

## **Big Data - Resumen**

Big data es el conjunto de datos extremadamente grandes y complejos que no pueden ser gestionados ni procesados mediante herramientas comunes de manejo de datos. Son también llamados Datos Masivos o Macro Datos.

Los Datos Masivos son caracterizados por tres v's, volumen, velocidad y variedad.

El volumen de datos quiere decir que los datos son enormes y pueden variar significativamente dependiendo de la aplicación y la organización. Los conjuntos de datos pueden llegar a terabytes, petabytes, exabytes e incluso zettabytes (un zettabyte equivale a un billón de gigabytes).

La velocidad se refiere a la velocidad a la que los datos son generados, capturados, procesados y analizados en tiempo real. A medida que la cantidad de datos generados aumenta exponencialmente, la capacidad de procesar y analizar estos datos en tiempo real se ha vuelto cada vez más importante.

Por último la variedad que se refiere a la diversidad de tipos y formatos de datos que se generan y se deben gestionar. A diferencia de los datos estructurados tradicionales, que suelen tener un formato predefinido y organizado en tablas, el big data incluye una amplia gama de tipos de datos no estructurados, semi-estructurados y estructurados. Los datos no estructurados son aquellos que no siguen un formato específico y no se pueden organizar fácilmente en una tabla o base de datos tradicional. Algunos ejemplos de los datos no estructurados son videos, fotos, archivos de audio, publicaciones de redes sociales, correos electrónicos. Los datos semi-estructurados son una combinación de datos estructurados y no estructurados. Estos datos tienen cierta organización, como etiquetas o metadatos, que permiten algún nivel de clasificación y búsqueda, por ejemplo archivos XML, JSON, archivos de registro y datos HTML. Los datos estructurados son datos organizados en tablas. Estos datos son más fáciles de procesar utilizando herramientas tradicionales, por ejemplo los datos CSV, JSON y hojas de cálculo.

Retos de los datos masivos - debido a la cantidad masiva de datos suelen haber datos incompletos o mal organizados, debe haber más seguridad ya que se puede infringir la privacidad de las personas o la información sensible de las empresas, finalmente los costos ya que por aun no tener la tecnología suficiente para procesar cantidades de datos tan grande, el costo es muy alto.

Análisis de los datos masivos - Existen 3 tipos de análisis en el Big Data, el predictivo que se usa con inteligencia artificial la cual predice acontecimientos futuros, el prescriptivo que recomienda posibles acciones en el cual se usa el aprendizaje automático para la toma de decisiones y por último el análisis descriptivo que ayuda a comprender acontecimientos pasados y de ahí detecta problemas o anomalías.

Herramientas y tecnologías - Hadoop, Spark, Cassandra y Map reduce

Big Data en el mundo actual - Ayuda para la sanidad ya que nos ayuda a hacer diagnósticos predictivos y patrones de enfermedad, el transporte ya que nos da transportes de tráfico, optimización de rutas y por último finanzas ya que nos da las tendencias del mercado.

Big Data en el futuro - Edge computing que permitirá analizar los datos más cerca de su fuente, blockchain que ofrece una manera segura de almacenar y compartir datos. Inteligencia Artificial que

logrará utilizar los datos, procesarlos y utilizarlos. Finalmente la computación cuántica que serán capaces de procesar grandes cantidades de información.