|  |  |  |
| --- | --- | --- |
| **Nombre**: José Guillermo Robles Sánchez | | **Matrícula**: 2905526 |
| **Nombre del curso:**  Infraestructura de big data | **Nombre del profesor**:  Miguel de Jesús Martinez Felipe | |
| **Módulo**:  2 | **Actividad**:  Avance 2 Evidencia 1 | |
| **Fecha**: 2023 | | |
| **Bibliografía**: | | |

Data lakes

Los data lakes son multiformatos de estructuras de almacenamiento que pueden almacenar distintos datos de formatos y fuentes ofreciendo bastantes ventajas como:

* Mayor flexibilidad y escalabilidad: los data lakes pueden almacenar grandes cantidades de todo tipo de datos sin importar lo cual ofrece mejor flexibilidad a todas las empresas permitiendo adaptarse a la empresa u organización que lo requiera
* Ayuda a tomar mejores decisiones: permiten que la empresa u organizacon que este haciendo uso de este data lake puedan analizar todo tipo de datos de distintas fuentes y formatos
* Reduce costos: estos permiten que se puedan reducir costos permitiendo centralizar el almacenamiento sin necesidad de invertir grandes cantidades de dinero

En hadoop todos los data lakes son ideales para el ecosistema de hadoop; esto debido a que ofrece mejor escalabilidad ya que hadoop es muy flexible y puede adaptarse a necesidades cambiantes; es muy bueno ante posibles fallos independiente si se llegan a presentar fallos tanto de hardware como de software y por ultimo permite procesar grandes cantidades de datos de manera eficiente, esto es muy importante ya que puede contener datos de distintos tipos

Dentro de las desventajas puede ser que es muy complejo ya que son un poco difíciles de gestionar ya que requieren mucha integración de información y datos; en cuanto a seguridad pueden ser un poco vulnerables ante ataques cibernéticos