6. Representación gráfica del proceso de construcción de un modelo	24
Figura 2.7. Ejemplos de Formalización de un modelo.	25
Figura 3.1. Fases propuestas para el Modelado del Sistema.....	28
Figura 3.2. Flujograma de la Tecnología Hardware.	31
Figura 3.3. Flujograma de la Tecnología Sistema Operativo	31
Figura 3.4. Flujograma de la Tecnología Base de Datos Relacional	32
Figura 3.5. Flujograma de la Tecnología Base de Datos No Relacional.....	35
Figura 3.6. Flujograma de la Tecnología Framework.....	36
Figura 3.7. Flujograma de la Tecnología Lenguaje de Programacion.....	36
Figura 3.8. Flujograma de la Tecnología Servicio en Nube.....	37

Índice de Pantallas

Pantalla 5.1. Inicio de la Aplicación Big Data Consulting.....	54
Pantalla 5.2. Históricos de consultorías en la aplicación Data Consulting.....	55
Pantalla 5.3. Datos del cliente en la aplicación Big Data Consulting.....	56
Pantalla 5.4. Tecnología Hardware para Big Data.....	56
Pantalla 5.5. Tecnología Sistema Operativo para Big Data	57
Pantalla 5.6. Tecnología Base de Datos Relacional para Big Data	57
Pantalla 5.7. Tecnología Base de Datos No Relacionales para Big Data.....	58
Pantalla 5.8. Tecnología Framework para Big Data	58
Pantalla 5.9. Tecnología Lenguajes de Programación para Big Data	59
Pantalla 5.10. Tecnología Base Servicio en la Nube para Big Data	59
Pantalla 5.11.Resultados de la Consultoría Big Data	60

www.bdigital.ula.ve

INTRODUCCIÓN

El ser humano desde tiempos remotos ha tenido la necesidad de procesar datos a lo largo de su vida, es por ello que hoy conocemos tanto de nuestro pasado. Con el transcurso del tiempo y el desarrollo tecnológico, el volumen de estos datos fue creciendo en gran medida; hoy en día, como consecuencia de la expansión y adopción de las nuevas tecnologías de la comunicación, se puede disponer de un volumen significativo de datos e información. Las plataformas sociales, los dispositivos móviles, los sensores digitales, los sitios de compra online e internet en general, han multiplicado significativamente la cantidad de información disponible. A finales del siglo pasado, se empieza a escuchar el término **Big Data**, para referirse a conjuntos de datos tan grandes y complejos que precisan de aplicaciones informáticas no tradicionales de procesamiento para tratarlos adecuadamente.

En la actualidad, la gestión y el análisis de grandes volúmenes de datos se ha convertido en uno de los mayores desafíos para las organizaciones. Con el creciente volumen de datos generados por usuarios de dispositivos y sistemas, las empresas han necesitado de nuevas tecnologías y herramientas para gestionar y analizar la gran cantidad de información y datos disponibles, de manera efectiva. En este contexto, las tecnologías de Big Data se han convertido en una solución clave para enfrentar estos desafíos.

Las tecnologías de Big Data permiten el procesamiento y análisis de grandes conjuntos de datos de manera rápida, eficaz y eficiente, lo que permite a las organizaciones tomar decisiones informadas y basadas en datos, en tiempo real. Además, estas tecnologías permiten la integración de datos de diversas fuentes internas en una empresa, con datos de fuentes externas incluyendo redes sociales, sensores y dispositivos móviles, lo que proporciona una visión más amplia de los datos y ayuda a identificar patrones y tendencias que de otra manera serían difíciles de detectar.

En el presente proyecto de grado, se explorarán las tecnologías de Big Data más demandadas por las empresas y reflejadas en las ofertas de empleo para profesionales de la ciencia de los datos tales como, ingenieros de datos, analistas de datos, arquitectos de datos y científicos de datos.

En este proyecto de grado se presenta el resultado de la investigación bibliográfica y webgráfica realizada, la información obtenida en el sondeo de ofertas de empleo, el modelo selectivo obtenido y detalles de la aplicación informática desarrollada. El contenido se ha estructurado de la siguiente manera:

Capítulo I: identificación del proyecto a través de los antecedentes de la investigación, definición del problema, justificación, objetivos, metodología a seguir y alcance.

Capítulo II: marco teórico sobre Big Data y sus tecnologías, lenguajes de programación, proveedores de base de datos, data warehouse, modelado del sistema.

Capítulo III: selección de variables y asignación de ponderación, aplicación del modelo selectivo ponderado.

Capítulo IV: desarrollo de la aplicación móvil y resultados.

Finalmente se presentan las conclusiones y recomendaciones.

www.bdigital.ula.ve

CAPÍTULO I

IDENTIFICACIÓN DEL PROYECTO

En el presente capítulo se realizará una descripción formal del proyecto de grado, partiendo de algunos antecedentes de investigación en el área que desarrolla el proyecto, seguido de un planteamiento de problema , justificación, formulación de objetivos generales y específicos, se indica también la metodología que se utilizó como guía durante el desarrollo del marco experimental y el alcance del proyecto.

1.1. Antecedentes de la investigación

A partir de la década de los sesenta, la compañía IBM comenzó a desarrollar lo que posteriormente se conocería como almacenamiento de datos masivos. No obstante, el término "Big data" fue acuñado por primera vez por investigadores de la NASA (Cox, M., Ellsworth, D., 1997) en su artículo Application-Controlled Demand Paging for Out-of-Core Visualización . Big data es un término que hace referencia a una cantidad de datos tal que supera la capacidad del software convencional para ser capturados, administrados y procesados en un tiempo razonable. El volumen de los datos masivos crece constantemente.

Luego de realizar una revisión bibliográfica se encontraron los siguientes antecedentes:

- Joyanes Aguilar, Luis (2013) publica en su libro titulado “Big Data: Análisis de Grandes Volúmenes de datos en organizaciones” un trabajo que muestra sus conocimientos esenciales y desempeño profesional en esta área, y que sin duda, lo atractivo que representa en el ámbito laboral el analista de datos y la arquitectura del Big Data.
- Rodríguez Sánchez, Francisco Miguel (2015) publica un trabajo de grado (Universidad Politécnica de Cartagena) titulado “Lenguajes y Sistemas Informáticos” en el cual tiene como objetivo presentar y estudiar la herramienta Apache Hadoop para el procesamiento de Big Data.

- Hernández-Leal, Emilcy J.; Duque-Méndez, Néstor D. y Moreno-Cadavid, Julián (2017) publican en su artículo de investigación “Big Data: una exploración de investigaciones, tecnologías y casos de aplicación” hablan sobre el mayor crecimiento de datos del mercado que lo envuelve y de las áreas de investigación asociadas, implementación y técnicas asociadas al Big data.
- Alava Cruzatty, José Efraín; Ramos Menéndez Ingrid Nathalia (2019) publican en su tesis de grado (Universidad Estatal del Sur de Manabí) titulado “Análisis Comparativo de Herramientas de Big Data para la Carrera de Ingeniería en Sistemas Computacionales” realizan un análisis comparativo de las herramientas de Big Data para la carrera de Ingeniería en Sistemas Computacionales, siendo este un medio de almacenaje de datos a gran escala y el medio de análisis de la información profunda que repercute en los mismos.

En la revisión bibliográfica, no se encontraron trabajos publicados similares o del estilo propuesto en el presente trabajo.

www.bdigital.ula.ve

1.2. Definición del problema

Actualmente encontramos en diferentes fuentes, como en sitios web de proveedores de servicios en la nube, en plataformas de código abierto, en revistas especializadas en tecnología, en eventos y conferencias de Big Data, entre otros, una gran cantidad de tecnologías y herramientas para utilizar en Big data, ofreciéndonos ciertas ventajas y beneficios que pueden adaptarse de mejor o peor manera a las necesidades de cada organización que quiera utilizar estas tecnologías.

La gran cantidad de opciones existentes para seleccionar es una gran ventaja, pero también crea un conflicto a la hora de tomar una decisión, es por ello, que el desarrollo o la adecuación de un modelo para ayudar con el proceso de decisión, utilizando un método selectivo que pueda adaptarse a las necesidades de cada institución, tiene una gran utilidad en la prestación de servicios de consultoría empresarial.

1.3. Objetivos

1.3.1. Objetivo general

Desarrollar o adecuar un modelo selectivo de tecnologías Big data, implementable en una aplicación informática, que pueda ser utilizado como herramienta de soporte en la consultoría empresarial.

1.3.2. Objetivos Específicos

1. Revisión bibliográfica y webgráfica sobre tópicos de la teoría de análisis de decisiones como técnica de Investigación de Operaciones.
2. Desarrollo o adecuación de un modelo matemático selectivo, fundamentado en algún método de Investigación de Operaciones, como herramienta de soporte en la selección de tecnología Big Data.
3. Investigación sobre tecnologías de Big Data más demandadas en el área empresarial.
4. Pruebas del modelo obtenido, mediante una aplicación informática, con datos obtenidos en la investigación sobre tecnologías más demandadas.

1.3 . Alcance del estudio

Como producto de la presente investigación, se pretende, una vez validado el modelo selectivo obtenido, alimentar la aplicación informática desarrollada con información obtenida sobre ofertas de empleo en Big Data, donde se haya detectado las tecnologías asociadas más relevantes, con el fin de ofrecer a la Empresa Consultora una herramienta de soporte en la Prestación de Servicios de Consultoría, al momento de necesitar tomar decisiones sobre adquisición o contratación de tecnologías para el procesamiento de grandes volúmenes de datos.

1.4 . Justificación

El volumen de datos en el mundo crece de manera exponencial. Según algunas estimaciones de las Naciones Unidas (Naciones Unidas, 2020), el 90% de los datos en el mundo se ha creado en los últimos dos años y se predice un crecimiento de un 40% anual. En 2020 se crearon 64,2 zettabytes de datos, lo que supone un aumento del 314% respecto a 2015. Este crecimiento exponencial, ha generado una necesidad de tecnologías para el procesamiento de tantos datos, lo cual ha motivado a su vez, a grandes empresas del sector tecnológico de la información, como Google, IBM, Microsoft, Oracle, al desarrollo de herramientas computacionales Big Data. Esta cantidad de opciones disponibles, generalmente de alto costo, hace que tomar una decisión al momento de seleccionar una de estas herramientas computacionales sea realmente complicado, es por ello, que el desarrollo de un modelo selectivo, será de gran utilidad en las actividades de consultoría para la toma de decisiones y satisfacción de las necesidades del cliente.

www.bdigital.ula.ve

1.5. Metodología utilizada

La propuesta de desarrollo o la adecuación del modelo selectivo de tecnologías Big Data para servicios de consultoría requiere la realización de diferentes fases. Estas se detallan a continuación:

Fase de investigación bibliográfica. Se realizó una revisión bibliográfica y webgráfica, con la finalidad de obtener conocimientos necesarios para resolver el problema.

Fase de determinación del método. Con los conocimientos necesarios obtenidos por la investigación bibliográfica se procedió a realizar una comparación entre algunos de los métodos de la teoría de decisiones y un enfoque propuesto por el Profesor tutor, para elegir el más adecuado para la obtención del modelo selectivo.

Fase de selección de tecnologías. En esta fase se indagó exhaustivamente la demanda de tecnologías de Big Data más utilizadas en el sector empresarial, mediante la consulta de ofertas de empleo publicadas en páginas web en diferentes partes del mundo.

Fase de selección y ponderación de variables. En esta fase se seleccionaron las variables a considerar en el modelo selectivo, así como, los criterios para la ponderación de las mismas.

Fase de desarrollo de la aplicación informática. En esta fase se desarrolló una aplicación móvil Android donde se implementó el modelo selectivo obtenido.

Fase de pruebas. Luego de la fase de selección y ponderación de variables, se procedió a una fase de pruebas alimentando la aplicación informática desarrollada, para confirmar que el modelo funciona adecuadamente y puede ser utilizada como herramienta tecnológica en la Empresa Consultora en la prestación de servicios de consultoría empresarial.

www.bdigital.ula.ve

CAPÍTULO II

MARCO TEÓRICO

En este capítulo se desarrolla brevemente el marco teórico inherente de Big Data con énfasis en las tecnologías utilizadas en el área, abarcando información relevante para el proyecto de grado, se abordan temas como lenguajes de programación distinguidos y emergentes, bases de datos relacional y no relacional, frameworks para el manejo de grandes volúmenes de datos y servicios en la nube, se puntualiza sobre el modelo y secuencia para la construcción de pasos a seguir.

2.1. Qué se entiende por Big Data

Big Data se refiere a la gestión y análisis de grandes volúmenes de datos, que son demasiado grandes, complejos o cambiantes para ser procesados por herramientas tradicionales de gestión y análisis de datos. Este término se refiere no solo a la cantidad de datos, sino también a la variedad de fuentes de datos, la velocidad a la que se generan los datos y la complejidad de los datos.

Los datos de Big Data provienen de diversas fuentes, como redes sociales, sensores, dispositivos móviles, transacciones financieras y registros de clientes, entre otros. Estos datos pueden ser estructurados, semi-estructurados o no estructurados, y pueden ser procesados en tiempo real para proporcionar información útil para la toma de decisiones.

El análisis de Big Data permite a las empresas y organizaciones identificar patrones y tendencias en los datos, lo que puede ayudar a tomar decisiones informadas y basadas en datos. El análisis de Big Data se utiliza en diversos campos, como marketing, finanzas, salud, transporte y seguridad, entre otros. (Bello. E, 2022).

Para caracterizar el Big data, inicialmente se habló inicialmente de tres características: Volumen, Velocidad y Variedad, pero posteriormente se han ido incorporando características como Veracidad, Variabilidad, Visualización y Valor, pero podemos considerar agregar Viabilidad (Dávila, L 2021). Las características del Big Data son:

- **Volumen**, se refiere a cómo almacenar, procesar y tratar una grandísima cantidad de datos. Es la principal característica con que suelen asociar al Big Data, por hacer referencia a la cantidad masivas de datos que se recopilan y almacenan con la finalidad de procesarlos y transformarlos en información.
- **Velocidad**, se refiere a la rapidez con que son creados los datos, las interconexiones que realizamos, almacenamos y procesamos en tiempo real, o cómo conseguir manejar nuestra ingente cantidad de datos lo más rápido posible.
- **Variedad**, se refiere a los diferentes tipos de datos que se registran, así como, las formas y fuentes de recolección. La diversidad de datos ha explotado, pasando de ser datos almacenados y estructurados, guardados en bases de datos empresariales, a ser no estructurados, y con distintos formatos (documentos de texto, vídeos, imágenes, audios, etc.). Se estima que el 90% de los datos son generados de manera no estructurada.
- **Variabilidad**, se refiere a los cambios que ocurren frecuente y rápidamente en los datos. Esto es relevante a la hora de analizarlos e inferir resultados para un momento específico.
- **Visualización**, se refiere al proceso de adecuación, presentación de los datos de manera comprensible, fácil de leer o interpretar para la obtención de información relevante.
- **Valor**, se refiere a la rentabilidad resultante de la gestión de los datos, o como extraer provecho de ellos. La clave del Big Data no es la cantidad ingente de información, sino el uso y/o manejo que se haga de ella.
- **Viabilidad**, se refiere a la factibilidad tecnológica de que un proyecto pueda llevarse a cabo, considerando factores, económicos, legales, planificación y técnicas.

2.2. Modelos Big Data para procesamiento de datos

Hay varios modelos de Big Data que se utilizan para procesar y analizar grandes conjuntos de datos en la Nube. A continuación, se describen algunos de los modelos más comunes (Mell P y Grance T. The Nist Definition,2019):

- **Batch processing (procesamiento por lotes):** En este modelo, los datos se procesan y analizan en grandes lotes. Los datos se recopilan durante un período de tiempo y luego se procesan juntos en un lote. Este modelo se utiliza para el análisis de datos históricos y para la generación de informes.
- **Stream processing (procesamiento de flujo):** En este modelo, los datos se procesan en tiempo real a medida que se generan. Los datos se procesan en pequeñas unidades a medida que se transmiten, lo que permite tomar decisiones en tiempo real. Este modelo se utiliza en aplicaciones que requieren una respuesta inmediata, como la detección de fraudes o el análisis de redes sociales.
- **Data lake (lago de datos):** Este modelo se utiliza para almacenar grandes cantidades de datos en su forma original. Los datos se almacenan en un depósito de datos centralizado, que se puede utilizar para el análisis posterior. Este modelo es útil para empresas que necesitan almacenar grandes cantidades de datos, pero no tienen un plan específico para su uso.
- **MapReduce:** Este modelo se utiliza para procesar grandes conjuntos de datos distribuidos en múltiples servidores. El modelo divide los datos en partes más pequeñas y las procesa en paralelo en varios servidores. Este modelo se utiliza comúnmente para el análisis de datos a gran escala, como el análisis de la web o la minería de datos.
- **NoSQL:** Este modelo se utiliza para almacenar y recuperar grandes conjuntos de datos no estructurados o semiestructurados. La base de datos NoSQL es escalable y flexible, lo que la hace ideal para el almacenamiento de datos no relacionales. Este modelo es utilizado en aplicaciones como el análisis de texto, video, audio y el procesamiento de imágenes.

Estos son solo algunos de los modelos de Big Data que se utilizan en la actualidad. Cada modelo tiene sus propias ventajas y desventajas, y la elección del modelo depende del tipo de datos que se esté procesando y de las necesidades específicas de la empresa.

2.3. Lenguajes de Programación

Un lenguaje de programación es un conjunto de reglas, símbolos y convenciones que se utilizan para escribir programas de computadora. Un lenguaje de programación permite a los programadores comunicarse con las computadoras y darles instrucciones para realizar una tarea específica. Los lenguajes de programación pueden ser de bajo nivel o de alto nivel.

Los lenguajes de programación de bajo nivel se utilizan principalmente para programar hardware y dispositivos de bajo nivel, como sistemas operativos, controladores de dispositivos y firmware. Estos lenguajes suelen ser muy detallados y complejos, y requieren que el programador tenga un conocimiento profundo de la arquitectura del hardware.

Por otro lado, los lenguajes de programación de alto nivel son más fáciles de leer y escribir, y se utilizan para desarrollar aplicaciones de software y sistemas de información. Estos lenguajes se centran en los conceptos de programación y permiten a los programadores enfocarse en la lógica de la aplicación, en lugar de preocuparse por los detalles del hardware. (Equipo Editorial, Etecé, 2021).

Algunos de los lenguajes más utilizados a nivel mundial y que mejores resultados ofrecen para el manejo de Big data son: Java, Python, Scala, Golang, R, JavaScript.

2.3.1. Java

Java es un lenguaje de programación orientado a objetos. Uno de los objetivos principales de Java es que los programas realizados con este lenguaje puedan ejecutarse en cualquier ambiente o contexto, siendo su portabilidad uno de sus principales características. Fue desarrollado por Sun Microsystems y posteriormente adquirido por Oracle. Puede usarse totalmente gratuito y se consiguen bastantes paquetes para desarrolladores, también puede ser modificado por cualquier persona, algo que le da la denominación de “código abierto”. (Doug L y Burd B., s.f)

2.3.2. Python

Python es un lenguaje de programación de propósito general. Es independiente de plataforma y orientado a objetos. Es un lenguaje interpretado, es decir, no se necesita compilar el

código para ejecutarlo, algo que permite una mayor rapidez de desarrollo. También es importante mencionar que la filosofía del lenguaje es contar con una sintaxis que haga sencillo entender el código. Una de las grandes ventajas de Python es que está preparado para realizar cualquier tipo de programa, servidores de red, páginas web o aplicaciones. (Morris S., s.f)

2.3.3. R

R se trata de un tipo de lenguaje de programación con software libre, que, a su vez, es un tipo de lenguaje interpretado. Es decir, los comandos o instrucciones son ejecutados directamente sin ningún tipo de software externo o compilación de datos de lenguaje máquina.

Este tipo de lenguaje en R consiste en la planificación o creación de programas, aplicaciones o herramientas de medición estadística y gráfica.

Del mismo modo, su capacidad permite que se puedan desarrollar distintas técnicas de medición, como lineales o no lineales, análisis de series de tiempo, agrupaciones, estadísticas clásicas o clasificaciones.

El código R es desarrollado actualmente por R Development Core Team, y su proyecto es colaborativo y abierto, por lo que otros usuarios programadores pueden ser partícipes del proyecto. Pudiendo publicar paquete de códigos que extienden su configuración básica. (Santaella J,2022).

2.3.4. JavaScript

JavaScript es un lenguaje de programación de scripts (secuencia de comandos) orientado a objetos, que los desarrolladores utilizan para hacer páginas web interactivas.

El programador norteamericano **Brendan Eich** es el creador de **JavaScript**. Eich desarrolló el lenguaje con el nombre de **Mocha** para el navegador **Netscape** a mediados de la década de **1990**. Con el paso de los años, **JavaScript** fue expandiéndose hasta convertirse en uno de los lenguajes más utilizados en la Web. (Pérez Porto, J., Gardey, A,2018).

2.3.5. Golang (GO)

Golang (también conocido como Go) es un lenguaje de programación de código abierto inspirado en C, compilado, concurrente y con un enfoque de programación orientada a objetos un poco inusual. Fue creado por Google en el año 2009 por Google, para solventar deficiencias que

presentaban otros lenguajes tales como Java, JavaScript, Python, pero fue liberado en 2007. Actualmente está siendo utilizado por empresas como Docker, UBER, Dropbox, Facebook, Twitter, entre otros. (Watson T., s.f).

Una de las grandes ventajas del lenguaje de programación es que permite crear la aplicación y luego compilar para ser utilizada en diferentes sistemas operativos. También es conocido por ser un lenguaje fácil de aprender por contar con una gran cantidad de herramientas que ayudan en el desarrollo.

A pesar de ser un lenguaje bastante nuevo, el lenguaje de programación en la nube Go tiene muchas razones que invitan a pensar que puede convertirse en un referente para los próximos años (cuando sea más reconocido por desarrolladores y tenga mayor soporte).

2.4. Mejores proveedores de base de datos en la nube

Existen muchos proveedores de bases de datos en la nube en el mercado, cada uno con sus propias características y precios. A continuación, se presentan algunos de los proveedores de bases de datos en la nube más demandados:

2.4.1. Amazon Web Services (AWS) - Amazon RDS

Amazon RDS es un servicio de bases de datos en la nube totalmente administrado por AWS que admite múltiples motores de bases de datos, como MySQL, PostgreSQL, Oracle, SQL Server y MariaDB. Amazon RDS ofrece alta disponibilidad, escalabilidad y seguridad, y se puede configurar en minutos. (Amazon Web Services, s.f).

2.4.2. Microsoft Azure - Azure SQL Database:

Azure SQL Database es un servicio de bases de datos en la nube completamente administrado por Microsoft que admite múltiples motores de bases de datos, como SQL Server, PostgreSQL y MySQL. Azure SQL Database ofrece escalabilidad, disponibilidad, seguridad y rendimiento. (Microsoft Azure, s.f).

2.4.3. Google Cloud Platform - Cloud SQL:

Cloud SQL es un servicio de bases de datos en la nube completamente administrado por Google que admite múltiples motores de bases de datos, como MySQL, PostgreSQL y SQL Server. Cloud SQL ofrece alta disponibilidad, escalabilidad y seguridad, y se puede configurar en minutos. (Google Cloud, s.f).

2.4.4. Oracle Cloud Infrastructure - Autonomous Database:

Autonomous Database es un servicio de bases de datos en la nube completamente administrado por Oracle que admite múltiples motores de bases de datos, como Oracle Database, MySQL y PostgreSQL. Autonomous Database ofrece automatización, escalabilidad, alta disponibilidad y seguridad. (Oracle Cloud, s.f).

2.4.5. IBM Cloud - IBM Cloud Databases:

IBM Cloud Databases es un servicio de bases de datos en la nube totalmente administrado que admite múltiples motores de bases de datos, como PostgreSQL, MongoDB y Redis. IBM Cloud Databases ofrece alta disponibilidad, escalabilidad y seguridad. (IBM Cloud, s.f).

2.5. Data Warehouse (almacén de datos)

Un Data Warehouse es un conjunto de bases de datos organizadas, relacionadas y centralizadas que se utiliza para almacenar grandes cantidades de datos empresariales de diversas fuentes y en múltiples formatos.

Una vez que los datos están en el Data Warehouse, se pueden utilizar para realizar consultas y análisis complejos, ya que están en un formato estructurado y organizado de manera que sea fácil de entender y utilizar.

Su finalidad es proporcionar a los usuarios, el rápido acceso a datos históricos y actuales que se utilizan para el análisis y la toma de decisiones. Es decir, su objetivo es mejorar la toma de decisiones a través del análisis de datos.

Esta puede contener datos de una variedad de fuentes de datos, incluyendo sistemas transaccionales, bases de datos departamentales, datos de sensores, y mucho más. Al centralizar

estos datos, el Data Warehouse permite a los usuarios obtener una visión completa y detallada de las operaciones, lo que puede ayudar en la toma de decisiones informadas. (Ondho, s.f).

Existen diferentes tipos de Data Warehouse, los cuales se diferencian principalmente en la forma en que los datos son almacenados y organizados. Aquí te presento algunos de los tipos más comunes:

- **Enterprise Data Warehouse**

Es un almacenamiento de datos centralizado, unifica toda la información de una organización para que toda la empresa pueda tener acceso. Ofrece un servicio en el que apoya la toma de decisiones en la empresa. Se enfoca principalmente en organizar y representar los datos, también se obtiene la capacidad de clasificar los datos según el usuario y dar acceso de acuerdo a las restricciones internas. Una esquematización de este tipo de Data Warehouse, puede apreciarse en la figura (2.1)

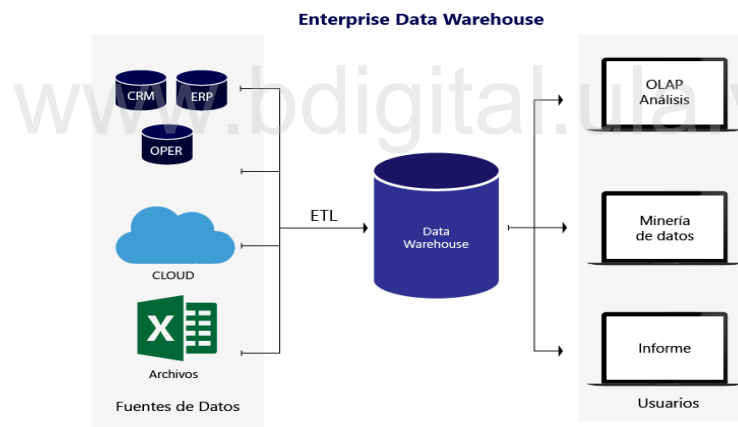


Figura 2.1 Enterprise Data Warehouse
Fuente: Mendoza A,2020

- **Operational Data Store (ODS)**

También conocido como ODS, es un almacén de datos, que cuando el almacenamiento de datos y los sistemas OLTP no admiten las necesidades de los informes de las organizaciones. En ODS, todo el almacenamiento de datos se actualiza en tiempo real y/o con baja latencia de actualización, y por eso mismo se utiliza habitualmente para actividades rutinarias, como es el almacenamiento de registros de la operación de la empresa y transacciones de venta. En pocas palabras, es un tipo de base de datos que se utiliza habitualmente como un área lógica provisional para un almacén de datos.

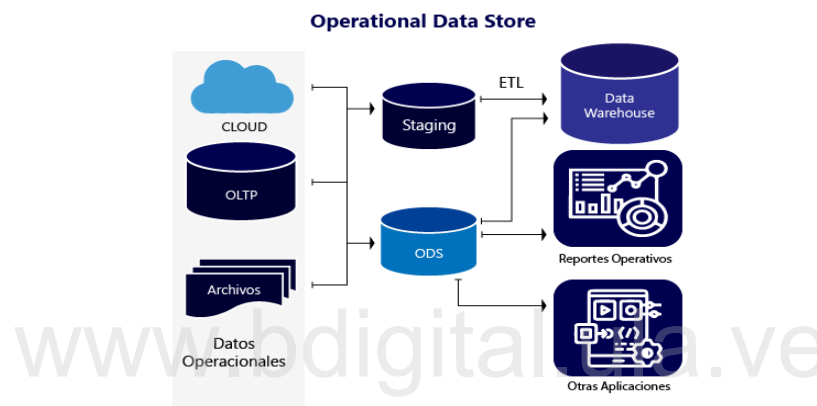


Figura 2.2 Operational Data Store.

Fuente: Mendoza A,2020,

- **Data Mart**

Un Data Mart es un subconjunto del almacenamiento de datos orientado a un área específica, también conocido como base de información departamental. Está especialmente diseñado para una línea de negocio particular, como ventas o finanzas. En un Data Mart independiente, los datos pueden recopilarse directamente de las fuentes. Tiene diversas funciones como organizar la información para posteriormente analizarla, realizar indicadores (KPI), evaluar los objetivos del sector que se está analizando, etc. El objetivo es realizar un análisis detallado sobre lo que ocurre en un negocio.

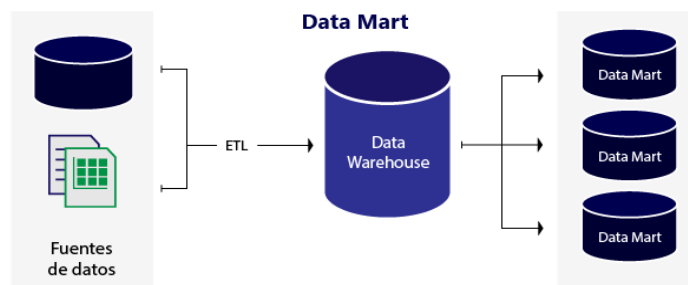


Figura 2.3 Data Mart.
Fuente: Mendoza A,2020,

- **Data Lake**

Es un repositorio centrado en almacenar gran cantidad de datos estructurados y sin estructurar sin importar su escala. Permite almacenar los datos tal cual vienen sin necesidad de ser estructurados. Ejecuta diferentes tipos de análisis, dashboards y visualizaciones, además de generar procesos de Big Data, análisis en tiempo real y de Machine Learning que facilitan la toma de decisiones.

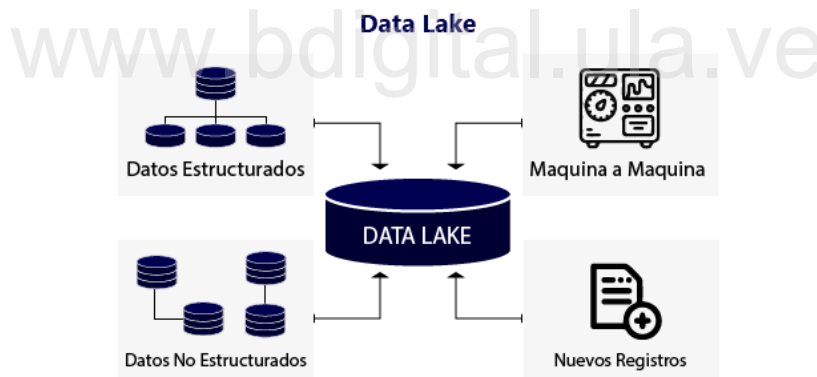


Figura 2.4 Data Lake.
Fuente: Mendoza A,2020,

Todas las estructuras de datos son importantes y cuentan con una gran funcionalidad según sus respectivos propósitos y perfiles de usuario. Por lo que es necesario comprender las diferencias y similitudes de cada una, para así poder identificar cuál se adapta mejor a las necesidades de cada empresa o bien si se requiere de una estructura de Data Warehouse íntegra para satisfacer los requerimientos. (Mendoza A,2020).

Los diferentes tipos de Data Warehouse ofrecen diferentes enfoques para almacenar y procesar datos y se adaptan a diferentes necesidades.

2.6. Herramientas de Big Data

Las herramientas de Big Data son un conjunto de instrumentos y técnicas utilizadas para recolectar, procesar, analizar y visualizar grandes cantidades de datos en tiempo real. Aquí te presento algunas de las tecnologías más populares utilizadas en el mundo del Big Data:

2.6.1. Apache Hadoop:

Es un framework de código abierto para el procesamiento distribuido de grandes conjuntos de datos en clústeres de servidores. Hadoop se compone de varias herramientas, incluyendo Hadoop Distributed File System (HDFS) y MapReduce, que permiten el procesamiento y análisis de grandes cantidades de datos. (PowerData, s.f)

2.6.2. Apache Spark:

Es un motor de procesamiento de datos de código abierto utilizado para el procesamiento en memoria de grandes conjuntos de datos. Spark es compatible con múltiples lenguajes de programación, como Python, Java y Scala, y es utilizado en aplicaciones de Big Data como análisis en tiempo real y aprendizaje automático. (Google Cloud, s.f).

2.6.3. Elasticsearch

ElasticSearch es una herramienta de búsqueda y análisis de datos de código abierto que se utiliza para buscar, analizar y visualizar grandes cantidades de datos en tiempo real. Es un motor de búsqueda y análisis distribuido optimizado para velocidad y relevancia en cargas de trabajo de producción a gran escala, que se basa en Apache Lucene y se utiliza para almacenar y buscar grandes conjuntos de datos. Se utiliza en una variedad de aplicaciones, como análisis de registros, análisis de seguridad, análisis de negocios y búsqueda de texto completo. (Elasticsearch,s.f).

2.6.4. Apache Hive

“Es una herramienta de Data Warehouse open source que funciona con Hadoop y permite consultar grandes cantidades de datos estructurados con un dialecto de SQL llamado HiveQL (HQL)”. (Fernández, O, 2023).

2.6.5. Impala

Una herramienta de procesamiento de datos en Hadoop. Tiene licencia open source. Fue desarrollada inicialmente por Cloudera y más tarde incluida en la Apache Software Foundation. Está incluida en las distribuciones de Cloudera. (Fernández, O, 2023).

2.6.6. MongoDB

MongoDB es una base de datos de documentos que ofrece una gran escalabilidad y flexibilidad, y un modelo de consultas e indexación avanzado. (MongoDB, s.f).

2.6.7. Apache Cassandra

“Cassandra es un sistema de código abierto distribuido de gestión de base de datos diseñado para manejar grandes cantidades de datos a través de muchos servidores de conveniencia, proporcionando alta disponibilidad sin ningún punto único de fallo.” (Power Data, 2016).

2.6.8. Apache Kafka

Es una plataforma de transmisión de datos en tiempo real utilizada para el procesamiento de grandes cantidades de datos de forma rápida y eficiente. Kafka se utiliza en aplicaciones de Big Data como el procesamiento de eventos en tiempo real y la integración de sistemas. (Red Hat, 2022).

2.6.9. Presto

Presto es un motor de consulta de SQL distribuido de código abierto creado por los desarrolladores de Facebook para ejecutar analíticas interactivas de grandes volúmenes de datos. (Google Cloud.s.f).

2.7. Modelado del sistema

El modelado de sistemas se refiere al proceso de crear modelos que representen un sistema con las características y comportamientos más relevantes del mismo, este proceso es significativo

para diferentes disciplinas entre ellas la ingeniería, permitiendo analizar y predecir el desempeño de un sistema en diferentes escenarios. El modelado de sistemas permite comprender cómo interactúan los componentes del sistema y cómo se ven afectados por diferentes variables.

2.6.1. Modelo

El modelo es una representación parcial de la realidad; esto se refiere a que no es posible explicar una totalidad, ni incluir todas las variables que esta pueda tener, por lo que se refiere más bien a la explicación de un fenómeno o proceso específico, visto siempre desde el punto de vista de su autor (Aguilera, 2000).

“El modelo es una representación o abstracción de una situación u objeto reales, que muestra las relaciones (directas o indirectas) y las interrelaciones de la acción y la reacción en términos de causa y de efecto. Como un modelo es una abstracción de la realidad, puede parecer menos complicado que las misma” (Theierauf & Grosse. 1997, p. 24).

“Un modelo de datos es una estructura abstracta que documenta y organiza la información para la comunicación entre el personal del departamento técnico y el resto de los empleados, según la acepción del concepto en el mundo de los negocios. En la informática, difiere en cuanto a su enfoque, el cual se centra en el planeamiento del desarrollo de aplicaciones y la decisión de cómo se almacenarán los datos y cómo se accederá a ellos.” (Pérez Porto, J., Gardey, A. 2008).

“El término modelo proviene del concepto italiano de modelo. La palabra puede utilizarse en distintos ámbitos y con diversos significados. Aplicado al campo de las ciencias sociales, un modelo hace referencia al arquetipo que, por sus características idóneas, es susceptible de imitación o reproducción. También al esquema teórico de un sistema o de una realidad compleja.” (Pérez Porto, J., Gardey, A. 2008).

Según la RAE un modelo es un esquema teórico, generalmente en forma matemática, de un sistema o de una realidad compleja, como la evolución económica de un país, que se elabora para facilitar su comprensión y el estudio de su comportamiento.

En base a las definiciones anteriores podemos decir que, un modelo es la abstracción de la realidad, donde se toman en cuenta características principales que aportan suficiente información para realizar el estudio deseado, todo esto desde el enfoque del autor, con el fin de facilitar

comprensión del estudio, su comportamiento y al ser una abstracción de la realidad resulta menos complicado que la misma

2.6.2. Modelado

“El modelado permite al usuario comprender mejor el problema y presenta un medio para manipular la situación con el fin de analizar los resultados de varias entradas (análisis "qué pasaría si") sometiéndolas a un conjunto cambiante de suposiciones.” (Kyriaki Kosmidou y Constantin Zopounidis, s.f).

“Los modelos para el computador se pueden elaborar tan complejos y realistas como lo permitan las teorías que utilicemos para construirlos, las técnicas de simulación nos capacitan para obtener las consecuencias de los modelos, sin que tenga que ver el grado de complejidad involucrado en la formulación de tales modelos, en consecuencia, se pueden modelar teorías económicas en modelos precisos, sin que existan distorsiones en los significados y conceptos contenidos en las teorías y, por otra parte, se puede determinar las descripciones del mundo real implicadas por dichas teorías.”(Box g., 1982 como se citó en Dávila L,A.2021).

“El proceso de observar algún sistema empresarial, industrial o económico en la realidad, formular una o más hipótesis relativas a su funcionamiento y reducirlo a un nivel de abstracción que permita la formulación de un modelo matemático para estudiar su comportamiento, de ningún modo constituye un proceso directo, debido a que los hechos económicos son extremadamente complicados, no obstante, se considera que la matemática constituye la forma más adecuada para investigarlos profundamente.” (Naylor, Th., 1982, como se citó en Dávila, L.A.,2021).

Podemos decir que un modelado, es una herramienta importante en la comprensión y el estudio de una amplia gama de disciplinas. Ayuda a simplificar la realidad y aporta claridad y estructura a los sistemas y fenómenos que se investigan o diseñan. Los modelos pueden ser muy simples o altamente complejos, dependiendo de los objetivos y las necesidades específicas del proceso de modelado.

2.6.3. Tipos de modelo

Aunque existen diferentes tipos de modelos para efectos de la presente investigación nos enfocaremos en los siguientes:

- **Modelo verbal o descriptivo** “Los modelos verbales utilizan palabras para representar algún objeto o situación que existe o podría existir en la realidad. Los modelos verbales pueden variar desde una simple presentación en palabras de un escenario descrito en un libro hasta un complejo problema de decisión empresarial (descrito en palabras y números).” (Kyriaki Kosmidou y Constantin Zopounidis, s.f).
- **Modelo Lógico** “aquel en el cual la representación de una estructura o proceso se formula mediante las leyes de la lógica. Estas últimas suelen acudir al auxilio de la simbología matemática.” (Ibanez, J .2008).
- **Modelo Matemático** “Son representaciones numéricas o algebraicas que permiten construir fórmulas y expresiones matemáticas para expresar las relaciones presentes en el mundo real.” (Ondarse, D. 2022).
- **Modelo de Programación** “provee un marco abstracto en el que entender la sintaxis de los lenguajes concretos que siguen el modelo, y representa la semántica del mismo, así como una filosofía de uso de dicha semántica. Es una abstracción de un mecanismo o infraestructura existente. Es un conjunto de bibliotecas abstractas de ejecución y de lenguajes de programación que forman un modelo de procesamiento. Este nivel de abstracción puede ser de bajo nivel, como el lenguaje de máquina de los ordenadores, o muy alto, como en lenguajes de programación de alto nivel.” (Garcia,F.s.f)

2.6.4. Enfoque metodológico para el modelado

Iniciando con el enfoque presentado por Aracil J, 1978, en su libro *Dinámica de Sistemas*, p.136, donde partiendo de una situación, en el mundo real, debe realizarse una fase de *Conceptualización*, donde se obtiene la perspectiva y comprensión de cierto fenómeno del mundo real, para obtener un *Modelo Mental*, a partir del cual, en una siguiente fase de *Formulación*, se trata de representar los elementos manejados en la fase anterior y así obtener un *Modelo Formal* sometiéndose a una *Evaluación*, que consiste en el análisis del mismo, así como, someterlo a varios criterios.

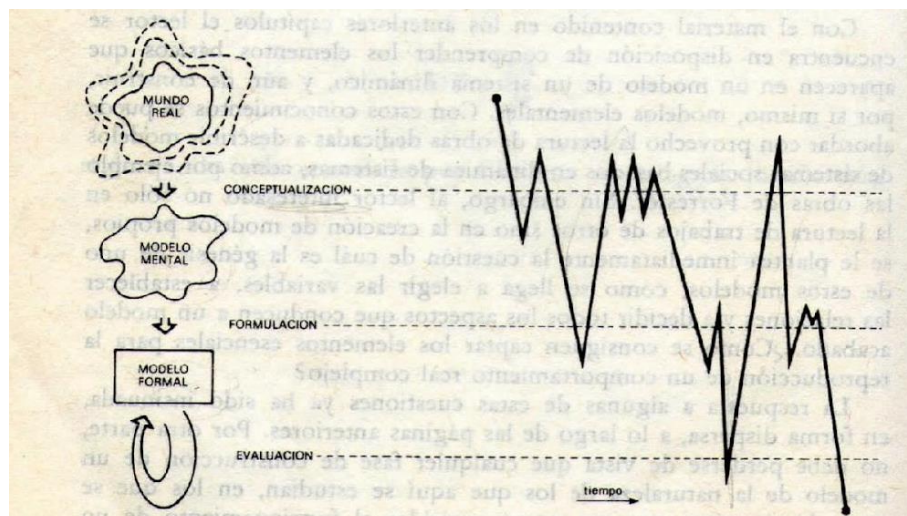


Figura 2.5 Carácter iterativo de las fases de construcción de un modelo.
Fuente: Introducción a la dinámica de sistemas (Aracil. J, p 136).

Cuando se está modelando situaciones, se debe tomar en cuenta que el proceso de construcción del mismo no es lineal, sino que algunos de sus pasos se repiten varias veces, construyendo varios modelos que son perfeccionados. Es sabido que pueden existir numerosos modelos capaces de satisfacer el problema, esto sucederá siempre que el modelo sea no lineal.

Es interesante dividir el modelo en dos etapas, una primera etapa donde se enjuicia, bajo la óptica del conocimiento que se tenga del sistema real, con el objetivo de modificarlo y mejorarlo a medida que avanza en los pasos; y la segunda etapa es el perfeccionamiento sucesivo, para alcanzar un modelo significativo. Un modelo inicial debe ser tan simple como sea posible, para luego, en una etapa de perfeccionamiento se simplifique de manera que el modelo se aproxime cada vez más a la realidad.

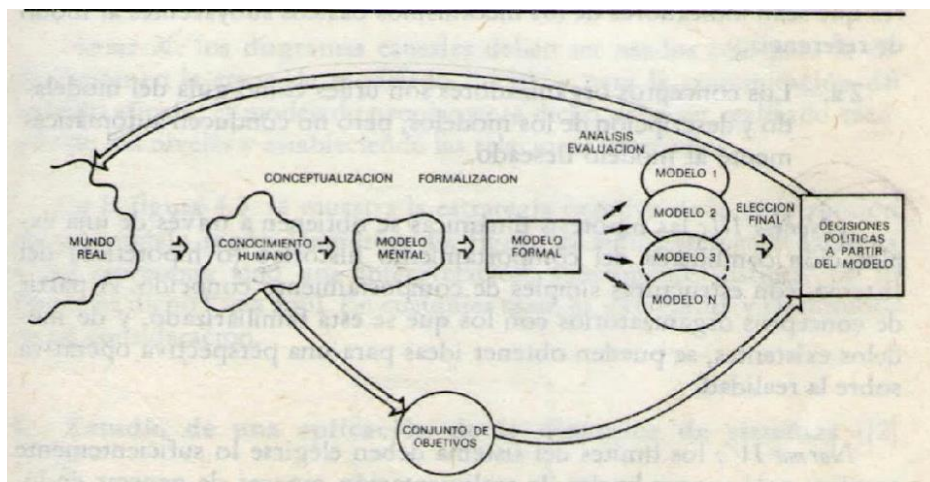


Figura 2.6 Representación gráfica del proceso de construcción de un modelo
Fuente: Introducción a la dinámica de sistemas (Aracil, J., p.141).

En la figura 2.6. se puede observar la representación gráfica de un modelo: del *mundo real* obtenemos el problema de caso de estudio, luego en la fase de *conceptualización*, con el conocimiento humano que se tenga del sistema en el mundo real y respetando los objetivos que se tengan establecidos, se puede obtener el modelo mental o descriptivo. En la fase de *formalización* se obtienen diversos modelos formales, de los cuales, al final, en una etapa de análisis y evaluación, debe seleccionarse el más adecuado. La implementación de este modelo final, facilitará la toma de decisiones.

Por otra parte, tenemos el enfoque presentado por Montgomery. D y Urban.G 1977 en el libro *Marketing Científico*, p.27, el cual puede apreciarse en la figura 2.7.

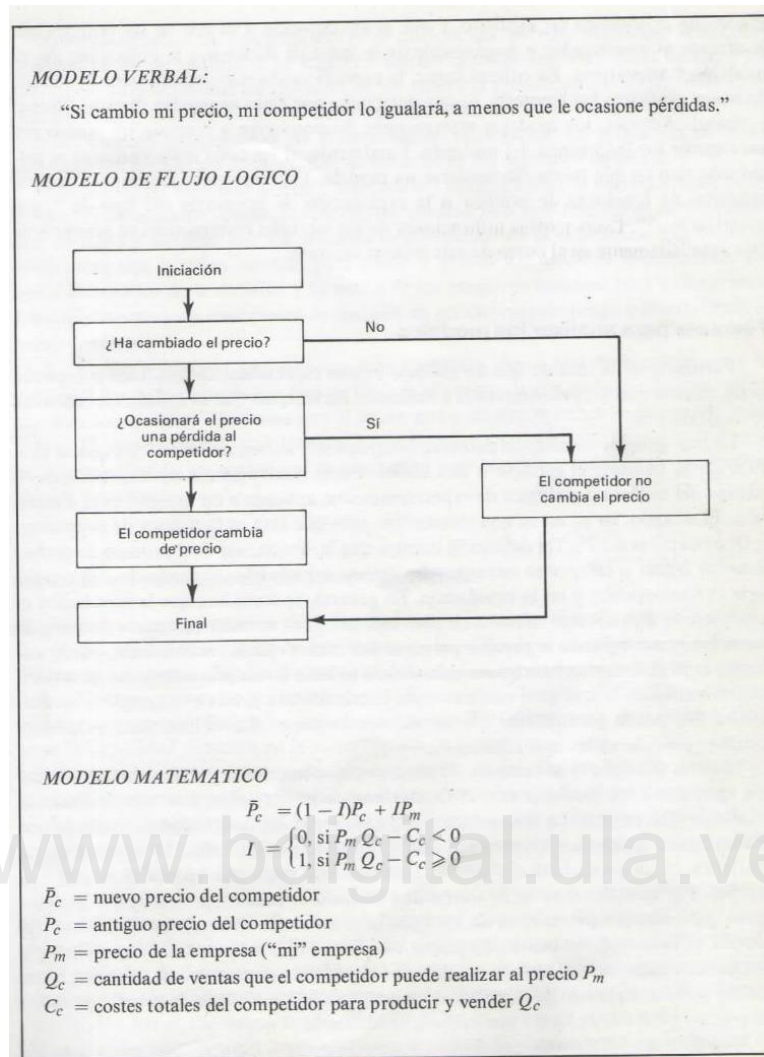


Figura 2.7. Ejemplos de Formalización de un modelo.
 Fuente: Marketing Científico por Montgomery. D y Urban.G, p. 27.

Partiendo de que un modelo es una simple representación de todas o algunas de las propiedades de un sistema mayor y que su representación puede ser física o abstracta, Montgomery, propone un proceso para la formalización de un modelo partiendo de su representación verbal. El siguiente paso es la obtención de un *modelo lógico*, que constituye una ampliación del *modelo verbal* mediante un diagrama de flujo ,en el cual de forma explícita, se realizan una serie de preguntas necesarias para avanzar que deberán contestarse de forma que estén relacionadas, por último, contamos con el *modelo matemático*, capaz de cuantificar las diferentes variables; los modelos matemáticos se pueden escribir de diferentes formas y estos pueden ser: estocásticos y determinísticos o estáticos y dinámicos.

3.1. Aportes de la Investigación de Operaciones para la Construcción de Modelos Selectivos.

Un modelo selectivo es un enfoque analítico que implica seleccionar ciertas variables o características específicas para analizar o evaluar una situación o problema en particular. En lugar de examinar todas las variables disponibles, se identifican aquellas que son más relevantes o significativas para el problema en cuestión y se les da mayor consideración. Este enfoque puede ser útil en situaciones donde hay una gran cantidad de datos o información, y se requiere una manera eficiente de centrarse en los aspectos más importantes o críticos del problema. Una revisión bibliográfica sobre métodos de análisis de decisiones, estudiados en Investigación de Operaciones, conllevó a considerar los siguientes para el desarrollo del modelo matemático:

3.1.1. Proceso Analítico Jerárquico

El proceso analítico Jerárquico (también llamado proceso de jerarquía analítica) (AHP) es uno de los métodos de toma de decisiones de criterios múltiples que fue desarrollado originalmente por el profesor Thomas L. Saaty (de la universidad de Pittsburgh). En resumen, es un método para derivar escalas de relación de comparaciones pareadas. La información puede obtenerse a partir de mediciones reales, como el precio, el peso, etc., o de una opinión subjetiva, como sentimientos de satisfacción y preferencia. (Piqueras V, 2019)

AHP permite una pequeña inconsistencia en el juicio porque los humanos no siempre son consistentes.

Esta técnica es utilizada ampliamente en la sociedad actual para la toma de decisiones gubernamentales, de negocios e incluso en la industria de la salud y la educación.

AHP opera de acuerdo con tres principios:

Descomposición jerárquica: el problema se estructura a modo de jerarquía, con el objetivo en la cúspide, las alternativas en el escalón inferior, y los criterios en el nivel (o niveles) intermedios.

Juicio comparativo: todos los elementos de la jerarquía se evalúan por parejas, comparándolos en relación a los elementos del nivel inmediatamente superior de la jerarquía; estas comparaciones se traducen en juicios codificados en una escala convencional que usualmente toma valores entre 1 y 9.

Síntesis: los juicios se agregan en vectores de prioridades dentro de cada nivel, y finalmente se integran para dar lugar a una clasificación general final.

3.1.2. Árboles de decisión

Un árbol de decisiones es una herramienta de soporte de decisiones que utiliza un gráfico o modelo de decisiones en forma de árbol y sus posibles consecuencias, incluidos los resultados de eventos fortuitos, los costos de recursos y la utilidad. Es una forma de mostrar un algoritmo que sólo contiene sentencias de control condicional.

Un árbol de decisión es una estructura similar a un diagrama de flujo en el que cada nodo interno representa una "prueba" en un atributo (por ejemplo, si una moneda lanza cara o cruz), cada rama representa el resultado de la prueba y cada nodo hoja representa una etiqueta de clase (decisión tomada después de calcular todos los atributos). Los caminos de raíz a hoja representan reglas de clasificación.

Los algoritmos de aprendizaje basados en árboles se consideran uno de los mejores y más utilizados métodos de aprendizaje supervisado. Los métodos basados en árboles potencian los modelos predictivos con alta precisión, estabilidad y facilidad de interpretación. A diferencia de los modelos lineales, mapean bastante bien las relaciones no lineales. Son adaptables para resolver cualquier tipo de problema (clasificación o regresión).

3.3.3. Modelo Selectivo Ponderado

El Modelo Selectivo Ponderado (propuesto por el profesor tutor) es una técnica utilizada en la toma de decisiones multicriterio, donde se asignan ponderaciones a diferentes criterios para evaluar y seleccionar la mejor opción entre varias alternativas. Este enfoque es útil cuando las decisiones implican múltiples factores que deben ser considerados simultáneamente. En pocas palabras, esto significa que la importancia (llamada ponderación) de cada variable será seleccionada por quien utilice el modelo, esto hace posible que la modificación del modelo (para diferentes casos de estudios) sea rápida, sencilla, efectiva y que se ajuste fácilmente a cualquier caso que se pueda presentar.

CAPÍTULO III

MODELADO DEL SISTEMA

Este capítulo explica las diferentes etapas del proyecto de grado partiendo del mundo real donde se originó la idea pasando por los diferentes tipos de modelo como lo son el modelo mental, modelo verbal, modelo lógico, modelo matemático y culminando con el modelo programático.

Combinado el enfoque metodológico de Aracil J, y el enfoque metodológico de Montgomery - explicado en la sección 2.6.4 - a continuación, se presenta el enfoque que se seguirá, en el presente proyecto, para la obtención del modelo selectivo de tecnologías Big Data.

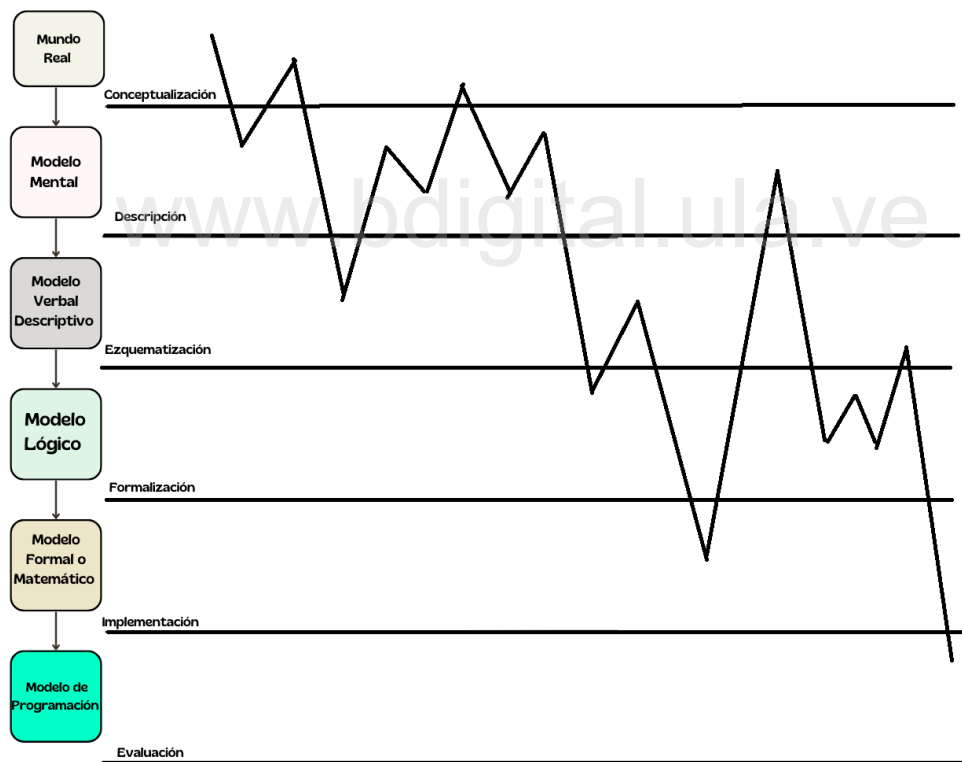


Figura 3.1. Fases propuestas para el Modelado del Sistema

Mundo real: La evolución tecnológica para el almacenamiento y transformación de datos ha conllevado a un crecimiento casi exponencial del volumen de los datos, estimando que la última

década este crecimiento alcanza 97 zettabits (para el 2022) a nivel empresarial aportados por el uso de dispositivos electrónicos, información compartida como consecuencia del incremento de la interacción entre factores externos, clientes, proveedores con el sector productivo empresarial , aportados a través de medios electrónicos pero que han podido ser almacenados en diferentes formatos, que gracias al avance tecnológico de la prestación de servicio en la nube, hoy en día puede ser organizados, depurados, analizados y visualizados para dar soporte a la toma de decisiones empresariales , utilizando tecnologías de Big Data.

Aunque una empresa en particular, podría abocarse en la aplicación de tecnologías Big Data utilizando sus propios recursos humanos y tecnológicos, lo más conveniente es que recurra a los servicios de consultoría empresarial para la selección de las diferentes tecnologías requeridas para el tratamiento de sus datos.

3.1. Modelo Mental

Partiendo de la apreciación que se haya obtenido del Mundo Real, de los objetivos de la empresa, tipo y características de los datos disponibles y necesidades de procesamiento de datos, se debe obtener una “abstracción mental” de la operatividad e interrelaciones existentes entre las diferentes áreas de la empresa donde se generan datos e información, así como, los requerimientos tecnológicos para el desarrollo de un Proyecto Big Data. Esta actividad la realiza generalmente el Consultor Big Data mediante un proceso de conceptualización, que aportará información al resto del equipo de trabajo para la realización de las siguientes etapas.

3.2. Modelo Verbal o descriptivo

Partiendo del Modelo Mental obtenido por el Consultor Big Data, se obtuvo una descripción verbal de los requerimientos tecnológicos de la Empresa. De esta descripción verbal, en lenguaje natural, se definió el proceso selectivo de las tecnologías que debían ser evaluadas.

En *primer lugar*, debe seleccionarse el Hardware más adecuado para el procesamiento de gran volumen de los datos existentes y futuros en el tiempo. Para ello, debe evaluarse las últimas generaciones de hardware tomando en cuenta un conjunto de variables o características que los diferencia entre sí, pero que son influyentes en el rendimiento del equipo. Si la generación seleccionada no está disponible en la empresa, esta selección debe ser adquirida.

El *siguiente paso* debe seleccionarse el sistema operativo más idóneo; dado que ha sido ampliamente demostrado que Big Data puede realizarse mejor en el sistema operativo de Linux, se procede a evaluar y seleccionar la distribución de Linux más eficiente entre las siguientes: Ubuntu, Mint, Manjaro, Fedora y Debían. Si el sistema operativo seleccionado no está disponible en la empresa, debe adquirirse e instalarse.

Como *tercer paso*, se debe seleccionar el gestor de base de datos más adecuado, tomando en cuenta que podemos encontrar datos estructurados, semiestructurados y no estructurados. Para los datos estructurados puede utilizarse gestores de base de datos relacionales SQL como Oracle, Sybase, DB2, mientras que, para los datos no estructurados y semiestructurados, existen en el mercado gestores de base de datos no relacionadas o No SQL como Casandra, Neo4j, HBase, MongoDB. Si el gestor seleccionado no está disponible, este debe adquirirse e instalarse en el hardware previamente seleccionado.

El *siguiente paso* conlleva a seleccionar el Framework (entorno de desarrollo) más adecuado para trabajar los datos. Entre los principales entornos, más demandados en las ofertas de empleo, se encontró a Hadoop, Spark, Presto y ElasticSearch. Se toma en consideración un conjunto de variables o características de cada Framework, que son influyentes en el procesamiento de los datos. Si el framework seleccionado no está disponible en la empresa, este debe ser adquirido e instalado.

Una vez seleccionado el framework, se procede a seleccionar el lenguaje de programación compatible con la selección anterior. Aunque generalmente los lenguajes de programación más utilizados son de código abierto, en el presente estudio se preseleccionaron los lenguajes Python, Go, R, Java, JavaScript, por ser los más compatibles con los entornos de desarrollo (framework) antes citados. Si el lenguaje seleccionado no está disponible, este debe ser adquirido e instalado.

Como *último paso* nos queda seleccionar el Servicio en la Nube, evaluando como candidatos a selección, los cuatro servicios más demandados en las ofertas de trabajo, entre los que destacan servicios ofrecidos por Amazon Web Service, Microsoft Azure, Google Cloud o IBM Cloud. Si el servicio seleccionado en la nube no está contratado, debe contratarse el servicio.

3.3. Modelo lógico

Partiendo del Modelo Verbal o Descriptivo, se diseñó el Diagrama de Flujo que se expone a continuación, el cual representa el Modelo Lógico Selectivo de las tecnologías Big Dat, iniciando con la opción de selección del Hardware y finalizando con la opción de selección del servicio en la Nube.

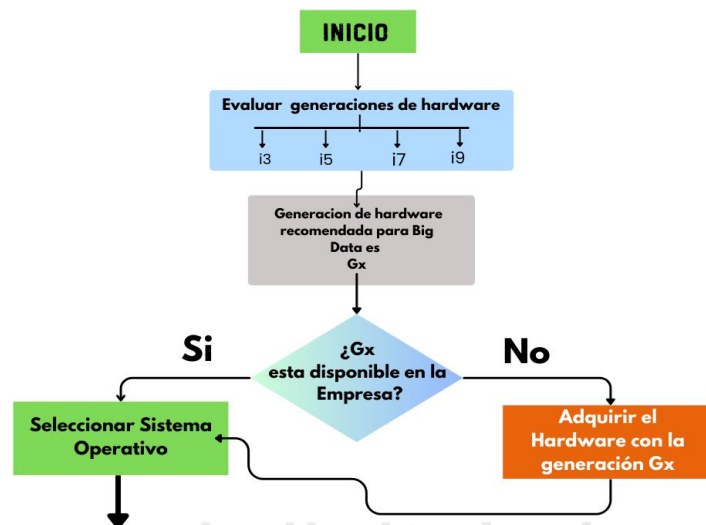


Figura 3.2. Flujograma de la Tecnología Hardware.

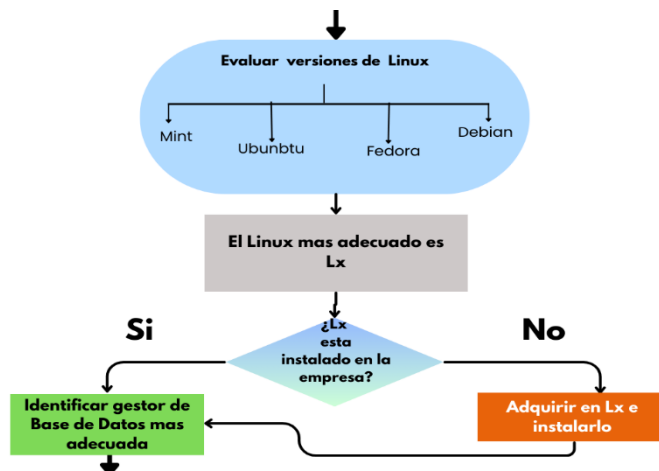


Figura 3.3. Flujograma de la Tecnología Sistema Operativo

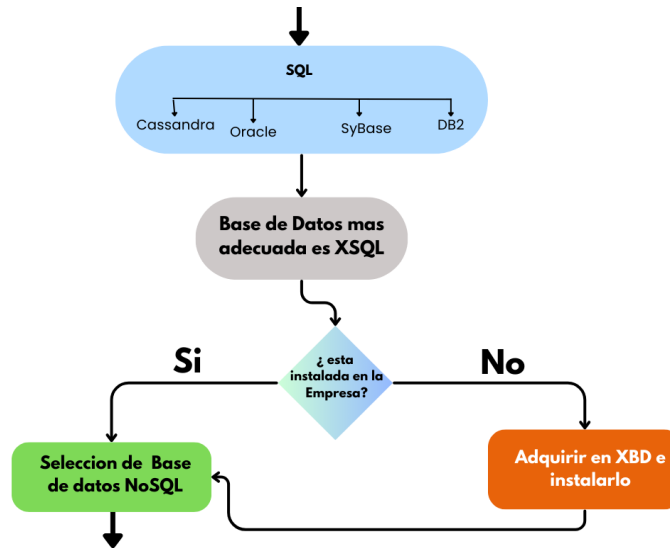


Figura 3.4. Flujograma de la Tecnología Base de Datos Relacional

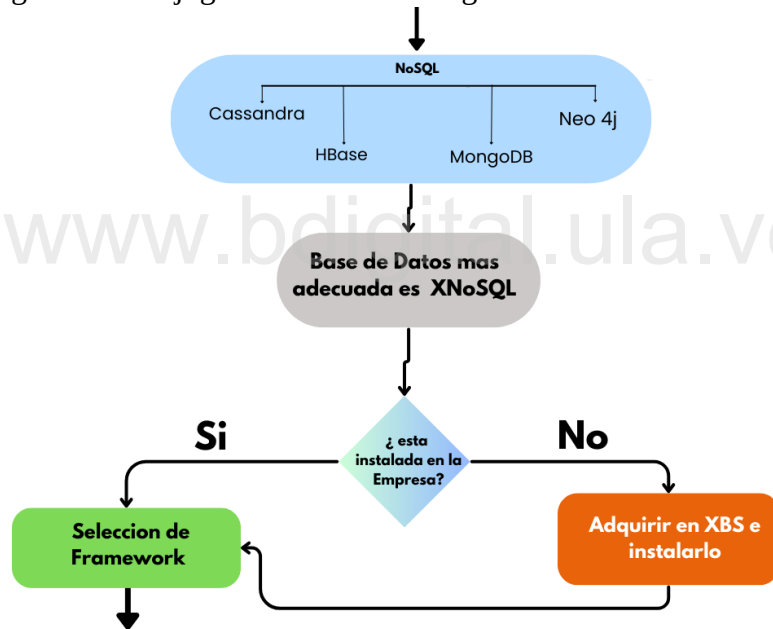


Figura 3.5. Flujograma de la Tecnología Base de Datos No Relacional

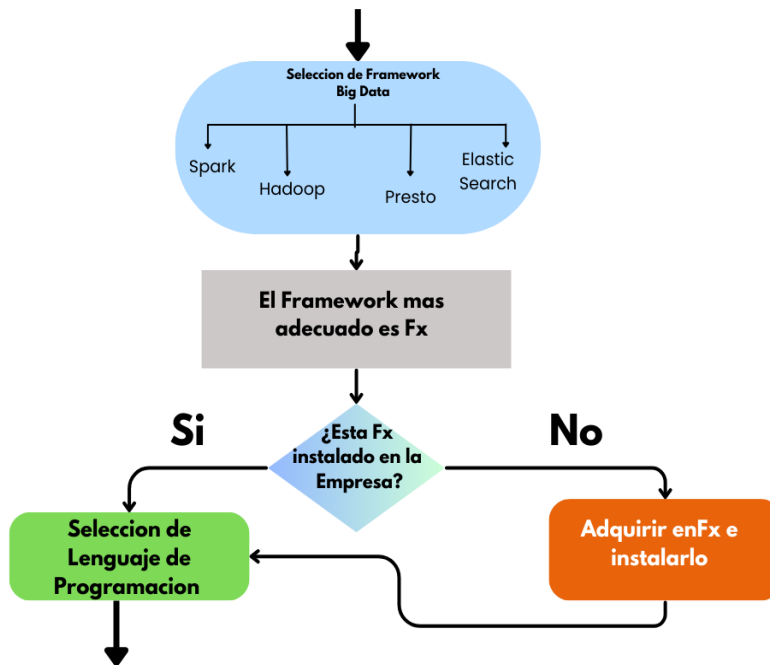


Figura 3.6. Flujograma de la Tecnología Framework

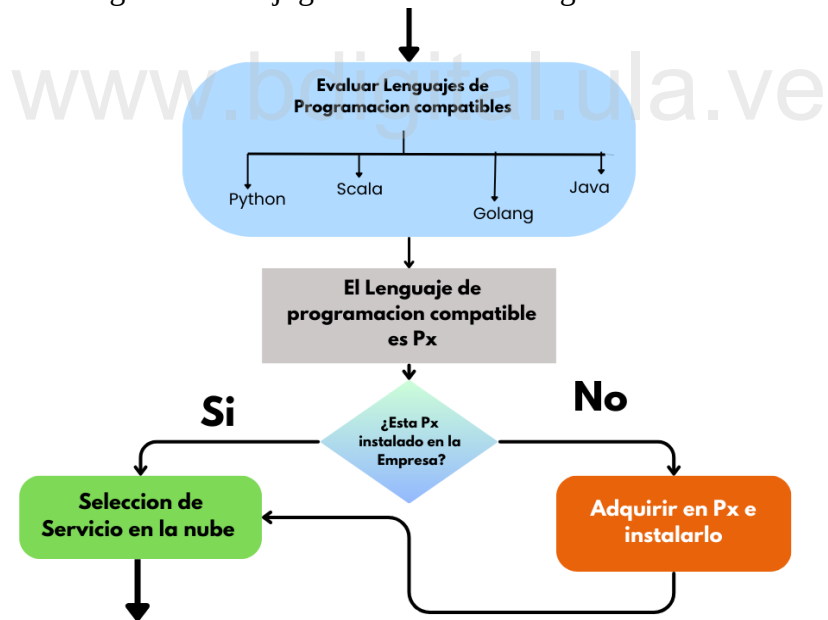


Figura 3.7. Flujograma de la Tecnología Lenguaje de Programación.

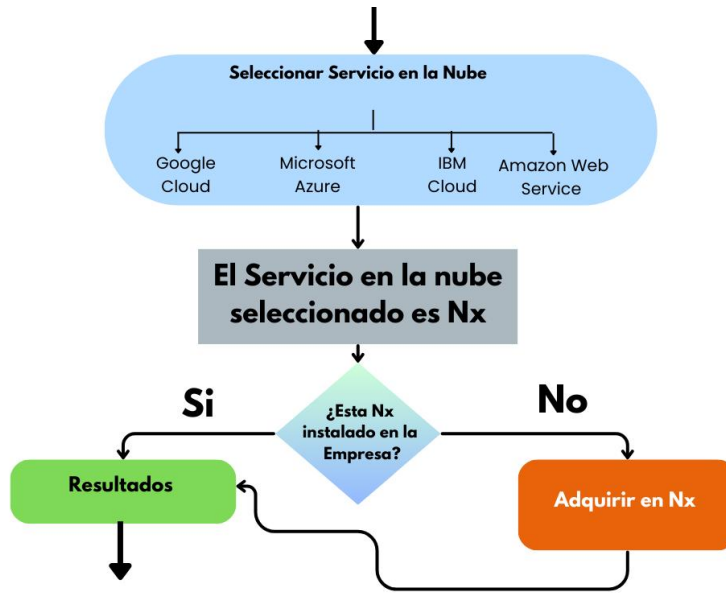


Figura 3.8. Flujograma para Servicio en la nube.

3.4. Modelo formal o matemático

Una vez teniendo claro los pasos a realizar, reflejados en el modelo lógico, se procedió a la “formalización” para obtener el modelo matemático, que permitirá obtener, en forma cuantitativa, un valor o puntuación ponderada para cada producto de la tecnología evaluada. El Modelo Selectivo Ponderado (propuesto por el profesor tutor) es una expresión matemática que utiliza las ponderaciones numéricas asignadas a cada variable selectiva y los valores de estas variables en cada caso. La expresión general para calcular la puntuación obtenida de cada producto tecnológico **k**, en cada tecnología **j** (hardware, sistema operativo, servicio en la nube), tomando en cuenta las respectivas ponderaciones **i** en cada tecnología, se expresa de la forma siguiente:

$$R_{k,j} = \sum_{i=1}^{n1} \sum_{j=1}^{n2} x_{kij} \cdot p_{ij}$$

Para: **j = 1, ..., n2**, representa las 7 tecnologías a evaluar (hardware, ..., servicio en la nube),

$k = 1, \dots, n3$, representa los 4 productos incluidos en cada una de las tecnologías a evaluar.

$i = 1, \dots, n1$, representa las 8 ponderaciones asignadas a las variables selectivas consideradas en cada tecnología a evaluar.

Donde

R_{kj} = Puntuación obtenida por el producto k en la tecnología j

x_{ikj} = Valor de la variable i en el producto k de la tecnología j .

p_{ij} = Ponderación de la variable i en la tecnología j .

- **Ponderación de variables**

Para la asignación de un valor, cuantitativo del 0 al 20, a las ponderaciones de las variables correspondientes a cada tecnología, se utilizó la Escala de Likert, criterio publicado por el psicólogo Rensis Likert (Likert. R., 1986) en su libro *Nuevas Formas para Solucionar Conflictos*. Simplificando, puede decirse que su criterio se basa en evitar asignar calificaciones drásticas como muy mal, mal, regular, bueno, excelente, pues por ejemplo, una calificación pudiera estar entre *moderado* y *bueno*, pero más cerca de *bueno* que de *regular*. Este investigador, propone asignar puntuaciones entre 0 y 20 a cada una de estas calificaciones. Por lo general se crean con 5 o 6 opciones de respuesta, pero hay ocasiones en donde puede haber más opciones, dependiendo las necesidades específicas del estudio y el nivel de medición.

de forma tal que una calificación de alguien o algo que pudiera estar entre *bueno* y *excelente*, pero más cerca de *bueno* que de *excelente*, pudiera ser 16 o 17.



3.5. Modelo programado

Una vez obtenido el modelo matemático se procedió a desarrollar el modelo programado, utilizando el entorno de desarrollo **MIT APP INVENTOR**, creado en el Instituto Tecnológico de Massachusetts para Google, la cual se encuentra alojada en la nube del MIT, obteniendo una aplicación móvil implementada en Sistema Operativo Android para celulares smartphone o tablet, que hemos bautizado Big Data Consultor. Detalles de las bondades y funciones de exponen en el Capítulo V: Desarrollo de la Aplicación Móvil.

www.bdigital.ula.ve

CAPÍTULO IV

APLICACIÓN DEL MODELO SELECTIVO PONDERADO

En este capítulo asignaremos las variables a los diferentes productos de cada tecnología y realizaremos la ponderación correspondiente, indagando en la web y consultando algunos expertos sobre el tema.

4.1. Investigación y selección de las variables para el modelo selectivo

Para la selección de las variables que se tomaron en cuenta en el modelo selectivo se realizó una investigación previa sobre los elementos más importantes que conviene tomar en cuenta a la hora de seleccionar tecnologías de Big Data. Durante esta investigación, se indagó en la web, criterios de selección de tecnologías, comparativas, y características (variables) que deben ser consideradas, así como también, la opinión de algunos expertos en la temática. A continuación, se presentan en un conjunto de tablas contentivas de las variables o características seleccionadas para cada tecnología:

4.1.1. Variables tomadas en cuenta para la selección de hardware para Big Data

Tabla 4.1. Variables seleccionadas para la tecnología Hardware

Características de la tecnología Hardware para Big Data
Escalabilidad
Cantidad de núcleos
Hilos del procesador
Eficiencia
Costo
Gama

4.1.1.1. Escalabilidad

La escalabilidad en hardware se refiere a la capacidad de un sistema para adaptarse a mayores demandas de recursos mediante la adición o actualización de componentes físicos.

4.1.1.2. Cantidad de núcleos

La cantidad de núcleos en un procesador es un factor clave que determina su rendimiento y capacidad para manejar múltiples tareas simultáneamente. Cuantos más núcleos tenga un procesador, mayor será su capacidad para procesar múltiples tareas en paralelo, lo que se traduce en un mejor rendimiento general.

4.1.1.3. Hilos del procesador

Los hilos de un procesador, también conocidos como subprocesos, son unidades virtuales de ejecución que permiten dividir y procesar tareas de forma simultánea dentro de un núcleo físico de la CPU. La función principal de los hilos es mejorar la eficiencia y el rendimiento del procesador al permitir el procesamiento paralelo de múltiples tareas en un mismo núcleo.

4.1.1.4. Eficiencia

La eficiencia se refiere a su capacidad para realizar tareas de manera óptima, consumiendo la menor cantidad de energía y recursos posibles del hardware.

4.1.1.5. Costo

El costo del hardware en términos de adquisición y mantenimiento, así como el costo de la implementación.

4.1.1.6. Gama

La gama de un procesador se refiere a la línea o serie de productos a la que pertenece dentro de la oferta del fabricante, lo cual indica su nivel de rendimiento y capacidades.

4.1.2. Variables tomadas en cuenta para la selección de Sistema Operativa para Big Data

A pesar de que existen varios sistemas operativos como Windows, Mac OS, Unix y Linux, la investigación arrojó que el 66% de las ofertas de trabajo demandaban saber y utilizar Linux, por ende, el estudio se hizo en base a el sistema operativo Linux.

Tabla 4.2. Variables seleccionadas para la tecnología Sistema Operativo

Características de la tecnología de Sistema Operativo para Big Data
Compatibilidad de software y aplicaciones
Personalización
Seguridad contra malware
Rendimiento del sistema
Políticas de actualización y soporte a largo plazo
Adaptabilidad a preferencias individuales
Consumo de hardware
Costos de soporte

4.1.2.1. Compatibilidad de software y aplicaciones

La compatibilidad de software y aplicaciones con los sistemas operativos es crucial para garantizar un funcionamiento adecuado en diferentes plataformas.

4.1.2.2. Personalización

La personalización del sistema operativo se refiere al proceso de configurar y adaptar el sistema operativo según las necesidades y preferencias específicas del usuario o la organización. Esto implica realizar ajustes y modificaciones en diversos aspectos del sistema, como la apariencia, las herramientas, las aplicaciones preinstaladas, las políticas de seguridad, entre otros.

4.1.2.3. Seguridad contra malware

La seguridad contra los malware es simplemente que tan seguro es el sistema operativo y las diferentes medidas de seguridad del proveedor para evitar la contaminación con los malware.

4.1.2.4. Rendimiento del sistema

Se refiere a su capacidad para ejecutar tareas y procesos de manera eficiente, aprovechando al máximo los recursos del sistema.

4.1.2.5. Políticas de actualización y soporte a largo plazo

Se refiere al período de tiempo durante el cual el fabricante o desarrollador brinda actualizaciones, parches de seguridad y asistencia técnica para esa versión específica del sistema operativo.

4.1.2.6. Adaptabilidad a preferencias individuales

Se refiere al poder adaptar y personalizar los elementos del sistema operativo para la satisfacción personal para un entorno de trabajo amigable.

4.1.2.7. Consumo de hardware

Se refiere a la estimación de los requisitos de hardware de un sistema operativo, es un proceso importante para garantizar un rendimiento óptimo y una experiencia de usuario satisfactoria.

4.1.2.8. Costo de soporte

Variable que considera el costo que puede tener el sistema operativo, actualización de versiones o parches del mismo.

4.1.3. Variables tomadas en cuenta para la selección de Base de Datos para Big Data

Tabla 4.3. Variables seleccionadas para la tecnología Base de Datos

Características de la tecnología de Base de Datos para Big Data	
Compatible con Cloud	
Capacidad de almacenamiento	
Escalabilidad	
Velocidad	
Compatible con Linux	
Costo	

4.1.3.1. Compatible con Cloud

Se basa en la compatibilidad de trabajar con la base de datos en la nube.

4.1.3.2. Capacidad de almacenamiento

La capacidad de almacenamiento de una base de datos es la cantidad de espacio en disco que ocupa físicamente.

4.1.3.3. Escalabilidad

Se refiere a la capacidad de una base de datos para adaptarse y mejorar su rendimiento y disponibilidad a medida que aumentan las demandas y el volumen de datos, tanto como su escalabilidad horizontal y vertical.

4.1.3.4. Velocidad

Se refiere a la rapidez con la que puede realizar operaciones de lectura, escritura, actualización y consulta de datos.

4.1.3.5. Compatible con Linux

Se refiere a la compatibilidad de una base de datos con un sistema operativo, es la capacidad de la base de datos para funcionar correctamente con un sistema operativo determinado.

4.1.3.6. Costo

Variable considerada para calcular el monto de pago por el uso del servicio.

4.1.4. Variables tomadas en cuenta para la selección de Framework para Big Data

Tabla 4.4. Variables tomadas en cuenta para seleccionar el Framework

Características de la tecnología de Framework para Big Data
Escalabilidad
Tolerancia a fallos
Eficiencia
Procesamiento en memoria
Procesamiento en tiempo real
Compatibilidad con Machine Learning
Soporte
Seguridad

4.1.4.1. Escalabilidad

Se refiere a la capacidad de adaptación y respuesta de un sistema con respecto al rendimiento del mismo a medida que aumentan de forma significativa el número de datos.

4.1.4.2. Tolerancia a fallos

Es la capacidad de un sistema para continuar funcionando de manera adecuada incluso cuando uno o más de sus componentes fallan, es la habilidad de un sistema para mantener su funcionamiento o para recuperarse de un fallo de manera automática, sin que el usuario final experimente una interrupción significativa en el servicio.

4.1.4.3. Eficiencia

Describe la capacidad de lograr los resultados deseados con el mínimo posible de recursos.

4.1.4.4. Procesamiento en memoria

Es la capacidad de realizar operaciones y manipular datos directamente en la memoria RAM (Random Access Memory), en lugar de tener que acceder constantemente al disco duro o almacenamiento secundario.

4.1.4.5. Procesamiento en tiempo real

Capacidades de procesamiento en tiempo real para gestionar flujos de datos continuos y realizar análisis en tiempo real.

4.1.4.6. Compatibilidad con Machine Learning

La compatibilidad con Machine Learning que es una rama de la inteligencia artificial (IA) que permite a las máquinas aprender de los datos y mejorar con la experiencia sin ser programadas explícitamente.

4.1.4.7. Soporte

Se refiere a la guiatura o soporte del uso de los frameworks

4.1.4.8. Seguridad

Describe la protección de la información con el objetivo de evitar la manipulación de datos y procesos por personas no autorizadas.

4.1.5. Variables tomadas en cuenta para la selección de lenguajes de programación para Big Data

Tabla 4.5. Variables seleccionadas para la tecnología Lenguaje de Programación

Características de la tecnología Lenguajes de Programación para Big Data
Flexibilidad
Fiabilidad
Escalabilidad
Velocidad
Compatibilidad
Seguridad

4.1.5.1. Flexibilidad

Se refiere a la capacidad de trabajar en diferentes industrias y roles, crear aplicaciones para múltiples dispositivos y sistemas operativos, elegir el enfoque de programación más adecuado y adaptarse a las nuevas tecnologías emergentes.

4.1.5.2. Fiabilidad

Esta variable define la capacidad de un lenguaje para ser confiable, estable, seguro y adecuado para el desarrollo de aplicaciones críticas y de alto rendimiento

4.1.5.3. Escalabilidad

Se refiere a la capacidad de adaptación y respuesta de un sistema con respecto al rendimiento del mismo a medida que aumentan de forma significativa el número de datos

4.1.5.4. Velocidad

Es importante decir que la velocidad de un lenguaje de programación es la rapidez con la que se puede escribir código y a la rapidez con la que el procesador ejecuta las instrucciones.

4.1.5.5. Compatibilidad

Esta variable describe la compatibilidad de los lenguajes de programación con diferentes sistemas y herramientas es un aspecto crucial en el desarrollo de software.

4.1.5.6. Seguridad

Describe la protección de la información con el objetivo de evitar la manipulación de datos y procesos por personas no autorizadas.

4.1.6. Variables tomadas en cuenta para la selección de servicio en la nube para Big Data

Tabla 4.6. Variables seleccionadas para la tecnología servicio en la nube

Características de la tecnología de servicio en la nube para Big Data
Precio mes (\$)
Espacio máximo ofrecido (TB)
Ancho de banda
Cantidad de conexiones permitidas
Accesibilidad de la base de datos
Calidad de la seguridad ofrecida
Portabilidad de la nube
Rendimiento del servicio informático

4.1.6.1. Precio por mes (\$)

La variable precio por mes se refiere al costo que tendrá el servicio contratado por cada mes que sea utilizado. El precio se verá reflejado en dólares (\$).

4.1.6.2. Espacio máximo ofrecido (TB)

Se refiere al espacio ofrecido de terabytes en la base de datos, ya que, al trabajar con grandes volúmenes de datos, se debe considerar un buen almacenamiento

4.1.6.3. Ancho de banda

Con ancho de banda se busca resaltar la velocidad del sistema y la cantidad de usuarios que puedan formar parte del sistema al mismo tiempo sin sufrir ningún tipo de caída o problemas en su funcionamiento.

4.1.6.4. Calidad de la seguridad ofrecida

Son los procesos, herramientas y controles implementados para proteger los datos almacenados contra amenazas accidentales e intencionales, garantizando la confidencialidad, integridad y disponibilidad de la información.

4.1.6.5. Accesibilidad de la base de datos

Se refiere a la capacidad de acceder y utilizar los datos almacenados en ella de manera eficiente y segura.

4.1.6.6. Calidad de la seguridad ofrecida

La seguridad del servicio es simplemente que tan seguros son los cortafuegos y las diferentes medidas de seguridad del proveedor para evitar que personas externas al cliente puedan ingresar, mirar y modificar lo que se tiene almacenado en el servidor.

4.1.6.7. Portabilidad de la nube

Se refiere a la capacidad de mover, transferir o migrar fácilmente los datos y aplicaciones de una base de datos en la nube a otra plataforma o proveedor de servicios en la nube.

4.1.6.8. Rendimiento del servicio informático

El rendimiento de las bases de datos es un aspecto crucial en los servicios informáticos, especialmente en entornos de Big Data.

4.2. Asignación del valor y la ponderación para cada variable

Luego de tener tanto las variables, como las tecnologías, el siguiente paso consiste en encontrar y asignar los valores y la ponderación a cada variable, para cada uno de las tecnologías.

Para la selección de las variables se tomaron algunos valores estandarizados y algunos valores fueron estimados tomando en cuenta diferentes factores como la opinión de expertos, información obtenida durante la investigación, y haciendo uso de la escala de Likert se les asignó la ponderación correspondiente, con un valor entre 0 y 20.

Las tablas con los valores asignados para cada una de las variables, respecto a cada producto en cada tecnología, se presentan a continuación:

HARDWARE	Intel i3	Intel i5	Intel i7	Intel i9
Escalabilidad	10	15	17	18
Cantidad de núcleos	4	10	20	24
Hilos del procesador	4	16	28	24
Eficiencia	10	15	18	19
Costo	112,87	209	446	670
Gama	5	10	15	19

Tabla 4.7. Valores de las variables para cada producto de la tecnología Hardware

SISTEMA OPERATIVO	Fedora	Debian	Manjaro	Ubuntu
Compatibilidad de software y aplicaciones	18	15	17	19
Personalización	16	16	16	16
Seguridad contra malware	18	17	15	17
Rendimiento del sistema	18	17	17	18
Políticas de actualización y soporte a largo plazo	18	18	16	18
Adaptabilidad a preferencias individuales	17	15	17	16
Consumo de hardware	19	15	17	17
Costo de soporte	4	8	5	12

Tabla 4.8. Valores de las variables para cada producto de la tecnología Sistema Operativo

BASES DE DATOS RELACIONALES	PostgreSQL	Sybase	Oracle	SQL Server
Compatibilidad con Cloud	17	16	18	13
Capacidad de almacenamiento	16	1	5	10
Escalabilidad	17	15	18	16
Velocidad	14	18	17	18
Compatibilidad con linux	16	18	17	10
Costo	10	5	1	18

Tabla 4.9. Valores de las variables para cada producto de la tecnología Base de Datos Relacional

BASE DATOS NO RELACIONALES	Cassandra	Hbase	MongoDB	Neo4j
Compatibilidad con Cloud	15	18	19	17
Capacidad de almacenamiento	2	10	4	3
Escalabilidad	17	16	19	14
Velocidad	16	18	19	18
Compatibilidad con Linux	18	16	17	16
Costo	16	15	16	18

Tabla 4.10. Valores de las variables para cada producto de la tecnología Base de Datos No Relacional

FRAMEWORK	Hadoop	Spark	Presto	Elasticsearch
Escalabilidad	17	19	16	19
Tolerancia a fallos	18	18	17	17
Eficiencia	17	19	17	18
Procesamiento en memoria	16	19	16	14
Procesamiento en tiempo real	17	19	16	14
Compatibilidad con Machine Learning	17	18	10	17
Soporte	18	18	14	17
Seguridad	19	17	18	16

Tabla 4.11. Valores de las variables para cada producto de la tecnología Frameworks

LENGUAJES	Python	Golang	Java	Scala
Flexibilidad	18	16	18	15
Fiabilidad	18	15	18	14
Escalabilidad	16	18	18	16
Velocidad	16	18	17	17
Compatibilidad	18	16	18	17
Seguridad	17	17	18	16

Tabla 4.12. Valores de las variables para cada producto de la tecnología Lenguajes de Programación

SERVICIO EN LA NUBE	Amazon Web Service	Microsoft Azure	Google Cloud	IBM Cloud
Precio mensual (\$)	240\$	560\$	500\$	1000\$
Espacio máximo ofrecido (TB)	40	64	64	20
Ancho de banda	17	18	17	16
Cantidad de conexiones permitidas	16	15	17	16
Accesibilidad de la base de datos	16	16	19	18
Calidad de la seguridad ofrecida	19	17	19	18
Portabilidad de la nube	17	18	18	16
Rendimiento del servicio informático	18	16	18	16

Tabla 4.13. Valores de las variables para cada producto de la tecnología Servicio en la Nube

CAPÍTULO V

DESARROLLO DE LA APLICACIÓN MÓVIL

En este capítulo se explicará de manera simplificada cómo se estructuró el desarrollo del sistema el sistema programado para obtener una aplicación móvil, el entorno de trabajo utilizado, las bondades del sistema y las pruebas realizadas para la validación de los resultados.

5.1. Entorno de Desarrollo

Como herramienta de desarrollo, se consideraron los lenguajes Kotlin, desarrollado por la sociedad JetBrains, de origen ruso, Flutter de Google y el MIT App Inventor II de Google. Kotlin es un lenguaje de programación de código abierto que se ha popularizado gracias a que se puede utilizar para programar aplicaciones Android. Flutter es un Kit de desarrollo de software (SDK) desarrollado por Google para crear aplicaciones móviles tanto para Android como para iOS, fue desarrollado como un software para uso interno dentro de la compañía, pero vieron el potencial que tenía y decidieron lanzarlo como proyecto de código libre. MIT App Inventor es un entorno de desarrollo, creado inicialmente en su primera versión, en el Google Labs, por los investigadores Mark Friedman de Google y Hal Abelson del Instituto Tecnológico de Massachusetts, La versión II, fue desarrollada y es mantenida actualmente en el MIT.

De estas tres herramientas, se seleccionó MIT App Inventor, por ser un entorno de desarrollo en la nube, donde se deseaba adquirir experiencia, y se contaba con el conocimiento previo del profesor tutor y suscripción al servicio en la nube del MIT.

5.2. Estructura de la Aplicación Móvil

La aplicación móvil fue estructurada en diez pantallas:

- Una pantalla de entrada, que identifica la aplicación.
- Una pantalla para el registro y selección de clientes.
- Una pantalla para el registro de datos de la Empresa Cliente seleccionada y datos de la Empresa Consultora.
- Una pantalla para cada una de las tecnologías evaluadas (hardware, ..., servicio en la nube) en las cuales puede registrarse las ponderaciones de las variables, los valores de las

variables para cada producto y aplicar el modelo matemático y seleccionar el producto con mayor puntuación.

- Finalmente, una pantalla que muestra los resultados obtenidos en cada una de las siete tecnologías evaluadas.

5.3. Prueba de la Aplicación Informática

Para probar la funcionalidad, amigabilidad y resultados de la aplicación informática se cargaron datos que identifican como Empresa Consultora a 3L Consulting, C.A. y a la Empresa SOS Tech Consulting, C.A como Empresa Cliente. Así mismo, se cargaron los datos correspondientes a las ponderaciones de las variables en cada tecnología, los valores de estas variables en cada producto. Los resultados obtenidos en las pruebas de cada sección o pantalla, operando la aplicación como un todo, se explican en las siguientes secciones.

5.3.1. Pantalla de ingreso

La pantalla de entrada da la bienvenida a la aplicación, en la que se muestra una imagen alusiva al tema de Big Data y contiene el nombre de la aplicación.



Pantalla 5.1 Inicio de la aplicación Big Data Consulting

5.3.2. Pantalla históricos de consultorías

Muestra en pantalla el histórico de las empresas que han solicitado consultoría para, registrando un número de cliente, el nombre de la empresa, ciudad y país donde se encuentran ubicados.



Cliente	Empresa	Ciudad	País	Detalles
C1	SOS Tech Consulting, C.A.	Panamá	Panamá	Ver
C2				Ver
C3				Ver
C4				Ver
C5				Ver
C6				Ver

Pantalla 5.2 Históricos de consultoría en la aplicación Big Data Consulting

5.3.3. Pantalla consultorías de Big Data

Al pulsar el *botón ver*, nos lleva a la siguiente pantalla la cual, cuenta con dos secciones, una en la parte superior con datos de la Empresa Cliente y una sección, en la parte inferior, con los datos de la Empresa Consultora.

CONSULTORIA BIG DATA

Datos de la Empresa Cliente

Razón Social: SOS Tech Consulting, C.A.

Ciudad: Panamá **País:** Panamá

Objeto de la Empresa

Desarrollo y consultoría de sistemas logísticos empresariales con tecnología SAP

Contacto: ING Raúl Colmenares **Móvil:** +507 1234 1234

Datos Consultoría

CONSULTORIA BIG DATA

Datos de la Empresa Consultora

Razón Social: 3L Consulting, C.A.

Ciudad: Valencia **País:** Venezuela

Objetivo de la Consultoría

Acompañamiento al cliente en la toma de decisiones para la adquisición de tecnologías para el desarrollo de un proyecto Big Data

Consultor Big Data: ING Luis A Dávila **Ingeniero de Datos:** Laura Rosa

Inicio: 01/07/2024 **Finalización:** 01/08/2024

Aplicar Modelo **Ver Resultados**

Pantalla 5.3 Datos del cliente en la aplicación Big Data Consulting

5.3.4. Pantallas de Selección de Hardware

Permite cargar y modificar las variables selectivas del modelo, los valores de las variables para los productos Intel i3, Intel i5, Intel i7 e Intel i9, y ver en *Resultados* el producto seleccionado.

APLICACIÓN DEL MODELO SELECTIVO

Empresa Cliente

SOS Tech Consulting, C.A.

Ciudad: Panamá **País:** Panamá

Hardware

Ponderación de Variables

Variable	Ponderación
Escalabilidad	17
Cantidad de núcleos	20
Hilos del procesador	16
Eficiencia	18
Costo	10
Gama	15

APLICACIÓN DEL MODELO SELECTIVO

Empresa Cliente

SOS Tech Consulting, C.A.

Ciudad: Panamá **País:** Panamá

Hardware

Aplicación del Modelo Selectivo

	Procesador Intel i3	Procesador Intel i5	Procesador Intel i7	Procesador Intel i9
Escalabilidad	10	15	17	18
Cantidad de núcleos	4	10	20	24
Hilos del procesador	4	16	28	24
Eficiencia	10	15	18	19
Costo	112.87	209	446	670
Gama	5	10	15	19

APLICACIÓN DEL MODELO SELECTIVO

Empresa Cliente

SOS Tech Consulting, C.A.

Ciudad: Panamá **País:** Panamá

Hardware

Resultados Ponderados

Tecnología Evaluada:	Puntuación
Intel i3	1697.7
Intel i5	3221
Intel i7	6146
Intel i9	8497

Decisión de la Consultoría

Selección: Intel i9

Comentario: En la empresa existe un Intel i7 y es recomendable adquirir un Intel i9

Pantalla 5.4. Tecnología Hardware para Big Data.

Permite cargar y modificar las variables selectivas del modelo, los valores de las variables para los productos Fedora, Debian, Manjaro, Ubuntu, y ver en *Resultados* el producto seleccionado.

Pantalla 5.5. Tecnología Sistema Operativo para Big Data

Permite cargar y modificar las variables selectivas del modelo, los valores de las variables para los productos PostgreSQL, Sybase, Oracle, SQL Server y ver en *Resultados* el producto seleccionado.

Pantalla 5.6. Tecnología Base de Datos Relacional para Big Data

5.3.7. Pantallas de Selección de Bases de Datos No Relacionales

Permite cargar y modificar las variables selectivas del modelo, los valores de las variables para los productos Cassandra, Hbase, MongoDB, Neo4j y ver en *Resultados* el producto seleccionado.

APLICACIÓN DEL MODELO SELECTIVO

Empresa Cliente

SOS Tech Consulting, C.A.

Ciudad: Panamá
País: Panamá

NoSQL

Base Datos NoSQL

Ponderación de Variables	
Variable	Ponderación
Compatibilidad con Cloud	17
Capacidad de Almacenamiento	17
Escalabilidad	19
Velocidad	18
Compatibilidad con Linux	17
Costo	13

APLICACIÓN DEL MODELO SELECTIVO					
Empresa Cliente					
SOS Tech Consulting, C.A.					
Ciudad:		Panamá	País:		Panamá
					
Aplicación del Modelo Selectivo					
Sistema de Datos NoSQL	Gestor NoSQL	Gestor NoSQL	Gestor NoSQL	Gestor NoSQL	
Método de Almacenamiento Selectivo	Cassandra	HBase	MongoDB	Neo4j	
Eficiencia y Escalabilidad en la Nube	15	18	19	17	
Capacidad de almacenamiento masivo	2	10	4	3	
Seguridad y Disponibilidad	17	16	19	14	
Flexibilidad	16	18	19	18	
Compatibilidad con Linux	18	16	17	16	
Tipo de licencia	16	15	16	18	

APLICACIÓN DEL MODELO SELECTIVO			
Empresa Cliente			
SOS Tech Consulting, C.A.			
Ciudad:	Panamá	País:	Panamá
			
Resultados Ponderados			
Tecnología Evaluada:			BD NoSQL
Bases de Datos NoSQL Evaluadas			Puntuación
Cassandra			10
HBase			10
MongoDB			10
Neo4j			10
Decisión de la Consultoría			
Selección:	MongoDB		
Comentario:	La empresa cliente no tiene experiencia con ninguna, se recomienda adquirir y entrenar a MongoDB		

Pantalla 5.7. Tecnología Base de Datos No Relacionales para Big Data

5.3.8. Pantallas de Selección de Frameworks

Permite cargar y modificar las variables selectivas del modelo, los valores de las variables para los productos Hadoop,Spark, Presto,ElasticSearch y ver en *Resultados* el producto seleccionado.

APLICACIÓN DEL MODELO SELECTIVO

Empresa Cliente

SOS Tech Consulting, C.A.

Ciudad:	Panamá	País:	Panamá
----------------	---------------	--------------	---------------





FRAMEWORKS






Ponderación de Variables

Variable	Ponderación
Escalabilidad	18
Tolerancia a fallos	16
Eficiencia	19
Procesamiento en Memoria	15
Procesamiento en tiempo real	17
Compatibilidad con Machine Learning	15
Soporte	17
Seguridad	18

APLICACIÓN DEL MODELO SELECTIVO

Empresa Cliente

SOS Tech Consulting, C.A.

Ciudad:	Panamá	País:	Panamá
----------------	--------	--------------	--------





FRAMEWORKS






Aplicación del Modelo Selectivo

FRAMEWORKS	Framework	Framework	Framework	Framework
Modelo Selectivo	Hadoop	Spark	Presto	Elasticsearch
Escalabilidad	19	19	16	19
Tolerancia a fallos	18	18	17	17
Flexibilidad	19	17	17	18
Consumo de Recursos / Memoria	17	19	16	14
Consumo de tiempo real	19	17	16	14
Compatibilidad con Machine Learning	17	18	10	17
Coste de soporte	18	18	14	17
Seguridad	19	17	18	16

APLICACIÓN DEL MODELO SELECTIVO			
Empresa Cliente			
SOS Tech Consulting, C.A.			
Ciudad:	Panamá	Páfs:	Panamá
			
Resultados Ponderados			
Tecnología Evaluada:			FRAMEWORK
Frameworks Evaluados			Puntuación
Hadoop			2472
Spark			2409
Presto			2107
Elasticsearch			2236
Decisión de la Consultoría			
Selección:	Hadoop		
Comentario:	Se recomienda adquirir el Framework Apache Hadoop		

Pantalla 5.8. Tecnología Frameworks para Big Data

Políticas de Selección de Servicio en la Nube

Permite cargar y modificar las variables selectivas del modelo, los valores de las variables para los productos Amazon Web Service, Microsoft Azure, Google Cloud, IBM Cloud y ver en *Resultados* el producto seleccionado.



Pantalla 5.9. Tecnología Lenguajes de Programación para Big Data

Políticas de Selección de Servicio en la Nube

Permite cargar y modificar las variables selectivas del modelo, los valores de las variables para los productos Amazon Web Service, Microsoft Azure, Google Cloud, IBM Cloud y ver en *Resultados* el producto seleccionado.



Pantalla 5.10. Tecnología Servicio en la Nube para Big Data

5.4. Pantalla de resultados

Por último, contamos con una pantalla de resultados donde se muestran los productos seleccionados por el modelo selectivo ponderado para cada tecnología.



RESULTADOS DE LA CONSULTORIA	
Empresa Cliente	
SOS Tech Consulting, C.A.	
Ciudad:	Panamá
País:	Panamá
Tecnologías Seleccionadas	
Tecnología	Selección
Hardware	Intel i9
Sistema Operativo	Ubuntu
Base Datos SQL	Oracle
Base Datos No SQL	MongoDB
Framework	Hadoop
Lenguaje	Java
Servicios Nube	Google Cloud

Pantalla 5.11. Resultados de la Consultoría Big Data Consulting

CONCLUSIONES

Al finalizar el presente proyecto se considera haber alcanzado los objetivos propuestos, logrando desarrollar exitosamente un modelo selectivo programado de tecnologías en el ámbito del Big Data. La investigación bibliográfica realizada, aunada a la opinión de algunos expertos en el tema y la indagación profunda sobre los diferentes productos asociados a cada tecnología, resultó fructífera arrojando las bases para la identificación de las variables más relevantes y que se deben considerar en cada tecnología.

El enfoque propuesto para el desarrollo del modelo, producto de la combinación de los enfoques de Aracil y Montgomery, conllevó a la obtención del modelo matemático ponderado, para finalmente obtener el modelo programado. El modelo selectivo ponderado ha demostrado ser una herramienta eficaz para la selección de los productos convenientes en las diferentes tecnologías de Big Data. Mediante el análisis de datos e información provenientes de investigaciones previas, se ha logrado asignar las ponderaciones a las diferentes variables propuestas. Estos resultados tienen implicaciones significativas para el uso de este modelo en general.

Es importante destacar que el modelo propuesto no se limita al dominio específico de las tecnologías de Big Data. Los principios subyacentes del modelo se pueden aplicar a una variedad de otras áreas de estudio que involucran la toma de decisiones entre múltiples alternativas. Por ejemplo, el modelo podría usarse para seleccionar los métodos más adecuados para un estudio de investigación en particular, identificar las oportunidades de inversión más efectivas o elegir elementos más apropiados para crear un negocio desde cero. La versatilidad del modelo lo convierte en una herramienta valiosa para investigadores y profesionales en una amplia gama de campos. El modelo propuesto ofrece un enfoque robusto y flexible para seleccionar las opciones más relevantes de un conjunto de alternativas. Su eficacia en el dominio de las tecnologías de Big Data demuestra su potencial de aplicación en una variedad de otras áreas de estudio.

RECOMENDACIONES

- Profundizar el caso de estudio, para brindar no solamente una consultoría sino también una asesoría más personalizada-
- Ahondar en la consulta especializada a expertos, para afinar los valores de las ponderaciones de las tecnologías, ya que estas pueden se pueden prestar a subjetividades.

www.bdigital.ula.ve

REFERENCIAS BIBLIOGRÁFICAS

Joyanes Aguilar, Luis (2013) “*Big Data: Análisis de Grandes Volúmenes de datos en organizaciones*” .disponible en:

https://books.google.es/books?hl=es&lr=&id=1GywDAAAQBAJ&oi=fnd&pg=PT6&dq=herramientas+de+big+data&ots=XV4N51g_S&sig=z8YUr5TeEDEA8n3mKvoAs80L-pw#v=onepage&q=herramientas%20de%20big%20data&f=false

Rodríguez Sánchez, Francisco Miguel (2015). *Lenguajes y Sistemas Informáticos* disponible en :
<https://repositorio.upct.es/handle/10317/4402>

Emilcy J. Hernández-Leal, Néstor D. Duque-Méndez y Julián Moreno-Cadavid (2017) . *Big Data: una exploración de investigaciones, tecnologías y casos de aplicación* , recuperado de :
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-77992017000200002

Alava Cruzatty Jose Efrain, Ramos Menéndez Ingrid Nathalia, “*Análisis Comparativo De Herramientas De Big Data Para La Carrera De Ingeniería En Sistemas Computacionales*” (marzo, 2019), disponible en: <http://repositorio.unesum.edu.ec/handle/53000/1537>

Bello, E (9 marzo 2022). *Big Data: qué es, para qué sirve y por qué es importante*, disponible en :
<https://www.iebschool.com/blog/valor-big-data/>

PowerData, *Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad*, disponible en:
<https://www.powerdata.es/big-data>

Dávila, Luis A (2023). *Tecnologías Big data Para Consultoría Gerencial*. 3L Consulting, C.A., Valencia, Venezuela

Infinitia Industrial Consulting. (28 de octubre de 2021). *Estudio de viabilidad de un proyecto ¿Cómo realizarlo?* disponible en: <https://www.infinitiaresearch.com/noticias/estudio-de-viabilidad-de-un-proyecto-como-realizarlo/#:~:text=Un%20estudio%20de%20viabilidad%20es,completar%20el%20proyecto%20con%20%C3%A9xito.>

Angel M.Rayo. Tipos de Big Data: clasificación por categoría y por origen disponible en:
<https://netmind.net/es/tipos-de-datos-en-big-data-2/>

Bismart. *Las 9 mejores herramientas de análisis de datos para data management*. disponible en <https://blog.bismart.com/9-mejores-herramientas-analisis-datos->

McAfee, A., & Brynjolfsson, E. (2012). *Big data: the management revolution*. *Harvard Business Review*, 90(10), 60-68. Recuperado de <https://hbr.org/2012/10/big-data-the-management-revolution>

Pascual Parada Torrealba / IEBS. *Tendencias Big Data 2023 para que el futuro no te pille de sorpresa*. Recuperado de <https://www.iebschool.com/blog/tendencias-big-data/>

Doug L y Burd B. ¿What is Java and why it is so great?. Recuperado en Octubre de 2019, disponible en: <https://www.dummies.com/programming/java/what-is-java-and-why-is-it-so-great/>

Chand M. ¿What is C# and what is C# used for? ,disponible en: <https://www.c-sharpcorner.com/article/what-is-c-sharp/>

Morris S. What is Python. ,disponible en: <https://skillcrush.com/2019/06/21/what-is-python/>

Santaella Jesus, ¿Qué es la programación en R? , (04 enero 2022), disponible en: <https://talently.tech/blog/programacion-en-r/#:~:text=R%20quiere%20decir%20que%20se,de%20datos%20de%20lenguaje%20m%C3%A1quina>

Alvarez Miguel , *Qué es HTML*, (01 de enero 2001), disponible en: <https://desarrolloweb.com/articulos/que-es-html.html>

Pérez Porto, J., Gardey, A. (19 de septiembre de 2018). *JavaScript - Qué es, definición y concepto*. Definicion.de. Última actualización el 16 de agosto de 2019. Recuperado el 16 de septiembre de 2023 de <https://definicion.de/javascript/>

Amazon Web Service (2023). ¿Qué es JavaScript? en: <https://aws.amazon.com/es/what-is/javascript/>

Watson T. *What is GO programming and when to use it?*, disponible en: <https://skywell.software/blog/what-is-go-programming-language/>

Amazon Web Services, sitio oficial., disponible en: <https://aws.amazon.com/es/>

Microsoft Azure, sitio oficial., disponible en: <https://azure.microsoft.com/es-es/>

Google Cloud, sitio oficial. disponible en: <https://cloud.google.com/>

Oracle Cloud, sitio oficial., disponible en: <https://www.oracle.com/cloud/>

IBM Cloud, sitio oficial. ,disponible en: <https://www.ibm.com/cloud>

Ondho, Branding y Marketing online, *Data Warehouse*, disponible en: <https://ondho.com/diccionario-de-marketing/term/data-warehouse/>

Mendoza Aranza , (10 diciembre 2020), *Tipos de Data Warehouse* , disponible en : <https://gravitar.biz/datawarehouse/data-warehouse-tipos/>

Google Cloud. *¿Qué es Presto?*. disponible en: <https://cloud.google.com/learn/what-is-presto?hl=es#:~:text=Presto%20es%20un%20motor%20de,de%20grandes%20vol%C3%BAmenes%20de%20datos.>

Elasticsearch. *El corazón del Elastic Stack, gratuito y abierto.* disponible en : <https://www.elastic.co/es/elasticsearch>

PowerData (28 febrero 2020) *¿Qué es el Apache Hadoop?* disponible en : <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/397377/qu-es-el-apache-hadoop#:~:text=Apache%20HBase%3A%20sistema%20de%20almacenamiento,tiempo%20real%20para%20cualquier%20aplicaci%C3%B3n>

Google Cloud (s.f), *¿Qué es Apache Spark?* ,disponible en: <https://cloud.google.com/learn/what-is-apache-spark?hl=es>

IBM (s.f), *¿Qué es una base de datos NoSQL?* , disponible en: <https://www.ibm.com/es-es/topics/nosql-databases>

Fernandez Oscar (10 julio 2023),*Apache Hive: Introducción* , disponible en : <https://aprenderbigdata.com/apache-hive/>

Fernandez Oscar (30 agosto 2023), *Introducción a Apache Impala*, disponible en : <https://aprenderbigdata.com/apache-impala/>

MongoDB sitio oficial, ¿Qué es MongoDB?, disponible en : <https://www.mongodb.com/es/what-is-mongodb>

Power Data (11 marzo 2016), *Cassandra base de datos: agilidad y rendimiento a prueba de fallos* , disponible en: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/cassandra-base-de-datos-agilidad-y-rendimiento-a-prueba-de-fallos#:~:text=Cassandra%20es%20un%20sistema%20de,ning%C3%BAAn%20punto%20%C3%BAnico%20de%20fallo>

Red Hat (10 de Octubre de 2022), *¿Qué es Apache Kafka?* , disponible en : <https://www.redhat.com/es/topics/integration/what-is-apache-kafka>

Aguilera, (2000), *CONCEPTO DE MODELO*, eumed.net, https://www.eumed.net/tesis-doctorales/2012/lsg/concepto_modelo.html

Ibáñez Juan Jose (10 MAYO 2008), *Concepto y Tipos de Modelos Científicos* , disponible en : <https://www.madrimasd.org/blogs/universo/2008/05/10/91441#:~:text=Un%20modelo%20%C3%B3gico%20es%20aquel,auxilio%20de%20la%20simbolog%C3%ADa%20matem%C3%A1tica>.

Ondarse Alvarez Dianely (12 de marzo 2022), *Modelo científico*, disponible en : <https://www.ejemplos.co/modelo-cientifico/#:~:text=Son%20copias%20o%20representaciones%20tridimensionales,los%20prototipos%20y%20las%20maquetas>

Garcia Francisco (s.f), *Modelos de programación de Big Data* , disponible en : <http://spaceanalytics.blogspot.com/2016/07/modelo-programacion-big-data.html>

Davila, L (2021), *¿ es factible crear un gerente autómatas?* , Editorial Independently Published

Eugene Bloom (enero 10, 2024), *Escala de Likert: qué es y ejemplos de cómo utilizarla*, disponible en : <https://www.ispring.es/blog/escala-de-likert>

Piqueras V. Proceso Analítico Jerárquico (AHP). Recuperado en noviembre de 2019, disponible en: <https://victoryepes.blogs.upv.es/2018/11/27/proceso-analitico-jerarquico-ahp/>

Gutierrez,L.,(1988), *Sistema de Apoyo Gerencial para la Diversificación de la Producción en FORAMET C.A*, EISULA, Mérida, Venezuela.

DaSilva,J.,(20202), *Modelo Selectivo de Tecnologías de Computación en la Nube para la Presentación de Servicios de consultoría* , EISULA,Mérida,Venezuela

Likert,R.(1986), *Nuevas Formas para Solucionar Conflictos*, Editorial Trillas,México.

Montgomery,D.B & Urban,G.L(1977), *Marketing Científico*, Editorial Pirámides,Madrid.

Aracil,J.(1978), *Introducción a la dinámica de sistemas*, Editorial Alianza, Madrid.

www.bdigital.ula.ve