

Asignatura: OPC13 – Cloud Computing

Ensayo de resultados de aprendizaje de la **semana 9**

Temas: Transformación de los Datos, Modelos de Datos, Macrodatos, Análisis de Datos

Integrantes:

Diego Pérez Prieto
Matrícula: 365341
a365341@uach.mx

Jared Alejandro Rosas Molina
Matrícula: 365337
a365337@uach.mx

1. Ensayo Tema “Transformación de Datos”

Los datos dentro de una computadora, se guardan principalmente en bits, este es el bloque de construcción de los datos digitales usando unos y ceros, que, con ocho bits se construye un Byte, con el que se representa todas las letras, números y otros caracteres, los cuales si se juntan son almacenados en diferentes campos siendo organizados en registros, que representan una entidad separada a una población observada. Un cumulo de registros de datos es considerado como una base de datos, la cual, si es demasiado grandes la cantidad de datos y estos tienen distintos orígenes de donde fueron sacados se les llama “almacenes de datos” (Data Warehouse), siendo estas últimas la base principal del Big Data.

Este tipo de datos pueden ser guardados en distintos lugares, donde muchas empresas optan por guardarlos en la nube ya que implica múltiples beneficios, siendo uno de ellos la posibilidad de almacenar y recuperar varios y múltiples datos desde casi cualquier parte con acceso a internet, otro beneficio importante de esto, es que tiene una elasticidad en los datos, permitiendo que mientras los datos aumentan en cantidad también aumenta la potencia informática y la capacidad de almacenamiento, junto con estos beneficios también tiene la posibilidad de transformar los datos, ya que si se extraen de distintas fuentes se deben de transformar en distintos atributos para crear un registro con la información extraída.

El Big Data, es una gran recopilación de distintos tipos de datos como lo es texto, imágenes, videos, entre otros, por lo que provienen de fuentes completamente diferentes, y tiene características de que permite un gran volumen de datos, bastante velocidad de procesamiento y una gran variedad de los datos. El volumen de las bases de datos dentro del big data son de distintos tamaños, donde las más pequeñas se encuentran en un rango de megabytes a gigabytes mientras que las más grandes tienen un rango de tera a petabytes, la variedad de la información es bastante debido a que todo aquello que se puede digitalizar

es posible almacenarlo, por lo que el tipo de dato siempre cambia o crece, donde la recuperación de esta información es crucial para los usuarios, por lo que se tiene que recopilar, almacenar, procesar y analizar dicha información en el menor tiempo posible.

2. Ensayo Tema “Modelos de datos”

Los datos son esenciales para la toma de mejores decisiones, el problema llega cuando para existen tantos datos que puede llegar a ser abrumador al principio decidir qué sí usar y qué no usar. Para evitar este problema lo que se tiene que hacer es filtrar todos los datos no necesarios para obtener lo que necesitas y responder la problemática.

A veces observar un conjunto de datos sin ningún tipo de organización puede llegar a ser confuso para algunas personas, es por eso que para simplificar la tarea se puede utilizar herramientas de visualización de datos, como lo son tablas de datos, modelos informáticos y gráficas para ayudar a comprender mejor con qué se está trabajando.

Como era de esperarse y a sorpresa de nadie, la nube es muy útil para procesar conjuntos de datos enormes, en el sector empresarial se utiliza para hacer que las empresas sean más eficaces, además, los datos en la red no paran de crecer a pasos agigantados, por lo que es necesario mucho cómputo para poder procesarlos, y con la nube los programas creados pueden crecer proporcionalmente con la cantidad de datos que se recopilan.

En resumidas cuentas, necesitas saber qué buscas y qué preguntas tratas de responder para encontrar los datos correctos, y no estamos limitados a mostrar los datos en forma de tablas, también se pueden representar como mapas, simulaciones y gráficos

3. Ensayo Tema “Macrodatos”

Los datos son muy útiles cuando se usan para resolver algún problema, pero para lograr lo anterior necesitamos saber cómo usar los datos, para hacer esto existen varias técnicas o decisiones que forman parte de un proceso llamado “ciclo de vida de la gestión de datos”, el cual consiste en recopilar y almacenar los datos, analizarlos, visualizarlos y por último generar más datos.

Una vez que hayamos recopilado los suficientes datos de distintas fuentes, tendríamos lo que se conoce como un conjunto de datos y ahora toca observar la calidad de los datos para saber si es conveniente trabajar con ellos o se debería buscar nuevos datos, ya que, si se trabaja desde un principio con datos de poca calidad, los resultados tendrán esa misma calidad. Además, se tiene que transformar los datos para que estos se adapten a las necesidades del problema, ya que puede haber datos que se repitan o que no sirvan para el contexto del problema.

Todo el proceso anterior se le llama modelado de datos, que se refiere a la forma en la que los elementos de datos se terminan organizando. Lo que se espera de la salida de datos del modelo es el resultado de la recopilación de los distintos tipos de datos que se recopilaron de múltiples fuentes. Una vez que se tiene todo el proceso organizado, se muestran los datos para tener una mejor idea de con qué se está trabajando.

Se pueden mantener los datos en la nube a través de un proveedor de servicios en la nube, para almacenar y administrar los datos de manera eficiente. Los entornos que trabajen con una cantidad grande de datos, como Big Data, deben de proteger su seguridad, además ofrecer la flexibilidad que permita que su volumen aumente y permitir a los usuarios acceder a los datos y procesarlos

4. Ensayo Tema “Análisis de Datos”

El Big Data es un gran desafío para la administración de datos, gracias a su velocidad, su gran volumen y variedad de datos, no existen soluciones posibles con las bases de datos tradicionales. Gracias a la gran variedad de los datos existentes dentro del Big Data, muchos de estos pueden ser completamente útiles proporcionando distinta información para la solución de un problema, o simplemente no logran serlo debido al enfoque que se le esta dando al problema a solucionar, por lo que se debe de conocer bien dicho ámbito del problema para poder dar tener una buena elección de la información que se escogera, por ejemplo, dentro de las clases a los profesores les sirve conocer si los alumnos llegan tarde o no a las clases, y a cuales sí o cuales no.

Para un análisis de los datos, es complicado o casi imposible hacerlo de manera manual, yendo dato por dato para tener un registro y poder encontrar un patrón o una tendencia dentro de estos, con lo que existen multiples plataformas que ayudan a solucionar esto permitiendo dar un análisis rápido y mas confiable para que se pueda tomar una decisión en base a estos, donde la información recopilada y analizada es variada, ya que por ejemplo, se recopila información de los usuarios de un servicio, como lo son sus gustos, opiniones, reseñas, entre otros, donde ésta información se almacena y se procesa para que cuando un nuevo usuario necesite datos de un servicio, se pueda tomar una recopilación de todos los usuarios para mostrarle información que se adapte a las necesidades del cliente nuevo.

Para dar una gran velocidad de búsqueda, existen distintos algoritmos, como lo es la función hash, que permite que una búsqueda no le de importancia a los elementos duplicados, otros mediante la clasificación datos mediante o algoritmos de búsqueda que muestran distinta información al usuario según lo que ingrese, lo cual se logra mediante que el mismo programa le de un nombre o categoria a cada pieza de dato. La programación dinámica permite que se den mejores búsquedas, debido a que permite que el programa tome decisiones en base a lo que el usuario coloca, otra forma de una mejor búsqueda es la coincidencia de la cadena de caracteres, coincidiendo elementos de una lista con los de otra lista.