|  |  |  |
| --- | --- | --- |
| **Nombre**: Diego Ronaldo Sanchez Pichardo | | **Matrícula**: 05077303 |
| **Nombre del curso:** Ciencia de datos | **Nombre del profesor**: Kristel Cano Galán | |
| **Módulo**: 2do | **Actividad**: Actividad 1 | |
| **Fecha**: 24 de Marzo del 2025 | | |
| **Bibliografía**:  Robinson, S., & Gillis, A. S. (2023, 17 noviembre). 5V’s of big data. Search Data Management. <https://www.techtarget.com/searchdatamanagement/definition/5-Vs-of-big-data>  Gyata Ai (2024, 19 marzo). Roles de Ciencia de Datos. <https://www.gyata.ai/es/data-science/data-science-roles> | | |

1. *Menciona los perfiles de ciencia de datos que debe contratar la empresa para solucionar el problema. Argumenta el porqué de la contratación de cada perfil.*

* Científico de Datos: Responsable de analizar grandes volúmenes de datos, aplicar modelos predictivos y generar conclusiones clave para la toma de decisiones.
* Ingeniero de Datos: Encargado del diseño, implementación y mantenimiento de la infraestructura de almacenamiento y procesamiento de datos, garantizando escalabilidad y eficiencia.
* Analista de Datos: Especialista en la interpretación de datos mediante herramientas de visualización, proporcionando información clara y útil a los equipos de negocio.
* Arquitecto de Datos: Diseña soluciones de almacenamiento y procesamiento optimizadas para soportar grandes volúmenes de datos de manera segura y eficiente.

1. *Como sabes, los proyectos de big data deben cumplir con las cinco “V”. Tu tarea en este punto es justificar cómo se relaciona cada característica con el caso planteado. Por ejemplo: Velocidad: Los datos de las ventas se generan en tiempo real, lo que requiere un procesamiento y análisis rápido para tomar decisiones oportunas.*

* **Volumen:** La tienda genera grandes cantidades de datos provenientes de ventas, clientes y productos, lo que requiere una infraestructura escalable.
* **Velocidad:** Las ventas se registran en tiempo real, lo que demanda procesamiento ágil para generar reportes y análisis oportunos.
* **Variedad:** Se manejan diferentes tipos de datos, desde estructurados (ventas, clientes) hasta no estructurados (reseñas, interacciones en redes sociales).
* **Veracidad:** La calidad y limpieza de los datos es crucial para evitar errores en la toma de decisiones.
* **Valor:** La correcta interpretación de los datos permitirá mejorar la experiencia del cliente y aumentar las ventas.

1. *Realiza un análisis y define qué tipo de arquitectura/arquitecturas de almacenamiento de datos es óptima o adecuada para la empresa. Deberás justificar el porqué de la selección de dichas arquitecturas de almacenamiento, además de plantear y sustentar qué tipo de base de datos NoSQL es la más factible de usar en la empresa.*

Para DeportivaMX, se recomienda una arquitectura híbrida que combine almacenamiento en la nube con bases de datos NoSQL y SQL (A continuacion se recomiendan los casos de uso exactos):

* Data Lake en la Nube: Permite almacenar grandes volúmenes de datos en formatos estructurados y no estructurados, asegurando escalabilidad y bajo costo.
* Base de Datos Relacional (SQL - PostgreSQL): Para gestionar datos transaccionales de clientes y ventas con integridad referencial.
* Base de Datos NoSQL (MongoDB): Ideal para almacenar datos no estructurados como opiniones de clientes, historial de navegación y preferencias de compra.

1. *Genera el análisis y plasma las estructuras de las posibles colecciones en formato JSON que deberá tener la base de datos NoSQL para el almacenamiento de datos no estructurados en MongoDB a través de su interfaz gráfica.*

**Colección: clientes**

{

"\_id": "ObjectId",

"nombre": "Juan Pérez",

"correo": "juanperez@email.com",

"preferencias": ["fútbol", "tenis", "ciclismo"]

}

**Colección: productos**

{

"\_id": "ObjectId",

"nombre": "Balón de fútbol",

"categoría": "Fútbol",

"precio": 500.00,

"stock": 150,

"reseñas": [

{

"cliente\_id": "ObjectId",

"comentario": "Excelente calidad",

"puntuación": 5

}

]

}

**Colección: ventas**

{

"\_id": "ObjectId",

"cliente\_id": "ObjectId",

"productos": [

{ "producto\_id": "ObjectId", "cantidad": 2, "precio\_unitario": 500.00 }

],

"total": 1000.00,

"fecha": "2025-03-20"

}

1. Conclucion

La implementación de esta arquitectura y estrategia de almacenamiento permitirá a DeportivaMX manejar su crecimiento de manera eficiente, mejorando la toma de decisiones y la experiencia del cliente a través del análisis de datos en tiempo real. Además, la optimización en la gestión de datos permitirá a la empresa adaptarse rápidamente a cambios en las tendencias del mercado, asegurando una ventaja competitiva en el sector.

El uso de bases de datos escalables y herramientas avanzadas de análisis facilitará la identificación de patrones de compra y preferencias del consumidor, lo que permitirá personalizar la oferta de productos y mejorar la experiencia del usuario en la plataforma.

La seguridad de los datos será un aspecto clave en la arquitectura implementada, asegurando la protección de la información sensible de los clientes y la confidencialidad de los registros de ventas.