

Actividad [#2] – [Complementación minería de datos] [Minería y analisis de datos]

Ingeniería en Desarrollo de Software

**Tutor: Elizabeth Díaz**

**Alumno: Manuel Enrique Ramirez Lopez**

**Fecha: 21/02/2022**

***Indice***

*[Software de data mining](#_Toc23486)* [1](#_Toc23486)

*[Proceso del proyecto](#_Toc20739)* [4](#_Toc20739)

*[Bibliografía](#_Toc7028)* [6](#_Toc7028)

# *Software de data mining*

# IMG_256

***Teradata***

¿Por qué consideras el uso de este software para el desarrollo del proyecto?

Por su gran versatilidad de uso, es decir su multi plataforma en la nube que se unifica y crea un ecosistema de analisis empresarial.

¿Qué procesos de minería de datos te permite realizar este software?

-Redes neuronales

-Árbol de decision

¿Qué software utilizarías como alternativa? ¿Por qué?

Netezza gracias a su velocidad que ahorra mucho tiempo de espera para los resultado.

¿Qué gestor de base de datos crees que es la mejor opción?¿Por qué?

MySQL

Por que es de los software mas confiables para el almacenaje de datos y para tener el “ecosistema” compatible para su analisis desde casi cualquier dispositivo.



***Netezza***

¿Por qué consideras el uso de este software para el desarrollo del proyecto?

Por su velocidad de analisis de grandes cantidades de datos y que ahorra tiempo en configurarse ya que viene con pre-configurado por defecto.

¿Qué procesos de minería de datos te permite realizar este software?

Asymmetric Massively Parallel Processing

¿Qué software utilizarías como alternativa? ¿Por qué?

Teradata ya que su confiabilidad o seguridad de almacenaje de datos, resumido con una frase “lento pero seguro”.

¿Qué gestor de base de datos crees que es la mejor opción?¿Por qué?

SQLite ya que al ser un programa “ligero” proporciona una ventaja al momento de realizar consultas.



***DATAllegro***

¿Por qué consideras el uso de este software para el desarrollo del proyecto?

Por su alta compatibilidad con software de Microsoft

¿Qué procesos de minería de datos te permite realizar este software?

-Redes neuronales

-Árbol de decision

¿Qué software utilizarías como alternativa? ¿Por qué?

Netezza ya que es el competidor directo por que es un software de comunidad y tiene amplia gama de módulos para actualizar.

¿Qué gestor de base de datos crees que es la mejor opción?¿Por qué?

Microsoft access ya que gracias a la colaboración y que la misma pertenece a la compañía Microsoft trabajan en un buen ecosistema.

***Perfiles y roles***

Los roles que manejaremos son:

-Data Analyst (DA)

Encargado de ser pionero en el analisis de datos, su rol o funciones que implementara son la minería de datos, obtención y/o recuperación de datos así como su procesado, estudio avanzado y representación de la informacion.

-Data Engineer

Enfocado en ele almacenamiento y procesado de datos se encarga de extraer los datos de la big data y procesarlos y mostrarlo ya procesado listo para su analisis.

-Data Scientist

Su rol primordial es aplicar técnicas de representación de los datos y “dirigir” el rumbo con los resultado ya analizados y pre-veer acontecimientos futuros.

Todos estos roles se manejarian ya que desde mi punto de vista de lider estos son los mas importantes (todos los demas roles son importantes) y a su vez son los que necesita mi proyecto, adjunto a esto con su experiencia en los rubros que manejan desempeñaran un trabajo y manejo de situaciones adecuadas.

# *Proceso del proyecto*

***Costos de las distintas licencias que se propuso:***

**Teradata**

|  |  |
| --- | --- |
| Teradata Temporal | $ 300  por CCU |
| Teradata Data Mover | $ 160  por CCU |
| TPT Stream o Tpump | $ 240  por CCU |
| Laboratorio de datos Teradata | $ 1,750 por sistema |

**Netezza**

El costo es variable segun la magnitud de la base de datos y flujo de datos.

**DATAllegro**

El costo es variable segun la magnitud de la base de datos y flujo de datos.

− ¿Qué tipo de infraestructura vas a requerir para poder llevar a cabo el proyecto? ¿Qué tipo de equipos son mejores para este tipo de procesos y cuántos vas a utilizar? ¿Son necesarios servidores para almacenar la base de datos?

-Infraestructura tradicional

En la infraestructura tradicional, las empresas son las propietarias de todos los elementos (como los centros de datos, los sistemas de almacenamiento de datos, entre otros), a los cuales gestionan en sus propias instalaciones. El funcionamiento de esta infraestructura suele considerarse costoso y requiere grandes cantidades de sistemas de hardware (por ejemplo, servidores), así como energía eléctrica y espacio físico.

Los equipo mas adecuados son los de “gama media” es decir cada uno con su propio poder de computo para tareas complejas de analisis y extracción de datos enfocado en la velocidad de transferencia y extracción de datos.

-Se utilizaran alrededor de 4 equipos uno por cada rama o rol dentro del proyecto y otro para exclusivo para gestionar la base de datos.

Los servidores para almacenar los datos va a ser de suma importancia dentro del proyecto tanto así que para ahorrar a futuro se planea usar NAS para los diferentes proyecto que se planee usar a su vez este es altamente actualiza-ble para la situación que requiera.

− ¿Cuánto cobra el personal que decidiste contratar para el desarrollo de este proyecto? Recuerda que debes pagar por cada persona que consideraste.

-Data Analyst (DA) $270,000 anual $22,500 mensualmente $138 por hora.

-Data Engineer $282,000 anual, $23,500 mensualmente, $145 por hora.

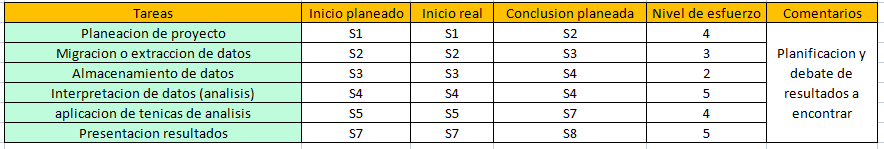
-Data Scientist $372,480 anual, $31,040 mensualmente, $192 por hora

− ¿Cuánto cuesta mantener los equipos? (el costo es por equipo que hayas decidido poner). ¿Cuánto cuesta mantener un servidor?

Alrededor si se usa 24/7 todos los equipos gasta alrededor de $12 al día por los cuatro equipos y otros $6 diarios por el almacenamiento es decir el costo de mantener el equipo durante un mes es de alrededor de *$950 mensualmente* por parte de la luz entre el equipo de computo y servidores junto a ello con la luz del establecimiento y ventilación del mismo.

− Con base en los costos de lo anterior, ¿cuánto tiempo va a tardar el desarrollo del proyecto? ¿Cuánto va a costar el desarrollo del proyecto en general?

Con la infraestructura ya generada, es decir el establecimiento ya funcionando junto con la adecuada configuración, armado y generada la red a usar se presenta el tiempo neto del proyecto el cual tiene una duración de 2 meses divididos por fases (semanas) para cada tarea correspondiente.



# *Bibliografía*

Enrique Lopez. (2021). DATAllegro. 2022, de Privada Sitio web: <https://en.wikipedia.org/wiki/DATAllegro>

Enrique Lopez. (2021). Teradata. 2022, de Privada Sitio web: <https://es.wikipedia.org/wiki/Teradata>

Enrique Lopez. (2021). Netezza . 2022, de Privada Sitio web: <https://en.wikipedia.org/wiki/Netezza>

Enrique Ramirez. (2021). ¿Cuáles son los roles clave dentro del universo Big Data?. 2022, de Privada Sitio web: <https://www.paradigmadigital.com/techbiz/cuales-son-los-roles-clave-dentro-del-universo-big-data/>