



INSTITUTO POLITECNICO NACIONAL. "UPIICSA"

Nombre del Alumno:

- Izquierdo Espinoza Angélica Lizbeth
 - Maldonado Velázquez César Irvin.
 - Salinas López Rosa Abigail
- Sosa Hernández César Manuel
 - García Rivera Juan Pablo
- González Carranza Jhonatan
 - Hernández Cecilio Martha



Nombre del Profesor:

GUTIERREZ GONZALEZ DR.. ANGEL.

Unidad de aprendizaje:

"TECNOLOGÍAS DE INFORMACIÓN"

Temario:

Unidad temática 1.

Secuencia:

"3AM35"

1.3 FUENTES DE BIG DATA.

HISTORIA DEL BIG DATA.

El término "big data" se refiere a los datos que son tan grandes, rápidos o complejos que es difícil o imposible procesarlos con los métodos tradicionales. El acto de acceder y almacenar grandes cantidades de información para la analítica ha existido desde hace mucho tiempo. Pero el concepto de big data cobró impulso a principios de la década de 2000 cuando el analista de la industria, Doug Laney, articuló la definición actual de grandes datos como las tres V:



Volumen: Las organizaciones recopilan datos de diversas fuentes, como transacciones comerciales, dispositivos inteligentes (IO), equipo industrial, vídeos, medios sociales y más. En el pasado, su almacenamiento habría sido un problema - pero el almacenamiento más barato en plataformas como los data lakes y el Hadoop han aliviado la carga.

Velocidad: Con el crecimiento del Internet de las Cosas, los datos llegan a las empresas a una velocidad sin precedentes y deben ser manejados de manera oportuna. Las etiquetas RFID, los sensores y los medidores inteligentes están impulsando la necesidad de manejar estos torrentes de datos en tiempo casi real.

Variedad: Los datos se presentan en todo tipo de formatos: desde datos numéricos estructurados en bases de datos tradicionales hasta documentos de texto no estructurados, correos electrónicos, vídeos, audios, datos de teletipo y transacciones financieras.

Consideramos otras dos dimensiones cuando se trata del big data:

Variabilidad: Además de las crecientes velocidades y variedades de datos, los flujos de datos son impredecibles, cambian a menudo y varían mucho. Es un reto, pero las empresas necesitan saber cuándo algo está de moda en los medios sociales, y cómo gestionar los picos de carga de datos diarios, estacionales y desencadenados por eventos.

Veracidad: La veracidad se refiere a la calidad de los datos. Debido a que los datos provienen de tantas fuentes diferentes, es difícil vincular, comparar, limpiar y transformar los datos a través de los sistemas. Las empresas necesitan conectar y relacionar las relaciones, las jerarquías y los múltiples vínculos de datos. De lo contrario, sus datos pueden salirse de control rápidamente.

1. Volumen
2. Velocidad
3. Variedad
4. Veracidad de los datos
5. Viabilidad
6. Visualización de los datos
7. Valor de los datos

DEFINICIÓN.

Big Data o datos masivos es un término que describe a los conjuntos de gran volumen de datos, tanto estructurados como no estructurados, que inundan los negocios cada día. Pero no es la cantidad de datos lo que es importante.



Consiste en desarrollar mecanismos capaces de procesar y gestionar datos masivos que provienen de diversas fuentes y se emplea para encontrar patrones repetitivos, modelos predictivos o estadísticas más precisas dentro de esos millones de datos. En definitiva, el objetivo es procesar esos datos para

convertirlos en información capaz de ser interpretada por el ser humano y que le ayude a tomar decisiones.