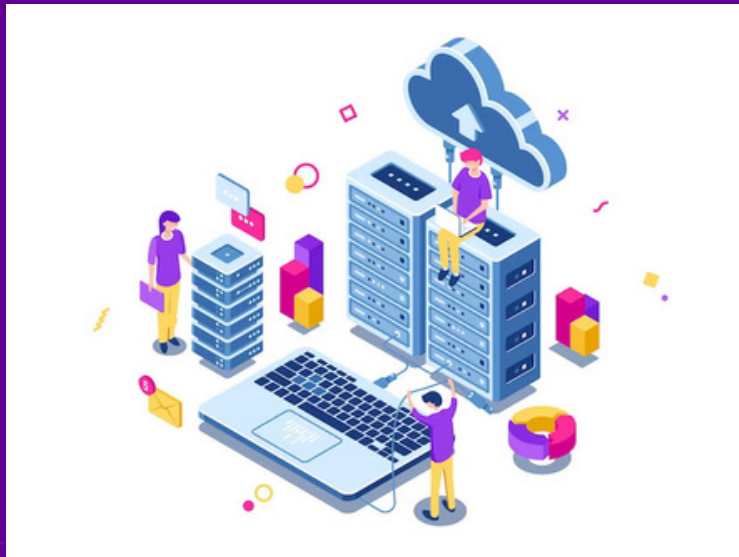


Big Data Tools



BIG DATA



Big Data adalah kumpulan data besar dan kompleks yang mencakup data seperti transaksi bisnis, interaksi media sosial, sensor digital, dll. Data ini dapat berupa data struktur atau data tidak struktur, dan dapat berisi teks, gambar, suara, atau data numerik.

Untuk mengelola dan menganalisisnya, kita membutuhkan tools.





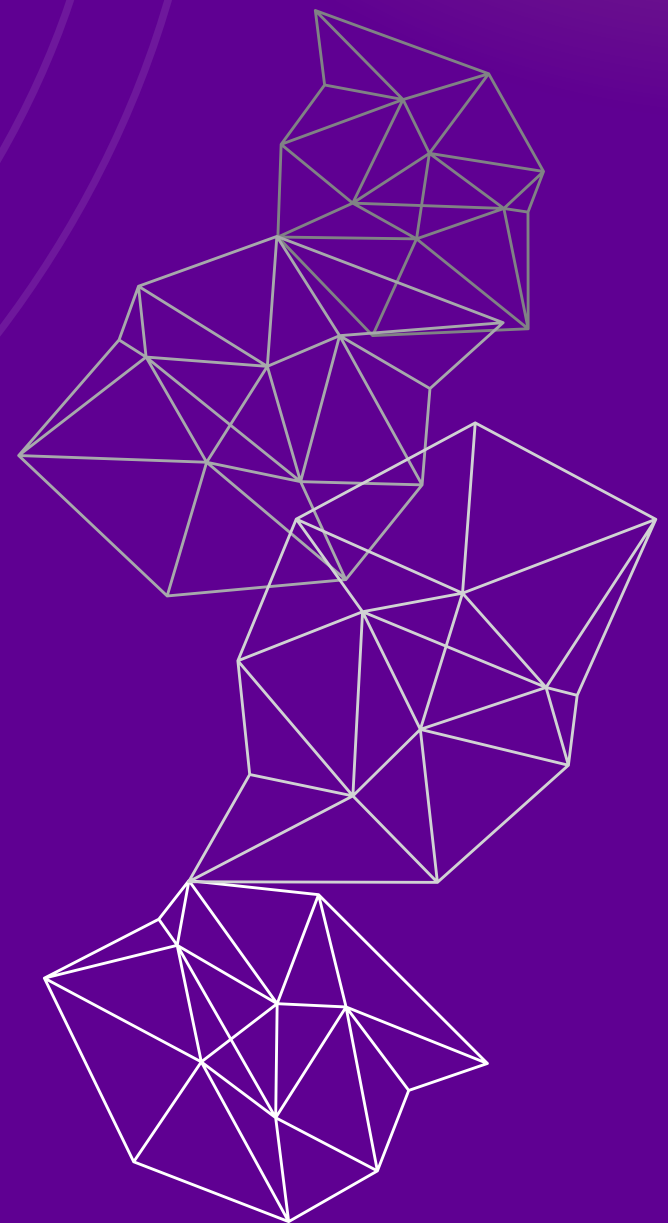
Hadoop adalah open source framework yang digunakan untuk penyimpanan terdistribusi dan pemrosesan kumpulan data besar.

- Sangat skalabel.
- Dapat menangani petabyte data di ribuan node.
- Hadoop telah menjadi standar de facto untuk pemrosesan big data dan digunakan oleh perusahaan seperti Facebook, Yahoo, dan Twitter.



Spark adalah open source framework big data sumber terbuka lainnya yang dirancang untuk kecepatan dan kemudahan penggunaan.

- Dibangun di atas Hadoop dan menyediakan antarmuka untuk pemrograman di Java, Scala, dan Python.
- Dikenal karena kemampuannya dalam menangani data streaming real-time, serta pemrosesan batch.





cassandra

- Sangat skalabel dan dapat menangani petabyte data dengan mudah.
- Cassandra digunakan oleh perusahaan seperti Netflix, eBay, dan Twitter.

Cassandra adalah database NoSQL terdistribusi yang dirancang untuk menangani jumlah data besar di beberapa server.





elastic

Elasticsearch adalah mesin pencari dan analitik terdistribusi yang dirancang untuk pencarian real-time dan analisis kumpulan data besar.

- Sangat skalabel dan dapat menangani terabyte data dengan mudah.
- Elasticsearch digunakan oleh perusahaan seperti Wikipedia, GitHub, dan Yelp.

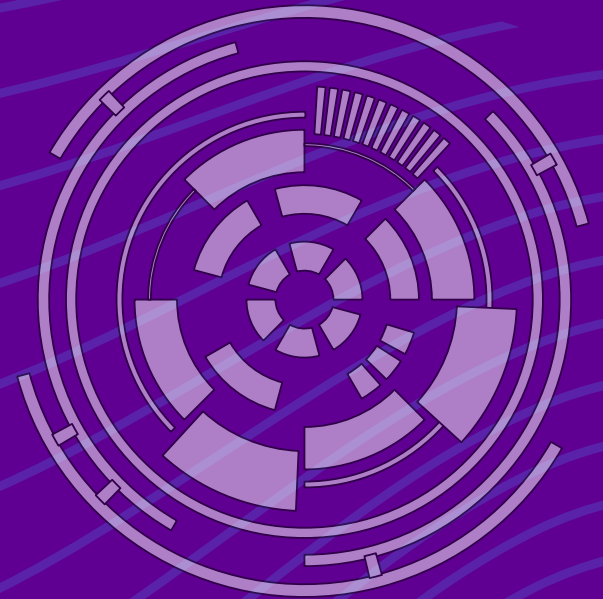




+ a b l e a u[®]

Tableau adalah alat visualisasi data yang memungkinkan pengguna membuat visualisasi interaktif, laporan, dan dashboard dari kumpulan data besar.

- Memungkinkan user membuat visualisasi, laporan, dan dasbor yang interaktif dan dinamis.
- Tableau digunakan oleh perusahaan seperti Airbnb, Coca-Cola, dan Verizon.



Splunk adalah tools analitik data yang digunakan untuk pemantauan, pencarian, dan analisis real-time data yang dihasilkan oleh mesin.

- Splunk dirancang untuk membantu organisasi mendapatkan wawasan dari big data mereka secara real-time.
- Splunk digunakan oleh perusahaan seperti Adobe, Domino's Pizza, dan Uber.



Apache Storm adalah sistem komputasi real-time terdistribusi yang dirancang untuk memproses aliran data.

- Dapat memproses dan menganalisis data dalam jumlah besar dan memiliki waktu pemrosesan latensi rendah hanya dalam milidetik.
- Sangat terukur dan toleran terhadap kesalahan, membuatnya cocok untuk memproses data dalam jumlah besar di lingkungan terdistribusi.

Referensi

- Apache Hadoop. (2022). Retrieved from <https://hadoop.apache.org/>
- Apache Spark. (2022). Retrieved from <https://spark.apache.org/>
- Apache Cassandra. (2022). Retrieved from <https://cassandra.apache.org/>
- Elasticsearch. (2022). Retrieved from <https://www.elastic.co/>
- Tableau. (2022). Retrieved from <https://www.tableau.com/>
- Splunk. (2022). Retrieved from <https://www.splunk.com/>
- Apache Storm. (2022). Retrieved from <https://storm.apache.org/>



KELOMPOK 21

Akbar Maulana - 1512620077

Anna Maulidita - 1512620074

Rafly Adhyra I - 1512620068

Sulthon Abdillah - 1512620096



MongoDB merupakan salah satu jenis database NoSQL, penyimpanan datanya menggunakan dokumen format JSON, berbeda dengan database jenis SQL yang menggunakan relasi tabel.



Apache Spark adalah *framework* yang digunakan untuk menganalisis big data. Pemrosesan datanya melalui in-memory yang dinilai lebih cepat daripada framework lain.



Apache Spark adalah platform cloud computing yang menyediakan berbagai layanan *cloud*, seperti komputasi, tools analisis, ruang penyimpanan data, hingga *networking*.