



SOBRE CÓMO SE MANEJAN LOS GRANDES VOLÚMENES DE DATOS

Actividad de transferencia

Suponga que Ud. trabaja en una organización que se llama “Smart Contact”. Esta organización es una empresa dedicada a las telecomunicaciones e internet que cuenta con más de 500.000 clientes en la actualidad.

Los productos que comercializa esta compañía son:

- líneas de teléfonos (abonados) fijos y móviles (celulares)
- internet wifi
- cable de TV

La empresa ha notado que sus ventas no están siendo óptimas y ha intentado con ofertas y promociones, pero aun así no logra superar los objetivos de negocio (los cuales fueron definidos muy claramente), como por ejemplo, aumentar la rentabilidad en un 30% durante los próximos 12 meses, sin reducir la estructura del personal. Lo cierto es que esta empresa dispone de toda la tecnología necesaria para brindar una excelente atención, y todos los clientes y posibles clientes (leads) se encuentran en un sistema CRM (la información de todos ellos es muy completa).

Para la gestión diaria de incidentes, se tiene integrado al CRM un sistema de Mesa de ayuda. Si bien los tiempos de resolución de problemas son variados, en general las encuestas de satisfacción son buenas.

El personal que trabaja en la empresa está acorde a lo esperado. Todos trabajan muy bien y en general los objetivos se cumplen, excepto el de las ventas, que no logra llegar a la meta propuesta: por algún motivo no se está pudiendo vender lo que los clientes desean, demandan o necesitan.

Usted fue contratado como Analista de Negocio para pensar, modelar e idear alguna solución que utiliza Big Data de manera que pueda usar los datos a fin de tomar la mejor decisión de negocio posible, en cuanto a qué vender, a quién venderle y a cuánto venderle.

El dueño de la empresa escuchó que la moda actual es tener Big Data y desea comprar una solución de USD 1.500.000 que acaban de ofrecerle.

Se pide que haga lo siguiente:

1. Enumere al menos 5 pasos que inicialmente realizaría si fuera el responsable de Big Data dentro de la organización.
 - 1.1. Análisis de requerimientos y objetivos de negocio: realizar un acercamiento a los principales stakeholders para comprender los



objetivos de negocio y los desafíos específicos que enfrenta la empresa en términos de ventas. Además de identificar qué datos son críticos para alcanzar estos objetivos.

- 1.2. Evaluar la infraestructura existente: Realizar un inventario de los sistemas y bases de datos existentes para determinar la disponibilidad y calidad de los datos. Evaluar la capacidad de TI actual para manejar grandes volúmenes de datos.
- 1.3. Diseño de la arquitectura: Desarrollar una arquitectura de Big Data que se alinee con los objetivos de negocio y aproveche al máximo los recursos disponibles. Identificar las herramientas y tecnologías necesarias para la captura, almacenamiento, procesamiento y análisis de datos.
- 1.4. Selección de la solución de Big Data: Evaluar distintas soluciones y escoger la que mejor se adapte a las necesidades y presupuesto; así como, considerar factores como escalabilidad, facilidad de uso y soporte para fuentes de datos variadas.
- 1.5. Planificación de implementación y capacitación: Desarrollar un plan detallado de implementación que incluya actividades como la migración de datos, configuración de la infraestructura y pruebas de rendimiento. Capacitar al personal en el uso de la nueva solución y en la interpretación de los análisis generados.
2. Arme un diagrama de arquitectura en formato Word o Excel (tomando como modelo el de referencia aportado en el libro) e indique cuáles serían las fuentes de información interna, y si utilizaría alguna fuente externa, dónde alojaría la información.
 - 2.1. Componentes:
 - 2.1.1. Fuentes de Información internas:
 - 2.1.1.1. CRM de Smart contact.
 - 2.1.1.2. Mesa de ayuda integrada al CRM
 - 2.1.1.3. Bases de datos de ventas, facturación y servicios contratados.
 - 2.1.2. Fuentes externas:
 - 2.1.2.1. Datos demográficos y de comportamiento del mercado de telecomunicaciones.
 - 2.1.2.2. Comentarios, reacciones e impresiones en redes sociales.
 - 2.1.3. Hadoop Cluster:
 - 2.1.3.1. HDFS
 - 2.1.3.2. MapReduce
 - 2.1.3.3. Hive
 - 2.1.3.4. Hadoop YARN
 - 2.2. Conexiones:
 - 2.2.1. Las fuentes de información internas y externas están conectadas al Hadoop Cluster para ingestión de datos.



- 2.2.2. El Hadoop Cluster utiliza HDFS para el almacenamiento distribuido de datos.
- 2.2.3. MapReduce se utiliza para el procesamiento de datos en lotes.
- 2.2.4. Hive permite realizar consultas SQL-like sobre los datos almacenados en el Hadoop Cluster.
- 2.2.5. Hadoop YARN se encarga de la gestión de recursos y la programación de tareas en el clúster.
- 2.3. Alojamiento de la información:
 - 2.3.1. Los datos se almacenan en el Cluster utilizando HDFS.
 - 2.3.2. Los datos externos pueden alojarse en un sistema de almacenamiento en la nube, como Amazon, google cloud, etc. antes de ingresar en el Cluster de Hadoop.
 - 2.3.3. Hive se utiliza para consultas SQL-like sobre los datos almacenados en el Cluster de Hadoop.
- 3. Mencione qué componentes de Hadoop utilizaría para una prueba conceptual, siendo que el modelo de procesamiento a implementar sería Batch.
 - 3.1. HDFS (Hadoop Distributed File System): Para el almacenamiento distribuido de datos.
 - 3.2. MapReduce: para procesar grandes volúmenes de datos en lotes y realizar análisis complejos.
 - 3.3. Hive: Para consultas SQL-like sobre los datos almacenados.
 - 3.4. Hadoop YARN: para la gestión de recursos y la programación de tareas en el clúster de Hadoop.
 - 3.5. Spark: para el procesamiento de datos en memoria y análisis en tiempo real, en caso de que se requiera.