

TAREA 3

Alumno: Murrieta Villegas Alfonso

Base de Datos

Resumen - Opinión

① Big Data

② Introducción a Ciencia de datos

► Big Data

Antes de mencionar la relación del Big Data con Hadoop y Oracle, es necesario mencionar qué es cada elemento y a qué están destinados.

Hadoop es un sistema open source diseñado por Google mediante Java con el único objetivo de almacenar y procesar grandes volúmenes de información.

A la par de Hadoop existen frameworks y herramientas dedicados a la funcionalidad y complemento de manejo de datos. → "HDFS" y "MapReduce"

Sabiendo esto y considerando que el término de "Big Data" hace referencia a todo el conjunto de técnicas de computo dedicadas a almacenar, procesar y gestionar grandes volúmenes de información.

Oracle DataBase a pesar de no ser una herramienta de software dedicada a Big Data, mediante "Big Data" Appliance es como con software tipo Hadoop o R Server puede trabajarse con este paradigma.

NOTA | HDFS: Tiene una forma particular de almacenar datos que es mediante "Name-Nodes" y "Data-Nodes".

| MapReduce: Es la parte encargada de procesar los archivos y puede escribirse en Java, Python o R.

Opinión: | Oracle al igual que otras empresas, saben que las tendencias y necesidades del mercado requieren cumplir este nuevo nicho de mercado y ya sea que saquen su propio software o que se alien con otras empresas es como podrán hacerlo.

Extra: Es tal la necesidad que Microsoft con Azure tiene dentro de todos sus recursos para Big Data a Hadoop e incluso Spark.

► Ciencia de Datos

Sin duda, una de las "carreras del siglo" pues no solo tomará una gran importancia sino una gran demanda.

Prácticamente, Ciencia de Datos estudia todos los aspectos más relevantes y características de los datos, desde su generación, almacenamiento, análisis hasta su tratamiento.

Por otro lado, no solamente esta área tiene relación con áreas o temas matemáticos, a su vez encontramos otros recursos necesarios como el desarrollo de software, visión e interpretación de modelos e incluso creatividad.

Por último, y como bien se menciona en el artículo, dentro de este gran campo laboral no solo encontramos a los científicos de datos sino a una variedad de integrantes como Ingenieros de Datos y Analistas de Datos que a pesar de compartir una formación relativamente similar, cada uno desempeñará una labor distinta.

Opinión: Ciencia de Datos es sin duda una de las ramas que tendrán mayor importancia en este siglo. Sin duda lenguajes como R y Python además de ser o tener curvas de aprendizaje fáciles, notamos que al ser ayudados por la comunidad open source poco a poco han crecido enormemente.

NOTA: A pesar de que el artículo menciona de la "escalabilidad" de Python respecto a R, yo creo que debería especificarse al momento de tratar datos particularmente NO estructurados.

Referencias

- 1) Recuperado el 11 de febrero de 2020, de www.cracktutorial.com/technetwork/es/articles/database-performance/big-data-oracle-hadoop-2313460-030.html.
- 2) Recuperado el 11 de febrero de 2020, de <https://dzone.com/articles/an-introduction-to-data-science>