**Sistema de Big Data para CompraGamer**

Objetivo general: Implementar un sistema de Big Data que permita a “CompraGamer” tomar decisiones informadas mediante el análisis de datos de comportamiento de clientes, optimización de inventarios y personalización de ofertas a través de machine learning.

1. Análisis de Comportamiento del Cliente

Objetivo Específico: Recopilar y analizar datos de comportamiento de usuarios en el sitio web y la app en tiempo real para identificar patrones de compra y optimizar la experiencia del usuario.

Tecnologías Propuestas:

* **Gestión del Flujo de Datos**: Se utilizará Apache Kafka para la captura y transmisión de eventos de usuario en tiempo real. Esto permitirá recopilar datos de clics, visitas a productos, y tiempo de permanencia en diferentes secciones de la web.
* **Almacenamiento**: Los datos recopilados se almacenarán en un Data Lake utilizando Amazon S3 para asegurar la capacidad de escalar con los volúmenes de datos. Esto permitirá almacenar datos de usuario en su formato nativo para futuros análisis.
* **Procesamiento de Datos**: Apache Spark se utilizará para el procesamiento y análisis de estos datos, usando Spark Streaming para el análisis en tiempo real de la información de comportamiento.
* **Visualización**: Se utilizarán Tableau y Power BI para crear dashboards que visualicen patrones de uso, segmentos de usuarios, y áreas de mejora en la navegación web.

1. Gestión de Inventarios

Objetivo Específico: Integrar datos de ventas y niveles de inventario para prever la demanda y optimizar el stock, evitando faltantes y sobreabastecimiento.

Tecnologías Propuestas:

* **Almacenamiento**: Los datos históricos de ventas y niveles de inventario se almacenarán en un Data Warehouse utilizando Amazon Redshift o Snowflake para un acceso rápido y consultas optimizadas.
* **Procesamiento y Análisis:** Con Apache Spark y Apache Hive, se realizarán consultas y análisis de datos históricos, complementando con Spark MLlib para desarrollar modelos predictivos de demanda que utilicen técnicas de series temporales.
* **Visualización:** Power BI y Looker serán utilizados para crear dashboards que muestren el inventario en tiempo real, los niveles de stock y las predicciones de demanda, ayudando en la toma de decisiones sobre reposición de inventario y ajustes de stock.

1. Personalización de Ofertas

Objetivo Específico: Utilizar algoritmos de machine learning para generar recomendaciones personalizadas para los usuarios según su historial de navegación y compra.

Tecnologías Propuestas:

* **Almacenamiento**: Se creará un Data Mart en MongoDB para almacenar perfiles de usuario y sus interacciones, optimizando el acceso a la información que los algoritmos de recomendación requieren.
* **Machine Learning**: Se usarán Apache Spark MLlib y TensorFlow Recommenders para entrenar modelos de filtrado colaborativo y sistemas de recomendación basados en contenido, utilizando tanto el historial de compras como la interacción en la web.
* **Integración de Recomendaciones**: Las recomendaciones se servirán mediante una API construida con FastAPI, integrada en el sitio web y la app móvil, lo que permitirá a los usuarios ver recomendaciones personalizadas en tiempo real.
* **Deployment**: Amazon SageMaker se puede utilizar para el desarrollo y despliegue de los modelos de machine learning en la nube, facilitando la integración continua de nuevos datos y el ajuste de los modelos.

1. Hardware Recomendado

Para garantizar el rendimiento de la infraestructura de Big Data, se propone:

* **Servidores de Almacenamiento:** Utilizar servidores como Dell EMC PowerEdge para almacenamiento local si se decide trabajar con una infraestructura on-premise, o aprovechar el escalamiento y flexibilidad de Amazon S3 en la nube.
* **Servidores de Procesamiento**: NVIDIA DGX Systems para tareas intensivas de machine learning y procesamiento paralelo en local, o bien instancias de alto rendimiento en la nube como AWS EC2 p4 para cargas de trabajo en la nube.
* **Conectividad de Red**: Se recomienda el uso de Cisco Nexus 9000 Series Switches y adaptadores de red de baja latencia Mellanox ConnectX para garantizar la transferencia de datos rápida y eficiente entre los diferentes nodos de procesamiento y almacenamiento.
* **Escalabilidad**: La nube proporciona flexibilidad, por lo que se pueden usar instancias escalables en AWS, Google Cloud, o Azure para ajustar recursos en función de la demanda de procesamiento.

1. Planificación del Proyecto

Organización del Equipo y Estructura de Trabajo:

**“Estructura horizontal para fomentar la colaboración”**

* Responsable 1: Análisis de Comportamiento del Cliente.
* Responsable 2: Gestión de Inventarios.
* Responsable 3: Personalización de Ofertas.
* Responsable 4: Infraestructura y redes.

**Tareas y Responsabilidades**:

* Configuración de Apache Kafka y Spark Streaming.
* Integración del Data Warehouse y desarrollo de modelos predictivos.
* Desarrollo de modelos de recomendación y API de personalización.
* Configuración de servidores y optimización de la red.
* Reuniones de Seguimiento: Cada dos semanas para revisar avances y ajustar tareas.
* Pruebas Iniciales: Al cabo de un mes para asegurar que los flujos de datos y los análisis se están realizando correctamente.

Fecha Límite de Entrega: 3 meses desde el inicio del proyecto.

**Resumen Final**

Con esta propuesta, CompraGamer obtendrá un sistema de Big Data robusto que permitirá:

Mejorar la experiencia de sus clientes a través de recomendaciones personalizadas.

Optimizar la gestión de inventarios, reduciendo el riesgo de faltantes o exceso de stock.

Tomar decisiones basadas en datos, lo que contribuirá a una mayor eficiencia operativa y un incremento en las ventas.