



**Kampus
Merdeka**
INDONESIA JAYA

BIG DATA (TFC303)

Pertemuan 1 – Konsep Analisis Big Data

ALIFIA REVAN PRANANDA

Department of Information Technology
Faculty of Engineering
Universitas Tidar



Before we begin, let's hear
what is your thought

LEARNING OBJECTIVE

Setelah menyelesaikan perkuliahan ini, diharapkan mahasiswa mampu:

- CPMK 0701** : Mahasiswa mampu menggunakan pemikiran logis dan kritis dalam menelaah informasi dan data dalam ruang lingkup bidang Teknologi Informasi
- CPMK 0702** : Mahasiswa mampu menganalisis informasi dan data secara sistematis untuk pengambilan keputusan secara tepat dalam penyelesaian masalah di bidang Teknologi Informasi
- CPMK 1003** : Mahasiswa mampu membuat dan mengelola basis data secara individu maupun tim

COURSE MATERIAL

SEBELUM UTS

1. Pengantar Analisis Big Data
2. Penyiapan Big Data Stack
3. Pola Big Data
4. NoSQL
5. Akuisisi Data
6. Big Data Storage
7. Batch Analysis

SETELAH UTS

1. Real-time analysis
2. Interactive Query
3. Analisis Algoritma
4. Visualisasi Data

COURSE ASSESSMENT



50%



Activeness & Group
Project



5%



Individual/group
assignment



5%



Quizzes



40%



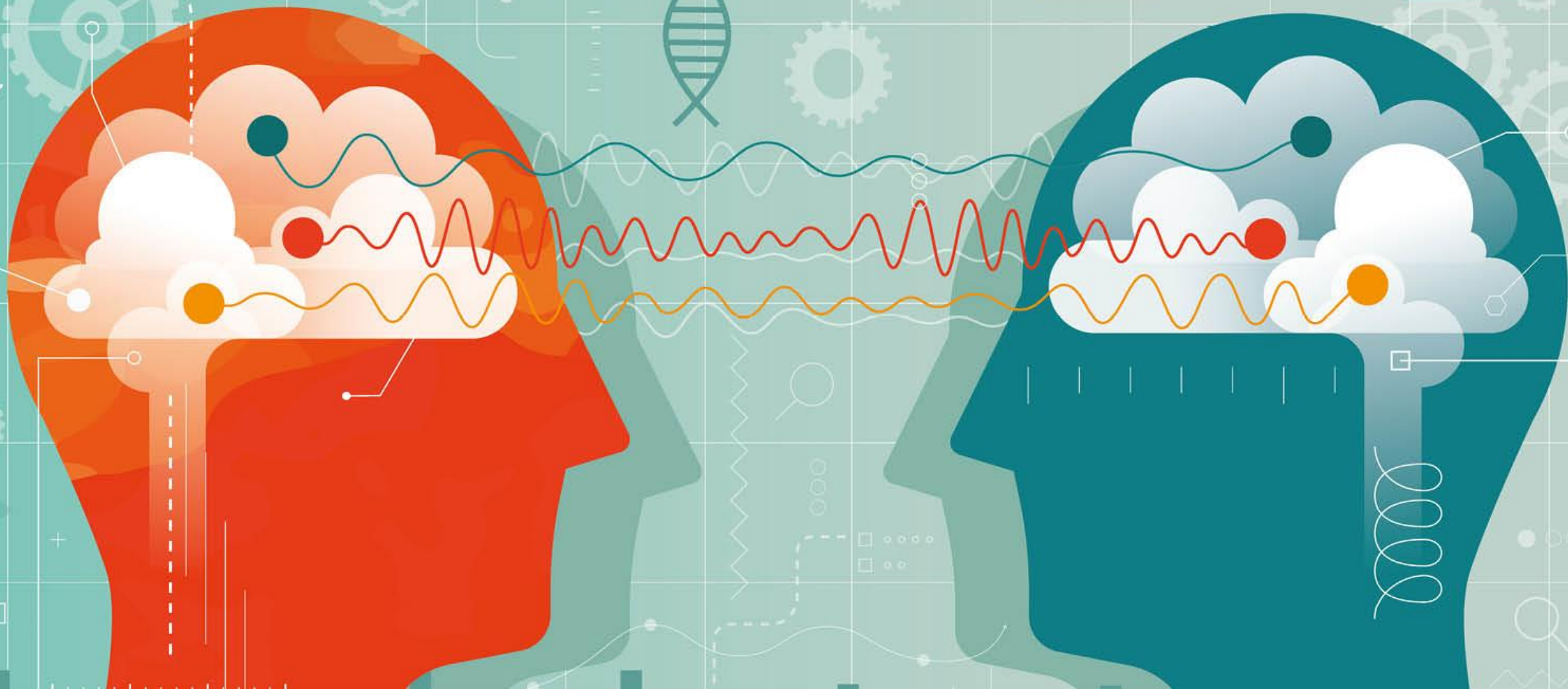
Mid-term & Final
exam

All information about this course will be announced in **ELITA**.
Please register in "**elita.untidar.ac.id**" using your SSO account and
enroll to the "**Big Data**" course

TODAY'S MATERIAL

- ✓ Definisi Big Data
- ✓ Karakteristik Big Data
- ✓ Alur Analisis Big Data
- ✓ Big Data Stack
- ✓ Pola Analisis

WHAT IS BIG DATA?



QUIZ

1. Jelaskan definisi “**BIG DATA**” menurut pemahaman Anda.
2. Menurut Anda, bagaimana perkembangan big data saat ini serta berikan contoh pemanfaatan big data yang Anda ketahui.

Kerjakan kuis di atas dalam waktu 10 menit.



“

Data is the new oil”

Clive Humby



INTRODUCTION TO BIG DATA

✓ What is Data?

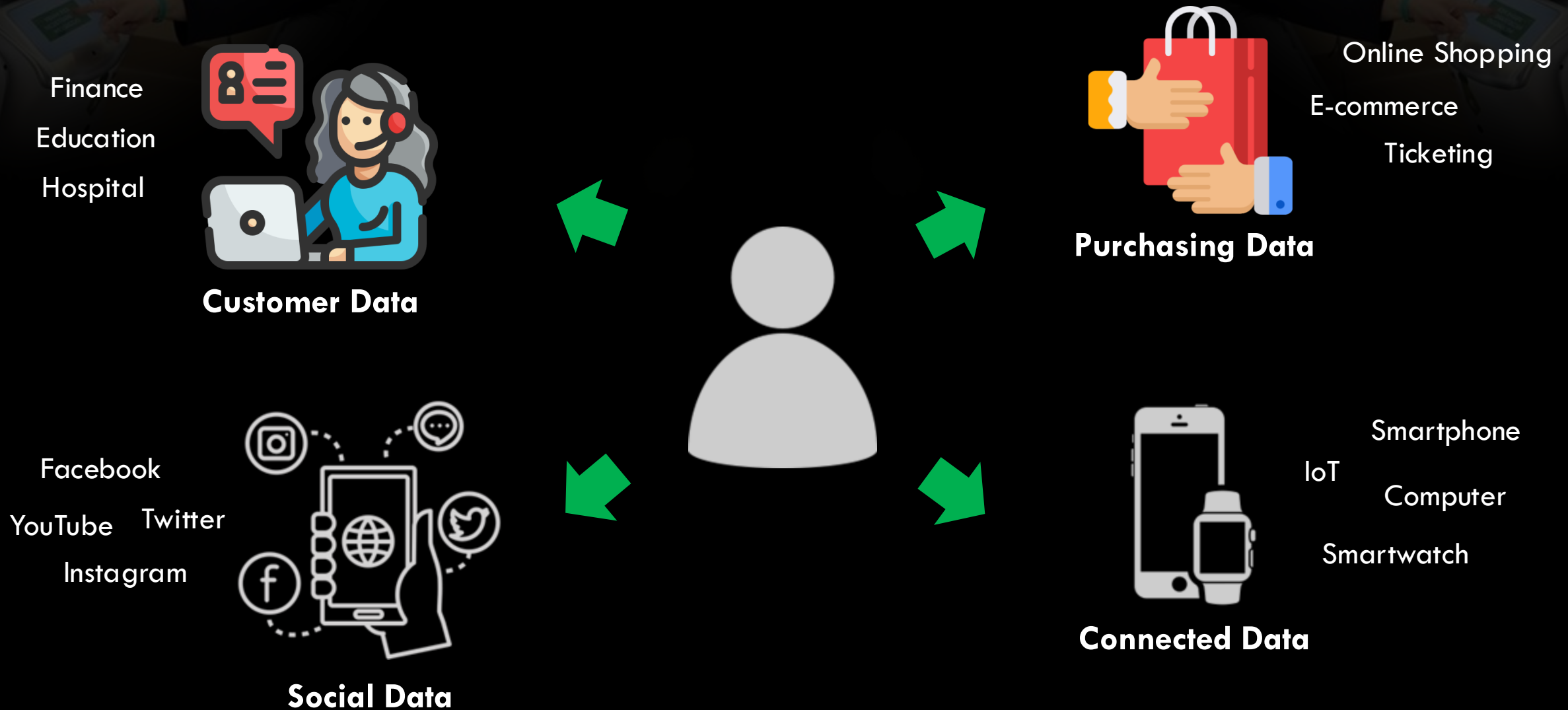
The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.

✓ What is Big Data?

Big Data is also data but with a huge size. Big Data is a term used to describe a collection of data that is huge in volume and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.

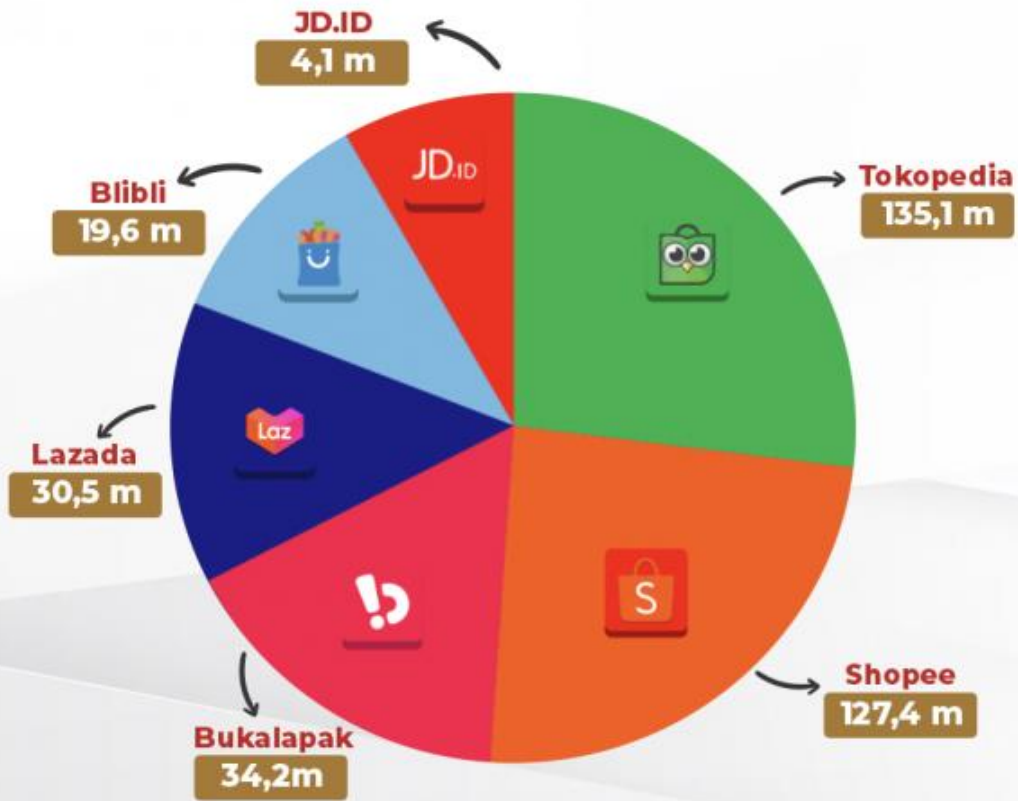
➔ Extremely large data sets that may be analyzed computationally to reveal patterns, trends and association, especially relating to human behavior and interaction are known as Big Data.

BIG DATA IN DAILY LIFE



BIG DATA IN DAILY LIFE

**Pengguna Aktif E-Commerce
Di Indonesia 2021**



Source: Genee.com

POTENTIAL REACH
OF ADVERTISING ON
FACEBOOK: NUMBER OF
USERS AGED 13 TO 17



we
are
social

113.3
MILLION

POTENTIAL REACH
OF ADVERTISING ON
TWITTER: NUMBER OF
USERS AGED 13 TO 17



20.2
MILLION

POTENTIAL REACH
OF ADVERTISING ON
INSTAGRAM: NUMBER OF
USERS AGED 13 TO 17



52.9
MILLION

**Penduduk
Indonesia**



274,9
juta jiwa

**Pengguna
Gawai**



125.6 %
(345.3 juta jiwa)

**Pengguna
Internet**



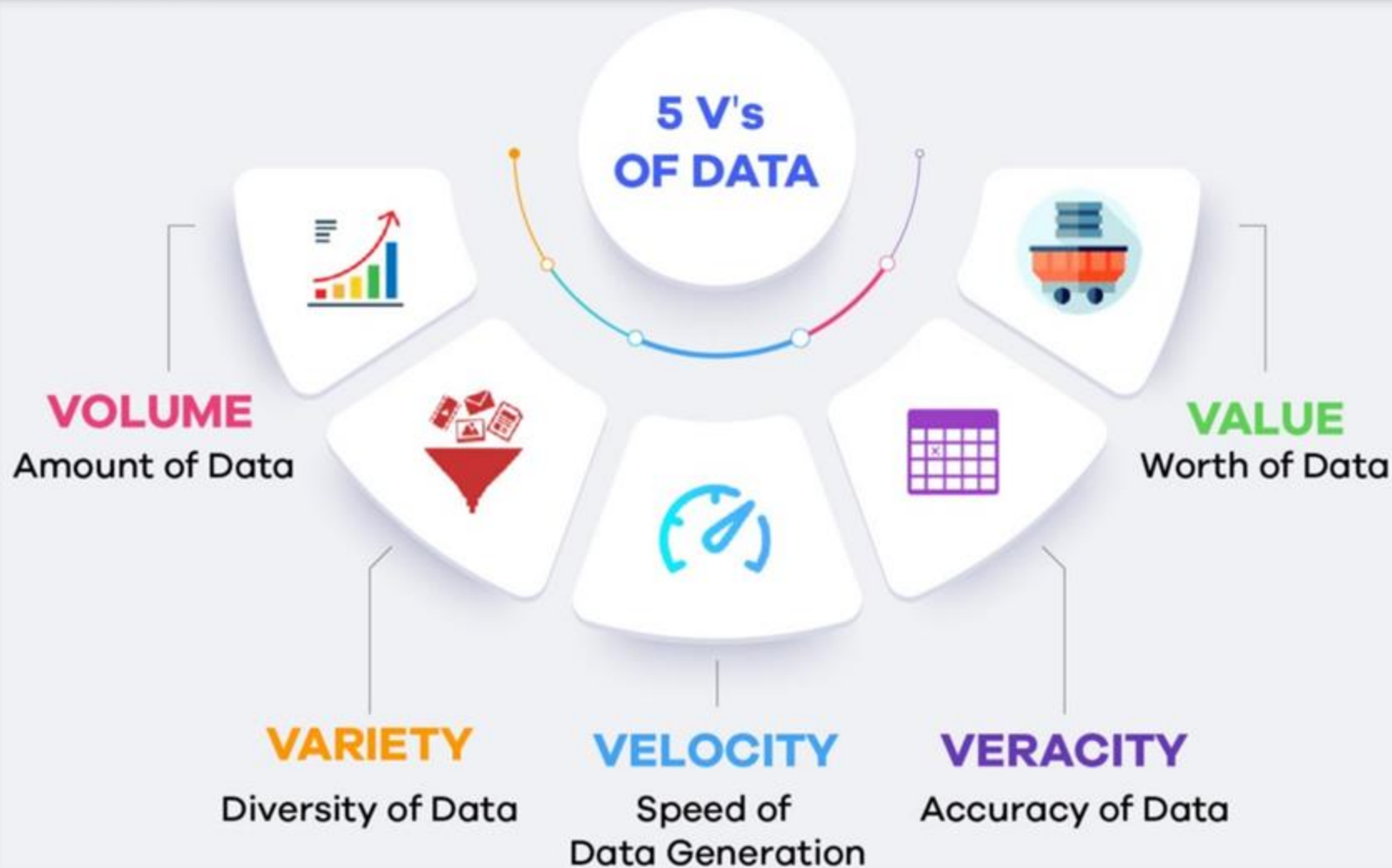
73.7 %
(202.6 Juta)

**Pengguna
Aktif Media Sosial**



61.8%
(170 juta jiwa)

KARAKTERISTIK BIG DATA



Data can be classified by Volume, Variety, Veracity, Value, and Velocity.

This classification is also known as the 5 V(s) of Big Data.

The 5Vs provide a taxonomy for classifying data into manageable categories.

It simplifies the process of understanding big data and its business value.

KARAKTERISTIK BIG DATA

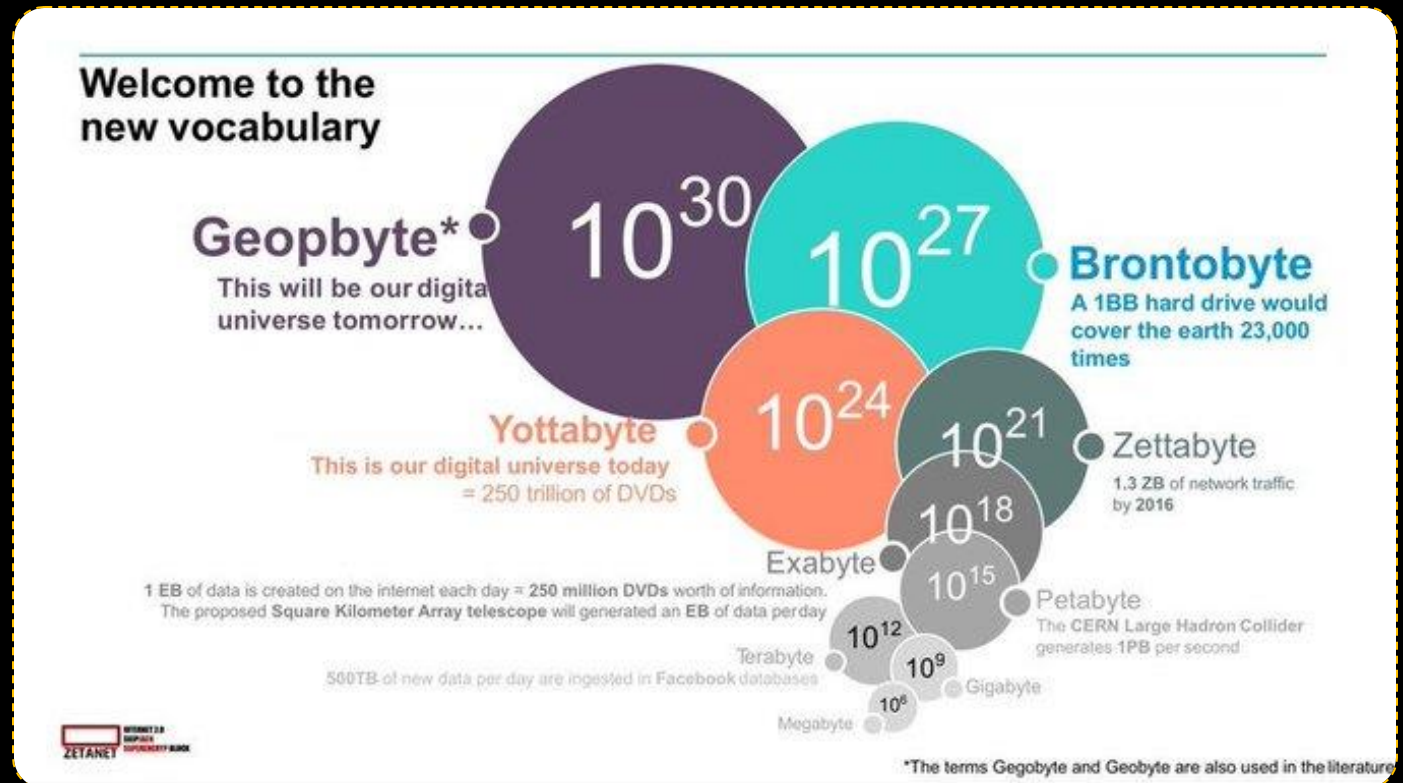


VOLUME

Big Data menunjukkan '**VOLUME**' data yang sangat besar yang dihasilkan setiap hari dari berbagai sumber seperti platform media sosial, proses bisnis, mesin, jaringan, interaksi manusia.

Istilah Big Data digunakan untuk **data skala besar** yang sulit untuk disimpan, dikelola, dan diproses menggunakan database tradisional dan arsitektur pemrosesan data. Big Data memiliki volume sangat besar sehingga **tidak dapat ditampung pada satu mesin**.

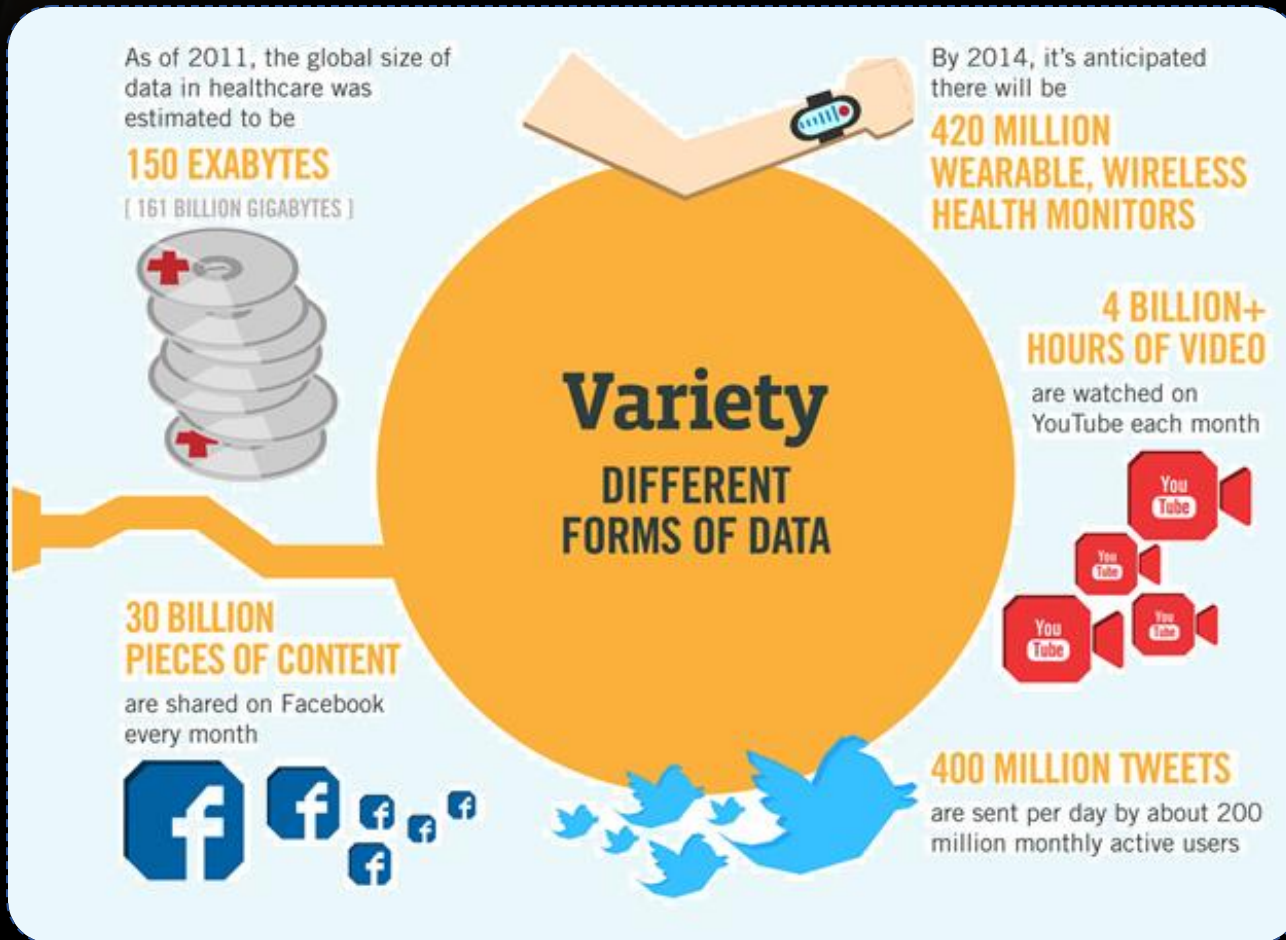
Karenanya, dibutuhkan **alat khusus dan kerangka kerja khusus** untuk melakukan pengelolaan dan analisis.



KARAKTERISTIK BIG DATA



VARIETY



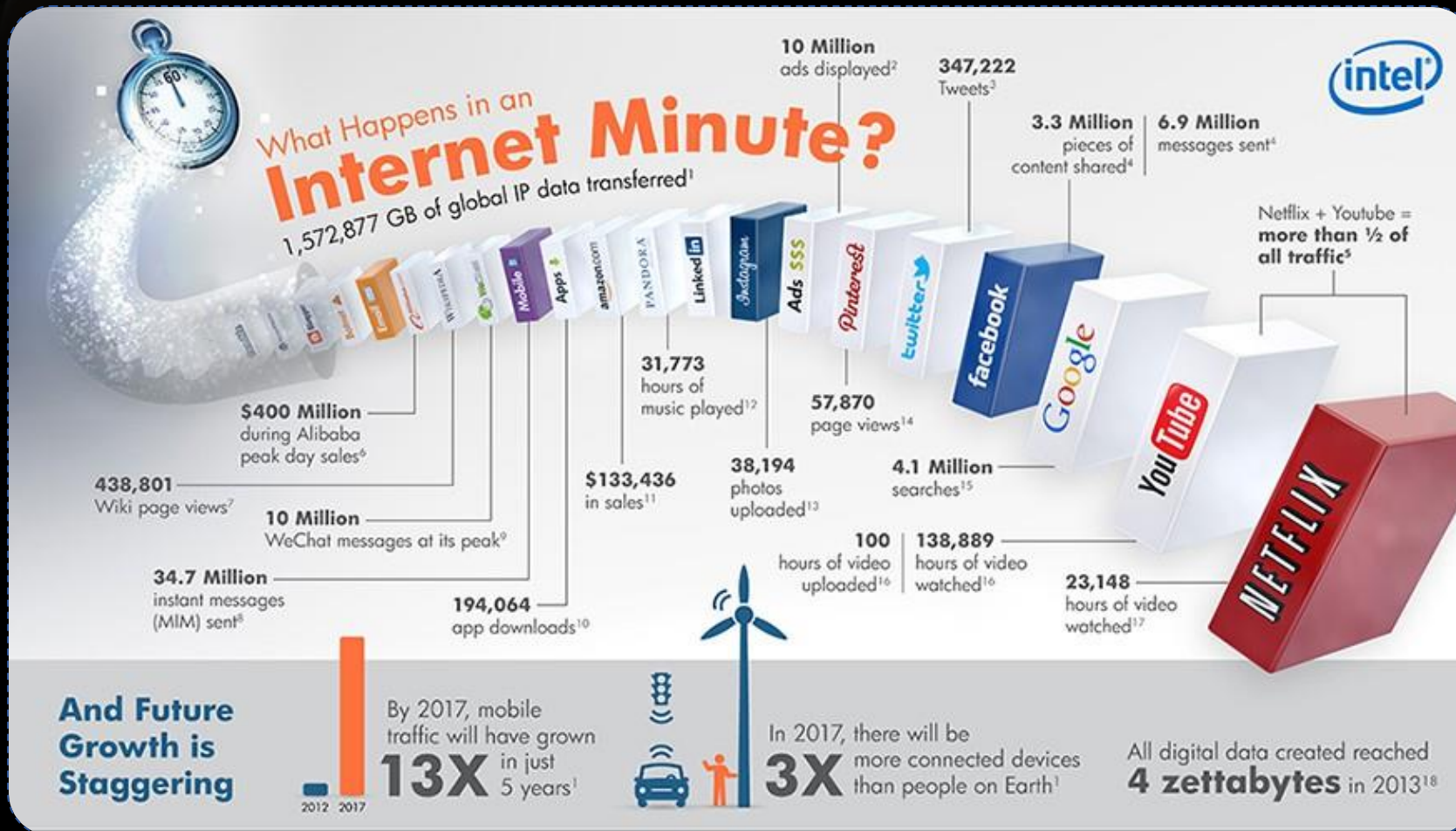
Proses globalisasi yang melibatkan berbagai macam teknologi telah menghasilkan **variasi data** yang sangat beragam.

Model Big Data dirancang mengacu pada **data terstruktur, tidak terstruktur, dan semi terstruktur** yang telah dikumpulkan dari berbagai sumber. Hal ini agar proses pengelolaan big data dapat dilakukan dengan baik.

KARAKTERISTIK BIG DATA



VELOCITY



VELOCITY atau kecepatan mengacu pada seberapa cepat data dihasilkan secara *real-time*.

Kecepatan merupakan karakteristik penting dari big data dan alasan utama terjadinya **pertumbuhan data secara eksponensial**.

Kecepatan data yang tinggi mengakibatkan volume data yang terkumpul menjadi sangat besar

KARAKTERISTIK BIG DATA



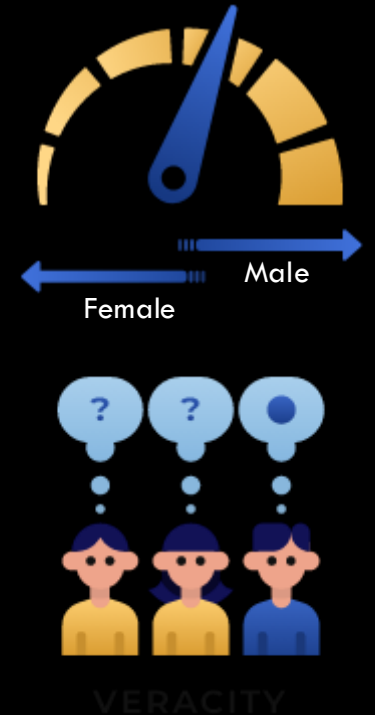
VERACITY

Veracity menyangkut tentang ke-valid-an sebuah data apakah bisa dipercaya atau tidak. Dengan banyaknya bentuk data, kebenaran tentang sebuah informasi menjadi kurang terkontrol.

Sebagai contoh:

sistem informasi akademik suatu program studi mendefinisikan jenis kelamin dengan simbol “L” (untuk laki-laki) dan “P” (untuk perempuan) sedangkan di sistem informasi akademik program studi lain menggunakan simbol “P” (pria) dan “W” (wanita). Pada saat dilakukan sinkronisasi maka data menjadi tidak akurat.

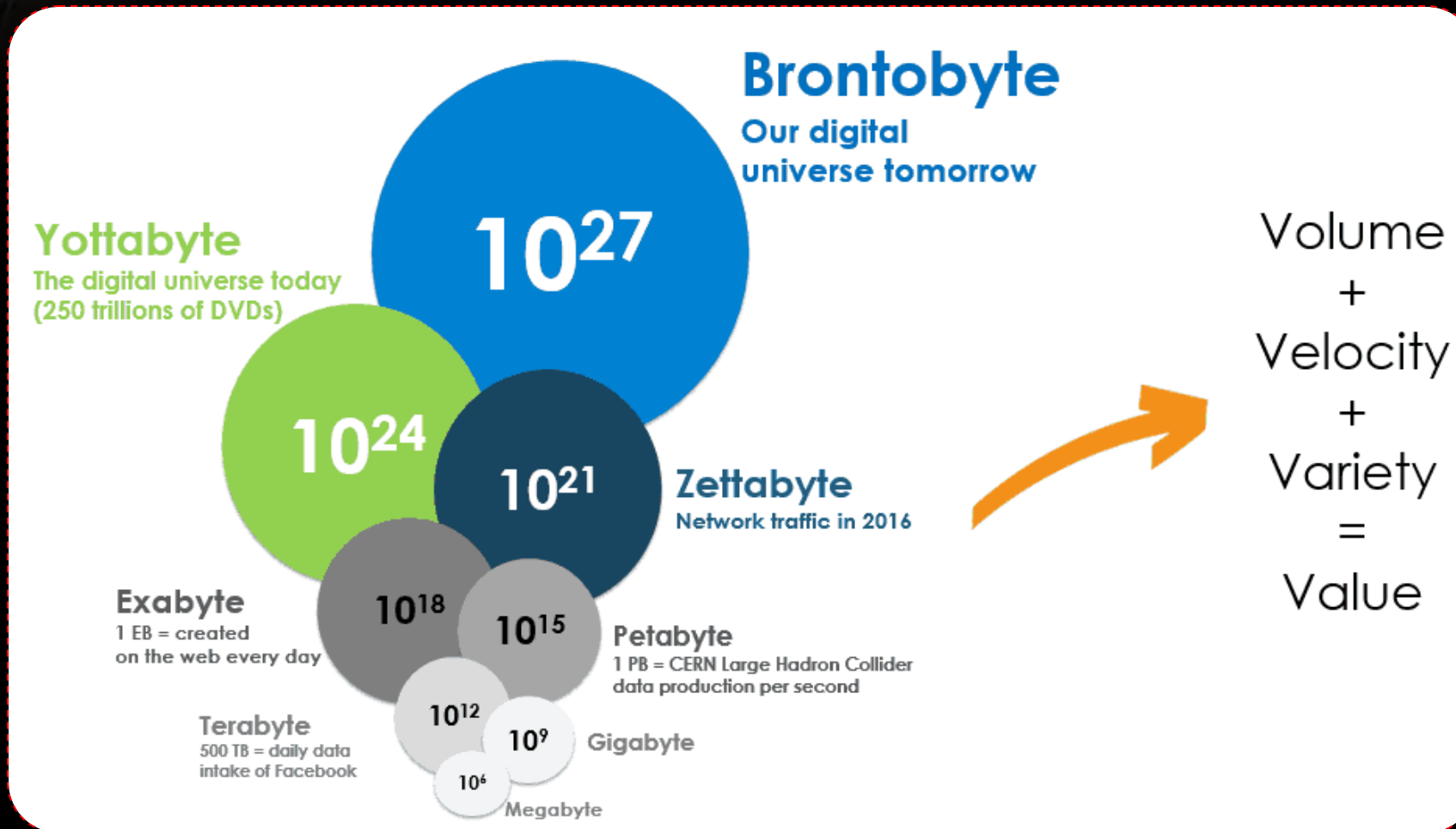
Big data dengan teknologi analisis membantu kita untuk dapat bekerja dengan data tersebut melalui hasil analisis, karena semakin besar volume suatu data maka akan semakin tidak akurat data tersebut.



KARAKTERISTIK BIG DATA



VALUE



VALUE atau nilai data mengacu pada proses dimana data diekstrak untuk **menghasilkan informasi** sesuai dengan tujuan.

Sebagai contoh, data cuaca yang telah dikumpulkan selama beberapa periode dapat digunakan untuk memprediksi cuaca di masa mendatang.



BIG DATA YOUR FUTURE LIFE



Manfaat Big Data bagi Dunia Kesehatan

Menurut Ir. Beno K. Pradekso Founder iDBigData



QUIZ

1. Buat kelompok dengan anggota maksimal 3 orang.
2. Berikan contoh implementasi big data pada bidang-bidang berikut: (satu kelompok cukup mengerjakan 1 bidang).
 - a. Bidang militer
 - b. Bidang pendidikan
 - c. Bidang pertanian
 - d. Bidang ekonomi
 - e. Bidang industri
 - f. Bidang logistic dan transportasi
3. Presentasikan hasil diskusi kelompok Anda.





HOW TO ANALYZE BIG DATA TO GET INFORMATION?

DEFINISI ANALISIS

Pengertian **ANALISIS** adalah aktivitas yang memuat sejumlah kegiatan seperti mengurai, membedakan, memilah sesuatu untuk digolongkan dan dikelompokkan kembali menurut kriteria tertentu kemudian dicari kaitannya dan ditafsir maknannya.

Tujuan analisis antara lain :

- (1) untuk memprediksi sesuatu
- (2) untuk menemukan pola dalam data
- (3) menemukan hubungan dalam data

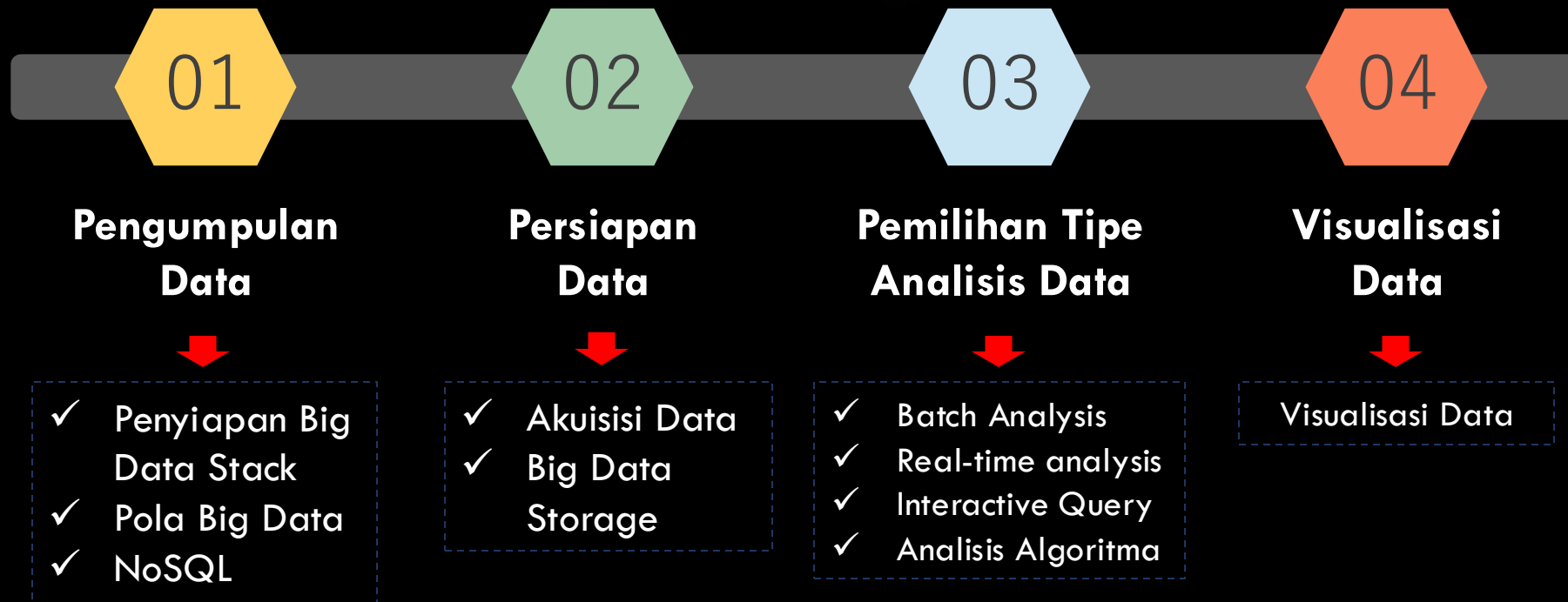


TIPE-TIPE ANALISIS



DEFINISI ANALISIS BIG DATA

Analisis Big Data berkaitan dengan pengumpulan, penyimpanan, pemrosesan, dan analisis data berskala besar ini. Analisis Big Data melibatkan beberapa langkah mulai dari pembersihan data, data munging (atau pertenggaran), pemrosesan dan visualisasi data.



POLA ANALISIS



POLA ALPHA

- Pola Alpha biasanya digunakan untuk analisis data batch.
- Pola ini dapat digunakan untuk menyerap volume data yang besar ke dalam sistem file terdistribusi (seperti HDFS) atau database NoSQL (seperti HBase) menggunakan konektor sumber-sink (seperti Flume) dan konektor SQL (seperti Sqoop).
- Setelah data dipindahkan ke tumpukan, data dapat dianalisis dalam mode batch dengan kerangka kerja analisis batch termasuk MapReduce (menggunakan Hadoop), kerangka kerja skrip (seperti Pig), kerangka grafik asiklik terdistribusi (seperti Spark), kerangka kerja pembelajaran mesin (seperti Spark MLlib).
- Hasil analisis disimpan baik dalam database relasional maupun nonrelasional.

POLA ANALISIS



POLA BETA

- Pola Beta biasanya digunakan untuk analisis analisis real-time.
- Pola ini dapat digunakan untuk menyerap data streaming menggunakan framework perpesanan publish-subscribe, antrean, dan konektor khusus. Untuk analisis real-time, kita dapat menggunakan kerangka kerja pemrosesan aliran (seperti Storm) atau kerangka kerja pemrosesan dalam memori (seperti Spark).

POLA ANALISIS



POLA GAMMA

- Pola Gamma menggabungkan pola analisis batch dan real-time.
- Pola ini dimaksudkan untuk menyerap data streaming ke dalam Big Data Stack dan menganalisis data baik secara real-time maupun dalam mode batch. Untuk analisis batch, data dikumpulkan dan dianalisis selama interval tertentu.

Contoh

Bagaimana pola ini dapat digunakan untuk sistem yang mendeteksi kebakaran hutan berdasarkan data sensor yang dikumpulkan dari sejumlah besar perangkat IoT yang diterapkan di hutan.

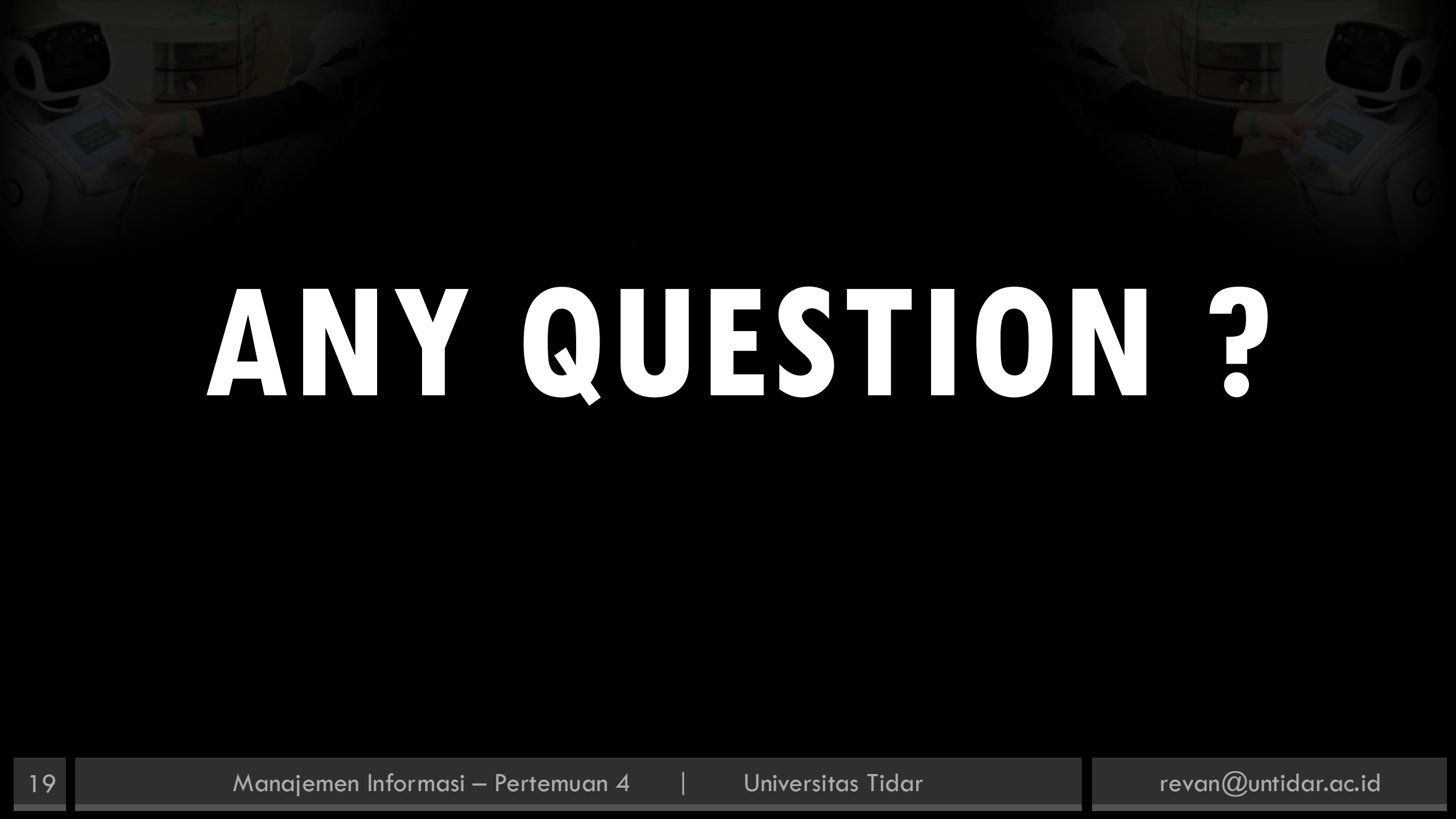
- ✓ Blok analisis real-time dalam pola ini dapat menyaring dan menganalisis data secara real-time dan membuat prediksi menggunakan model pembelajaran mesin terlatih.
- ✓ Sedangkan blok analisis batch dapat menganalisis data yang dikumpulkan selama interval tertentu (seperti per jam, harian, bulanan, atau tahunan)

POLA ANALISIS



POLA DELTA

- Pola Delta biasanya digunakan untuk analisis kueri interaktif.
- Pola ini menggunakan, konektor sumber-sink (seperti Flume) atau konektor SQL (seperti Sqoop) untuk menyerap data massal ke dalam Big Data Stack.
- Setelah data dipindahkan ke sistem file terdistribusi, Anda dapat menggunakan kerangka kerja kueri interaktif (seperti Hive atau Spark SQL) untuk membuat kueri data dengan kueri seperti SQL dalam mode interaktif.
- Pola Delta dapat digunakan oleh aplikasi seperti analisis web, penargetan iklan, manajemen inventaris, perencanaan dan kontrol produksi, dan berbagai jenis aplikasi perusahaan



ANY QUESTION ?

TUGAS

1. Sebutkan dan jelaskan yang dimaksud dengan **10 V's of Big Data!**
2. Bagaimana langkah-langkah yang dilakukan dalam menyiapkan Big Data Stack?
3. Sebutkan beberapa platform untuk mengelola Big Data Stack dan jelaskan karakteristik masing-masing! (minimal 3 platform)

Dikerjakan secara individu. Deadline 26 Agustus 2024 pukul 23:59 melalui ELITA. Hindari plagiarisme dan sertakan sumber referensi jika ada.