|  |  |  |
| --- | --- | --- |
| **Nombre**: José Guillermo Robles Sánchez | | **Matrícula**: 2905526 |
| **Nombre del curso:**  Infraestructura de big data | **Nombre del profesor**:  Miguel de Jesús Martinez Felipe | |
| **Módulo**:  1 | **Actividad**:  ACTIVIDAD 2 | |
| **Fecha**: 30-08-2023 | | |
| **Bibliografía**: <https://blog.es.logicalis.com/analytics/biginsights-grandes-oportunidades>  <https://www.ibm.com/mx-es/products/storage-insights>  <https://aprenderbigdata.com/introduccion-cloudera-hadoop/>  <https://www.cloudera.com/products/hdp.html>  https://es.wikipedia.org/wiki/Hortonworks | | |

1. Investiga, en al menos tres fuentes confiables, tres diferentes distribuciones de Hadoop.
   * Incluye el mercado objetivo de cada una.
   * Incluye al menos dos ventajas y desventajas de cada una.
   * Señala aspectos sobresalientes de cada una, por ejemplo, características principales, antecedentes, costos, etc.
   * Incorpora una sección titulada “Referencias bibliográficas”, en la cual citarás correctamente las referencias utilizadas en el desarrollo de tu actividad en formato APA.
2. Posteriormente, con tu investigación concluida, realiza la grabación de un video con las siguientes características:
   * Su duración debe ser de al menos tres minutos y un máximo de cinco.
   * Debe incluir cada uno de los elementos solicitados en la investigación.
3. Sube tu video a YouTube.
4. Prepara un documento electrónico en el que incluyas la liga de tu video, además de las referencias bibliográficas utilizadas en el desarrollo de tu video.
5. Entrega el documento electrónico a tu profesor.

HADOOP

Hadoop es una estructura de software de código abierto para almacenar datos y ejecutar aplicaciones en clústeres de hardware comercial, tiene una gran cantidad de procesamiento y mucha capacidad de procesar tareas o trabajos concurrentes ilimitados

1. Cloudera

El mercado objetivo de esta es hacer que los datos analíticas sean fáciles y accesibles para todos permitiendo mejorar el acceso a las habilidades y software impulsando la sustentabilidad

Ventajas de cloudera

* Se puede aplicar analítica avanzada sobre datos ilimitados

Desventajas

* Su latencia va disminuyendo con El Paso del tiempo en comparación a otras
* Dependiendo el plan que se elija puede llegar a ser muy costoso

¿Qué es cloudera?

Es una firma especializada en big data, permitiendo añadir funciones a Hadoop ara mejor seguridad, control para poder dar una solución empresarial. Dentro de sus características Cloudera incluye Cloudera Distribution Hadoop (CDH) y licencia por un año a Cloudera Manager y soporte técnico dentro de esto existen 3 modos: Basic, Flex y DataHub

1. Hortonworks

El objetivo de esta es poder crear proveedor de plataformas de datos de próxima generación y logre establecer el estándar de la industria para la gestión datos híbridos en la nube.

Ventajas

* Ya cuenta con muchos socios reconocidos incluyendo proveedores de DBMS
* Con la plataforma Open Data es compatible con el crecimiento de nuevos proyectos de la fundación apache

Desventajas

* Tal cual no creo que presente alguna desventaja ya que puede servir como expansión o como ayuda para migrar datos en caso de que ya contemos con uno

¿Qué es hortonworks?

Es un marco de trabajo que es de código abierto ara poder almacenar y que pueda procesar conjuntos grandes de datos y en varias fuentes, Hortonworks tiene modernizada su infraestructura y mantener todos los datos seguros, ayuda a acortar tiempos de implementación reducir costos, también acelera tiempo de análisis para decisiones mas acertadas y se puede tener control y acceso metadatos para una seguridad de mayor nivel, por ultimo hortonworks se. Puede asociar con compañías de software para administración de servicios empresariales y automatización

1. IBM BigInsights

Permite alcanzar un alto rendimiento en el trabajo con grandes conjuntos de datos al producir sus resultados a través de un sistema de procesamiento por lotes, altamente escalable y distribuido

Ventajas

* Mejora la panificación de la capacidad ayudando a eliminar las conjeturas de la planificación de la capacidad dando mayor visibilidad sore las tasas de crecimiento de datos y capacidad disponible
* Ayuda a resolver problemas complejos ya que da una visión mas completa y contextualizada de los datos que se pueden trabajar en condiciones de mayor agilidad

¿Qué es IBM Insights?

IBM da una solución completa de Hadoop, incluyendo spark, para escalar analítica de forma rápida y sencilla, combinando software de código abierto y capacidades de grado empresaria.

IBM puede ofrecer: proporcionando distintas capacidades para descubrir y analizar oportunidades empresariales; Combina diferentes tecnologías y características para apoyar la gestión de los datos desde el momento en que entra a la empresa y por ultimo puede ampliar el marco de código abierto de Hadoop con seguridad empresarial y disponibilidad de integración dando herramientas que simplifican la productividad de los desarrolladores.