

Introduction to Big Data

oleh Randi Eka Sanjaya



Apa itu data?

- Data merupakan **koleksi nilai diskrit maupun kontinyu yang mengungkapkan informasi.**
- Data mendeskripsikan kuantitas, kualitas, fakta, nilai statistik, serta unit basis lainnya, atau serangkaian simbol yang dapat diinterpretasikan lebih jauh secara formal.

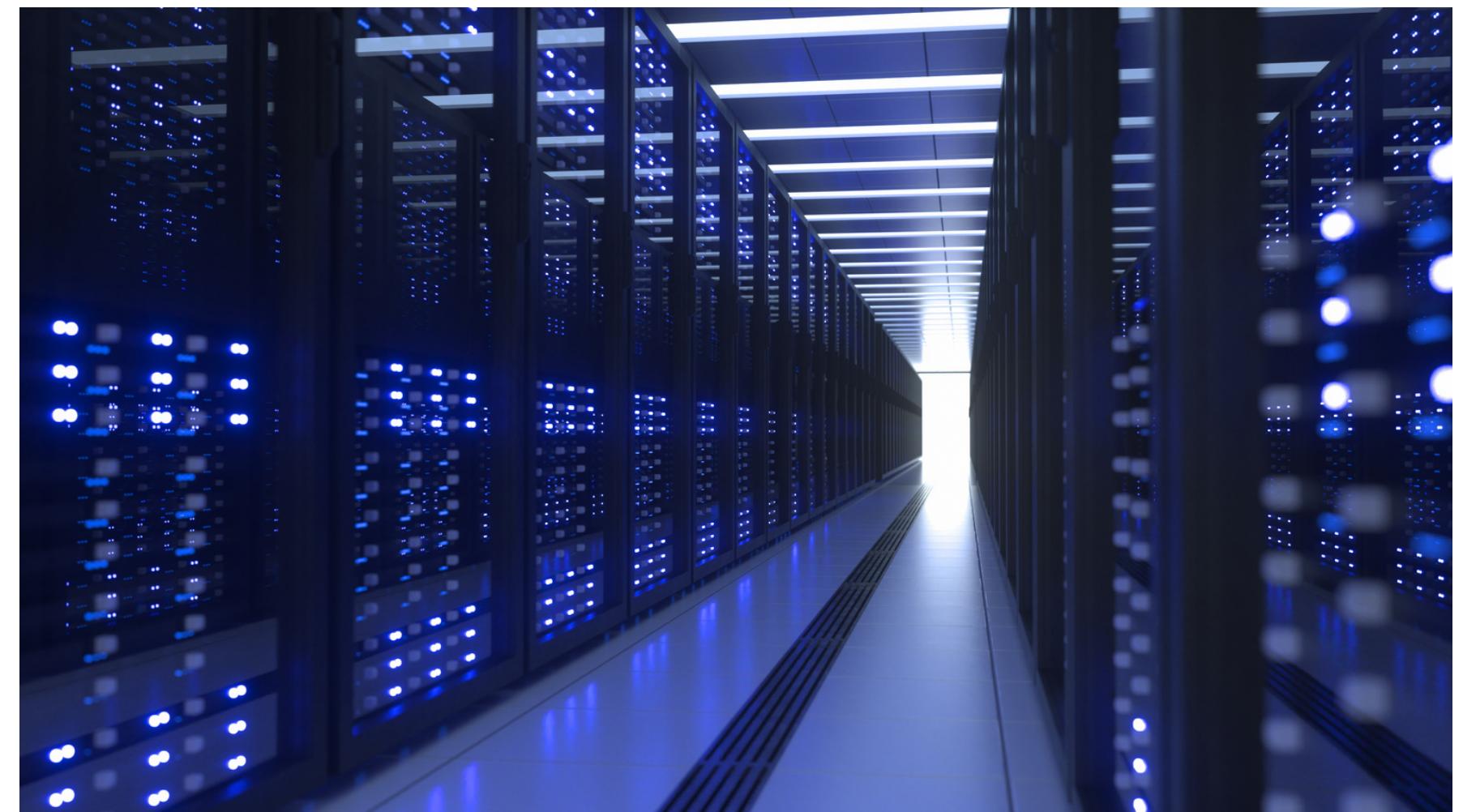
Attributes (Dimension; Features; Variables)

Objects (Samples, Individuals)

ID	Height	Weight	Age
Student 1	189 cm	81 kg	24
Student 2	210 cm	90 kg	26
Student 3	191 cm	92 kg	27
...
Student N	162 cm	71 kg	21

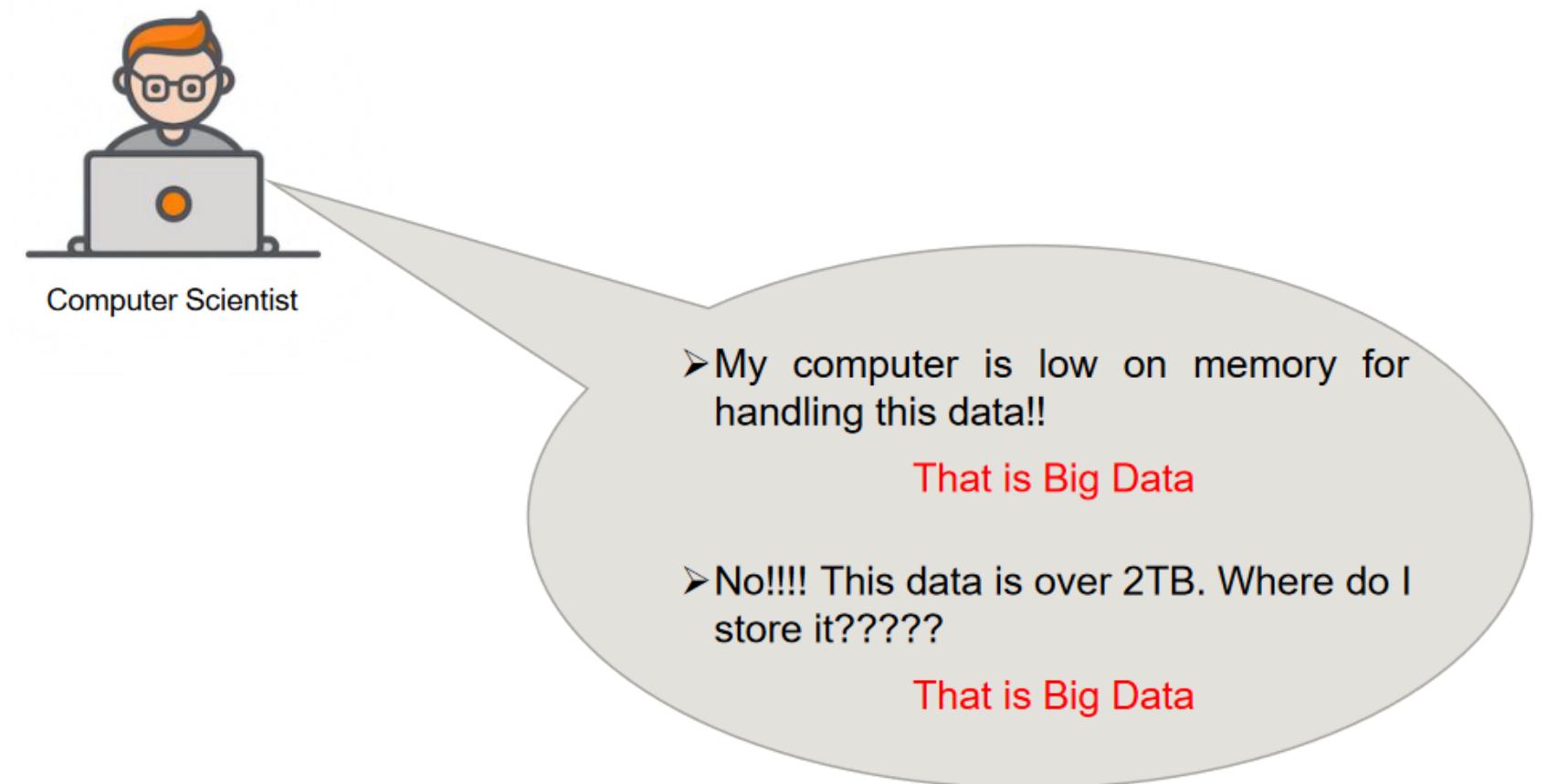
Kemudian, bagaimana dengan Big Data?

- Big Data secara sederhana dapat diartikan sebagai **data yang memiliki jumlah besar dan beragam**.
- Big data ditujukan pada semua hal tentang data, dari bagaimana data dikumpulkan, diproses, di analisis, untuk menghasilkan nilai tambah melalui data-driven insights dan keputusan.
- Trend terhadap data yang semakin besar dikarenakan terdapat informasi tambahan yang dapat diperoleh ketika menganalisis suatu set data yang besar.

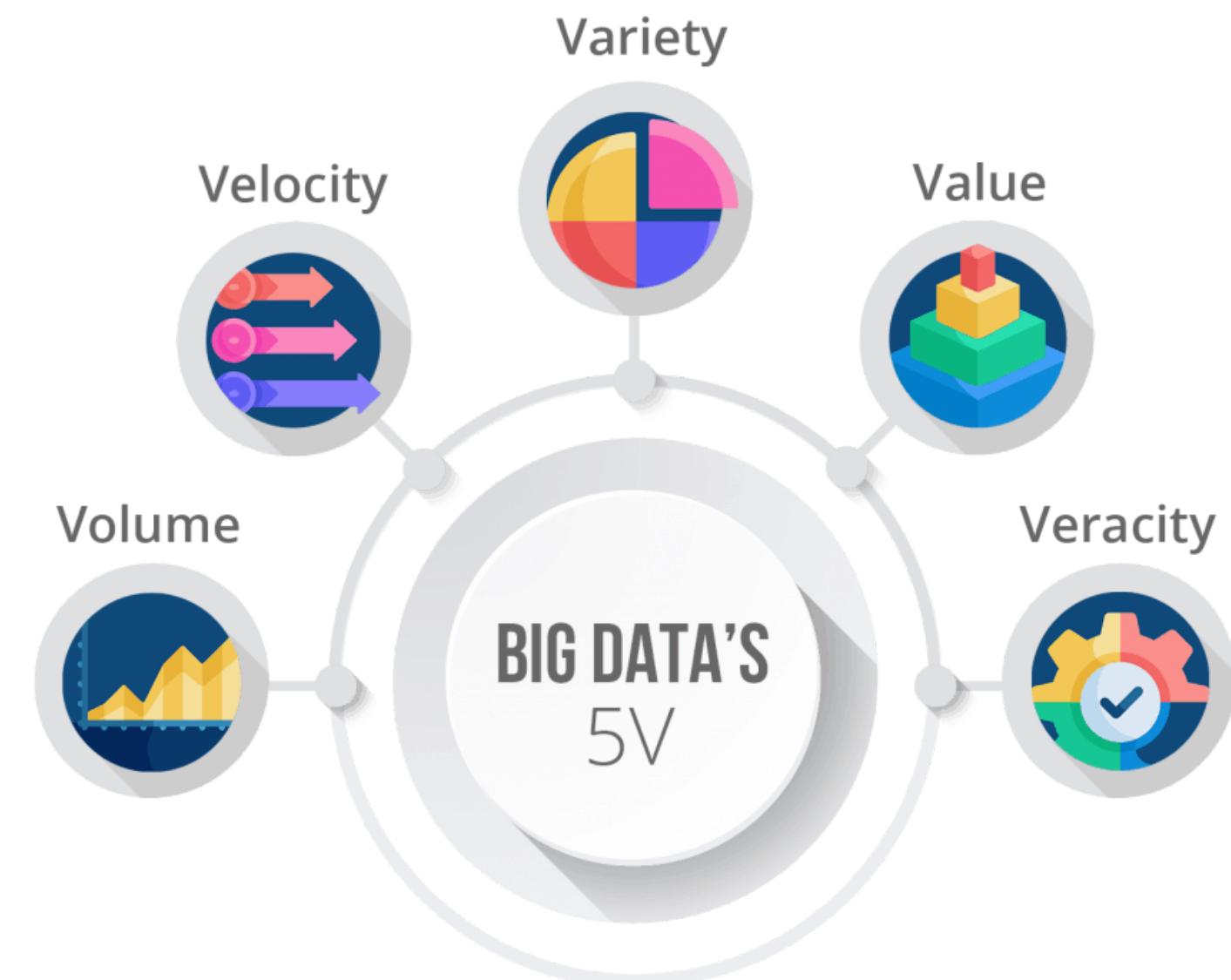


Bagaimana kita tahu data kita merupakan Big Data?

- Big Data merupakan istilah untuk koleksi set data yang sangat besar dan kompleks, hingga sulit untuk diolah menggunakan database management tools ataupun aplikasi pemrosesan data tradisional.
- Singkatnya, **jika anda mengalami kendala dalam melakukan pemrosesan dan analisis data dengan komputer anda, hal ini dapat disebabkan oleh Big Data**

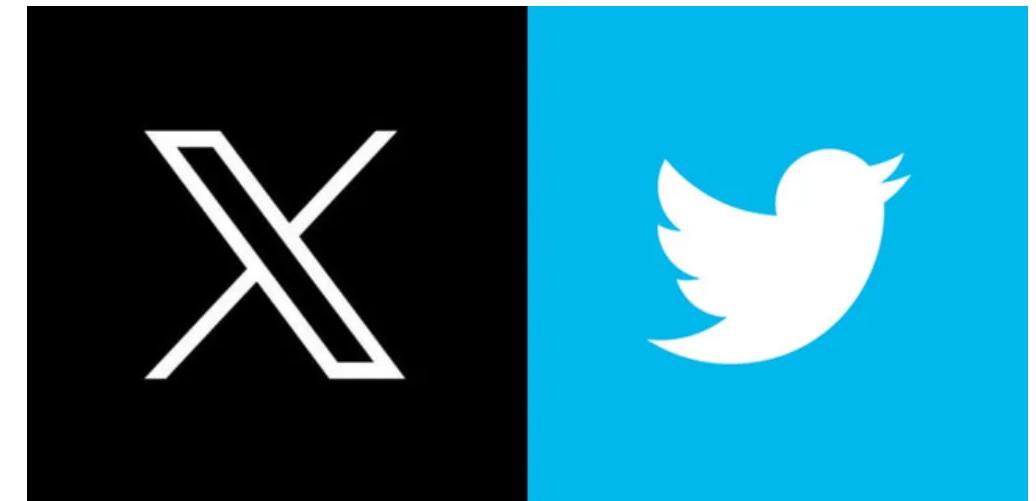


Apa Saja Karakteristik Big Data?



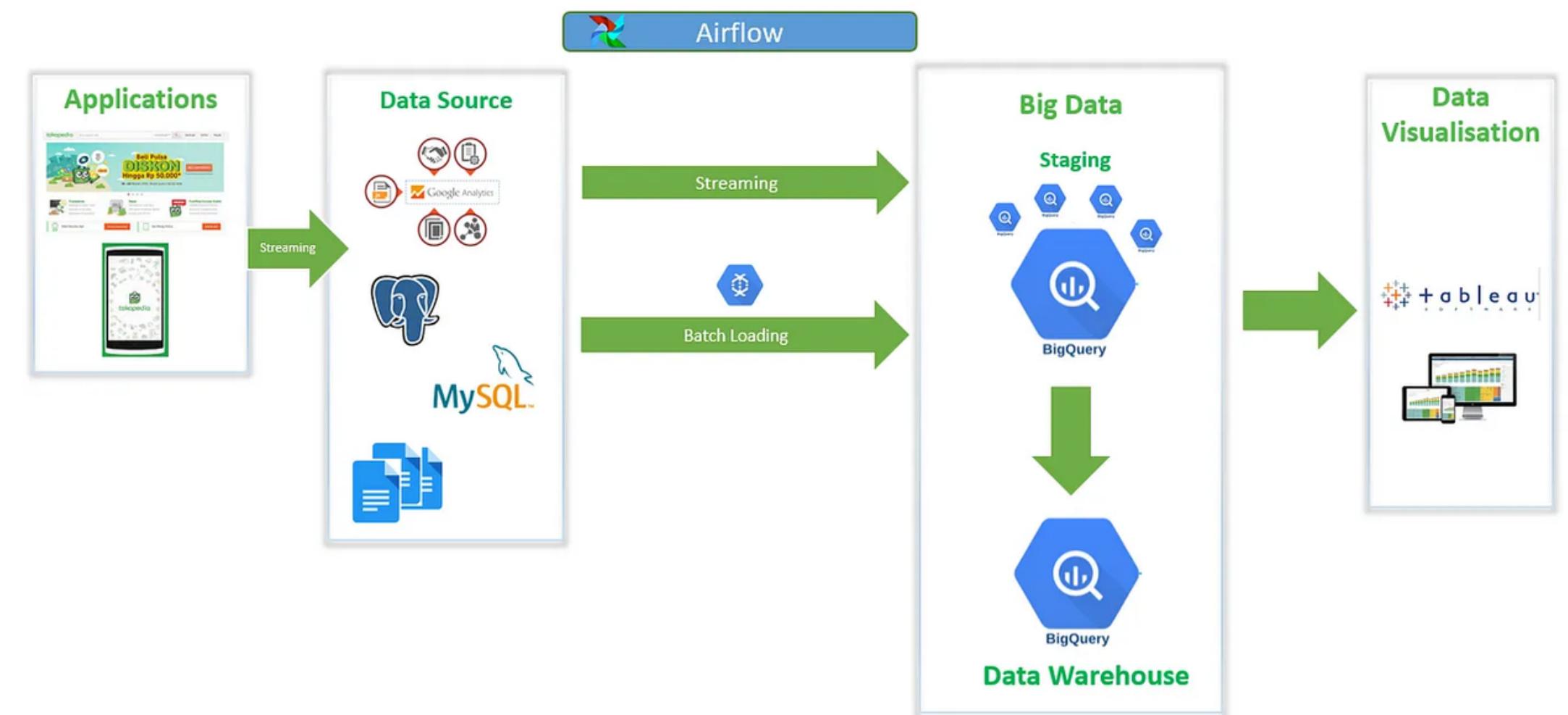
Apa Saja Karakteristik Big Data?

- **Volume** terkait dengan jumlah data yang dikumpulkan. Ukuran dari data yang kita miliki dapat menentukan nilai dan wawasan yang kita peroleh, dan apakah data yang kita miliki dapat dikategorisasikan sebagai Big Data. Ukuran Big Data yang umum dijumpai mencapai terabyte hingga petabyte.



Apa Saja Karakteristik Big Data?

- **Velocity** terkait dengan seberapa cepat data kita hasilkan dan proses untuk kebutuhan dan tantangan saat ini. Big Data sering kali kita jumpai dalam bentuk real-time. Dibandingkan dengan data kecil, Big Data dihasilkan secara lebih berkelanjutan. Dua jenis velocity pada konteks Big Data meliputi frekuensi data dihasilkan serta frekuensi data diolah, direkam, dan di publikasikan.



Apa Saja Karakteristik Big Data?

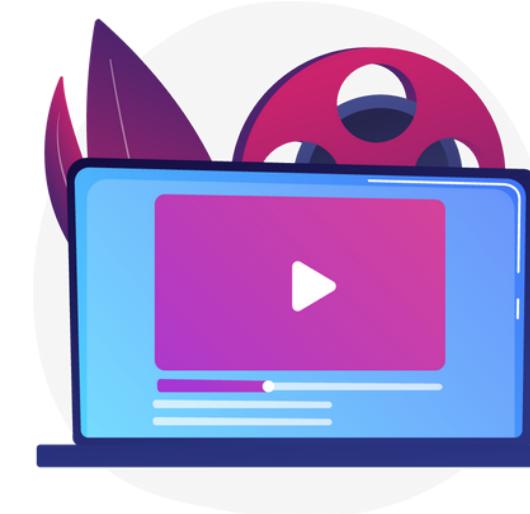
- **Variety** terkait dengan berbagai macam bentuk data yang digunakan untuk analisis. Teknologi Big Data berkembang untuk meng capture, menyimpan, dan memproses structured, semi-structured, dan unstructured data.

Objects (Samples, Individuals)

Attributes (Dimension; Features; Variables)			
ID	Height	Weight	Age
Student 1	189 cm	81 kg	24
Student 2	210 cm	90 kg	26
Student 3	191 cm	92 kg	27
...
Student N	162 cm	71 kg	21

```
[  
  {  
    "description": "quarter",  
    "mode": "REQUIRED",  
    "name": "qtr",  
    "type": "STRING"  
  },  
  {  
    "description": "sales representative",  
    "mode": "NULLABLE",  
    "name": "rep",  
    "type": "STRING"  
  },  
  {  
    "description": "total sales",  
    "mode": "NULLABLE",  
    "name": "sales",  
    "type": "INTEGER"  
  }  
]
```

It was the best of
times, it was the worst
of times, it was the age
of wisdom, it was the
age of foolishness...



Apa Saja Karakteristik Big Data?

- **Veracity** terkait dengan kualitas dari data yang dapat diperoleh dari beragam sumber. Hal ini berkaitan erat dengan kebenaran dan reliabilitas dari data yang kita miliki

Data Source	Availability	Veracity/Kebenaran	Owner
Sensor	Closed	High	Producer
Transactional/Register Database	Closed	High	Producer
Medical Record	Closed	High	Individu
Jurnal paper	Open	High	Publik
Reputable online news and web	Open	Medium	Producer
Social Media dan general web	Open	Low	Producer

Apa Saja Karakteristik Big Data?

- **Value** terkait dengan nilai informasi yang dapat dicapai dengan pemrosesan dan analisis pada Big Data.

Frequently Bought Together

Total price: \$83.09

This item: Structure and Interpretation of Computer Programs - 2nd Edition (MIT Electrical Engineering and... by Harold Abelson Paperback \$50.50

The Pragmatic Programmer: From Journeyman to Master by Andrew Hunt Paperback \$32.59

Customers Who Bought This Item Also Bought

The Little Schemer - 4th Edition
Daniel P. Friedman
★★★★★ 64 Paperback \$36.00 Prime

Instructor's Manual t/a Structure and Interpretation of Computer Programs...
Gerald Jay Sussman
★★★★★ 5 Paperback \$28.70 Prime

The Pragmatic Programmer
Andrew Hunt
★★★★★ 328 Paperback \$32.59 Prime

Introduction to Algorithms, 3rd Edition (MIT Press)
Thomas H. Cormen
★★★★★ 313 Paperback \$66.32 Prime

An Introduction to Functional Programming Through Lambda Calculus
Greg Michaelson
★★★★★ 23 Paperback \$20.70 Prime

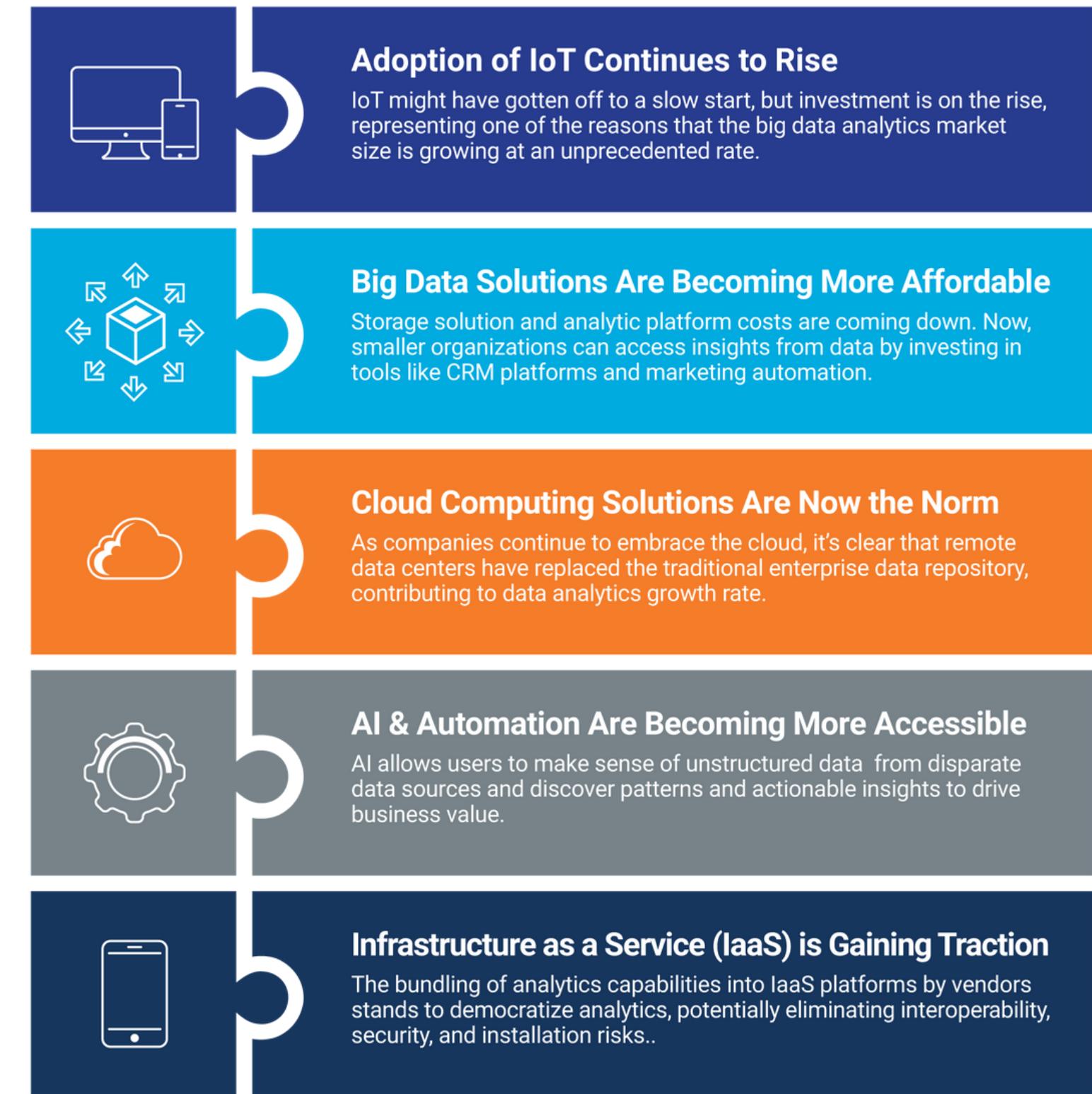
Purely Functional Data Structures
Chris Okasaki
★★★★★ 19 Paperback \$40.74 Prime

Code: The Hidden Language of Computer Hardware and Software
Charles Petzold
★★★★★ 334 Paperback \$31.78 Prime

The Little Prover (MIT Press)
Daniel P. Friedman
★★★★★ 4 Paperback \$17.99 Prime

Page 1 of 13

Apa yang mendorong pertumbuhan Big Data?



Big Data Tools

“Big data is where parallel computing tools are needed to handle data.”



Akses hands-on [disini](#)



Apa keuntungan dan kelemahan Big Data?

Keuntungan

- Meningkatkan kualitas pengambilan keputusan
- Meningkatkan efisiensi
- Menambah sumber pendapatan
- Memperoleh keunggulan kompetitif

Kelemahan

- Permasalahan privasi
- Risiko pelanggaran data
- Tantangan dari segi teknis meningkat
- Kesulitan dalam mengintegrasikan sumber data
- Tidak cocok digunakan untuk data kecil
- Kompleksitas Analisis

Big Data Best Practice

Guna secara efektif mengelola dan menutilisasi Big Data, organisasi dapat mengikuti beberapa kaidah berikut:

- Mendefinisikan tujuan bisnis dengan jelas.
- Hanya mengumpulkan dan menyimpan data yang relevan.
- Memastikan kualitas data terjaga.
- Menggunakan teknologi dan tools yang tepat.
- Menerapkan data security dan kebijakan privasi.
- Meningkatkan adaptasi terhadap Machine Learning dan Artificial Intelligence.
- Berfokus pada visualisasi data.



Aplikasi Big Data dan Studi Kasus dalam Teknik Industri

Area Supply Chain:

- Monitoring dan perencanaan stok
- Analisa anomali stok opname
- Segmentasi jenis transportasi
- Analisa kontrak transportasi
- Segmentasi gudang
- Evaluasi kinerja distributor / gudang
- Optimalisasi least cost distribution (Network Optimization)
- Mitigasi Risiko
- Estimasi Safety Stock
- Greenfield Analysis
- Digital Twins



Aplikasi Big Data dan Studi Kasus dalam Teknik Industri

Area Market Intelligence:

- Segmentasi konsumen
- Analisa perilaku konsumen
- Analisa profil 360 konsumen
- Analisa informasi permintaan dan dampak terhadap harga
- Prediksi harga komoditi
- Analisa market share
- Prediksi penentuan harga jual





Sumber

- [Introduction to Big Data by Korea University in 2019](#)
- [Introduction to Big Data by Joan Jarmack](#)
- [Introduction to Big Data by Kent State University](#)
- [What is Big Data? Introduction, Uses, and Applications by Analytics Vidhya](#)
- [Big Data: An Introduction](#)
- [Big data quality framework: a holistic approach to continuous quality management](#)
- [A Must-Read Guide on How to Work with PySpark on Google Colab for Data Scientists! by Analytics Vidhya](#)
- [How Tokopedia modernized its data warehouse and analytics processes with BigQuery and Cloud Dataflow](#)
- [\[Sore Berilmu\] Open Data Pemerintahan, Swasta dan Akademik, Tantangan dan Peluang. I Widyawan](#)



“Domain knowledge is your North Star in the world of data science.”

- Richard Warepam