

Herramientas Visual Analytics

Bases de Datos para el Big Data

(Aprende a elegir el mejor repositorio NoSQL para tus datos). Son cuatro los tipos de bases de datos NoSQL: Documental, Columnar, Grafo y de Clave valor.

MongoDB

Es una base de datos documental muy utilizada en entornos Big Data. En el máster es la base de datos más tratada.



Cassandra

Es una base de datos columnar (en modo columna). Se suele usar en proyectos muy específicos donde el modelo de datos es apropiado abordarlo con consultas concretas (necesidades concretas del cliente).



Neo4J

Es una base de datos de tipo grafo, utilizada para mostrar relaciones entre los datos. Se utilizan mucho cuando las relaciones de los datos son útiles para extraer conocimiento de ellas. El fraude de tarjeta de crédito se suele analizar con este tipo de base de datos, así como los datos de redes sociales.



Redis

Es una base de datos clave-valor en memoria, muy rápida y simple. Su uso más difundido es como un sistema de cache que evita o ahorra conexiones o peticiones a otros recursos.



Amazon, Google, Microsoft, entre muchos otros, la utilizan para optimizar consultas y muchas funcionalidades más.

Ingeniería para el procesamiento masivo de datos

(Aprende a crear infraestructuras para el Big Data)

Apache Hadoop

Permite crear entornos Big Data con la tecnología y herramientas necesarias para su manipulación y explotación.



Spark

Es una herramienta para tratar, analizar y procesar grandes volúmenes de datos en memoria. Está basada en Python y es muy utilizada en los proyectos Big Data y de Analítica Avanzada.



Spark MLlib

Es la librería de Machine Learning para Spark.



Spark Streaming

Es una librería de Spark utilizada para tratar datos en Streaming (datos que fluyen constantemente y deben ser tratados muchas veces en tiempo real).



Apache Hive

Es un software que forma parte del ecosistema de Hadoop. Es utilizado para gestionar enormes datasets almacenados bajo el HDFS. HDFS es un sistema de ficheros para sistemas distribuidos (varios ordenadores).



Apache Kafka

Es una plataforma de transmisión de eventos distribuida (Open Source). Es utilizada por miles de empresas para canalizar datos de alto rendimiento, análisis de transmisión, integración de datos y aplicaciones de misión crítica.



R

Es un entorno y lenguaje de programación con un enfoque al análisis estadístico.



Herramientas de Visualización

(Aprende a elegir el mejor repositorio NoSQL para tus datos)

HTML, CSS y JS

Necesarios para aprender a desarrollar páginas web.



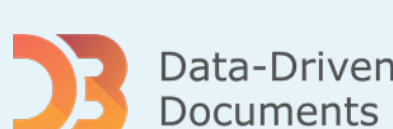
Google Developers Charts

La Suite de Google para crear gráficos dinámicos.



Data-Driven Documents

Es una biblioteca de JavaScript que sirve para manipular documentos basados en datos. Permite crear webs vivas para visualizar datos.



Power BI

Es un servicio de análisis empresarial de Microsoft, su objetivo es proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial con una interfaz lo suficientemente simple como para que los usuarios finales creen sus propios informes y paneles.



Tableau

Permite tomar datos de cualquier tipo y fuente y ponerlos a disposición de todos, en cualquier lugar y con formatos de presentación adecuados para su comprensión.



Carto

Es una plataforma Software como Servicio de computación en nube que proporciona herramientas SIG y de mapeo web para visualizar en un navegador de web.

