DECANATO DE INGENIERÍA E INFORMÁTICA

ESCUELA DE INFORMÁTICA

Título de la Monografía:

DISEÑO DE LAGO DE DATOS EN NUBE PARA COMPARTIR "INTELIGENCIA DELICTIVA" DE CIUDADANOS DOMINICANOS ENTRE INSTITUCIONES GUBERNAMENTALES.

Sustentada por:

Isaac Alfonso Sánchez Lora	A00084851
Mario Yunior Vargas Cabrera	A00099841
Engel De Jesús Rodríguez Gómez	A00102052

Asesores

Ing. Willis Ezequiel Polanco

Ing. Mario Antonio Luciano Recio

Monografía para optar por el título de:

Ingeniero de Sistemas de Computación

Santo Domingo, D.N.
Diciembre, 2021

DISEÑO DE LAGO DE DATOS EN NUBE PARA COMPARTIR "INTELIGENCIA DELICTIVA" DE CIUDADANOS DOMINICANOS ENTRE INSTITUCIONES GUBERNAMENTALES

Índice General

Índice de Figuras

Índice de Gráficos

Resumen

La población dominicana desde hace varios años ha estado lidiando con una alta tasa de criminalidad en la mayor parte del territorio, pero con la paralización de las distintas actividades de labor, debido a la situación sanitaria que inició en el año 2020, el país ha estado pasando por una crisis económica grave. Tal crisis también ha tenido lugar en varias partes del mundo, lo que ha provocado un aumento grave de los actos delictivos.

Esta situación ha provocado que los ciudadanos, cuando salen a las distintas calles y avenidas, lo hagan con cierto temor de ser asaltados por algún delincuente. A su vez, esto ha generado cierto miedo de transitar a horas de la noche por alguna de estas vías. Además, con la cierta lentitud en la que las instituciones responsables en preservar la seguridad y el orden resuelven los casos o querellas, afianzan aún más este temor.

Para tratar de solucionar estas problemáticas, la presente investigación desarrolla un diseño de un lago de datos. El mismo consiste en que las instituciones gubernamentales que tengan como objetivo preservar la seguridad y el orden de los ciudadanos puedan cargar la inteligencia delictiva que tengan almacenada en sus servidores, para posteriormente ser utilizada desde el lago de datos para realizar análisis y reportes por medio de herramientas especializadas en la analítica de datos e inteligencia de negocios, con la finalidad de agilizar el proceso de resolución de los casos o querellas.

Agradecimientos

No se puede empezar por otro punto que no sea agradeciendo a Dios siempre por permitirnos ser, por permitir nuestro accionar día tras día para cumplir con nuestros propósitos y con los suyos; Por todo lo que nos da y nos aleja. No puedo continuar en otra dirección que no sea agradecerle especialmente a mi madre, a mi abuela, a los ‘’capitaleños’’ y a toda mi familia en general (amigos incluidos) por todo el apoyo que siempre han depositado en mí para que hoy por hoy haya podido llegar hasta este punto y seguir con ganas de continuar. Es para mí un placer y un honor poder disfrutar de ellos, y espero que todo el tiempo que nos queda juntos sea de muchos más éxitos, para que también puedan seguir disfrutando de mí.

Además, y manteniendo igual grado de importancia, definitivamente debo agradecer a esos compañeros colegas con los que pude tener la dicha de compartir este trayecto de toda la carrera. Demostraron ser personas leales, serviciales, responsables y sobre todo humanos. Infinitas gracias, porque sin ustedes innegablemente hubiera sido mucho más difícil haberlo logrado. No voy a mencionar nombres, para evitar que el sueño que tengo en estos momentos me traicione y se me pase alguno, pero les enviaré este agradecimiento para que sepan que los tendré presente siempre.

Y por último, hay que mencionar a aquellos profesores que siempre trataron de guiarnos por caminos reales e ilustrativos; aquellos que nos mostraron, con sus acciones, las directrices para ser buenos profesionales. Antes de terminar, también quiero agradecerles a aquellos que, por sus mismas acciones, despertaban en nosotros eso que nos decía que así no debíamos ser. Todos ellos sirvieron de ejemplo; unos de ejemplo a seguir y otros de todo lo contrario. Gracias interminables para todos.

~ Engel Rodríguez

Agradecimientos

En primer lugar, agradezco a Dios por haberme dado la vida, salud y la familia con la que cuento. Me dio las fuerzas para lograr completar mi carrera universitaria.

A mi madre **Esperanza Cabrera** por haberme dado la vida y por siempre estar a mi lado dándome todo el apoyo para lograr mis metas. A mi padre **Mario Vargas Ayala** por siempre estar firme y ayudarme con su apoyo y consejos a lo largo de todo lo que he vivido, y con esto poder regalarles este logro. ¡Los amo!

A mi hermana **Ambar María** por siempre darme apoyo en todo lo que me he propuesto.

A los compañeros que conocí desde el día uno en mi travesía por la universidad y que actualmente siguen firmes, se me hace difícil mencionar solo a unos cuantos, pero para mí siempre fueron un gran apoyo y este es el resultado del final de ese viaje que empecé en 2018.

Gracias al **Politécnico OSCUS San Valero** quienes fueron los que me dieron la oportunidad de una beca para realizar mis estudios en la Universidad APEC.

A la **Universidad APEC** por haber sido la casa de estudios que me permitió formarme académicamente en sus aulas y a cada uno de los docentes que se empeñaron firmemente en transmitir de la mejor manera sus conocimientos y todas sus experiencias.

A mis compañeros de monográfico **Engel Rodríguez** e **Isaac Sánchez** por haberme aceptado para desarrollar este trabajo de investigación y por la buena vibra que tuvimos durante todo este proceso.

~ Mario Yunior Vargas Cabrera

Agradecimientos

La culminación de este monográfico ha sido uno de los desafíos y uno de los logros más significativos de mi vida. Quiero agradecer a todas las personas que, de una forma u otra, han contribuido con la exitosa finalización de esta investigación. En primer lugar, quiero agradecer a Dios, a quien todo se lo debo. Después a mi madre, Ceyda Luz Lora; a mi padre Leonel Sánchez y mis hermanos: Leonela y Daniel, por su comprensión y tolerancia. Y a todos mis demás familiares.

Agradezco a mis profesores, por su apoyo, tanto académico como personal. También, mi agradecimiento a la Universidad APEC por permitirme ser uno más de la familia de esta institución y haberme formado académicamente. Mi agradecimiento se extiende a todos y cada uno de mis compañeros de monográfico, principalmente a mis compañeros Mario Vargas y Engel Rodriguez. A mis amigos y amigas que siempre me animaron en los momentos difíciles. Además, a Randol Frías, por su amistad y su disposición a ayudarme en mi progreso personal y académico. A todos muchas gracias.

~ Isaac Alfonso Sanchez Lora

Dedicatoria

Este esfuerzo lo dedicaré a mi madre hermosa que la amo y la amaré siempre. Te lo dedico a ti mi amor, porque me das las fuerzas y motivación necesaria para querer seguir adelante siempre. Y también me lo dedico a mí, porque solo Dios y yo sabemos todo por lo que se ha pasado y lo que se ha sacrificado para poder lograrlo. Hay sacrificios que solo nosotros mismos comprendemos lo grande que es hacerlos, y por eso hay que aplaudirnos a nosotros mismos.

Por último, se lo dedico a todos esos seres importantes que permanecen en mi corazón y que aportaron de una u otra forma para que esto fuese un hecho.

~ Engel Rodríguez

Dedicatoria

Le dedico este trabajo a Dios por siempre darme las fuerzas para salir adelante y no caer. Además de bendecirme con la familia que tengo.

A mis padres Esperanza Cabrera y Mario Vargas, por ser las personas que se dedicaron a formarme, por enseñarme que los sueños se cumplen con dedicación, esfuerzo y trabajo duro, y por guiarme en los momentos en donde tuve más dudas y ayudarme a ver la luz. Este logro se lo dedico a ustedes.

Para mí el conocimiento siempre tendrá un gran peso en mi vida, y para mí esto es un sueño hecho realidad, como dijo mi rapero favorito Nach en su canción Manifiesto: “Mis sueños son mentiras que algún día dejarán de serlo”, gracias a esas palabras hoy puedo decir que una mentira se ha vuelto realidad.

~ Mario Yunior Vargas Cabrera

Dedicatoria

Esta monografía se la dedico en primer lugar a Dios, por haberme dado la sabiduría y fuerza necesaria para terminar la carrera. A mi madre Ceyda Luz Lora por su apoyo en todo momento, entiendo que sin ella no hubiera sido posible seguir adelante y culminar la carrera. También a mi abuela Altagracia Lora, aunque hoy no está entre nosotros, pero sé que desde el cielo podrá ver a su nieto graduarse.

~ Isaac Alfonso Sanchez Lora

Introducción

La República Dominicana ha alcanzado un crecimiento constante que ha traído consigo varias problemáticas que jamás habíamos imaginado. Un ejemplo de esto es la alta tasa de criminalidad que ha estado arropando a todo el país desde hace varios años, pero se ha intensificado desde el año 2020.

En la sociedad que vivimos contamos con innumerables herramientas destinadas a mejorar la eficiencia de muchos procesos y, en otros casos hasta tareas cotidianas con la finalidad de hacer estas de una manera más sencilla y rápida.

Las instituciones gubernamentales son un foco de varias problemáticas que afectan a la población dominicana. La poca eficiencia al momento de resolver un caso criminal entre otras razones es la que ha llevado a que la tasa de criminalidad siga en aumento, y esto desata el miedo en la población de no poder transitar las calles y avenidas con tranquilidad.

El diseño de un Lago de datos para las instituciones gubernamentales es una forma de ayudar a mitigar todas estas consecuencias, ya que permitirá utilizar los datos almacenados para realizar análisis y reportes de una manera más eficiente con el objetivo de resolver cada uno de los casos que se presenten.

El siguiente trabajo de investigación presenta su estructura de la siguiente manera:

El capítulo 1 - se desarrollan los aspectos generales de nuestra investigación como es el planteamiento del problema, los objetivos, justificaciones, entre otros más.

El capítulo 2 - se desarrolla cada uno de los términos y aspectos relacionados con los sistemas gestores de bases de datos y los lagos de datos.

El capítulo 3 - se desarrolla un diagnóstico sobre los antecedentes de un almacén de datos, su evolución y su estado actual junto con el lago de datos.

El capítulo 4 - se desarrolla un análisis sobre las instituciones gubernamentales y las distintas tecnologías con la que estos cuentan en la actualidad.

El capítulo 5 - se desarrolla la necesidad de diseñar un lago de datos, beneficios, infraestructura y algunos de los proveedores más confiables en la actualidad.

El capítulo 6 - se desarrolla nuestra propuesta de diseño de un lago de datos, con sus características, requerimientos, arquitectura y su modelo de despliegue. Además de ciertos factores a tomar en consideración antes de realizar una migración hacia este.

CAPÍTULO 1: ASPECTOS GENERALES

1. Problema de Investigación

1.1. Título

Diseño de Lago de Datos en Nube Para Compartir “Inteligencia Delictiva” de Ciudadanos Dominicanos Entre Instituciones Gubernamentales.

2. Planteamiento del Problema

Actualmente, no existe la armonía y consonancia necesaria en las operaciones de las instituciones que rigen las fuerzas del orden del país. Según su accionar, se aprecia que las operaciones de una y otra son aisladas y no sincronizadas, cuando en realidad deberían estar todas más alineadas y orientadas a propósitos más similares que produzcan como resultado la seguridad y el orden de la ciudadanía.

Por la forma de actuar, es evidente que no hay integración alguna en las plataformas digitales de las fuerzas del orden que manejan informaciones de naturaleza delictiva. Informaciones que son vitales y sirven para la toma de decisión y acción inmediata de los agentes en el campo de trabajo. De no ser el caso, la implementación de esta integración es bastante deficiente y no produce los resultados esperados, pudiendo ser por falta de imposición en la utilización de estos mecanismos, por falta de herramientas necesarias para realizar la identificación correspondiente, o porque fue una integración fallida.

De cualquier modo, si en alguna provincia del país se comente algún crimen o delito por un individuo o grupo de personas, y estas personas son fichadas por la justicia, no será posible para un agente de tránsito, por ejemplo, identificar el objetivo en caso de que establezca contacto con el/los implicados mientras regula el tránsito. Por otro lado, si los implicados intentasen desplazarse a otra provincia del territorio nacional, los agentes del ejército nacional colocados en los llamados ‘’chequeos militares’’, agentes de tránsito de esa otra provincia, o incluso el mismo cuerpo policial local no estaría en la capacidad de distinguir a los implicados del resto de ciudadanos de forma inmediata. Además, si una persona tiene antecedentes penales no puede considerarse como un candidato nato para ser político en nuestro país.

Esto denota una innegable brecha en los métodos para mantener y garantizar la seguridad y justicia dentro de nuestras fronteras. Sin un mecanismo para la identificación de aquellos que violentan las leyes constitucionales que luego intentan desplazarse para evitar su sometimiento ante la justicia, cada día

resulta más complicado garantizar la seguridad y el orden en el país.

3. Objetivos Generales y Específicos de la Investigación

3.1. Objetivo General

Elaborar un diseño de un Lago de Datos basado en Nube con la finalidad de compartir inteligencia delictiva de ciudadanos dominicanos, entre instituciones gubernamentales.

3.2. Objetivos Específicos

1. Analizar los requisitos técnicos necesarios para la implementación de un Lago de Datos en República Dominicana.
2. Diagnosticar los sistemas existentes.
3. Identificar debilidades y deficiencias de los sistemas actuales.
4. Determinar la plataforma basada en la nube que resulte idónea para la correcta implementación de un Lago de Datos.
5. Diseñar plan de implementación de herramientas que permitan la revisión, integración, migración y transformación de los datos existentes en los sistemas previamente analizados.

4. Justificaciones Teóricas, Metodológicas y Prácticas de la Investigación.

4.1. Justificación Teórica

Se ha identificado que el Estado Dominicano cuenta con un grave problema relacionado a la inteligencia delictiva de los ciudadanos. Por ende, la población se ha visto en una gran desventaja, cuando llegan a ser víctimas de algún delito cometido por alguien hacia su persona.

Por lo general, los agentes de las distintas instituciones encargadas de hacer cumplir las leyes no cuentan con una plataforma de gestión de datos que les permita manejar con flexibilidad toda la inteligencia delictiva de un ciudadano. El objetivo es diseñar una solución basada en un “Lago de Datos” que permita facilitar la gestión y manejo de toda la inteligencia delictiva y que, como consecuencia, sirva para reducir la desventaja que tiene la población frente a las amenazas para con su seguridad e integridad que representan aquellos que delinquen.

4.2. Justificación Metodológica

En virtud de que el objetivo de este estudio es diseñar una solución informática, enfocada en un lago de datos, para facilitar la gestión y manejo de la inteligencia delictiva de los ciudadanos dominicanos por parte de las

instituciones gubernamentales, se optó por implementar la investigación de tipo descriptiva. Por otro lado, se utilizará el análisis como método de investigación, que pretende identificar las características de cada una de las partes que conforman el tema en cuestión. Y, por último, la encuesta será el instrumento de recolección de información que se necesitará para el estudio.

4.3. Justificación Práctica

Esta investigación se realiza debido a que existe una alta probabilidad de que una persona que haya cometido algún hecho delictivo se dé a la fuga y pueda moverse libremente entre las distintas provincias y pueblos del país.

Con esto, se busca que las instituciones gubernamentales pertinentes puedan tener un manejo y gestión eficiente de la inteligencia delictiva de estas personas. Además, se busca también centralizar toda esta información en un único punto. Adicional a esto, la población en general obtiene como beneficio el rápido reconocimiento del o los implicados en algún crimen, con la finalidad de que las autoridades pertinentes realicen un buen trabajo.

5. Marco de Referencias

5.1. Marco Teórico

Hasta el día de hoy, existen distintas organizaciones a nivel internacional que brindan diferentes tipos de servicios basados en la nube, entre los cuales están los lagos de datos. Por esta razón, existe una amplia variedad de información relacionada a esta tecnología. Basado en esto, se tomará como referencia algunos de los conceptos más destacados para el desarrollo de nuestra investigación.

García (2017) describe que un lago de datos alberga una gran cantidad de datos primarios en su formato nativo hasta que resultan necesarios; mientras que un depósito de datos jerárquico (datawarehouse) almacena datos en archivos o carpetas. Un lago de datos utiliza una arquitectura plana para almacenar datos. Cada elemento de datos en un lago tiene asignado un identificador único y está marcado con un conjunto de etiquetas de metadatos extendidos. Cuando surge el análisis de un elemento, resulta posible efectuar una consulta al lago de datos en busca de datos relevantes y, al mismo tiempo, cabe la posibilidad de analizar dicho conjunto de datos más pequeño para ayudar a responder a la consulta.

Amazon (2021), en su página web “Arquitectura de Lake House | Amazon

Web Services", plantea que la razón por la cual las empresas deberían optar por un enfoque de Lake House es porque los volúmenes de datos aumentan a un ritmo sin precedentes, de terabytes a petabytes y a veces a exabytes. Los enfoques tradicionales de análisis de datos en las instalaciones no pueden gestionar estos volúmenes de datos, porque no son lo suficientemente escalables y resultan demasiado caros. Varias empresas se encuentran en el proceso de recopilar todos sus datos a partir de diversos silos, para agruparlos en un solo lugar, que muchos denominan lago de datos, para realizar análisis y ML (Machine Learning) directamente sobre esos datos. En otras ocasiones, estas mismas empresas almacenan otros datos en almacenes de datos creados para fines específicos y así poder analizar y obtener rápidamente información a partir de datos estructurados y no estructurados. Este traslado de datos puede ser "desde el interior hacia el exterior", "desde el exterior hacia el interior" o "alrededor del perímetro", porque los datos tienen gravedad.

Peternel (2021) plantea que una arquitectura lakehouse y los componentes internos de Delta Lake están diseñados para eliminar la necesidad de tener siempre una configuración de arquitectura de dos niveles de Data Warehouse / Data Lake. Con las transacciones ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad) en un data lake, los archivos de datos subyacentes vinculados a una tabla externa no se actualizarán hasta que una transacción se complete correctamente o falle por completo. Por lo tanto, esto simplifica la carga incremental, es decir, la adopción de datos. De esta forma, se evita que los trabajos fallen en el proceso (recuperación de ETL) y la modificación de datos (eliminaciones, actualizaciones e inserciones). Como resultado, las discrepancias de metadatos se reducen significativamente debido al registro de control de versiones y cumplimiento de ACID que Delta Lake trae a Data Lakes.

Aslett (2021), propone que un data lakehouse difumina las líneas entre los lagos de datos y el almacenamiento de datos al mantener las ventajas de costo y flexibilidad de los datos persistentes en el almacenamiento en la nube. Al mismo tiempo, permite que se aplique el esquema para subconjuntos de datos seleccionados en zonas conceptuales específicas del lago de datos, o una base de datos analítica asociada, para acelerar el análisis y la toma de decisiones comerciales.

5.2. Marco Conceptual

Datos

“Un dato se puede describir como la representación simbólica de un atributo o característica de una entidad en específico, ya sea por medio de números, letras, entre otros más” (UCLA, 2005).

Bases de datos

Como menciona Oracle (2021), una base de datos no es más que una recopilación organizada de información o datos estructurados, que por lo general suele ser almacenada de forma electrónica en un sistema informático.

Lago de datos

Según Red Hat (2021), un lago de datos es conocido como un tipo de repositorio que permite almacenar conjuntos grandes y diversos de datos sin procesar en su formato original, y que permite mantener una perspectiva general de ellos.

Instituciones gubernamentales

De acuerdo con la Dirección General de Impuestos Internos (DGII, 2021), las instituciones gubernamentales son cada una de las organizaciones fundamentales de un estado, nación o sociedad, que tienen como finalidad brindar un servicio público necesario para la ciudadanía.

Inteligencia delictiva

“Es simplemente cualquier tipo de información con valor adicional que puede ser utilizado por los agentes del orden para combatir la delincuencia” (Oficina de las Naciones Unidas contra la Droga y el Delito, 2010).

Nube (Cloud)

Es un término utilizado para referirse a un modelo informático que se basa en la entrega y consumo de servicios, que implica el uso de distintas tecnologías para satisfacer la necesidad de cualquier cliente o empresa sin importar su ubicación geográfica.

Información

“Es un conjunto organizado de datos, que constituyen un mensaje sobre un determinado ente o fenómeno” (UCLA, 2005).

Sistema

El Instituto de Nutrición de Centroamérica y Panamá (INCAP, 2021)

afirma que un sistema es un conjunto de elementos que permiten interactuar entre sí con un fin común, y permite que la información esté disponible para satisfacer las necesidades de una organización en cualquier momento.

Ley

"Es la norma o regla que se aprueba a través de procedimientos específicos por la autoridad de un Estado y que contiene regulación determinada de ciertos ámbitos de la vida de las personas" (Trujillo E., 2020).

Seguridad

Según el Institut National de Santé Publique du Québec, en español Instituto Nacional de Salud Pública de Quebec (INSPQ, 2021), es un estado en el cual los peligros y las condiciones que pueden provocar daños de tipo físico, psicológico o material pueden ser controlados para preservar la salud y el bienestar de los individuos y de la comunidad.

Justicia

"Es un principio universal que rige la aplicación del derecho para conseguir que se actúe con la verdad dando a cada uno lo que le corresponde" (Trujillo E., 2020).

Ciberseguridad

"Es el conjunto de procedimientos y herramientas que se implementan para proteger la información que se genera y procesa a través de computadoras, servidores, dispositivos móviles, redes y sistemas electrónicos" (InfoSecurity, 2020).

ETL

El término ETL, corresponde a las siglas en inglés de: Extract: extraer, Transform: transformar y Load: cargar. Es un proceso que extrae los datos de diferentes sistemas que sirven como fuente, luego los transforma y finalmente carga esa información en el sistema de gestión de datos, siendo la forma completa de ETL. (Tecnologías Información, 2018).

Bodega de Datos

"Es una base de datos que organiza y almacena una colección de información derivada directamente de los sistemas operativos y de algunos datos externos" (Wiley ,2005).

Datamart

"Es una base de datos departamental, especializada en el

almacenamiento de los datos de un área de negocio específica" (Sinnexus, 2021).

Cubos

La página web de Business Intelligence fácil (2008) afirma que los cubos son una estructura de datos organizada por jerarquías. En donde cada indicador puede ser evaluado en cualquiera de los niveles de las jerarquías.

OLAP

"Es una tecnología que organiza grandes bases de datos empresariales y proporciona análisis complejos. Se puede utilizar para realizar consultas analíticas complejas sin afectar negativamente los sistemas transaccionales" (Microsoft - OLAP, 2021).

OLTP

"Son sistemas que registran interacciones empresariales a medida que se producen en el funcionamiento diario de la organización y admiten consultas de estos datos para realizar inferencias" (Microsoft - OLTP, 2021).

MPP

García J. (2021) señala que es un tipo de arquitectura de procesamiento paralelo que utiliza más de mil microprocesadores de manera individual para así poder resolver distintos cálculos usualmente de altos niveles de complejidad.

NOSQL

"Estas bases de datos almacenan información de forma dinámica, superando las limitaciones de las estructuras SQL tradicionales y sin sufrir ningún tipo de problema en cuanto a escalabilidad. También entendemos que estas bases de datos almacenan la información de una manera alternativa sin recurrir a las típicas estructuras de tablas, sino optando por otros formatos distintos" (García R., 2020).

5.3. Marco Espacial

Esta investigación se ubica en las siguientes instituciones gubernamentales de la República Dominicana: Policía nacional, Sistema Nacional de Atención a Emergencias y Seguridad 9-1-1, Dirección General de Seguridad de Tránsito y Transporte Terrestre (DIGESETT), Dirección General de Migración (DGM) y Ministerio de Obras Públicas y Comunicaciones (MOPC). Estas instituciones están orientadas a registrar los delitos cometidos por ciudadanos dominicanos, en territorio nacional y colocarlos, de forma

centralizada, en una plataforma que sería utilizada por los agentes policiales.

5.4. Marco Temporal

Esta investigación se realizará en el periodo septiembre-diciembre 2021.

6. Aspectos Metodológicos

6.1. Tipo de investigación

Esta investigación tiene un enfoque cualitativo. De acuerdo con Hernández, Fernández y Baptista (2008), el enfoque cualitativo “se fundamenta más en un proceso inductivo (explorar y describir, y luego generar perspectivas teóricas). Van de lo particular a lo general” (p. 8). Esta investigación se enfoca en dar solución a un problema de falta de un sistema de información de inteligencia delictiva entre instituciones gubernamentales.

Investigación descriptiva

La presente investigación es de tipo descriptivo, pues permitirá realizar una descripción de las características fundamentales de un lago de datos, su descripción detallada y su gestión de diseño (Bernal,2006).

Investigación exploratoria

Esta investigación, también es de tipo exploratorio, porque se examinará un tema poco estudiado, como lo es un lago de datos en nube para compartir inteligencia delictiva en la República Dominicana. Se obtendrá información sobre el tema de la investigación para familiarizarnos con él y verlo desde una nueva perspectiva (Hernández, Fernández y Baptista, 2008).

6.2. Métodos de investigación

En esta investigación se aplicará el método analítico y el inductivo. De acuerdo con Bernal (2006), el método analítico permitirá descomponer el objeto de estudio en cada una de las partes, para estudiarlo de manera individual. Como es en este caso el diseño de un lago de datos en la nube para compartir inteligencia delictiva. El método inductivo se utilizará el razonamiento para obtener conclusiones partiendo de hechos particulares.

6.3. Población y Muestra

La población está dada por la cantidad de habitantes de Santo Domingo Este y el Distrito Nacional mayores de 18 años, y se considerará a todo el personal activo que opera las distintas cámaras de videovigilancia del 911 y personal capacitado en el tema que trabaja con la red.

Se tomará una muestra diversa o de máxima variación. Las muestras de

ese tipo se utilizan para presentar diferentes perspectivas e ilustrar la complejidad del fenómeno en estudio, para documentar la diversidad con el fin de localizar diferencias y coincidencias, patrones y particularidades (Hernández, Fernández y Baptista, 2008).

6.4. Técnicas de investigación

Las técnicas de análisis de datos son herramientas útiles para obtener información científica. Después de organizar los datos, es necesario analizarlos cualitativa o cuantitativamente, dependiendo de su naturaleza. Para obtener los resultados de la investigación, es necesario primero analizar la data (Vara Horna, 2012). En tal sentido para este trabajo de investigación, se utilizarán las siguientes técnicas: (a) análisis de documentos, (b) encuesta y (c) internet. Y el siguiente instrumento: cuestionario.

7. Tabla de contenido preliminar

CAPÍTULO 2: MARCO TEÓRICO

Introducción

Los sistemas gestores de bases de datos son aplicaciones destinadas para la gestión y el manejo de los datos que entran y salen de una organización, con la finalidad de preservar la integridad de la información que es almacenada en las bases de datos.

En la actualidad, estos sistemas son parte de la infraestructura de la mayoría de las empresas alrededor del mundo, debido a su gran utilidad en el ámbito tecnológico y en los negocios.

Los lagos de datos por su parte son repositorios de datos con una gran capacidad de almacenamiento, los cuales son utilizados para albergar cualquier tipo de dato sin importar el formato original de este. Han obtenido una gran popularidad entre los administradores de sistemas, gracias a los beneficios a nivel de usabilidad que estos pueden generar.

En esta investigación, se pretende plantear los beneficios que genera utilizar un lago de datos, sus características y cómo poco a poco ha logrado popularizarse entre las nuevas tecnologías para el almacenamiento de volúmenes enormes de información.

2.1 Sistema gestor de base de datos

Los sistemas gestores de base de datos, también conocidos como SGBD, y en inglés como DBMS (Data Base Management System) no son más que un conjunto de datos organizados y estructurados que, entre sí, guardan cierta relación y son accedidos por medio de un conjunto de programas especializados en su gestión. Cabe resaltar que a una colección de datos se le denomina Base de Datos o BD.

Al presente, estos sistemas son parte de la infraestructura de todas las empresas que día tras día, mueven y obtienen grandes cantidades de datos y estos son almacenados en sus servidores para su posterior uso en análisis o consultas específicas.

Antes de aparecer los SGBD (década de los setenta), la información se trataba y se gestionaba utilizando los típicos sistemas de gestión de archivos que iban soportados sobre un sistema operativo. Éstos consistían en un conjunto de programas que definían y trabajaban sus propios datos. Los datos se almacenaban en archivos y los programas manejaban estos archivos para obtener la información. Si la estructura de los datos de los archivos cambiaba, todos los programas que los manejaban se debían modificar; por ejemplo, un programa trabaja con un archivo de datos de alumnos, con una estructura o registro ya definido; si se incorporan elementos o campos a la estructura del archivo, los programas que utilizan ese archivo se tienen que modificar para tratar esos nuevos elementos. [1]

[1] Ramos, M. J., Ramos, A., & Montero, F. (2006). *Sistemas gestores de bases de datos* (Ángel Rodríguez Luengo ed.). Mc Graw Hill.

Figura 1 - 2.1: *Gestores de bases de datos*

Fuente: *Genbeta*

Sí bien por bastante tiempo los datos fueron manejados por medio de estos sistemas, a la hora de tratar enormes volúmenes de información se presentaban varios inconvenientes como la redundancia e inconsistencia, dependencia de datos, dificultad para el acceso de manera concurrente, problemas en la integridad de éstos, entre otros más. Es por esto por lo que surge la idea de separar los datos contenidos en los archivos de los programas, y de tal manera estructurar y organizar los datos con el objetivo de poder acceder a ellos independientemente del programa que los gestione.

Estos inconvenientes son los que impulsan al desarrollo de los SGBD, los cuales, gracias a la evolución y mejoras que las empresas desarrolladoras han ido implementando con el paso del tiempo, son sistemas cuyo objetivo principal es buscar la eficiencia y el mayor nivel de seguridad a la hora de almacenar, extraer o consultar información en un BD.

2.1.1 Importancia

Los SGBD son herramientas clave para la gestión y manejo de los datos dentro de las empresas debido a su énfasis en la eficiencia, integridad y seguridad con la que cuentan al momento de realizar alguna acción que pueda poner en riesgo la información almacenada en una BD.

Por lo tanto, estos sistemas cuentan con ciertos puntos que resaltan la importancia de su implementación en cualquier institución, los cuales son:

- Agrupan y almacenan la información de manera fiable y segura. [2]
- Permiten el acceso y la modificación de forma transaccional, incluso con operaciones masivas. [2]
- Ayudan a compartir la información de la empresa transversalmente. [2]
- Evitan que exista redundancia en la información, trabajando el concepto de dato único. [2]
- Permiten ser el soporte de todos los procesos de la empresa. [2]

Estos puntos son relevantes al momento de tomar en consideración la implementación de un SGBD, en virtud de que es importante para estos la eficacia y la optimización de muchos de los procesos internos de la organización, por tanto, para el negocio la adaptabilidad, flexibilidad y rendimiento que puede aportar un sistema de estos es clave para el éxito de este.

2.1.2 Características

Los SGBD son un producto que para los desarrolladores resulta rentable en la actualidad. Existen ciertas características que todo sistema gestor de base de datos debe contener para ser eficiente en las empresas y resulte idóneo su uso dentro de estas.

Características de un SGBD:

- **Administrar la redundancia**

En términos generales, la redundancia es la existencia de información duplicada en una base de datos, por tanto, el objetivo principal dentro de esta es que dicha redundancia sea mínima y controlada, ya que puede conducir al desarrollo de problemas relacionados con la integridad y la consistencia.

---

[2] Kyocera. (2020). *La importancia de la gestión de base de datos para el CIO*. Kyocera. https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/insights-hub/articles/la-importancia-de-la-gestion-de-base-de-datos-para-el-cio.html

● **Consistencia en la información**

La consistencia es la que verifica el uso correcto de las reglas establecidas en la empresa. Una base de datos no solo es capaz de almacenar datos, sino que también dentro de ella se pueden implementar ciertas restricciones con el único objetivo de presentar al usuario final información de calidad y que esta cumpla con las reglas pautadas por el negocio.

● **Información con integridad**

Una base de datos se compone de campos, por tanto, cada uno de esta espera almacenar información del tipo de dato que se le especificó. Así mismo, la integridad es la capacidad para mantener los datos con una congruencia global dentro de la base de datos.

● **Seguridad**

Esta característica, al tener concentrado en un único punto toda la información, es la que implica mantener seguros nuestros datos y administrar los privilegios de acceso. Por medio de la seguridad, se busca asegurar que solo los usuarios autorizados puedan ver o modificar en una base de datos.

2.1.3 Tipos

Los distintos tipos de SGBD se dividen en seis, tales como:

● **Jerárquica**

Este es el tipo de SGBD más antiguo que existe. En su arquitectura, relacionaba los datos por medio de una estructura con forma de árbol, en donde utilizaba segmentos y arcos para unir los puntos. Hoy en día, se encuentra obsoleta debido a su baja capacidad para reflejar relaciones complejas.

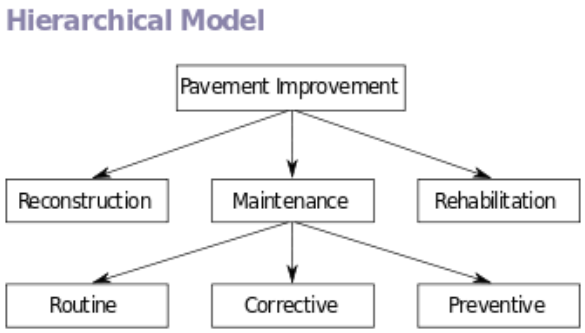

Figura 2 - 2.1.3: *SGBD Jerárquica*

Fuente: *Wikipedia*

- **En red**

En este tipo de SGBD se estructura y relaciona los datos usando nodos y enlaces. Además, es capaz de reflejar cualquier tipo de relación, aunque su manejo es bastante complejo y poco intuitivo.

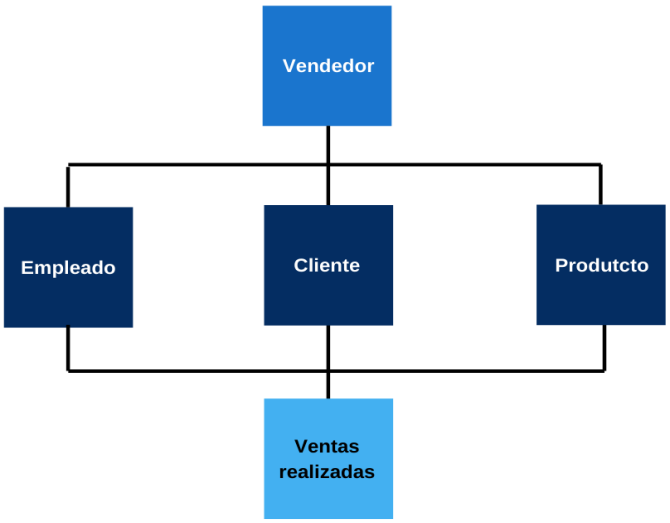

Figura 3 - 2.1.3: *SGBD en red*

Fuente: *AyudaLey*

- **Modelo relacional**

Este es el tipo de SGBD más utilizado en la actualidad, debido a que es muy conceptual e independiente de la física. En adición, permite utilizar tablas como estructura fundamental. El lenguaje utilizado para trabajar con éste es SQL, que es fácil de entender y manipular.

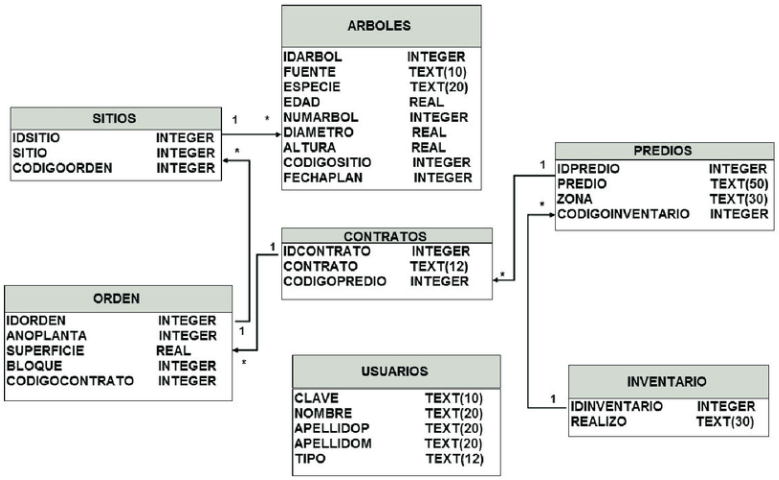

Figura 4 - 2.1.3: *Modelo relacional*

Fuente: *ResearchGate*

- **Orientadas a objetos**

Son basados en la programación orientada a objetos, en donde, como estructura fundamental, usan a los objetos para poder unir datos y operaciones, además de contar con la facilidad de asociarlos con lenguajes modernos como Java y C#.

- **Objetos relacionales**

Son el resultado de unir el modelo relacional con el orientado a objetos, los cuales utilizan características de este último. La mayor parte de las bases de datos son de este tipo, dígase Oracle, SQL Server, entre otros más.

- **NoSQL**

Este tipo de modelo no utiliza el lenguaje SQL para gestionar el manejo de los datos, sino que almacena la información por medio de archivos en formato XML y JSON.

2.1.4 Componentes

Básicamente un SGBD se compone de herramientas que contribuyen al buen funcionamiento del software y permiten que el usuario logre sus objetivos. Estos son:

Diccionario de datos: consiste en una lista de metadatos que reflejan las características de los diversos tipos de datos incluidos en la base de datos. Además, estos metadatos informan sobre los permisos de uso de cada registro y su representación física. [3]

Lenguaje de definición de datos: permite estructurar todo el contenido de una base de datos. Por medio de este, es posible crear, modificar y eliminar cualquier objeto ya sean individuales o relacionales, como referencias o derechos de usuario.

---

[3] Guide, D. (2020, March 16). *Introducción al sistema gestor de base de datos (SGBD)*. Ionos.

https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/sistema-gestor-de-base-de-datos-sgbd/

Lenguaje de manipulación de datos: por medio de este se pueden introducir registros nuevos, así como también eliminar, modificar y consultar los que tenga almacenados.

2.1.5 Funcionalidad

En general, un SGBD es un software cuya funcionalidad es crear, administrar y gestionar la información que se encuentra almacenada en una base de datos, además de ayudar en ciertos procesos que demanden el uso de información de la BD. Aquí algunas funcionalidades que son parte de estos sistemas:

- Proporcionar la posibilidad de que un usuario pueda ingresar datos, acceder a ellos y poder generar informes que le sean de uso.
- Contener un diccionario de datos, el cual permite que los datos sean almacenados según su tipo o estructura.
- Preservar la seguridad de los datos.
- Proteger la integridad de los datos.
- Ahorrar detalles al usuario, con la abstracción de información, acerca del almacenamiento físico de los datos, por tanto, da igual si una base de datos está conformada por uno o cientos de archivos, este detalle se hace transparente al usuario.

2.2 Lago de datos

Es un repositorio de datos de gran magnitud y fácil acceso que permite almacenar distintos conjuntos de datos en su formato original, además de poder visualizarlos a cualquier escala, ya sean estructurados o no.

Este tipo de repositorio permite conservar todo tipo de datos, sin eliminarlos ni filtrarlos antes de su almacenamiento. Gracias a esto, pueden ser analizados con brevedad y utilizados con algún propósito en específico.

2.2.1 Características

Un lago de datos representa una ubicación céntrica en la que podemos almacenar todos los datos generados, sin importar su fuente o el formato en el

que fueron creados. Por consiguiente, para considerar un repositorio de datos como un lago de datos, debe comprender las siguientes características:

- Comprender la lectura y escritura simultánea de datos.
- Acceder de manera directa a los datos de origen.
- Separar los recursos de almacenamiento y computación.
- Contener formatos de almacenamiento estandarizados.
- Contar con compatibilidad con los tipos de datos estructurados, no estructurados y semiestructurados.
- Transmitir de extremo a extremo.

2.2.2 Arquitectura

Un lago de datos puede presentar variaciones en su arquitectura física, ya que más que un servicio o un espacio físico en donde se almacena información es una estrategia que puede ser aplicada a múltiples tecnologías. Sin embargo, esta arquitectura cuenta con varios principios que hacen diferir a un lago de datos de otros métodos de almacenamiento de big data:

- Ningún dato es rechazado, estos se cargan desde distintos sistemas de origen y son conservados.
- Los datos son almacenados en distintos estados, ya sea sin transformar o casi transformados, tal cual como se reciben desde la fuente de origen.
- Estos datos son posteriormente transformados y llevados a un esquema basado en las necesidades del análisis.

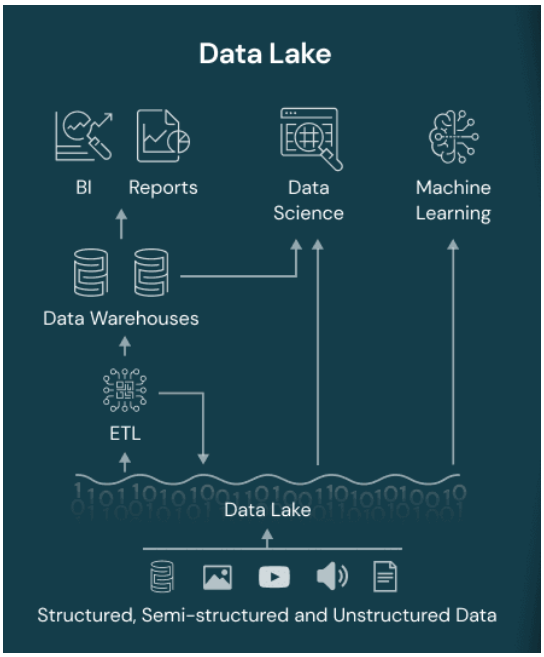

Figura 5 - 2.2.2: *Arquitectura de un lago de datos*

Fuente: *Databricks*

2.2.3 Componentes

Básicamente, un lago de datos se compone a nivel de usabilidad por cuatro herramientas que hacen posible lograr los objetivos: ingestión de datos, almacenamiento de datos, procesamiento de datos y acceso a los datos.

La ingestión de datos: aquí es en la donde los datos deben pasar por distintos chequeos antes de ser almacenados en el componente de almacenamiento, tales como: controles básicos de calidad, procesos de encriptación y registros de metadata. [4]

Almacenamiento de datos: en este, todos los datos, estructurados o no estructurados, son almacenados sin sufrir adaptaciones. Es un componente en el que se precisa de analistas expertos en data discovery mediante herramientas big data. [4]

Procesamiento de datos: una vez que los analistas de datos han realizado data discovery en el raw data, se puede ver la necesidad de procesar y adaptar determinados conjuntos de datos para alojarlos en una capa de uso recurrente. En este componente pueden tener lugar procesos avanzados de data

quality, integridad y otras adaptaciones para disponer de una capa de confianza de exploración de datos a la que tengan acceso otros usuarios. [4]

Acceso a los datos: este es un componente más avanzado donde, finalmente, los datos se ponen a disposición de los analistas de negocio. Estos analistas podrán generar informes y análisis para responder a preguntas de negocio y afianzar la toma de decisiones. [4]

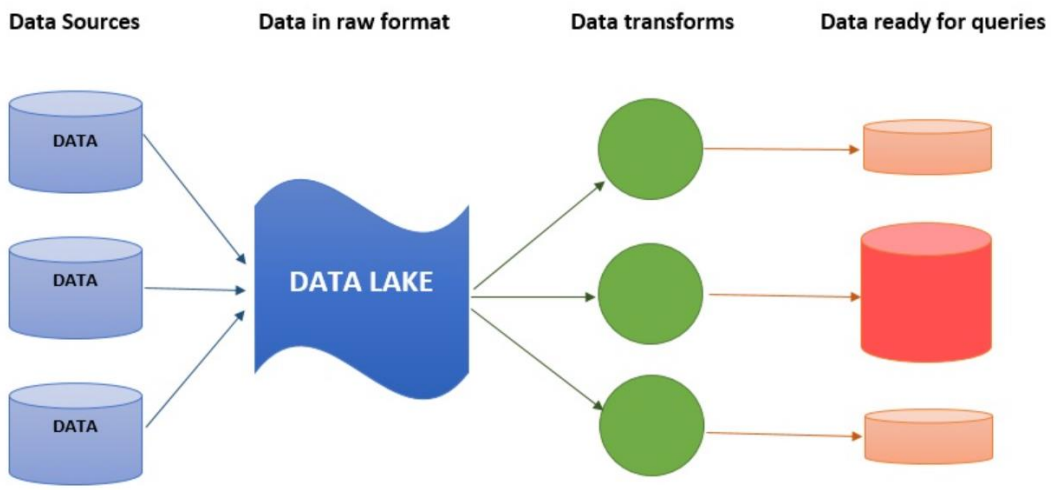

Figura 6 - 2.2.3: *Componentes de un lago de datos*

Fuente: *Data Analytics*

2.2.4 Ventajas y desventajas

Ventajas

1. Un lago de datos permite a los usuarios almacenar cantidades enormes de datos en su formato nativo, sin realizar ningún tipo de organización. Esto permite una mayor flexibilidad para analizar cosas como POS y Big Data, en donde los usuarios pueden acceder a toda la información mucho más fácil y en tiempo real.

[4] Dertiano, V., & Ríos, R. (2017, October 23). *Data Lake - Introducción | Características y Capas básicas*. BI Geek Blog. https://blog.bi-geek.com/arquitectura-bi-introduccion-al-data-lake/

2. Mantener la información en su formato original representa una gran ventaja, ya que el equipo que la utilizará no necesita especificar para qué lo usará, y así poder empezar a cargar la información en cuanto el lago esté listo. Además, poder subir información desde cualquier sistema de origen.

3. Otra ventaja más es la gran cantidad de usuarios que admite un lago de datos. Combinando esto con herramientas adecuadas, el cliente podrá responder a más preguntas y analizar más información.

Desventajas

1. En un lago de datos se puede almacenar información en su formato nativo, por ende, puede darse el caso de que se almacene información en formatos no estándar, y estos necesitan ser reformateados de manera manual.

2. Podrá necesitar en ciertos casos de herramientas expertas en el análisis científico de datos para poder sacar el máximo provecho del lago.

3. Un lago de datos simplemente recopila datos nativos en un repositorio. La manera en la que salen de ese lugar va a depender de la capacidad que se tenga para poder organizarlo y analizarlo.

2.2.5 Comparación entre Almacén de datos y Lago de datos

Los lagos de datos y los almacenes de datos son ampliamente utilizados para almacenar big data, pero no son términos tan similares como parece. Un lago de datos es un vasto conjunto de datos sin procesar, cuyo propósito aún no está definido, mientras que un almacén de datos es un repositorio de datos estructurados y filtrados que ya se han procesado para un propósito específico.

Estructura de datos

En el caso de los lagos de datos, los datos que tienen almacenados están sin procesar (en inglés se le conoce a esto como raw), mientras que en un almacén de datos ya están procesados y con un propósito definido. Debido a esto, por lo general un lago de datos requiere de mayor capacidad de almacenamiento frente a un almacén de datos.

Aunque los datos de un lago sean maleables y permitan analizar rápidamente cualquier propósito, existe el riesgo de que esos datos en bruto transformen a nuestro lago en un pantano de datos, o sea, sin medidas de calidad ni gobernanza.

Los almacenes de datos, al almacenar sólo los datos procesados, ahorran espacio de almacenamiento costoso al no mantener datos que nunca se pueden usar.

Propósito

En general, un dato almacenado en un lago no cuenta con un propósito específico. Muchas veces, estos son almacenados con algún plan en mente, pero otras veces simplemente para tenerlos a mano. Por consiguiente, esto significa que un lago de datos cuenta con menor organización y filtración de datos que un almacén de datos.

Como un almacén solo alberga datos procesados, o sea, que estos han sido utilizados al menos una vez con algún propósito específico dentro de una organización, esto significa que el espacio de almacenamiento suele ser más organizado y no se desperdicia en datos que no se vayan a utilizar.

Usuarios

Un lago de datos suele ser difícil de navegar para quienes no están familiarizados con los datos sin procesar, por lo tanto, estos requieren de un científico o analista de datos experto y herramientas especializadas que permitan la comprensión de los datos para su uso específico.

En el caso de los almacenes de datos, los usuarios pueden utilizarlos directamente desde el mismo almacén, ya sea para visualizar los datos en gráficos, tablas u hojas de cálculo para su debido análisis.

Accesibilidad

En el caso de la accesibilidad, el lago de datos muestra mayor facilidad cuando se necesita acceder y cambiar algo en su interior. Además, cualquier cambio que se realice a los datos se puede hacer rápidamente, ya que estos cuentan con muy pocas limitaciones.

Por otro lado, un almacén de datos, al tener un diseño en su arquitectura más estructurado, hace que el procesamiento y la estructura de los datos permita que los datos en sí sean más fáciles de descifrar, pero las limitaciones a nivel estructural hacen que estos sean más difíciles y costosos de manipular.

2.2.6 Ejemplificaciones

En la actualidad, existen muchos proveedores de servicios cloud que cuentan en su estructura con los lagos de datos. Ejemplo de esto es Amazon Web Services, que hoy en día es uno de los proveedores más grandes y completos de estos servicios; enfatizados en el procesamiento ágil, flexible y seguro de sus datos dentro de estos sistemas.

Por otro lado, se encuentra Google con una oferta bastante interesante sobre lagos de datos, enfocados en el tema de las analíticas y la IA. También, a esta ola se suman otros más como IBM y Microsoft, en donde cada quien muestra su oferta hacia los clientes que se encuentran en busca del que mejor se ajuste a su necesidad.

Figura 7 - 2.2.6: *Ejemplificaciones de lago de datos*

Fuente: *Dataprius*

Resumen

Los sistemas gestores de bases de datos (SGBD) están destinados al control y gestión de los datos que son generados día tras día en las distintas empresas alrededor del mundo, con la finalidad de poder realizar análisis y reportes que generen beneficios al negocio.

Los SGBD se caracterizan por tener una buena administración de la redundancia, ser consistentes con la información, preservar la integridad y la seguridad de los datos almacenados. Estos a su vez, constan de varios tipos como: Jerárquica, en red, modelo relacional, orientada a objeto, objetos relacionales y NoSQL.

Se componen por diccionario de datos, lenguaje de definición de datos y por el lenguaje de manipulación de datos.

Un lago de datos es un repositorio de datos de una amplia magnitud, que permite además poder visualizarlos a cualquier escala, ya sean estructurados o no, y facilita el acceso para los usuarios que lo utilicen. Estos se caracterizan por realizar una lectura y escritura simultánea, acceder a los datos de origen directamente, contener formatos de almacenamiento estándar, soporta datos estructurados, no estructurados y semiestructurados.

La arquitectura de los lagos de datos se basa en la estrategia que el cliente utilice para poder sacar el máximo provecho del servicio, aunque cuenta con varios puntos: ningún dato es rechazado, estos se almacenan en distintos estados y posteriormente son transformados para poder ser utilizados en algún plan específico.

Los lagos de datos están compuestos por cuatro herramientas para su uso, la ingestión, almacenamiento, procesamiento y acceso de datos.

CAPÍTULO 3: DIAGNÓSTICO DE LOS SISTEMAS EXISTENTES

Introducción

Hoy por hoy, los almacenes y lagos de datos han demostrado ser en muchos casos las soluciones idóneas cuando una empresa se plantea la idea de utilizar un sistema o establecer una estrategia que le permita manejar la información que entra y sale de sus servidores. Por esa razón, es importante conocer la evolución de estas tecnologías y el impacto que han tenido en las empresas.

Para la época de los 80, se empezó a popularizar el término de almacén de datos, el cual en esos tiempos fue utilizado como un simple almacén de información con una gran capacidad para el manejo del almacenamiento, pero con el paso del tiempo esta tecnología se fue refinando y evolucionando hasta convertirse en toda una infraestructura dedicada a la analítica de datos.

Debido a que los almacenes pueden manejar bastantes tipos de datos, se les hace complicado el trabajo cuando nos referimos a datos no estructurados y semiestructurados, y aquí es donde a mediados del 2010 conocemos el concepto de lago de datos, el cual permite albergar cualquier tipo de dato sin importar el sistema de origen. Gracias a esto, muchas empresas pueden cargar cualquier tipo de dato para su posterior uso en cualquier análisis o reporte.

En este capítulo se pretende mostrar la evolución que han tenido estas tecnologías con el paso del tiempo desde sus inicios hasta la actualidad. En adición, también se pretende comprender la necesidad de su utilización en casi cualquier entorno tecnológico de muchas empresas alrededor del mundo.

3.1 Antecedentes

Para finales de la época de los 80, se empieza a hablar del concepto de un almacén de datos, los cuales tenían como objetivo principal pasar los datos de los sistemas operativos hacia los sistemas de apoyo de decisiones (DSS), con la finalidad de que éstos sean analizados y utilizados para algún propósito en específico. En esos tiempos, las empresas solían utilizar varios entornos DSS para los distintos usuarios que tenían, provocando así que los procesos de recopilación, limpieza e integración se tuvieran que realizar de manera individual para cada entorno existente, siendo esto una tarea en muchos casos tediosa para los administradores.

A raíz de esto, los almacenes de datos se volvieron más eficientes; pasaron de ser almacenes de información, utilizados para plataformas de BI, a transformarse en infraestructuras analíticas con una amplia variedad de aplicaciones para el manejo y gestión de la información.

Para 1991, con la publicación de “The IRM Imperative” por parte de James M. Kerr, se planteó la idea de administrar y dar un valor en dólares a los recursos de datos de una organización y posteriormente informar ese valor como un activo en un balance general.

Kerr también describió una manera de rellenar las bases de datos relacionadas a las áreas temáticas a partir de datos derivados de sistemas basados en transacciones para crear un área de almacenamiento donde los datos resumidos podrían aprovecharse aún más para informar la toma de decisiones ejecutivas. Este concepto sirvió para promover un mayor pensamiento sobre cómo se podría desarrollar y administrar un almacén de datos de una manera práctica dentro de cualquier empresa.

Si bien los almacenes de datos sentaron las bases para el manejo de sistemas de soporte de decisiones y la inteligencia empresarial, los tiempos han cambiado y hay ocasiones en que las empresas deben manejar datos no estructurados, semiestructurados y con alta variedad, velocidad y volumen. Por tanto, en situaciones como estas, los almacenes de datos no son los más adecuados ni los más rentables.

A partir de esta situación, los arquitectos de datos comenzaron a imaginar un único sistema en donde se pudiera albergar datos para muchos productos analíticos y cargas de trabajo. Es por eso que, a mediados del 2010, muchas empresas alrededor del mundo empezaron a construir los lagos de datos: repositorios destinados para datos sin procesar en una variedad amplia de formatos. Si bien este sistema cuenta con ciertas desventajas, la verdad es que ha podido salir a flote y ser la solución para muchas empresas que contaban con situaciones tediosas al momento de manejar sus datos.

3.2 Situación actual

El futuro para los sistemas de repositorio de datos nunca había sido tan claro. Gracias a los distintos avances tecnológicos que año tras año van surgiendo, cada vez se hace más énfasis en la necesidad de tener un almacén o un lago de datos en las empresas, debido a la masiva cantidad de información que éstas mueven día a día.

En la actualidad, muchas organizaciones han sacado el máximo provecho de las ventajas que este tipo de estructuras pueden generar a los clientes. Empresas como Google, Amazon y Microsoft han optado por incluir a los lagos de datos en sus planes de tecnología cloud como servicio, tal y como han hecho con los almacenes de datos, no en búsqueda de sustituir a este último, sino que el cliente pueda optar por la tecnología que más se ajuste a su necesidad, y además poder ofrecer lo que más está en demanda a día de hoy.

Por tanto, gracias a los lagos de datos muchas empresas han podido sobrellevar la carga de información tan masiva de una manera un poco más ágil, como es el caso de empresas que cargan datos directamente desde la web como Facebook o Google. Esto ha permitido que surjan nuevos tipos de valores empresariales como estos:

1. Utilizar lagos de datos para almacenar datos web ha logrado incrementar de manera significativa la velocidad y la calidad de las búsquedas.[5]

2. Permiten almacenar datos de navegación para implementar métodos más eficaces de publicidad web.[5]

3. Los lagos de datos, en aplicaciones de análisis de interacciones y comportamientos de clientes, han permitido obtener una imagen más completa del perfil que puede tener cada cliente. [5]

4. Los lagos de datos en los comercios pueden obtener información rentable a partir de datos sin procesar de múltiples fuentes, como archivos de registro, consumo de audio y video online, entre otros más, para identificar el comportamiento de los consumidores en tiempo real y convertir acciones en ventas. [5]

5. Las empresas pueden utilizar lagos de datos para centralizar datos dispares de diferentes fuentes y ejecutar algoritmos de análisis y aprendizaje automático para ser los primeros en identificar oportunidades comerciales. [5]

De esta forma, los lagos de datos, junto a los almacenes, han tomado gran popularidad por las diversas ventajas a nivel empresarial que representan. Ya es un hecho que el futuro de estas tecnologías no puede ser más comprometedor, debido a que cada vez más nuevas herramientas de BI salen al mercado para mostrar las innovaciones que traen.

---

5 Tiwari, S. (2021, June 9). *¿Qué es un lago de datos? Ventajas y contras* | *OpenMind*. BBVA Openmind. https://www.bbvaopenmind.com/tecnologia/mundo-digital/un-lago-de-datos-una-oportunidad-o-un-sueno-para-el-big-data/

Resumen

Para finales de los 80, el término de almacén de datos se empezó a popularizar, puesto que para esa época muchas empresas deben utilizar varios entornos DSS para los distintos usuarios que tenían, provocando así que los procesos de recopilación, limpieza e integración se tuvieran que realizar de manera individual para cada entorno, convirtiendo estos procesos en tareas muy tediosas para los administradores.

Para 1991, con la publicación de “The IRM Imperative” por parte de James M. Kerr, se planteó la idea de administrar y dar un valor en dólares a los recursos de datos de una organización y posteriormente informar ese valor como un activo en un balance general.

Los almacenes de datos sentaron las bases para el manejo de sistemas de soporte de decisiones y la inteligencia empresarial, pero para el manejo de datos de distintas fuentes de origen y de un masivo volumen, un almacén no resulta como la solución más idónea a tomar. Es por eso por lo que, a mediados del 2010, muchas empresas alrededor del mundo empezaron a construir los lagos de datos: repositorios destinados para datos sin procesar en una variedad sumamente amplia de formatos.

Mirando hacia el futuro, para los sistemas de repositorio de datos, cada vez se hace más énfasis en la necesidad de tener un almacén o un lago de datos en las empresas debido a la masiva cantidad de información que estas mueven día a día. Es por esto por lo que muchas organizaciones como Google, Amazon y Microsoft han optado por incluir a los lagos de datos en sus planes de tecnología cloud como servicio, tal y como han hecho con los almacenes de datos, no en búsqueda de sustituir a este último, sino que el cliente pueda optar por la tecnología que más se ajuste a su necesidad, y además poder ofrecer lo que más está en demanda a día de hoy.

CAPÍTULO 4: TECNOLOGÍAS PARA EL MANEJO DE DATOS EN INSTITUCIONES GUBERNAMENTALES DE REPÚBLICA DOMINICANA

Introducción

Las instituciones gubernamentales son la base fundamental de cada país alrededor del mundo, ya que son las encargadas de brindar distintos tipos de servicios que, de una manera u otra, puedan solventar alguna situación o problemática, y además están estrechamente relacionadas con algún área ya sea económica, política, ambiental, entre otras más.

La función de éstas va relacionada con el área que manejan, o sea, pueden existir instituciones dedicadas a la sanidad, salud, economía, política, medio ambiente, educación, cultura, y así sucesivamente. Por tanto, cada una de estas debe elaborar algún plan estratégico o de acción que les permita realizar un trabajo que haga sentir satisfacción entre la población de un país o nación.

En este capítulo, se pretende mostrar la importancia, características, funciones y clasificaciones de estas instituciones, además de realizar un análisis sobre los distintos sistemas o al menos las herramientas tecnológicas que algunas de éstas utilizan para realizar sus trabajos en el día a día.

4.1 Instituciones gubernamentales

Las instituciones gubernamentales son las organizaciones fundamentales de un estado encargadas de administrar, regular y organizar la convivencia social de una sociedad, o sea, brindar un servicio que resulte idóneo para la ciudadanía, por medio de normas y leyes que los individuos que forman parte de ésta ejecuten, a través de obligaciones y derechos.

4.1.1 Importancia

En general, una institución gubernamental tiene el poder de generar un impacto en diversos entornos, como el empresarial y en las actividades que realiza un ciudadano. Debido a esto, son quienes mantienen el orden y permiten ciertas acciones por medio de la implementación de leyes.

Toda sociedad debe contar con un tipo de sistema político para que pueda asignar recursos y llevar a cabo proyectos de manera adecuada. Una institución política establece las reglas con las que una sociedad ordenada obedece y, en última instancia, revisa y administra las leyes para quienes buscan desobedecerlas. [6]

Figura 8 - 4.1.1: *Palacio Nacional de la República Dominicana*

Fuente: *Periódico El Caribe*

[6] Universidad Panamericana. (2021, February 26). *¿Qué son las instituciones políticas y cuál es su función?* Blog Universidad Panamericana. https://blog.up.edu.mx/topic/posgrados-de-gobierno-y-economia/que-son-las-instituciones-politicas-y-cual-es-su-funcion}

4.1.2 Características

Las instituciones gubernamentales, como organismos encargados de la administración y regulación de leyes y servicios, cuentan con las siguientes características:

1. Todas las instituciones están conformadas por un grupo de personas, o sea dos o más personas por su propia voluntad o por una ley en específico.

2. Este grupo de personas trabajan en conjunto para lograr determinados propósitos, ya sean organizacionales, grupales o individuales.

3. Se desarrollan tareas en conjunto estrechamente relacionadas unas con las otras, ya sea por su similitud o afinidad.

4. Al igual que una empresa, estas instituciones están divididas por niveles jerárquicos, así como también por el grado de responsabilidad acorde a cada uno de los cargos existentes.

5. Estas llevan un orden normativo, o sea, el ordenamiento interno que se requiere para el funcionamiento de una organización, los procedimientos y las normas que permiten el desarrollo de las actividades.

6. Aceptación cultural: las actividades que se desarrollan dentro de estas instituciones deben ser aceptadas culturalmente, ya que deben satisfacer alguna necesidad social que no vaya en contra de los principios morales ni religiosos del ambiente en el que actúe, y además se adecúe a las distintas costumbres.

4.1.3 Función

En general, las instituciones gubernamentales tienen diversas funciones en torno a las cuestiones a las que están orientadas. Estas funciones pueden ser en el ámbito económico, sanitario, administrativo, entre otros. Por tanto, para cada asunto que sea importante de atender, debería existir una institución dentro del estado.

Hay instituciones dedicadas a la preservación de la memoria histórica del país, otras para cuidar el medio ambiente, para mantener organizado el sistema educativo y de trabajo, y así sucesivamente para cada temática importante. Cada uno de éstas se dan la tarea de elaborar planes de acción, con la finalidad de realizar un trabajo que haga sentir satisfacción al estado o nación.

4.1.4 Clasificaciones

Como se ha mencionado anteriormente, para cada asunto que represente importancia, debería existir una institución capaz de dar solución a este. Por tanto, aquí algunos ejemplos de cómo podrían estar clasificadas las instituciones gubernamentales según el país:

1. Instituciones políticas: son aquellas instituciones dedicadas a la creación, aplicación y cumplimiento de las leyes de un país.

2. Instituciones económicas: están dedicadas a la regulación de leyes con el objetivo de mejorar aspectos concernientes a la economía nacional.

3. Instituciones jurídicas: son las que están dedicadas al plano judicial, dígase que buscan de cualquier manera que se cumpla un ideal de justicia.

4. Instituciones educativas: encargadas de llevar adelante el sistema educativo del país.

5. Instituciones ambientales: responsables de estudiar, monitorear y proteger el medio ambiente del maltrato, mal uso o degradación por parte de entidades físicas o jurídicas nacionales o internacionales.

4.2 Policía Nacional (PN)

La Policía Nacional es un cuerpo arma-do, técnico, profesional, de naturaleza policial, bajo la autoridad del Presidente de la República, obediente al poder civil, apartidista y sin facultad, en ningún caso, para deliberar. La Policía Nacional tiene por misión salvaguardar la seguridad ciudadana, prevenir y controlar los delitos, mantener el orden público para proteger el libre ejercicio de los derechos de las personas, entre otras cuestiones más.[7]

[7] Policía Nacional. (2019). *Quiénes Somos – Policía Nacional Dominicana*. Policía Nacional. https://www.policianacional.gob.do/sobre-nosotros/quienes-somos/

Según la Dirección de Comunicaciones Estratégicas, el pasado 22 de julio de 2021, se pusieron en marcha alrededor de ocho sistemas de información enfocados en fortalecer las áreas de: inteligencia, prevención, investigación y comunicación; El Sistema de Mando y Control de Acciones Preventivas basado en Tablas de Acciones Mínimas, Sistema Policial de Consultas Criminológicas. Por otro lado, implementaron mejoras en la logística relacionada con el Sistema de gestión de equipos de comunicaciones y telefónicas.

Figura 9 - 4.2: *Policía Nacional*
Fuente: *Policía Nacional*

4.3 Sistema Nacional de Atención a Emergencias y Seguridad 9-1-1

El Sistema Nacional de Atención a Emergencias y Seguridad 9-1-1 es la institución encargada de responder a los llamados de emergencia de la población, ya sea del área de la salud o una situación en donde deba intervenir la Policía Nacional o el Cuerpo de Bomberos.

Hoy en día, esta institución cuenta con ciertas herramientas tecnológicas como videovigilancia, drones y radiocomunicación, que permiten llevar a cabo con más facilidad los procesos que realizan para ayudar en situaciones de emergencia.

A estas herramientas se le suman distintos sistemas como es el caso del Monitoreo y supervisión de campo el cual permite dar seguimiento a las unidades de respuesta, conformadas en muchos casos por el personal militar.

Otro más es el sistema de Pronósticos y planificación del personal, el cual genera un estimado de la demanda del servicio, ya sea semanal, diario y en intervalos de tiempo de unos 15 minutos.

Figura 10 - 4.3: *Sistema Nacional de Atención a Emergencia y Seguridad 9-1-1*

Fuente: *Membreño Asesores*

4.4 Dirección General de Seguridad de Tránsito y Transporte Terrestre (DIGESETT)

La Dirección General de Seguridad de Tránsito y Transporte Terrestre (DIGESETT) es la institución encargada de viabilizar, fiscalizar, supervisar, ejercer el control y vigilancia en las vías públicas, y velar por el cumplimiento de estas leyes.

Actualmente, la DIGESETT cuenta con diversas herramientas apoyadas en la tecnología para eficientizar el trabajo que esta institución realiza, como es el caso de la utilización de drones con el objetivo de visualizar desde las alturas las distintas razones por las cuales se generan los embotellamientos vehiculares en las distintas avenidas y autopistas de la República Dominicana.

De igual forma, ésta cuenta con el sistema de multas o de recopilación de datos de incidentes y eventualidades en tiempo real con la finalidad de que los usuarios que tengan pendientes algunas que otra infracción, o hayan sido víctimas de alguna situación en su vehículo puedan visualizar su documentación, y además, en caso de tener alguna infracción, poder hacer el pago correspondiente para evitar alguna sanción por vencer el plazo del tiempo estipulado.

Figura 11 - 4.4: *Dirección General de Seguridad Tránsito y Transporte Terrestre*
Fuente: *25 horas de noticias*

4.5 Dirección General de Migración (DGM)

La Dirección General de Migración (DGM) es la institución encargada de salvaguardar la jurisdicción del territorio dominicano, por medio del control migratorio, en cumplimiento y aplicación de la ley 285-04, relacionada con la entrada y salida de ciudadanos dominicanos y extranjeros.

La DGM en el 2012, según el Portal de Data Center Dynamics, inauguró un centro de datos por medio de los servicios que ofrece el NAP del Caribe. Este centro tiene la finalidad de utilizarse como un nodo de conectividad. En otras palabras, todos los sistemas que funcionan para las operaciones de la DGM se alojan en el NAP, y por medio de los servicios de Altice proporciona la conectividad necesaria para las distintas oficinas o puntos de control distribuidos en la República Dominicana.

En estos tiempos, esta institución cuenta con ciertas aplicaciones y sistemas, aparte del data center, que permiten el cumplimiento efectivo de las operaciones que se realizan dentro de esta, como es el caso del sistema de control migratorio, que permite visualizar y analizar toda la información relacionada con los ciudadanos dominicanos en otros países y los extranjeros que actualmente están en el país.

Figura 12 - 4.5: *Dirección General de Migración*

Fuente: *Migración PNV*

4.6 Ministerio de Obras Públicas y Comunicaciones (MOPC)

El Ministerio de Obras Públicas y Comunicaciones (MOPC) es la institución encargada de fiscalizar y supervisar las obras de índole pública que están en ejecución, con la finalidad de garantizar que las normas y reglamentos establecidos se cumplan, y las especificaciones establecidas para la obra en cuestión sean implementadas.

En la actualidad, esta institución cuenta con un centro de datos especializado en recolectar información que está íntimamente relacionada con alguna obra o espacio público que se haya construido recientemente (o las que ya tengan un tiempo de haberse construido). Además, sirve para cargar información sobre estructuras como puentes, elevados, alcantarillas, entre otros más, que necesitan mantenimiento con urgencia para su continuo uso.

Figura 13 - 4.6: *Ministerio de Obras Públicas y Comunicaciones*

Fuente: *MOPC*

Resumen

Las instituciones gubernamentales son la base fundamental de un estado. Están encargadas de administrar, regular y organizar la convivencia social de una sociedad.

La importancia de estas instituciones es que tienen el poder de generar un impacto en diversos entornos, como el empresarial y en las actividades que realiza un ciudadano, debido a que estas son quienes mantienen el orden y permiten ciertas acciones por medio de la implementación de leyes.

Las funciones de las instituciones gubernamentales están orientadas por de las distintas áreas a las que pueden estar dirigidas, ya sean económicas, sanitarias, administrativas, entre otras más; así mismo con su clasificación, dígase que pueden existir instituciones económicas, políticas, ambientales, y demás.

La Policía Nacional es un cuerpo arma-do, técnico, profesional, de naturaleza policial, bajo la autoridad del Presidente de la República, obediente al poder civil, apartidista y sin facultad, en ningún caso, para deliberar. Dentro de esta, se pusieron en marcha alrededor de ocho sistemas de información enfocados en fortalecer las áreas de: inteligencia, prevención, investigación y comunicación.

El Sistema Nacional de Atención a Emergencias y Seguridad 9-1-1 es la institución encargada de responder a los llamados de emergencia de la población, ya sea del área de la salud o una situación en donde deba intervenir la Policía Nacional o el Cuerpo de Bomberos. Ésta cuenta con ciertas herramientas tecnológicas como videovigilancia, drones y radiocomunicación. A eso se le suman distintos sistemas como es el caso del Monitoreo y supervisión de campo y el Sistema de Pronósticos y Planificación del Personal.

La Dirección General de Seguridad de Tránsito y Transporte Terrestre (DIGESETT) es la institución encargada de viabilizar, fiscalizar, supervisar, ejercer el control y vigilancia en las vías públicas, y velar por el cumplimiento de estas leyes. Ésta cuenta con diversas herramientas, como es el caso de la utilización de drones y el sistema de multas o de recopilación de datos de incidentes y eventualidades en tiempo real, con la finalidad de que los usuarios

que tengan pendientes algunas que otra infracción, o hayan sido víctimas de algún incidente en su vehículo.

La Dirección General de Migración (DGM) es la institución encargada de salvaguardar la jurisdicción del territorio dominicano, por medio del control migratorio, en cumplimiento y aplicación de la ley 285-04, relacionada con la entrada y salida de ciudadanos dominicanos y extranjeros. Ésta cuenta con un centro de datos por medio de los servicios del NAP del Caribe, con la misión de utilizar éste como un nodo de conectividad; Todos los sistemas que funcionan para las operaciones de la DGM se alojan en el NAP.

El Ministerio de Obras Públicas y Comunicaciones (MOPC) es la institución encargada de fiscalizar y supervisar las obras de índole pública que están en ejecución, con el objetivo de garantizar que las normas y reglamentos establecidos se cumplan, y las especificaciones establecidas para la obra en cuestión sean implementadas. Dicha institución cuenta con un centro de datos especializado en recolectar información que está íntimamente relacionada con alguna obra o espacio público que se haya construido.

CAPÍTULO 5: TECNOLOGÍAS PARA EL DISEÑO DE LAGO DE DATOS

Introducción

Para muchas empresas, los lagos de datos se han convertido en su punto céntrico para el almacenamiento de información, sin importar la proveniencia de los datos. Por ende, estos representan el núcleo de todos los datos de una organización.

Actualmente, en República Dominicana, las instituciones gubernamentales necesitan de un repositorio que permita concentrar la información delictiva de los ciudadanos, con la finalidad de agilizar el proceso de resolución de casos criminales, ya que en comparación con hace varios años la tasa de criminalidad ha aumentado considerablemente siendo este uno de los motivos por los que se necesita de un sistema capaz de almacenar grandes volúmenes de información y poder procesar la necesaria para realizar análisis y dar con la solución de cualquier situación que se presente.

En este capítulo, se pretende mostrar de una manera más detallada la necesidad que existe de utilizar un lago de datos para las instituciones gubernamentales, además de sus beneficios, infraestructura, seguridad, proveedores y las problemáticas que pueden surgir al tratar con esta.

5.1 Necesidad

Nuestras instituciones gubernamentales hoy en día carecen de un repositorio de datos que pueda concentrar toda la inteligencia delictiva de los ciudadanos dominicanos almacenada en los distintos sistemas de estas organizaciones. Por ende, que una persona después de haber cometido algún crimen se dé a la fuga hacia otra provincia y que un oficial de tránsito lo detenga y no reconozca que es alguien implicado en un acto criminal, es el pan de cada día.

Un lago de datos trae consigo un conjunto de beneficios que podrían ayudar a mitigar las distintas problemáticas o consecuencias que podría traer la falta de uno de estos sistemas.

Utilizar esta tecnología permitirá concentrar todos los datos que sean cargados sin importar su formato ni el sistema de origen de éstos. También, ayudará en la realización de análisis con el objetivo de dar con la solución de algún problema en cuestión.

5.2 Beneficios

Al permitir que los datos permanezcan en su formato nativo, hay disponible un flujo de datos mucho mayor y más oportuno para el análisis. Además de poder generar reportes, cargar datos en diversos sistemas y poder consultar dichos datos en cualquier momento. Por consiguiente, esto produce mayor flexibilidad al momento de utilizar los datos almacenados. [8]

Otro de los beneficios que tiene la utilización de un lago de datos sería la masiva cantidad de almacenamiento que puede soportar, además del gran número de usuario que pueden interactuar con éste en tiempo real.

---

[8] Tecnologías Información. (2018). *Data Lake: Definición, Beneficios y Características*. tecnologias-informacion.com.

https://www.tecnologias-informacion.com/data-lake.html

Básicamente un lago de datos aporta:

1. Fácil adaptación a los cambios.
2. No rechaza ningún tipo de dato.
3. Al soportar todo tipo de datos, puede realizar diversos tipos de análisis.
4. Facilidad para agregar nuevos datos.
5. Puede dar soporte a distintos usuarios en la empresa.

5.3 Infraestructura

La infraestructura de los lagos de datos proporciona a los distintos usuarios y desarrolladores acceso a lo que tradicionalmente se conocía como información dispar o en silos. Con los desarrollos de la computación a lo largo del tiempo, las empresas y los equipos especializados en datos pueden medir el costo de la carga de trabajo individual para poder determinar si un proyecto debe ser ampliado o no.

El modelo de infraestructura de un lago de datos proporciona capacidades casi ilimitadas para los ciclos de vida analíticos de una empresa. Estas capacidades se componen en una sola infraestructura de datos, en donde se encuentran la ingesta, el almacenamiento, los procesos y el consumo. A partir de estos, se generan procesos u operaciones como análisis, soporte a clientes, desarrollo, productos, entre otros varios.

La ingesta comprende los datos que llegan en cualquier formato sin procesar y son almacenados para futuros análisis o recuperación ante cualquier desastre.

El almacenamiento es el que permite a las empresas administrar y organizar cantidades casi infinitas de información. En el caso de las organizaciones que ofrecen servicios en la nube ofrecen acceso de alta disponibilidad de big data, a un costo muy bajo.

Los procesos son, gracias a los avances en la computación, cada vez más ágiles y eficientes. Las APIs son un ejemplo de esto, donde ya no es necesario obtener un acceso directo a una infraestructura, sino que al utilizar estas APIs hacen la función de una llamada para utilizar lo que justamente se necesita.

El consumo es la parte en la que el cliente suele preocuparse más, debido a que quiere saber la capacidad que podrá tener para utilizar los datos. Para este punto, la necesidad debe ser medida para así poder poner a disposición los datos necesarios para realizar futuros análisis o reportes, según los usuarios lo requieran.

5.4 Seguridad

En general, la seguridad de un lago de datos se enfoca en que los usuarios solo tengan acceso a los datos que necesiten y archivos o datos específicos, según lo definido por las políticas de seguridad y acceso de la empresa. Estas políticas pueden ser conformadas por la influencia de la filosofía interna de la compañía con respecto al acceso de datos.

Para llevar a cabo un lago de datos con un buen nivel de seguridad es necesario realizar un plan que pueda ayudar a reforzar las debilidades o amenazas que pueden provocar alguna problemática a futuro. Este plan debe abordar cinco puntos importantes:

1. El control de acceso de datos. Es en este punto donde se hace énfasis en restringir el acceso a los datos y objetos de archivos, con la finalidad de solo permitir a quienes estén por orden jerárquico realmente autorizados a interactuar directamente con los datos. A los demás, solo permitirles la interacción con los datos que necesiten para realizar sus tareas.

Figura 14 - 5.4: *Plan para la seguridad de un lago de datos*

Fuente: *OKERA*

2. Protección de datos. Aquí se hace referencia a la utilización de herramientas que permitan proteger la información almacenada ocultando de alguna forma lo que realmente se almacenó, o sea, usando técnicas como el cifrado para codificar los datos almacenados y que algún hacker se le dificulte descifrar el valor original del dato.

3. Auditoría de uso del lago de datos. Es la parte que comprende tanto a los propietarios de una empresa como a los administradores de datos. La idea es que estos dos tengan una comunicación abierta y sigan las reglas del gobierno de datos empresariales antes de poner los datos a disposición de otros usuarios dentro de la organización.

4. Prevención de fugas de datos. La mayoría de las fugas se originan dentro de la propia organización, a veces inadvertidamente y otras veces intencionalmente. Por consiguiente, un control de acceso eficiente es necesario para evitar estas situaciones. Por ejemplo, si un analista de marketing desea acceder a un archivo en particular que contiene números de Seguro Social, entonces la información confidencial debe ofuscarse dinámicamente, en lugar de denegar el acceso a todo el archivo.

5. Gobierno de datos y cumplimiento. Cada empresa debe tratar con los datos de sus usuarios de manera responsable para evitar el daño a la reputación por una violación de datos importante. Además, muchas industrias tienen requisitos regulatorios en torno al manejo de datos, como PCI DSS para servicios financieros e HIPAA para atención médica.

5.5 Aplicativos tecnológicos

En la actualidad, los lagos de datos han sido adoptados como una tecnología que puede ser ofrecida como servicio por medio de la nube, al igual que los almacenes de datos y demás servicios. Por tanto, existe una palestra enorme de proveedores de servicios cloud que cuenta con lagos de datos en su lista.

5.5.1 Amazon Web Services

Amazon Web Services es la plataforma de servicios en la nube que ofrece servicios de infraestructura de tecnologías de la información para las empresas a través del cloud computing. En la actualidad, Amazon cuenta con un sin número de servicios que ofrece a través de su plataforma, en donde se pueden encontrar algunos como bases de datos, almacenamiento, análisis, redes, herramientas para desarrollo, IoT, entre otros más.

En el caso de los lagos de datos, Amazon cuenta con una arquitectura enfocada en que los clientes pueden almacenar datos en este y utilizar un conjunto de servicios de datos creados para fines específicos en torno al lago, con lo que es posible tomar decisiones con rapidez y agilidad, a una escala y un precio/rendimientos inigualables en el mercado. [9]

La arquitectura de Amazon para los lagos de datos busca satisfacer las necesidades de sus clientes lo más eficiente posible, implementando características como la escalabilidad, servicios de análisis de datos, seguridad para el monitoreo de sus datos y poder escalar sus sistemas a un bajo coste sin comprometer el rendimiento.

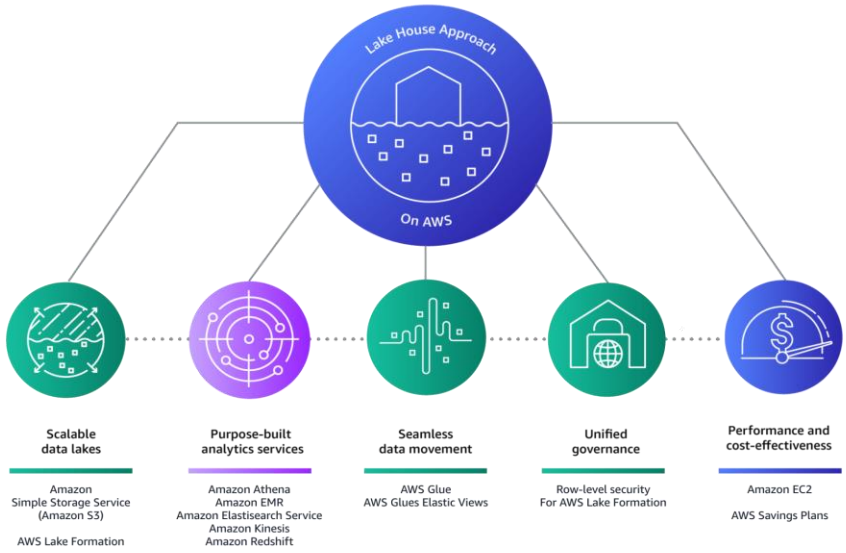

Figura 15 - 5.5.1: *Arquitectura de los lagos de datos en Amazon*

Fuente: *Amazon Web Services*

---

[9] Amazon Web Services. (2019). *¿En qué consiste el enfoque de Lake House?* Amazon. https://aws.amazon.com/es/big-data/datalakes-and-analytics/data-lake-house/?nc=sn&loc=11&refid=6e90e8fa-6bd8-4a6f-be4b-3bc9e717eb2e

5.5.2 Google Cloud

Google Cloud Platform es la suite de infraestructuras y servicios de Google, que utilizaban solamente de manera interna y, ahora, está disponible para cualquier empresa, con la finalidad de adaptar estos servicios a multitud de procesos empresariales.

En general, esta plataforma de Google provee herramientas para el diseño, testing, creación de aplicaciones, entre otras más, con altos niveles de seguridad y escalabilidad en cualquier herramienta.

Google Cloud ofrece una solución de lago de datos nativa de la nube, altamente escalable y segura que ofrece opciones e interoperabilidad a los clientes. Su arquitectura nativa de la nube reduce los costos y mejora la eficiencia para las organizaciones. Esta se basa en: [10]

1. Almacenamiento: proporciona opciones de almacenamiento a través del almacenamiento de objetos de bajo costo en Google Cloud Storage o el almacenamiento analítico altamente optimizado en BigQuery. [10]

2. Computación: proporciona diferentes motores para diferentes cargas de trabajo como BigQuery, Dataproc, Serverless Spark y Vertex AI. [10]

3. Administración: dataplex permite un tejido de administración de datos basado en metadatos a través de datos en Google Cloud Storage y BigQuery. Las organizaciones pueden crear, administrar, proteger, organizar y analizar datos en el lakehouse utilizando Dataplex. [10]

---

[10] Google Cloud. (2021, October 28). *Open data lake house on Google Cloud*. Google Cloud.

https://cloud.google.com/blog/products/data-analytics/open-data-lakehouse-on-google-cloud

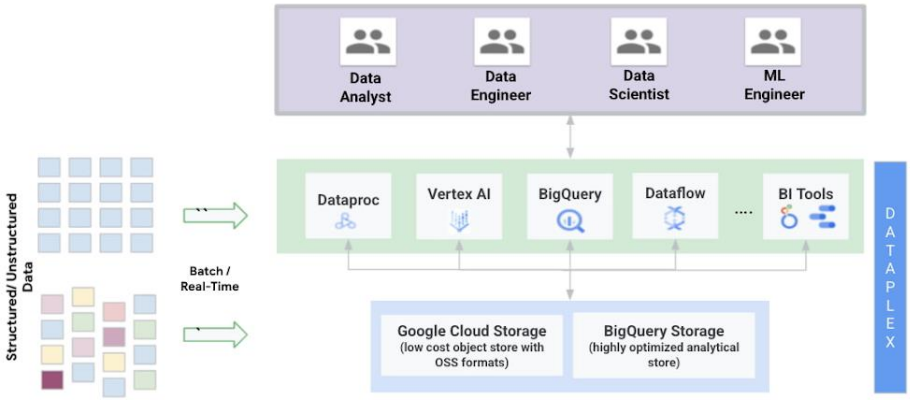

Figura 16 - 5.5.2: *Arquitectura de los lagos de datos en Google Cloud*

Fuente: *Google Cloud*

5.5.3 IBM Cloud

La plataforma de servicios cloud de IBM proporciona soluciones que permiten mayores niveles de cumplimiento, seguridad y gestión, con patrones de arquitectura probados y métodos de entrega rápida para ejecutar cargas de trabajo críticas, también cuenta con una plataforma de nube híbrida de última generación, capacidades avanzadas de datos e inteligencia artificial y una profunda experiencia empresarial en 20 sectores. [11]

IBM Cloud cuenta con una plataforma de lago de datos totalmente gestionada para los datos. El servicio central de esta es IBM Cloud SQL Query, el cual proporciona ingesta de datos, preparación de datos, análisis de datos y catálogo de tablas relacionales.

[11] IBM. (2018). *IBM Cloud Docs*. IBM Cloud Docs. https://cloud.ibm.com/docs/overview?topic=overview-whatis-platform&locale=es

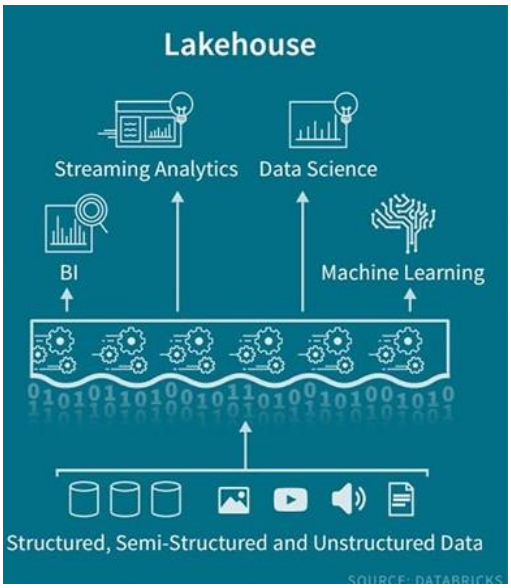

Figura 17 - 5.5.3: *Arquitectura de los lagos de datos en IBM Cloud*

Fuente: *IBM Cloud*

5.5.4 Microsoft Azure

Microsoft Azure es un servicio de nube pública de pago por uso, en donde se alojan distintos servicios como redes, inteligencia artificial, herramientas para desarrollo de aplicaciones, análisis de datos, bases de datos, entre otros más. Esta plataforma cuenta con alrededor de 200 productos y servicios en la nube diseñados con la finalidad de darle vida y escalabilidad a las nuevas implementaciones o soluciones que permitan resolver las dificultades presentes y a futuro poder mejorar.

Microsoft Azure cuenta con un lago de datos enfocado en la analitica, se puede desarrollar y ejecutar con mucha facilidad programas enfocados en el procesamiento y transformación de datos en paralelo de forma masiva; soporta lenguaje de programación como Python, .NET, R y U-SQL; puede procesar los datos a petición, escalar las unidades de análisis de forma instantánea y pagar solo por trabajo. [12]

---

[12] Microsoft Azure. (2019). *Data Lake Analytics*. Microsoft. https://azure.microsoft.com/es-mx/services/data-lake-analytics/?&ef_id=Cj0KCQiA7oyNBhDiARIsADtGRZZyHTGUHbOTlbhfks1Z4Y8ZRjqq2TS1QeTtgzDYlP8RGEpnG40D-BcaAgXBEALw_wcB:G:s&OCID=AID2200178_SEM_Cj0KCQiA7oyNBhDiARIsADtGRZZyHTGUHbOTlbhfks1Z4Y8ZRjqq2TS1QeTtgzDYlP8R

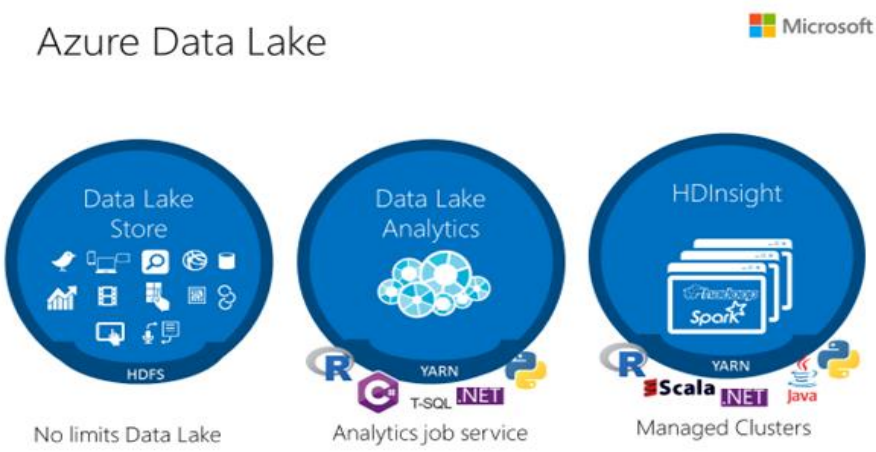

Figura 18 - 5.5.4: *Lago de datos en Microsoft Azure*

Fuente: *Microsoft Azure*

5.6 Problemáticas

En una encuesta realizada a finales del mes de noviembre de 2021 a 101 ciudadanos de distintas provincias del país, se pudo encontrar que alrededor del 94% de los ciudadanos, que están comprendidos entre las escalas de 3 a 5, afirman que la frecuencia con la que ocurren actos delictivos en sus respectivas provincias es muy alta siendo esta un promedio del 4.7 de 5.

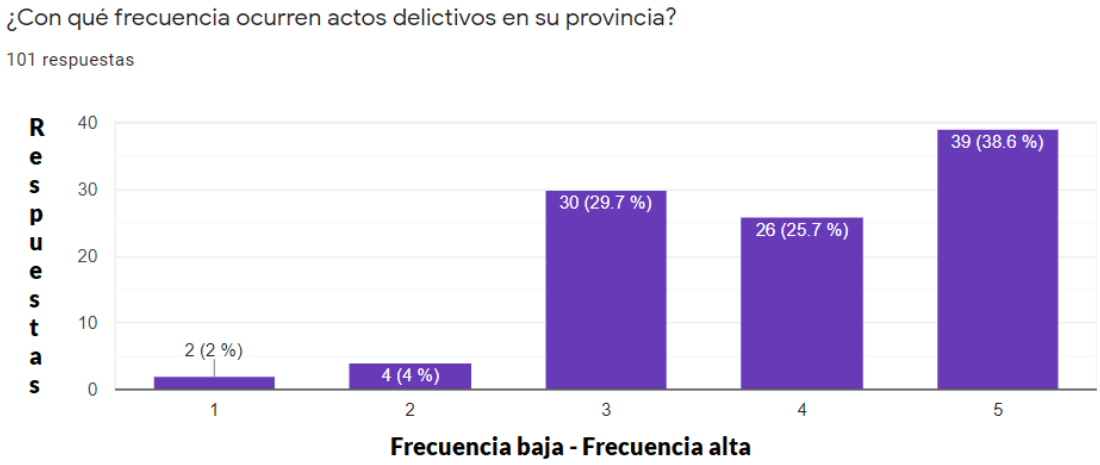

Gráfico 1 -5.6: *Frecuencia con la que ocurren los actos delictivos*

Fuente: *Autores*

A pesar de las extensas medidas de seguridad tomadas por la institución policial, se sigue reflejando por parte de la ciudadanía la situación de desconfianza para transitar tranquilamente por las calles y avenidas del país.

Una de las principales causas por la cual los ciudadanos se sienten tan inseguros al salir a las calles se debe a que, en muchas ocasiones en las que ocurre algún acto delictivo, no se captura al responsable. Esto es evidente, ya que, alrededor del 94.1% refleja un descontento con las autoridades encargadas de mantener el orden y la seguridad.

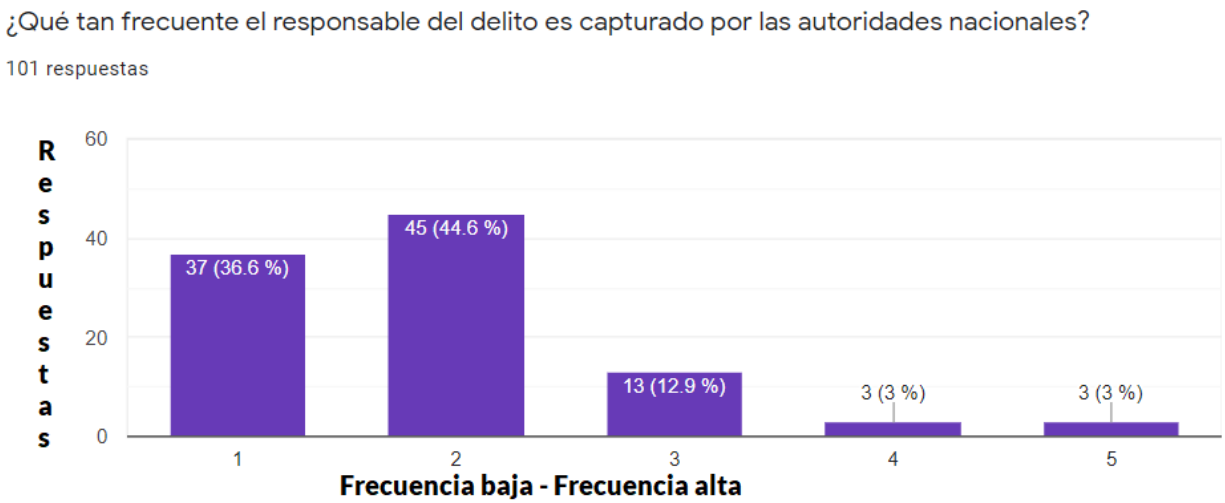

Gráfico 2 - 5.6: *Frecuencia con la que se atrapan a los responsables de algún acto delictivo*

Fuente: *Autores*

Otro factor más de riesgo que afecta gravemente a la seguridad de los ciudadanos es la falta de iluminación en muchas calles y avenidas del país, convirtiéndose esto en una vulnerabilidad al momento de transitar en una de estas.

Según la misma encuesta, los actos delictivos más frecuentes son por mucho los asaltos a transeúntes, carterismo, robos de bienes, violación de las leyes de tránsito, entre otros muchos, varios de los cuales se ven favorecidos cuando existe poca iluminación en las vías públicas. Siendo los asaltos a transeúntes y las violaciones a las leyes de tránsito los que con más frecuencia suceden, con unas 76 y 66 selecciones respectivamente de un total de 441.

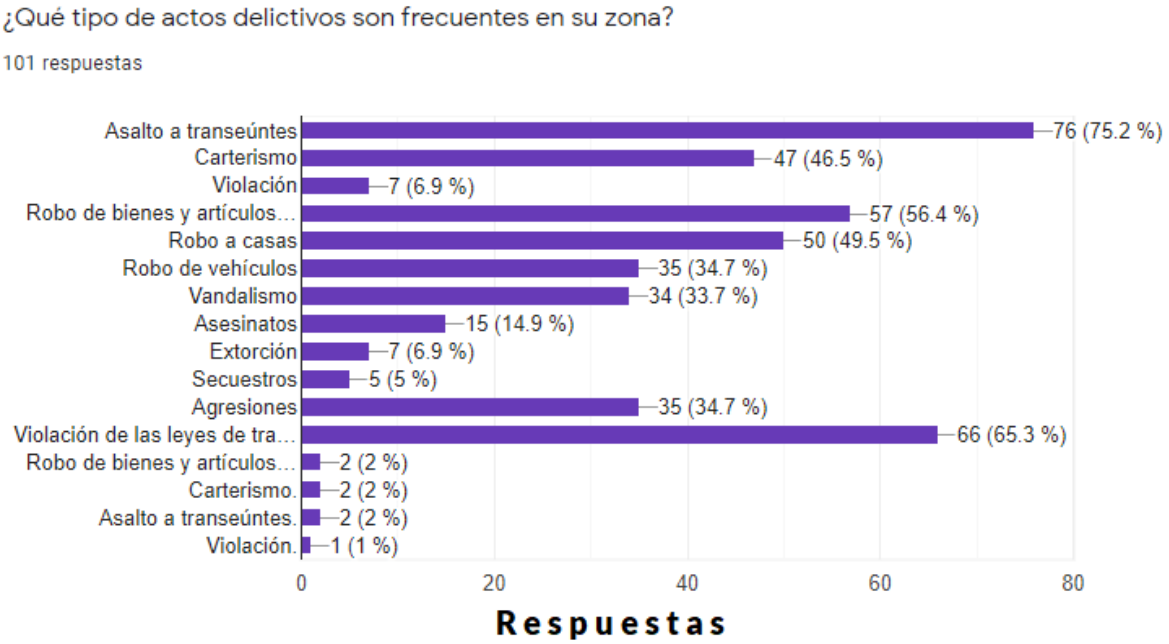

Gráfico 3 - 5.6: *Actos delictivos más frecuentes*

Fuente: *Autores*

De acuerdo con los datos recolectados, alrededor del 55.4% de los encuestados considera que las autoridades nacionales no cuentan con los recursos tecnológicos necesarios para realizar una captura eficiente y rápida del protagonista de algún hecho delictivo.

Estas estadísticas ponen en evidencia que las medidas tomadas por las autoridades nacionales al momento de pretender capturar a algún delincuente no son suficientes; se necesita del uso de recursos tecnológicos para realizar los distintos procesos de una manera más ágil y rápida.

Gráfico 4 - 5.6: *Número de ciudadanos que consideran que las autoridades nacionales no cuentan con recursos tecnológicos eficientes*

Fuente: *Autores*

Los resultados de la investigación confirman la necesidad de la implementación de recursos tecnológicos más eficientes, con un 95% de los encuestados de acuerdo con que en los procesos llevados a cabo para la captura de algún delincuente o criminal esté presente la tecnología como un aliado, y para con esto lograr hacer justicia y mantener una población segura.

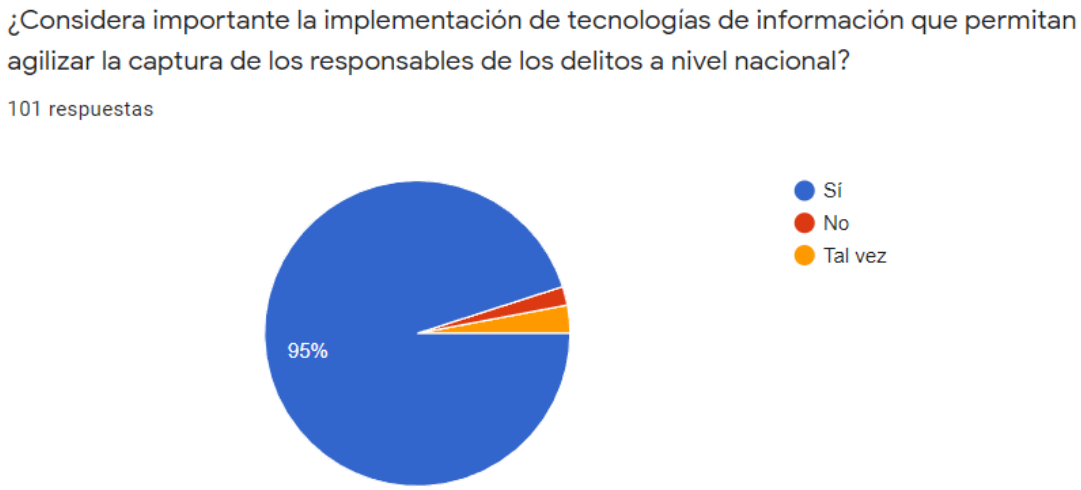

Gráfico 5 - 5.6: *Número de ciudadanos que consideran que las autoridades nacionales deberían contar con recursos tecnológicos eficientes*

Fuente: *Autores*

Según los resultados de la encuesta, se pudo analizar con qué frecuencia por provincia ocurren actos delictivos. Siendo Santo Domingo y el Distrito Nacional donde se encuentran los picos más altos de estos hechos, a estos le seguiría Valverde y Duarte mostrando picos medios.

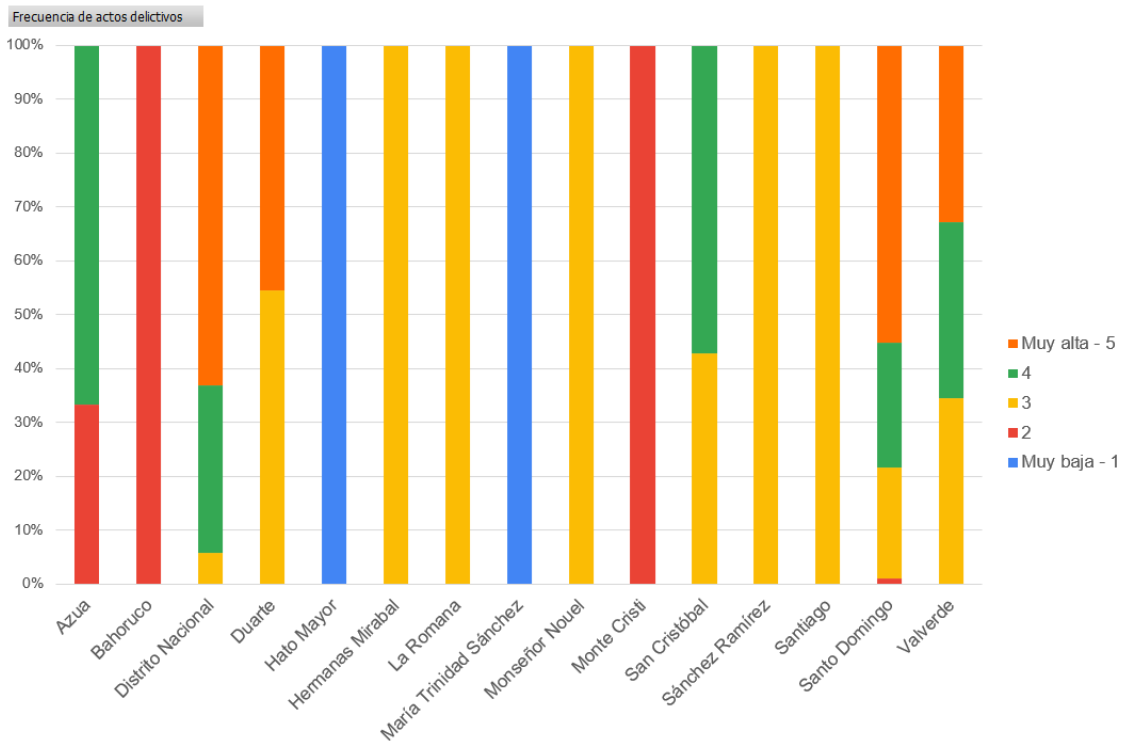

Gráfico 6 - 5.6: *Frecuencia de delitos por provincia*

Fuente: *Autores*

Un análisis realizado con los datos de la encuesta revela la frecuencia con la que los autores de los actos delictivos son atrapados por las autoridades, en donde en la mayoría de los casos muestran una baja tasa en cuanto a esto. Siendo Hato Mayor, María Trinidad y Sanchez Ramirez las provincias que más baja tasa muestran.

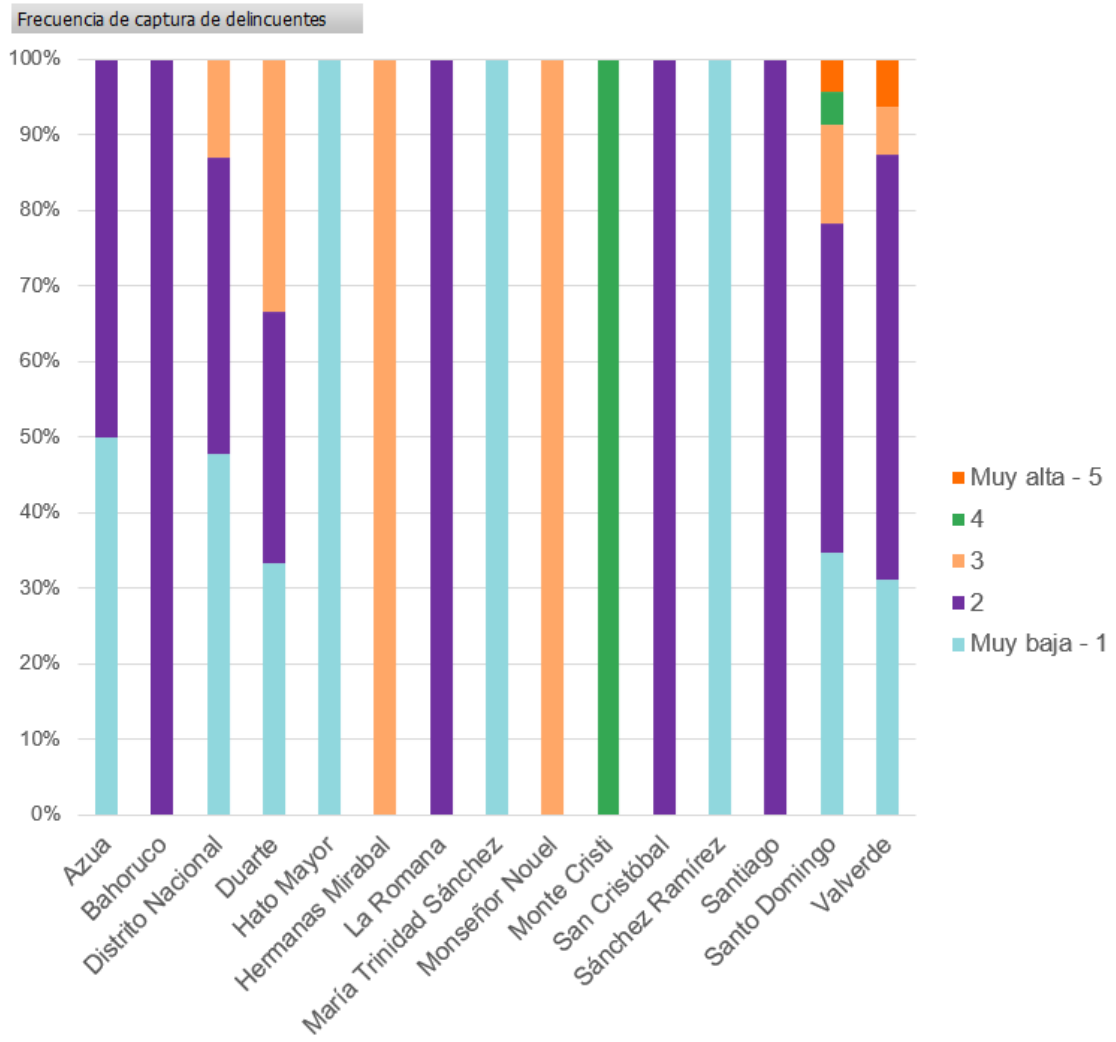

Gráfico 7 - 5.6: *Frecuencia de captura de delincuentes por provincia*

Fuente: *Autores*

Resumen

Los lagos de datos son una tecnología que, por medio de su uso, ha permitido a un sin número de empresas alrededor del globo concentrar todos los datos en un único punto, sin importar su formato ni el sistema de origen de estos. Esta tecnología cuenta con varios componentes dentro de su infraestructura, los cuales son: la ingesta, el almacenamiento, los procesos y el consumo.

Los lagos de datos en términos de seguridad se enfocan en que los usuarios solo tengan acceso a los datos que necesiten y archivos o datos específicos, según lo definido por las políticas de seguridad y acceso de la empresa.

Por otro lado, debido a los beneficios que posee un lago de datos, ciertas organizaciones han decidido optar por ofrecer en sus servicios cloud esta tecnología, como es el caso de Amazon, Google, IBM y Microsoft. En donde cada una pone suficiente carne en el asador para que el cliente sepa cuál opción tomar según sus necesidades.

CAPÍTULO 6: PROPUESTA DE DISEÑO DE LAGO DE DATOS EN INSTITUCIONES GUBERNAMENTALES DE REPÚBLICA DOMINICANA

Introducción

A continuación, se estará presentando la propuesta de diseño de un lago de datos para las instituciones gubernamentales, con el objetivo de que los analistas puedan darles distintas formas a los datos almacenados en este repositorio.

Esta propuesta busca erradicar los casos delictivos que nunca se resuelven, además de agilizar el proceso para resolver los casos más actuales y los que representen un mayor grado de urgencia. Por ende, el lago de datos pretende ser un repositorio que además de almacenar todo tipo de datos, pueda ayudar por medio de las herramientas internas a realizar análisis y reportes para cada uno de los casos y así optimizar en cierto sentido los procesos llevados a cabo actualmente.

En este capítulo, se pretende mostrar cada uno de los detalles sobre la propuesta, además de sus características, requerimientos técnicos, arquitectura, el modelo de despliegue y los distintos factores que se deben tomar en cuenta antes de migrar a una plataforma en la nube.

6.1 Propuesta de Lago de datos

De acuerdo con las problemáticas que por medio de encuestas y análisis se han detectado en las instituciones gubernamentales encargadas de preservar la seguridad de los ciudadanos dominicanos, hemos planificado una propuesta de diseño para un lago de datos para estos órganos del estado.

Esta propuesta está basada en un modelo de lago de datos en la nube, donde las instituciones podrán almacenar desde sus distintos sistemas todos los datos relacionados con delitos cometidos en el país sin importar el formato en el que se encuentren, y desde este poder realizar distintos tipos de análisis por medio de herramientas de analítica y creación de reportes, logrando así resolver casos que se generan día a día.

La idea de los lagos de datos ha sido implementada en múltiples empresas alrededor del mundo de manera exitosa, de este modo funciona como un repositorio de datos sin importar su formato de origen, además de poseer una gran capacidad de almacenamiento y procesamiento para el uso de múltiples usuarios al mismo tiempo.

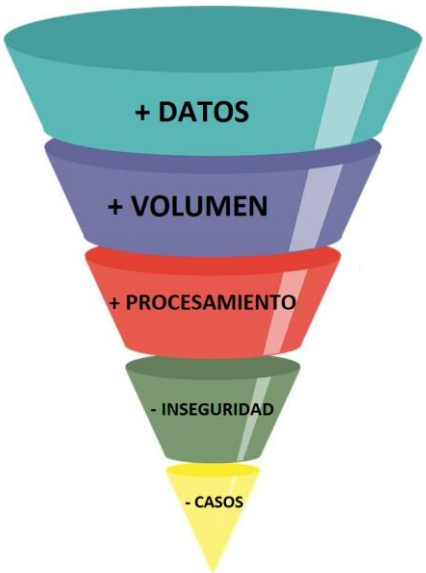

Gráfico 8 - 6.1: *Embudo de mitigación del Lago de datos*

Fuente: *Autores*

Basándonos en un modelo donde cada institución cargue los datos correspondientes al lago de datos, dentro de este se realiza la ingesta y posteriormente se almacenan los datos, luego se transforman, publican y distribuyen estos datos. De esta forma por medio de los beneficios que ofrece el lago se mitigarán los principales consecuentes de las problemáticas.

Algunas de las instituciones gubernamentales se encuentran situadas en las altas posiciones a nivel tecnológico reconocidas en muchos casos por instituciones internacionales, por tanto, añadiendo esta tecnología se lograría un gran impacto social en la seguridad de la población dominicana.

6.1.1 Características

El modelo estaría construido bajo seis pilares importantes:

1. Cargar la data desde los sistemas de origen.

2. Realizar la ingesta de los datos.

3. Almacenar los datos.

4. Transformar los datos.

5. Publicar los datos.

6. Distribuir los datos.

Los datos serán cargados desde los distintos sistemas de origen o centros de datos de las instituciones gubernamentales, ya sean la policía nacional, digesett, ministerio de obras públicas y comunicaciones, dirección general de migración y el 9-1-1.

La ingesta de datos es la encargada de realizar los distintos chequeos antes de almacenar los datos, tales como: controles de calidad, encriptación y registros de metadata.

Los datos son de tipo estructurado, semiestructurado o no estructurados y son almacenados sin sufrir ningún tipo de alteración. Se recomienda tener analistas expertos en data discovery para poder manejar estos.

Transformar los datos de un formato a otro que sea más manejable para el sistema que necesite de su uso. Por ende, con este pilar se pretende que los

datos sean utilizados de la mejor manera posible por la aplicación o sistema que los necesite para su funcionamiento.

La publicación de los datos implica alertar de que los datos están listos para ser utilizados en el sistema o aplicación que está a la espera de estos.

Distribuir los datos es repartir la información hacia los distintos sistemas que la pretenden utilizar para un objetivo específico.

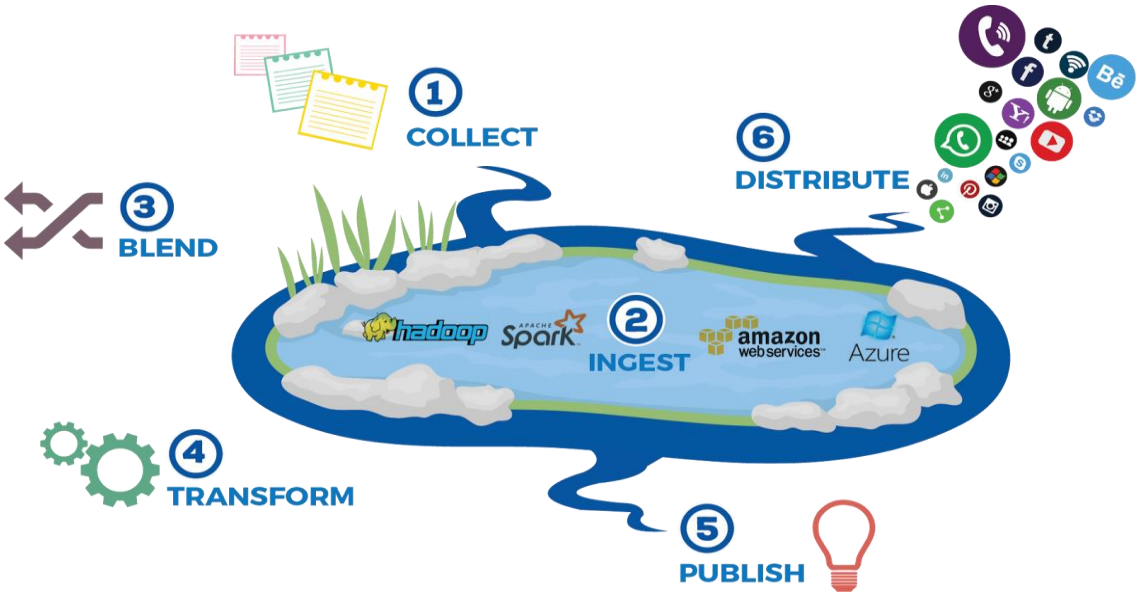

Figura 19 - 6.1.1: *Pilares para el Lago de datos*

Fuente: *Beristain*

6.2 Requerimientos técnicos

Los requerimientos técnicos son aquellos aspectos que deben de alguna manera u otra cumplirse con éxito en un proyecto. En este se pueden tratar aspectos como la fiabilidad, rendimiento y la disponibilidad.

6.2.1 Requerimientos de Red

1. Definir un ancho de banda de al menos 200 Mb/s.

2. El tipo de carga de trabajo debe ser definida, ya sea modalidad continua u orientada a transacciones.

3. Identificar el tamaño del MTU (Unidad máxima de transmisión) que se utilizara (al menos 1500).

4. Saber si el adaptador de ethernet compartido será utilizado en un entorno con o sin hebras.

5. Tener conocimiento de las velocidades de transferencia.

6.2.2 Requerimientos de Hardware

1. Procesador de 64 bits a 3 GHz con 8 - 16 núcleos.
2. 64 GB de Ram.
3. 2TB de almacenamiento interno (SSD o HDD).
4. Adaptador de red Ethernet de 1 gigabit por segundo.
5. Gráficos integrados que admitan al menos FULL HD (1920 x 1080).

6.2.3 Requerimientos de Software

1. Sistema operativo de 64 bits.
2. Soporte para tecnologías como .NET Framework.
3. Soporte para conexión a internet.
4. Soporte para IIS.
5. Soporte para VMs.
6. Soporte para usuarios y dispositivos ilimitados.

6.3 Arquitectura del Lago de datos

En general, la arquitectura es la representación de la correlación de funciones con los distintos componentes tanto de hardware como de software. Por ende, la arquitectura que se pretende mostrar para el lago de datos de nuestra propuesta contiene las siguientes características:

1. Escalabilidad. Es esa característica que permite a un sistema reaccionar y adaptarse a cualquier situación sin perder el nivel de eficiencia y calidad. Además de poder manejar el crecimiento continuo de las operaciones realizadas de manera fluida.

2. Uso de herramientas de analítica de datos. Son aplicaciones que permiten a los analistas realizar los reportes y análisis con la finalidad de explorar y encontrar la solución a diversas situaciones por medio de los datos que han sido transformados.

3. Acceso unificado a los datos. Esta característica comprende la capacidad de trasladar fácilmente una parte de los datos de un almacén de datos a otro.

4. Gobernanza. Es la capacidad de un lago de datos para que un cliente pueda autorizar, administrar y auditar el acceso a los datos, haciendo uso de políticas, controles de acceso, filtrado de datos y demás capacidades.

5. Rentable y alto rendimiento. Con esta característica se busca poder mantener el lago de datos con un alto rendimiento en los procesos diarios de las instituciones, pero que al mismo tiempo resulte rentable económicamente para las instituciones.

6.4 Modelo de despliegue

Los modelos de despliegue son un tipo de diagrama que permite visualizar la arquitectura de ejecución de un sistema, ya sean los componentes de hardware y software, además de los procesos realizados dentro de este.

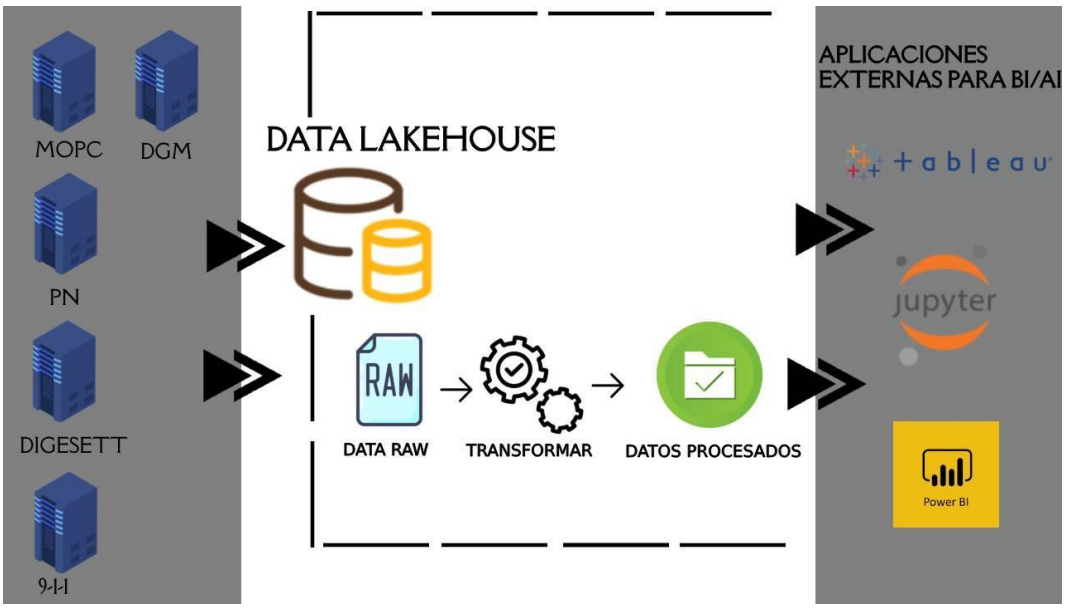

Figura 20 - 6.4: *Modelo de despliegue de la propuesta del lago de datos*

Fuente: *Autores*

Para esta propuesta, el modelo de despliegue consta de tres secciones para la realización de los procesos internos del lago de datos:

La primera sección consta de los servidores de las distintas instituciones gubernamentales especificadas en la Figura 20 - 6.4, desde donde se van a cargar todos los datos relacionados con la inteligencia delictiva de los ciudadanos dominicanos.

En la segunda sección, se encuentra el lago de datos, en donde éstos van a navegar por varios procesos de depuración y control de la calidad de la información, transformación y, por último, los datos son procesados y almacenados para posteriormente ser utilizados para algún propósito en específico.

En la tercera y última sección, los datos ya procesados y almacenados pasan a ser utilizados en distintas herramientas de analítica de datos e inteligencia de negocios, como es el caso de Power BI, Jupyter y Tableau. Aquí es donde los datos toman más sentido, ya que, gracias a los analistas cualquier persona puede apreciar la información de distintas maneras, ya sea por medio de tablas, gráficos, análisis y reportes, con la finalidad de darle un sentido a todos los datos recopilados.

Lo que se propone en este modelo es que los distintos servidores de las instituciones gubernamentales realicen una estrategia de carga que esté comprendida en horarios con menor tráfico de red. Se recomienda que esta carga de datos tenga lugar en horas de la madrugada, aprovechando así la tranquilidad en el flujo de la red por parte del ISP.

Si bien es cierto que hoy en día todavía no se cuenta con la infraestructura necesaria para sacarle el mayor provecho a un lago de datos, para complementar los datos registrados por los agentes del orden acerca de sucesos delictivos, el modelo propuesto cuenta con la capacidad de recibir y procesar datos directamente de dispositivos IoT (cámaras de seguridad, semáforos inteligentes, sensores en las vías de tráfico, etc.), afinando las conclusiones potenciales que puedan alcanzarse para determinar los detalles clave de estos sucesos.

Con este modelo de despliegue se pretende que las instituciones gubernamentales puedan utilizar, de la manera más eficiente y provechosa, todo lo que un lago de datos puede contener dentro de su arquitectura, con el objetivo de agilizar los distintos procesos relacionados a la criminalidad que está arropando al país en la actualidad.

El modelo de despliegue permite combinar los mejores puntos de integración, calidad y gestión de datos, además ayuda a las organizaciones a definir, administrar y gobernar el uso de las diversas tecnologías que se relacionan con el big data. Por tanto, utilizar un lago de datos permite la

realización de análisis de big data y otras aplicaciones de la ciencia de datos involucrando técnicas de análisis avanzadas, como la minería de datos, el modelado predictivo y el aprendizaje automático.

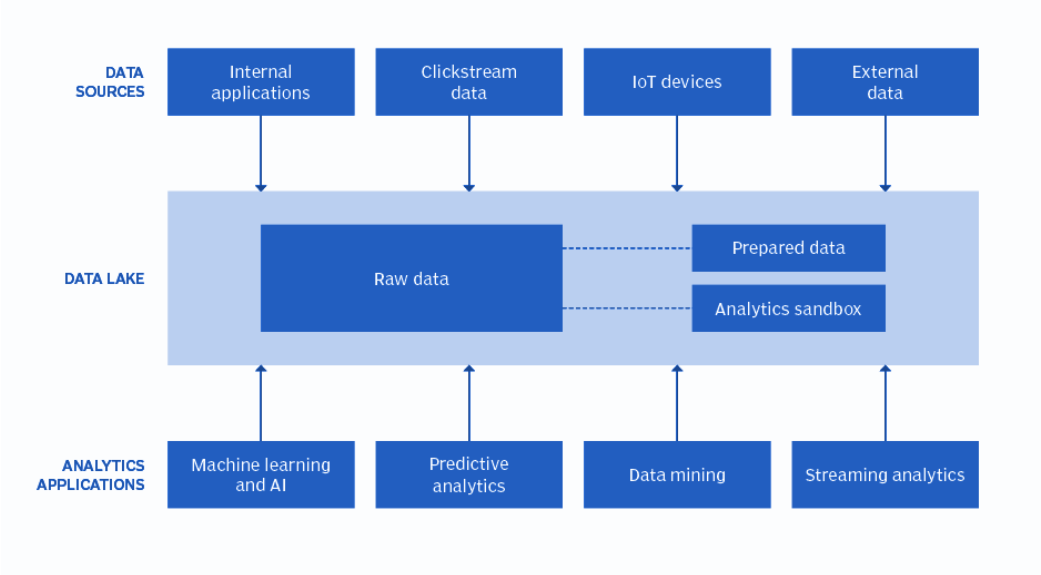

Figura 21 - 6.4: *Arquitectura de un lago de datos*

Fuente: *TechTarget*

6.5 Factores a tomar en cuenta antes de migrar a un Lago de datos

Las migraciones de bases de datos hacia la nube se han convertido en una estrategia muy utilizada por parte de las empresas en la actualidad, ya que esta acción puede ser impulsada por diversos factores como falta de actualización del software y hardware, las limitaciones del sistema, altos costos de uso, entre otros más. Por ende, las soluciones cloud suelen ser una vía por la cual desviar estos inconvenientes momentáneamente y obtener un mejor rendimiento en los procesos que se efectúan día a día en las empresas.

En base a esto, se deben tomar en cuenta los siguientes factores antes de migrar a un servicio cloud, en este caso, un lago de datos:

1. Descomprimir el lago de datos. Gran parte de los problemas que surgen en el proceso de migración es que las empresas deben conocer las diversas tecnologías y aplicaciones que operan dentro de un lago de datos, y posteriormente descubrir cómo desacoplarlas y migrarlas.

2. Identificar las cargas de trabajo y priorizar en función al caso de uso, rendimiento y costo. Dentro de un lago de datos puede existir mucha basura, por ende, si se está planeando hacer una migración se debe priorizar en función a la importancia de los distintos conjuntos de datos, ya sean para los informes ejecutivos, marketing, recursos humanos, entre otros más.

3. Automatizar la migración de la carga de trabajo al destino. Con los lagos de datos los volúmenes de datos pueden escalar a petabytes, por tanto, el riesgo y costo que existe es demasiado grande para desplegar un equipo de personas que codifiquen manualmente. Por ende, se debe considerar utilizar las herramientas que permitan realizar una migración de manera automática a las distintas plataformas en la nube.

Resumen

Proponemos un modelo de lago de datos en la nube, donde las instituciones podrán almacenar desde sus distintos sistemas todos los datos relacionados con delitos cometidos, y desde este poder realizar distintos tipos de análisis por medio de herramientas de analítica y creación de reportes, logrando así resolver casos que se generan día a día.

La idea se basa en la utilización de un lago de datos en la nube, en donde las instituciones gubernamentales puedan cargar la información delictiva de los ciudadanos y posteriormente dentro de este repositorio poder transformar esos datos y utilziarlos en aplicaciones de analitica e inteligencia de negocios con la finalidad de crear análisis y reportes para resolver los distintos casos delictivos que estén vigentes.

La arquitectura de estos sistemas permite obtener ciertos beneficios para el cliente que lo utilice, estos pueden ser la escalabilidad, el uso de herramientas de analitica de datos, acceso unificado a los datos, gobernanza y la rentabilidad y el alto rendimiento.

Esta propuesta de diseño presenta la solución a ciertos inconvenientes con los procesos realizados por parte de las instituciones para poder resolver los casos delictivos presentes de la manera más eficiente y rápida posible.

Conclusiones y Recomendaciones

El objetivo principal de un Lago de datos es permitir al cliente almacenar todos los datos sin importar su formato en el repositorio y posteriormente utilizarlos en análisis y reportes. El diseño de éste trae consigo múltiples beneficios, tales como: facilidad de adaptación a cambios, almacenamiento de cualquier tipo de dato, facilidad para agregar nuevos datos y poder dar soporte a múltiples usuarios en la empresa.

Otro punto importante, que es el principal que se quiere resaltar en este documento, es que se podrá agilizar los procesos para resolver los casos relacionados a algún hecho delictivo. Con esto, se lograría atrapar de manera más rápida al autor del hecho y se lograría hacer justicia por medio del sometimiento del responsable ante la ley.

Es por esto por lo que las instituciones encargadas de preservar el orden y seguridad deben cargar estos datos al lago de datos, para que así los analistas encargados puedan realizar los análisis y reportes correspondientes a cada caso que surja.

Por otro lado, hay que tener en cuenta que un lago de datos alojado en una nube pública es un costo, por tanto, el rendimiento debe ser sigilosamente calculado, con el objetivo de que sea lo más eficiente posible, pero que el costo permanezca dentro de un rango aceptable por la entidad contratante.

El proyecto podría tener un costo-beneficio factible siempre y cuando las instituciones gubernamentales pertinentes analicen los pros y contras, y puedan tomar una decisión según los datos que éstas tengan.

Por todos los datos que analizamos, se recomienda que la propuesta de diseño puede ser factible siempre y cuando las instituciones de turno evalúen su propia situación, y el impacto que tecnología de esta magnitud podría tener dentro de la sociedad dominicana.

Bibliografía

1&1 Ionos. (2020, 16 marzo). Introducción al sistema gestor de base de datos (SGBD). Recuperado de https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/sistema-gestor-de-base-de-datos-sgbd/

Amazon Web Services, Inc. (2021b). Arquitectura de Lake House | Amazon Web Services. Recuperado de https://aws.amazon.com/es/big-data/datalakes-and-analytics/data-lake-house/?nc=sn&loc=11&refid=6e90e8fa-6bd8-4a6f-be4b-3bc9e717eb2e

Banafa, A. (2021, 9 junio). ¿Qué es un lago de datos? Ventajas y contras. Recuperado de https://www.bbvaopenmind.com/tecnologia/mundo-digital/un-lago-de-datos-una-oportunidad-o-un-sueno-para-el-big-data/

Bernal, C. A. (2006). Metodología de la investigación (2da. ed.). Naucalpan, México: Pearson Educación.

Centros Culturales de México A.C. (2021, 26 febrero). ¿Qué son las instituciones políticas y cuál es su función? Recuperado de https://blog.up.edu.mx/topic/posgrados-de-gobierno-y-economia/que-son-las-instituciones-politicas-y-cual-es-su-funcion

Cortés, M. (2017, 4 agosto). Evite que su lago de datos se convierta en un pantano de datos. Recuperado de https://cio.com.mx/evite-que-su-lago-de-datos-se-convierta-en-un-pantano-de-datos/

DBA dixit. (2019, 16 junio). Características de un Sistema Gestor de Bases de Datos (SGBD o DBMS). Recuperado de http://dbadixit.com/caracteristicas-sistema-gestor-bases-datos-sgbd-dbms/

Dertiano, V. (2017, 23 octubre). Data Lake - Introducción | Características y Capas básicas. Recuperado de https://blog.bi-geek.com/arquitectura-bi-introduccion-al-data-lake/

Dirección General De Migración. (2021). ¿Quiénes Somos? : DIRECCIÓN GENERAL DE MIGRACIÓN. Recuperado de https://migracion.gob.do/quienes-somos/

Dirección General de Seguridad de Tránsito y Transporte Terrestre, DIGESETT. (2021). ¿Quiénes Somos? Recuperado de http://digesett.gob.do/index.php/sobre-nosotros/quienes-somos

EBM Software. (2021, 1 octubre). Data Lake vs Data Warehouse: Advantages and Disadvantages. Recuperado de https://ebmsoftware.com/data-lake-vs-data-warehouse-advantages-and-disadvantages/

Egremy, W. (2021, 10 abril). ¿qué son las instituciones gubernamentales y para qué sirven? Recuperado de https://aleph.org.mx/que-son-las-instituciones-gubernamentales-y-para-que-sirven

Equipo editorial, Etecé. (2021, 5 agosto). Institución - Concepto, clasificación y características. Recuperado de https://concepto.de/institucion/

Hernández, R., Fernández, C. & Baptista, P. (2008). Metodología de la investigación (4ta. ed.). Distrito Federal, México: McGraw-Hill Interamericana.

IBM. (2021, 7 julio). IBM Cloud Docs. Recuperado de https://cloud.ibm.com/docs/overview?topic=overview-whatis-platform&locale=es

IBM. (2019). Requisitos de Hardware. IBM. Obtenido de: https://www.ibm.com/docs/es/license-metric-tool?topic=requirements-disk-space-windows

Kyocera. (2021). La importancia de la gestión de base de datos para el CIO | Kyocera. Recuperado de https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/insights-hub/articles/la-importancia-de-la-gestion-de-base-de-datos-para-el-cio.html

Mathur, C. (2021, 18 junio). Top Three Considerations Before Migrating Your Data Lake To The Cloud. Recuperado de https://www.forbes.com/sites/forbestechcouncil/2021/06/18/top-three-considerations-before-migrating-your-data-lake-to-the-cloud/?sh=20df00053ba8

Ministerio de Obras Públicas y Comunicaciones. (2021). ¿Quiénes Somos? Recuperado de https://www.mopc.gob.do/nosotros/qui%C3%A9nes-somos/

Microsoft. (2021). Data Lake Analytics. Recuperado de https://azure.microsoft.com/es-mx/services/data-lake-analytics/?&ef_id=Cj0KCQiA7oyNBhDiARIsADtGRZZyHTGUHbOTlbhfks1Z4Y8ZRjqq2TS1QeTtgzDYlP8RGEpnG40D-

BcaAgXBEALw_wcB:G:s&OCID=AID2200178_SEM_Cj0KCQiA7oyNBhDiARIsADtGRZZyHTGUHbOTlbhfks1Z4Y8ZRjqq2TS1QeTtgzDYlP8R

Okera Marketing Team. (2021, 13 agosto). What Is Data Lake Security? Recuperado de https://www.okera.com/blogs/what-is-data-lake-security/

Oracle. (2019). Qué es un almacén de datos. Oracle. Obtenido de: https://www.oracle.com/ar/database/what-is-a-data-warehouse/

Pérez Porto, J., & Gardey, A. (2013). Definición de organismo gubernamental — Definicion.de. Recuperado de https://definicion.de/organismo-gubernamental/

Policía Nacional. (2021). Quiénes Somos –. Recuperado de https://www.policianacional.gob.do/sobre-nosotros/quienes-somos/

Qubole. (2021, 28 octubre). What is a Data Lake? Recuperado de https://www.qubole.com/what-is-data-lake/

Ramos, J., Ramos, A. & Montero, F. (2006). Sistemas gestores de bases de datos (1ra. ed.). Madrid, España, por McGraw-Hill / Interamericana. Recuperado de https://www.academia.edu/24311907/SISTEMAS_GESTORES_DE_BASES_DE_DATOS

Sanchez, J. (s. f.). Tipos de SGBD. Recuperado de https://jorgesanchez.net/presentaciones/bases-de-datos/introduccion-sgbd/tipos-sgbd-bn.pdf

Sistema Nacional de Atención a Emergencias y Seguridad 9–1-1. (2021, 19 mayo). SOBRE NOSOTROS. Recuperado de https://911.gob.do/sobre_nosotros/

Snowflake. (2021). What is a Data Lakehouse? Recuperado de https://www.snowflake.com/guides/what-data-lakehouse

Stedman, C., & Lutkevich, B. (2020, 24 enero). data lake. Recuperado de https://searchdatamanagement.techtarget.com/definition/data-lake

Talend. (2021). Data Lake vs Data Warehouse: Key Differences - Talend. Recuperado de https://www.talend.com/resources/data-lake-vs-data-warehouse/

Tecnologias-Informacion. (2018a). Data Lake: Definición, Beneficios y Características. Recuperado de https://www.tecnologias-informacion.com/data-lake.html

Tekiner, F., & Pierce, S. (2021, 28 octubre). Open data lakehouse on Google Cloud. Recuperado de https://cloud.google.com/blog/products/data-analytics/open-data-lakehouse-on-google-cloud

Universidad Nacional de la Plata. (2019). Las organizaciones, sus características y tipos. UNLP. Recuperado de: https://unlp.edu.ar/frontend/media/46/27446/e0adc2b174b18ae3438092b6d1e4748c.pdf

VARA HORNA, A. A. (2012). 7 PASOS PARA UNA TESIS EXITOSA (3ª edición). Lima: Universidad San Martin de Porres.

Villarrubia, C. (2012, 15 agosto). El organismo de Migración dominicano se apoya en el NAP del Caribe. Recuperado de https://www.datacenterdynamics.com/es/noticias/el-organismo-de-migraci%C3%B3n-dominicano-se-apoya-en-el-nap-del-caribe/

Wikipedia contributors. (2021, 30 noviembre). Data warehouse. Recuperado de https://en.wikipedia.org/wiki/Data_warehouse

Anexos

Anexo de Encuesta: Tecnología para cargar información delictiva para la identificación de criminales.

1. ¿Sexo?
2. ¿Edad?
3. ¿Nivel Educativo?
4. ¿De qué provincia eres?
5. ¿Con qué frecuencia ocurren actos delictivos en su provincia?
6. ¿Qué tan frecuente el responsable del delito es capturado por las autoridades nacionales?
7. ¿Qué tipo de actos delictivos son frecuentes en su zona?
8. ¿Cree usted que las autoridades nacionales cuentan con los recursos tecnológicos necesarios para capturar aquellos que delinquen?
9. ¿Considera importante la implementación de tecnologías de información que permitan agilizar la captura de los responsables de los delitos a nivel nacional?