

Nombre: Pablo Elías Ramírez Escalante		Matrícula: AL02883894
Infraestructura para Big Data	Nombre del profesor: Miguel de Jesús Martínez Felipe	
Módulo 1	Avance 1 de Evidencia 1	
Fecha: 19/02/2024		
Bibliografía: ¿Qué es el big data? (s. f.). https://www.oracle.com/mx/big-data/what-is-big-data/#:~:text=El%20big%20data%20le%20permite%20recopilar%20datos%20de%20redes%20sociales,como%20maximizar%20el%20valor%20ofrecido. Bantu Group. (s. f.). Las 5 mejores prácticas en gestión de datos que debes conocer. https://www.bantugroup.com/blog/mejores-practicas-en-gestion-datos		

Lee con atención plena el siguiente caso:

La empresa Bratt & Houston pertenece al giro aeroespacial en los Estados Unidos, teniendo operaciones a lo largo de los cinco continentes, siendo un referente a nivel mundial en el diseño, servicio y fabricación de motores aéreos. Su sede principal se localiza en Montgomery, Alabama (EE. UU). La empresa cuenta con más de 20,000 colaboradores para dar servicio a más 5,000 clientes localizados en 100 países.

Vanessa, quien es egresada, desde hace un año, de la carrera de Ingeniería en Computación Administrativa por parte de la Universidad Tecmilenio, no tuvo problemas para colocarse profesionalmente en la empresa Innova Analytics, dedicada al desarrollo de soluciones basadas en datos. Gracias a su desempeño y el desarrollo de sus competencias profesionales, ha logrado una plaza en la empresa Bratt & Houston, para trabajar como analista de datos en su sede central.

Uno de sus primeros proyectos asignados, consiste en implementar un ecosistema Hadoop utilizando la nube de Microsoft Azure, relacionado con el servicio de verificación del estado de las turbinas aéreas fabricadas en su sede central, mediante este servicio se logran recopilar una gran cantidad de datos de diferentes fuentes, con un énfasis especial en las rutas aéreas comerciales de las aeronaves que utilizan sus turbinas, además de conocer si estas rutas son cumplidas en tiempo y forma.

La importancia del proyecto radica, que, al día de hoy, miles de empresas están migrando sus infraestructuras a la nube, lo cual les permite, entre otras cosas, disponer de una mayor flexibilidad, agilizar sus procesos y reducir costos. Aunado a esto, la masiva generación de datos y su procesamiento vuelve primordial adquirir conocimientos en el uso, configuración y mantenimiento de estas herramientas.

Como resultado, se espera hacer una presentación con la implementación de la infraestructura para el procesamiento de datos masivos en la nube.

Para llevar a cabo este proyecto, imagina que tomarás el rol de Vanessa y deberás realizar cada una de las siguientes indicaciones, toma en cuenta que esta evidencia la entregarás en tres partes diferentes (avances).

Avance 1

1. En este primer avance, como paso previo al desarrollo del proyecto, se realizará una reflexión de arranque (con una extensión de al menos dos cuartillas), en la cual se estará dando respuesta a cada una de las siguientes preguntas:
 - ¿Cuál es la importancia de la infraestructura en proyectos de Big Data?
 - ¿Cuáles son las claves para construir una infraestructura adecuada, sólida y efectiva?
 - ¿Cuáles son las mejores prácticas en la creación de una infraestructura de Big Data?
 - ¿Qué habilidades iniciales son las necesarias para instalar, configurar y poner en marcha una infraestructura hardware y software para Big Data?
 - ¿Qué puntos débiles existen al momento de construir una infraestructura, además de propuestas para lograr resolverlos?

Reflexión de inicio sobre la infraestructura necesaria para los proyectos de Big Data.

En la revolución digital actual, la gestión y el análisis eficiente de grandes cantidades de datos se han vuelto esenciales para las empresas que buscan obtener información útil para la toma de decisiones. En este contexto, la infraestructura es fundamental para los proyectos de Big Data porque determina en gran medida el éxito y la eficacia de las iniciativas de procesamiento y análisis masivo de datos. En este primer paso, hablaremos sobre la importancia de esta infraestructura, las claves para su construcción, las mejores prácticas, las habilidades y los puntos débiles y soluciones.

La importancia de la infraestructura para los proyectos de Big Data es crucial:

Debido a la gran cantidad y complejidad de los datos generados en la actualidad, es necesaria una infraestructura robusta y escalable. La infraestructura es esencial para garantizar la recopilación, almacenamiento y procesamiento eficiente de datos en proyectos como el propuesto para Bratt & Houston, que busca implementar un ecosistema Hadoop en la nube de Microsoft Azure. Las organizaciones pueden obtener información valiosa, identificar patrones y tendencias y mejorar la toma de decisiones estratégicas al gestionar grandes conjuntos de datos.

Claves para Construir una Infraestructura Adecuada, Sólida y Efectiva:

1. **Escalabilidad:** La infraestructura debe tener la capacidad de crecer y adaptarse a la demanda de datos cambiante. En el caso de Bratt y Houston, la capacidad de expandir el ecosistema Hadoop en Azure permitirá manejar de manera eficiente grandes cantidades de datos.
2. **Flexibilidad:** La infraestructura debe ser capaz de adaptarse a una variedad de fuentes de datos y tecnologías. La flexibilidad es crucial en el proyecto de Bratt & Houston, que incluye datos desde rutas aéreas hasta el estado de las turbinas.
3. **Seguridad:** La seguridad de los datos debe ser una prioridad porque los datos son delicados. La protección y la privacidad de la información es crucial en el sector aeroespacial.
4. **Optimización de Recursos:** Para reducir costos y mejorar el rendimiento, se debe maximizar la eficiencia del uso de recursos hardware y software. En este sentido, elegir la nube correcta (Azure en este caso) es fundamental.

Mejores Prácticas en la Creación de una Infraestructura de Big Data:

1. **Entendimiento de Requerimientos:** Para diseñar una infraestructura que cumpla con los objetivos, es fundamental comprender los requisitos específicos del proyecto y las necesidades del negocio.
2. **Implementación gradual:** la implementación gradual permite una mejor gestión y reducción de riesgos. Iniciar con la implementación de Hadoop en la nube en el proyecto de Bratt & Houston es un enfoque sensato.

3. **Monitoreo Continuo:** Implemente sistemas de monitoreo continuo para evaluar el desempeño y hacer ajustes según sea necesario.

Habilidades Iniciales Necesarias:

1. **Conocimientos en Cloud Computing:** Es esencial comprender las tecnologías en la nube, especialmente si los proyectos como el de Bratt & Houston requieren la implementación de Azure.
2. **Programación y Configuración:** La instalación y puesta en marcha de herramientas Big Data como Hadoop requiere habilidades en programación y configuración.
3. **Seguridad Informática:** conocimientos profundos sobre cómo proteger la integridad de los datos.

Puntos Débiles y Propuestas de Solución:

1. **Complejidad de Integración:** La diversidad de fuentes de datos puede hacer que la integración sea más complicada. Utilizar un enfoque modular y herramientas de integración eficientes es la propuesta.
2. **Gastos:** La infraestructura en la nube puede costar mucho. Propuesta: establecer reglas de uso y optimización de recursos para controlar los gastos.
3. **Seguridad de Datos:** La seguridad de los datos es una gran preocupación. Propuesta: establecer estrictos protocolos de seguridad y realizar auditorías regulares.

En conclusión, la infraestructura de los proyectos de Big Data es un componente clave que determina el éxito y la eficacia de las iniciativas. La optimización de recursos, la escalabilidad, la flexibilidad, la seguridad y otros aspectos de la infraestructura requieren un enfoque integral. La programación, la seguridad y las habilidades de computación en nube son esenciales, y la gestión de puntos débiles requiere un enfoque proactivo y estratégico. Se detallarán los pasos concretos que Vanessa tomó para implementar el proyecto de Bratt & Houston en los siguientes avances.