

BIG DATA

Sirojul Munir | rojulman@nurulfikri.ac.id | @rojulman

Big Data

- ❖ SKS: 3 sks
- ❖ Tim Dosen: Dr Sigit W Djarot, Sirojul Munir S.SI M.KOM, Ahmad Rio Adriansyah M.SI, Henry Saptono S.SI, M.KOM
- ❖ Prerequisite: Data Warehouse
- ❖ Next Course : Data Mining, Visualisasi Data
- ❖ Penilaian:

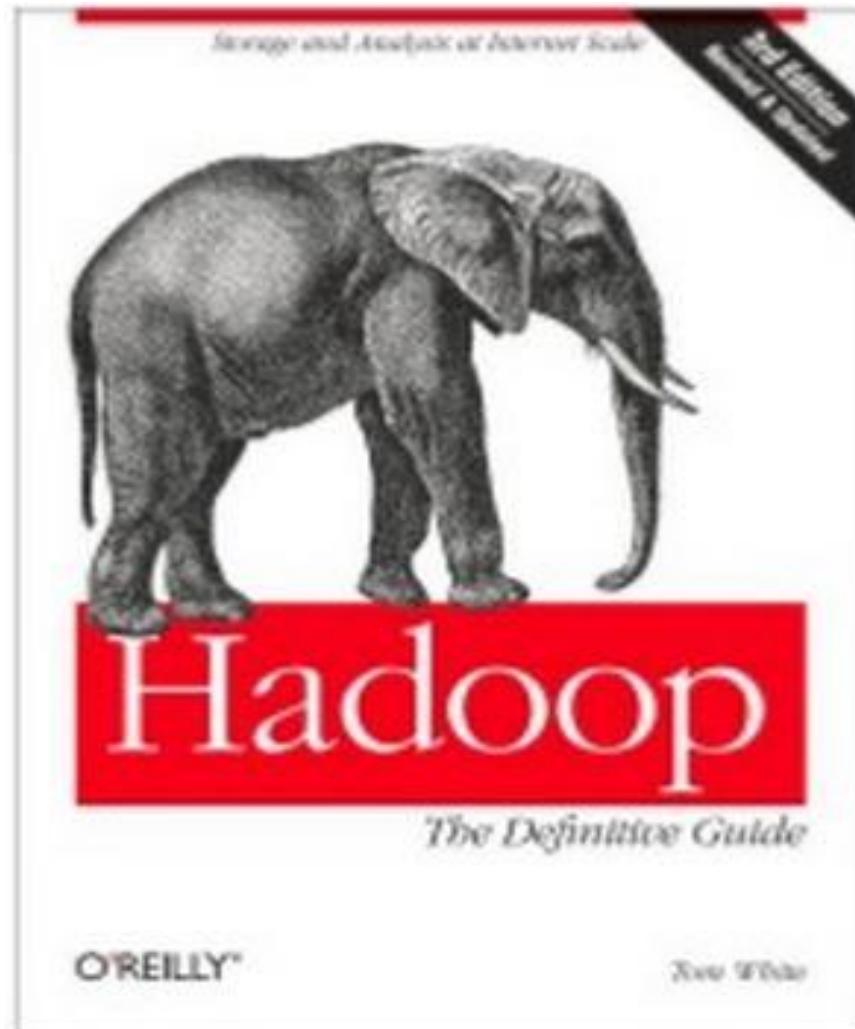
Kehadiran 5%

UTS : 25%

Tugas/Quiz/Lab : 40%

UAS : 30%

Buku Referensi



Materi

1. Pengantar Big Data
2. Big Data Technology & Infrastructure
3. Map Reduce
4. Hadoop
5. Proposal Project Big Data
6. NoSQL
7. Quiz 1 & Review *

✓ -----

1. MongoDB 1
2. MongoDB 2
3. Progress Project Big Data
4. Elastic Search 1
5. Elastic Search 2
6. Project Presentation
7. Quiz 2 & Review *

PENGANTAR BIG DATA

Sirojul Munir | rojulman@nurulfikri.ac.id | @rojulman

Definisi:: Data

- **Navathe dan Elmasri, 2000:** Data yaitu fakta yang dapat disimpan dan memiliki arti
- **Hoffer, Prescott, dan McFadden, 2005:** Data yaitu sesuatu yang mewakilkan objek dan peristiwa yang memiliki arti dan sangat penting bagi pemakai atau user
- **Data:** fakta, teks, hasil pengukuran, gambar, suara, dan video yang bernilai informasi.

Data: Raw Data

- Sumber data: event , hasil pengukuran, sistem informasi

Tabel Data Calon Ketua Kelas

No.	Nama	Turus	Banyak Pemilih
1.	Andi		10
2.	Ika		16
3.	Santi		8
4.	Rudi		6
Jumlah			40

Nama Toko Anda
Jl. Abcdefghijkl No. 1234567890
Telp. 123456789 Kota Anda
Sedia berbagai kebutuhan sehari-hari

Pelanggan : PLG0003
: Ida Ayu Setia
: Jl. Soekarno Hatta No 18
: 080000000000

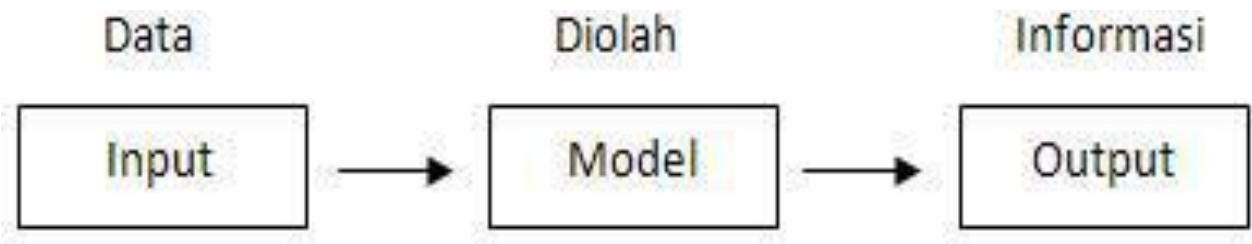
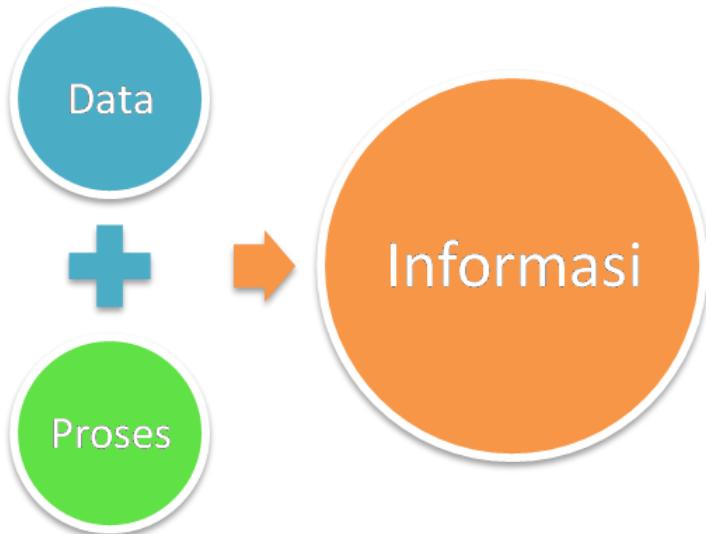
No. Transaksi : PJL20130827-0003-MAS-192944
Kasir : Master
Tanggal : 27-08-2013 19:29:44

P. Nama	Qty	Harga	Total
T Eskulin Gel 100ml Drea	5	6.468	32.340
Y Eskulin Gel 100ml Mond	6	6.600	39.600

Total-- 2 --item(s). Rp. 71.940,-
Pembulatan Rp. 60,-
GrandTotal Rp. 72.000,-
Tunai Rp. 80.000,-
Kembali Rp. 8.000,-

Belanja produk trntu mendptkan 1 poin tiap klprn 100rb
Dapatkan poin sebanyak-banyaknya
Tukarkan poin dengan hadiah yang keren!
Barang yg sudah dibeli tdk dapat ditukar/dikembalikan

Definisi:: Informasi

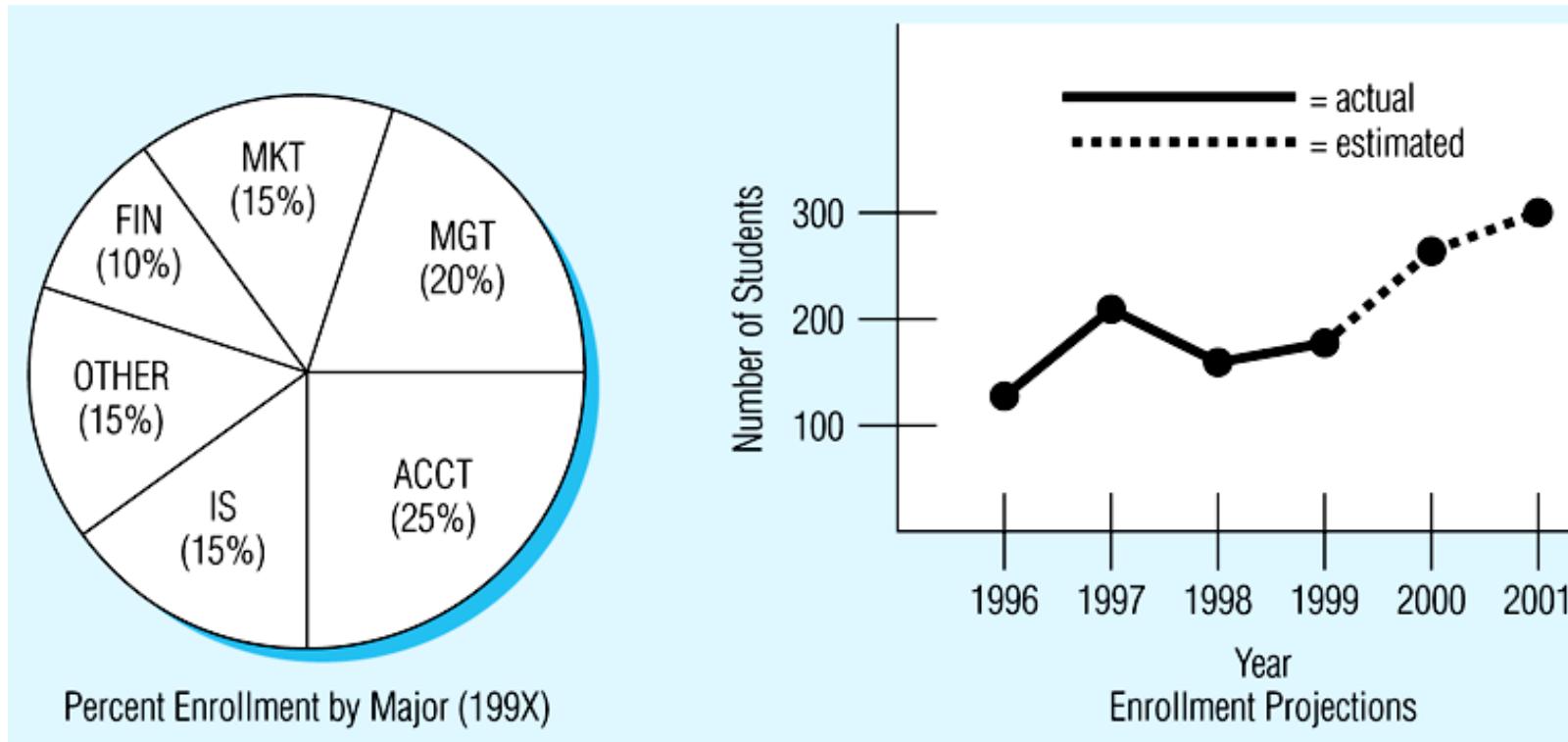


Gambar Sistem Pengolahan Data

- ❖ **Informasi:** data yang telah diproses sebagai bahan dalam proses pengambilan keputusan.

Informasi – Pengambilan Keputusan

Informasi - dapat dimanfaatkan sebagai dasar untuk pengambilan keputusan dan memahami permasalahan/situasi



Data vs Informasi

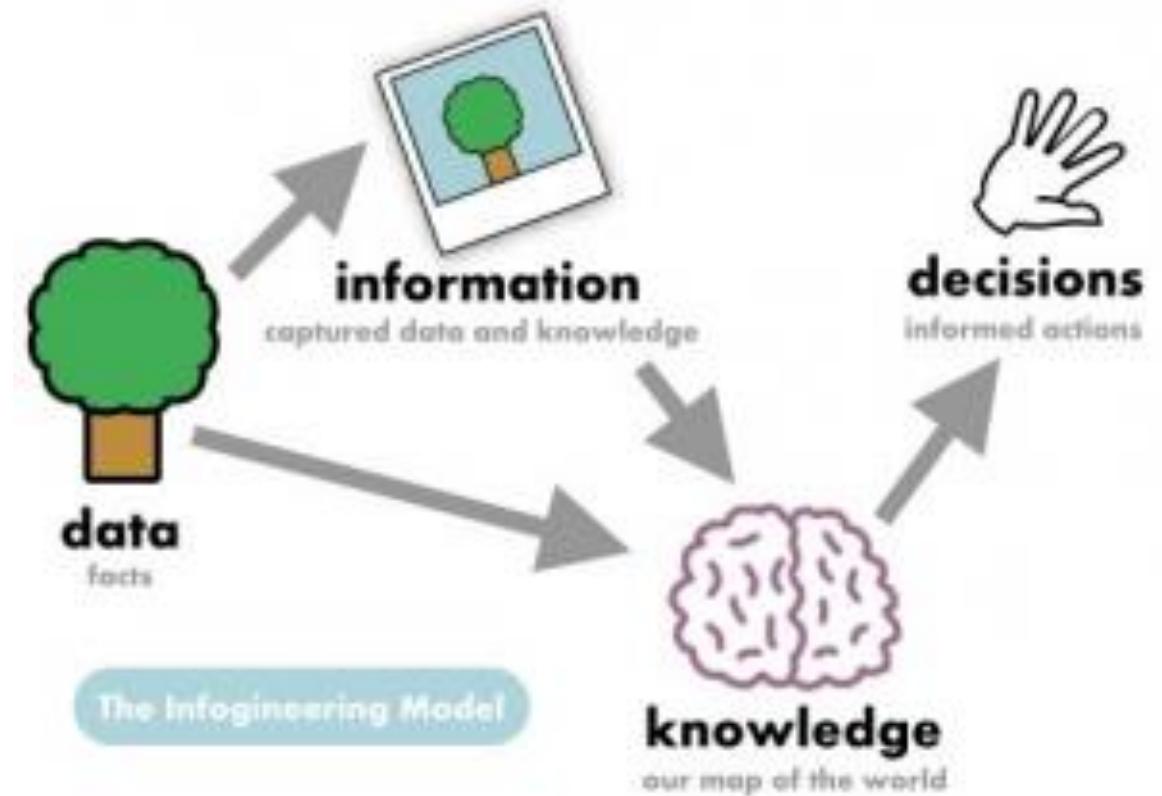
- The words "data" and "information" are often used as if they are synonyms. Nevertheless, they have different meanings.
- **Data:** Raw material from which you can draw conclusions; facts from which you can deduce new facts.
- **Information:** knowledge, intelligence, a particular piece of data with a special meaning or function. Information is often the result of combining, comparing, and performing calculations on data.

Knowledge

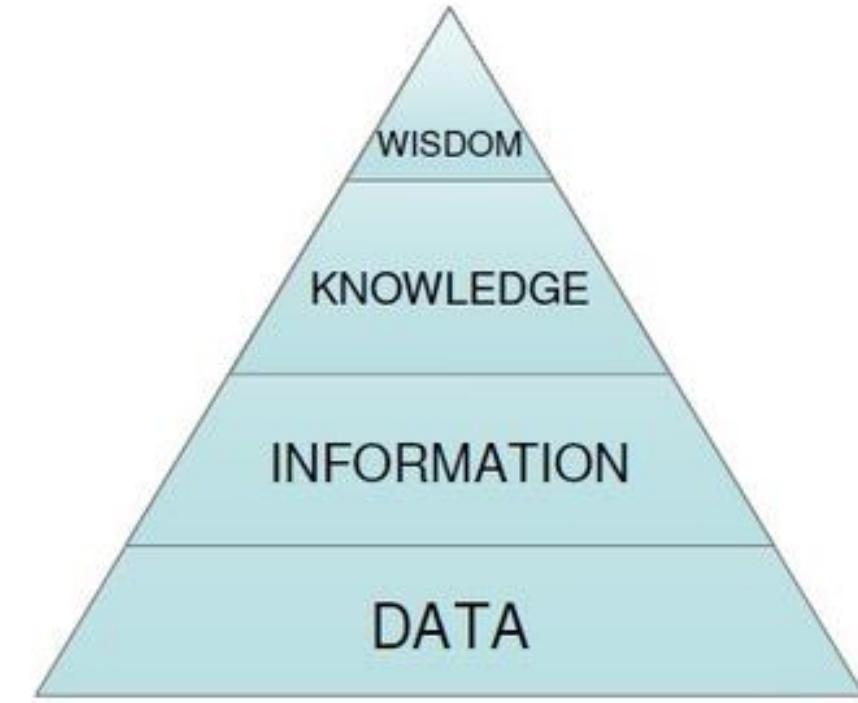
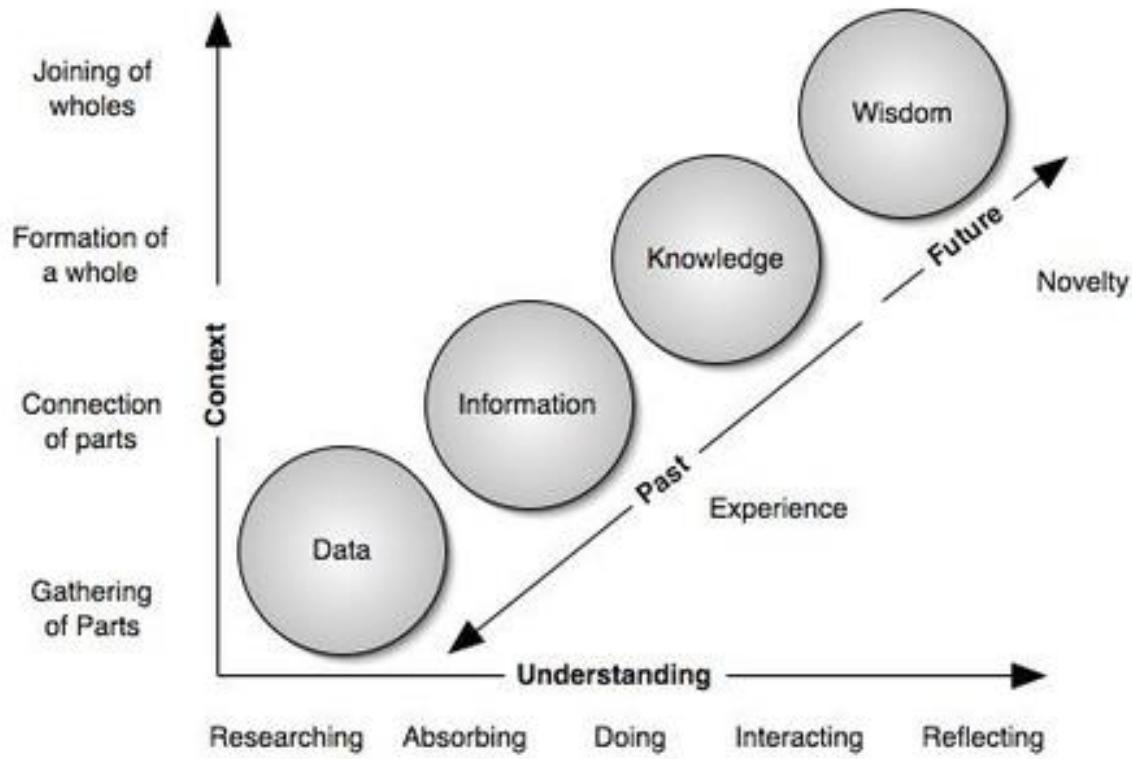
Knowledge: adalah informasi yang dilengkapi dengan pemahaman pola hubungan dari informasi disertai pengalaman, baik individu maupun kelompok dalam organisasi.

Fungsi Informasi: $I = i(D, S, T)$

- I : Informasi
- D: Data
- S : Pengetahuan awal
- T : Waktu

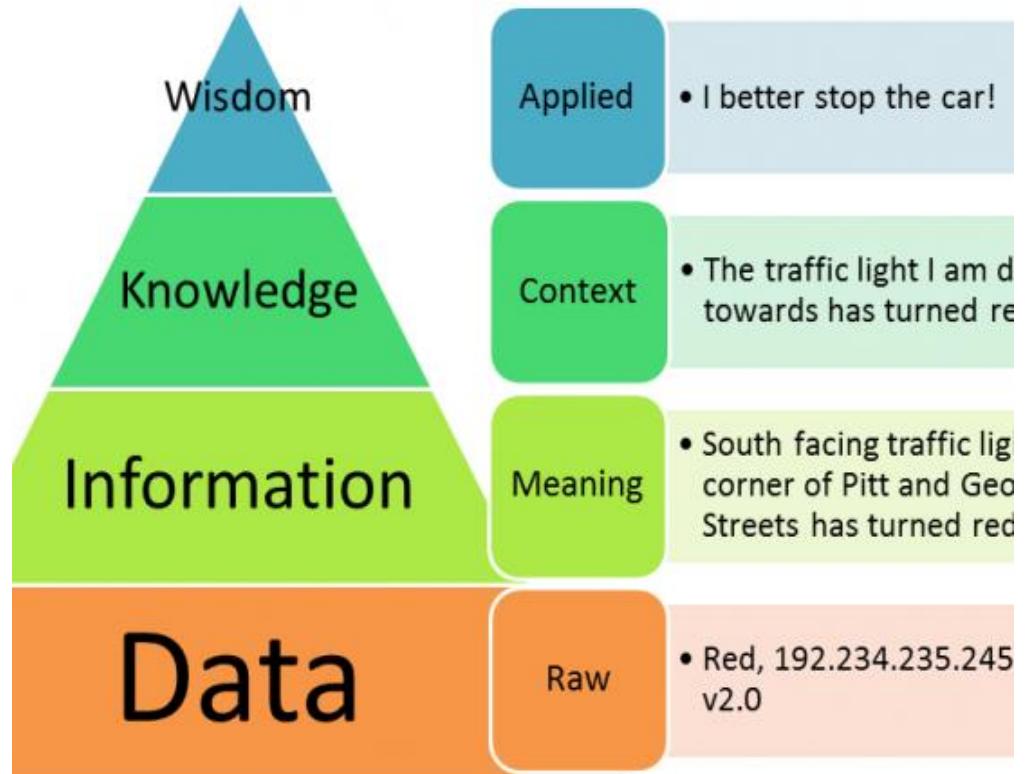


Data – Informasi – Knowledge - Wisdom



Wisdom: mengetahui kapan dan bagaimana bertindak berdasarkan pengetahuan (knowledge) untuk mencapai tujuan (goal) yang diinginkan

DIKW - Sample



Contoh lain !!?

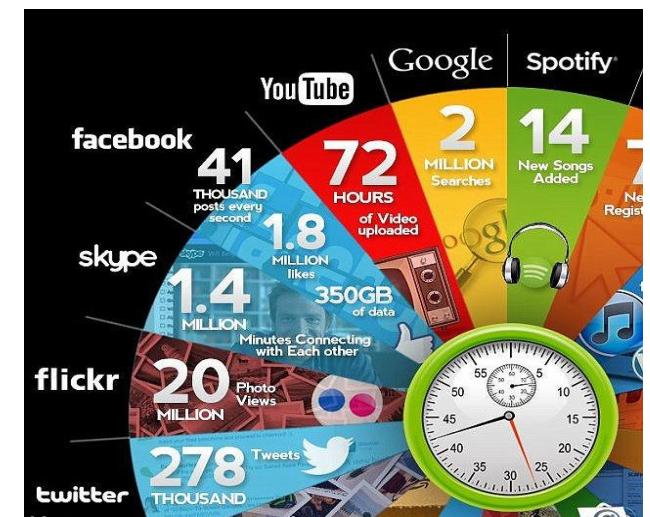
Data: Jenis Data

- *Structured data*

- *Tipe data yang dapat disimpan di database atau spreadsheet, diperlukan untuk dikelola sesuai dengan format penyimpanan standar dan ontologi, seperti : nama, alamat, telpon,*
- *Contoh : Aplikasi sistem informasi akademik, aplikasi work flow, aplikasi SDM dll*
- *Solusi Kelola: DBMS*

- *Unstructured data*

- *text, audio, imagery, video*
- *Contoh : data sistem email siswa, chat rooms, hasil questioner, video / audio di sistem e-elearning , RFID , barcode*
- *Solusi Kelola: BIGDATA (5V: Volume, Velocity, Varieaty, Veracity, Value)*



Big Data ?

- Big Data: istilah untuk **data sangat besar** dan kompleks yang tidak dapat dikelola (*capture, store, search, manage, analyze, visualize*) dengan software dan *tool* pemrograman database biasa/konvensional.
- Tidak cukup dengan SQL biasa saja (*Relational Database Management System*), sehingga butuh teknologi baru/tambahan **NoSQL (Not only SQL)**.
- Tidak hanya berisi data berstruktur/relational tapi juga (majoritas) tidak berstruktur (**unstructured**).

Where Is This “Big Data” Coming From ?

? TBs of
data every
day



12+ TBs
of tweet data
every day



25+ TBs of
log data
every day

76 million smart meters in 2009...
200M by 2014

30 billion RFID tags
today
(1.3B in 2005)



4.6
billion
camera
phones
world
wide

100s of
millions of
GPS
enabled
devices
sold
annually

2+ billion
people on
the Web
by end
2011



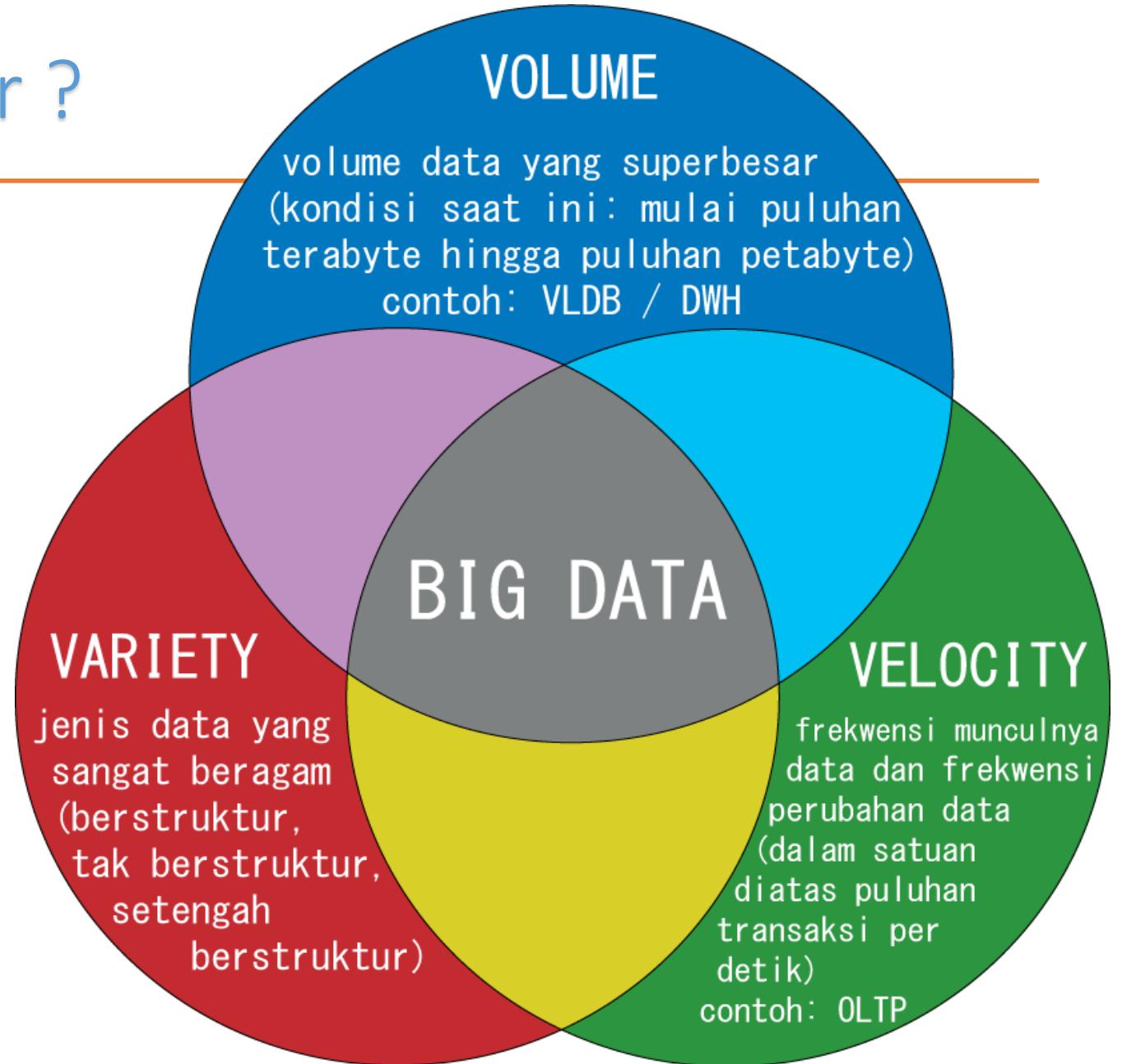
Sumber Big Data (1)

- ✓ Sebagian besar (80%, IBM) adalah *unstructured data*:
- ✓ Lebih dari 5 milyar telepon selular mengirim data sms, sekitar 1 milyar di antaranya berupa smartphone yang mengirim dan mengunduh berbagai jenis data seperti web blog, buku, email, chat, gambar, foto, video, gps, dsb.
- ✓ Lebih dari 6 milyar peralatan terhubung ke internet, a.l. server, pc, laptop, tablet, hp, peralatan kesehatan, satelit, jam tangan, sensor otomotif, dan *Internet of Things* lainnya.

Sumber Big Data (2)

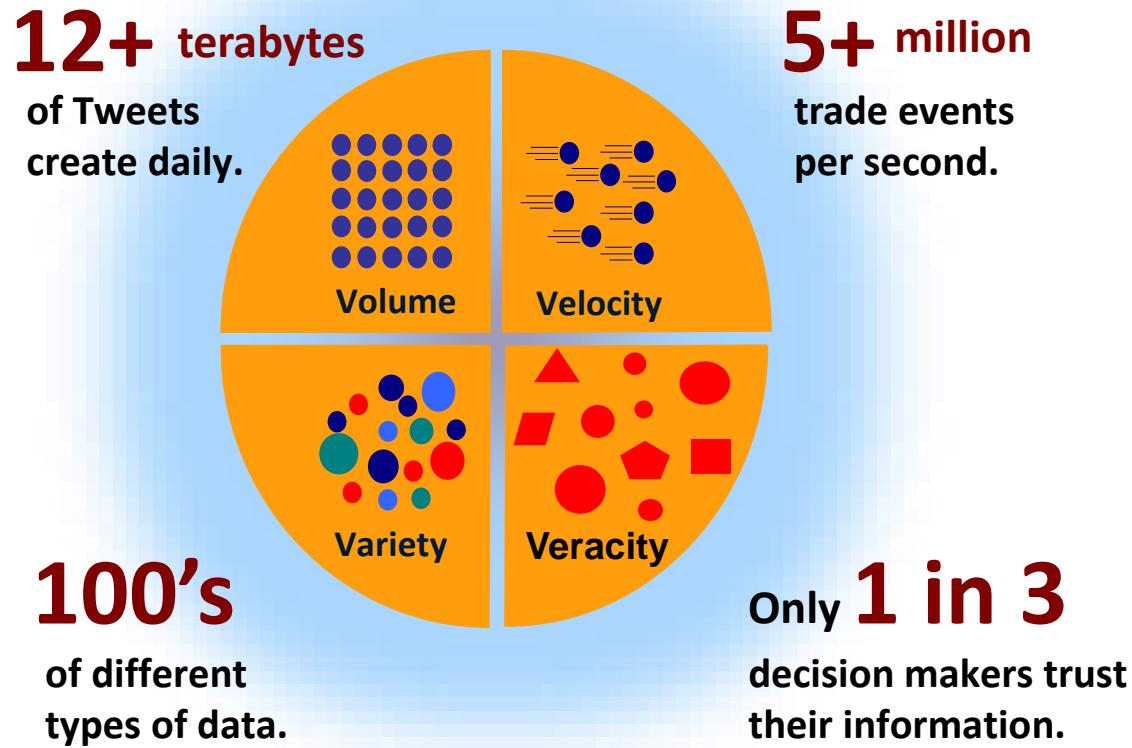
- ✓ Beberapa jenis aplikasi ini penyumbang Big Data:
 - ✓ Aplikasi web
 - ✓ e-commerce, dll.
 - ✓ Rekaman pembelian berbagai peralatan kantor dan transaksi lain di toko-toko, dll.
 - ✓ Transaksi perbankan, kartu kredit, dll.
 - ✓ Jejaring sosial: teks, gambar, suara, video, dll.
 - ✓ Streaming: audio dan video

Apa yang sangat besar ?



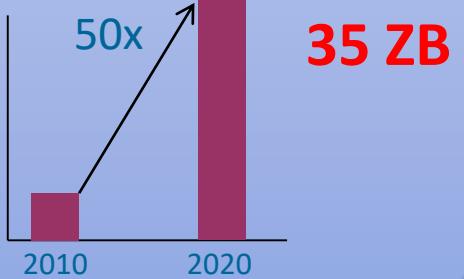
sumber: <http://vijjam.blogspot.com>

With Big Data, We've Moved into a New Era of Analytics



4 Karakteristik Big Data

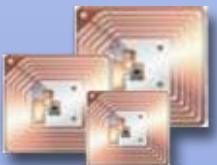
Cost efficiently processing the growing **Volume**



Responding to the increasing **Velocity**

30 Billion

RFID sensors and counting



Collectively Analyzing the broadening **Variety**



80% of the worlds data is unstructured



Establishing the **Veracity** of big data sources

1 in 3 business leaders don't trust the information they use to make decisions

يَأَيُّهَا الَّذِينَ ءَامَنُوا إِنْ جَاءَكُمْ فَاسِقٌ مُّبِينٌ فَتَبَيَّنُوا أَنْ تُصِيبُوْا قَوْمًا
بِجَهَلَةٍ فَنُصِيبُهُوْا عَلَىٰ مَا فَعَلْتُمْ نَدِيمِينَ



Wahai orang-orang yang beriman! Jika datang kepada kamu seorang fasik membawa sesuatu berita, maka selidikilah (untuk menentukan) kebenarannya, supaya kamu tidak menimpakan sesuatu kaum dengan perkara yang tidak diingini - dengan sebab kejahanan kamu (mengenainya) - sehingga menjadikan kamu menyesali apa yang kamu telah lakukan.

Kebutuhan Analisa Data



Survey Organisasi Memanfaatkan Analytics

**Studies show that organizations competing
on analytics outperform their peers**

substantially outperform

220%

1.6x
Revenue
Growth

2.5x
Stock Price
Appreciation

2.0x
EBITDA
Growth

3V – 4V – 5V

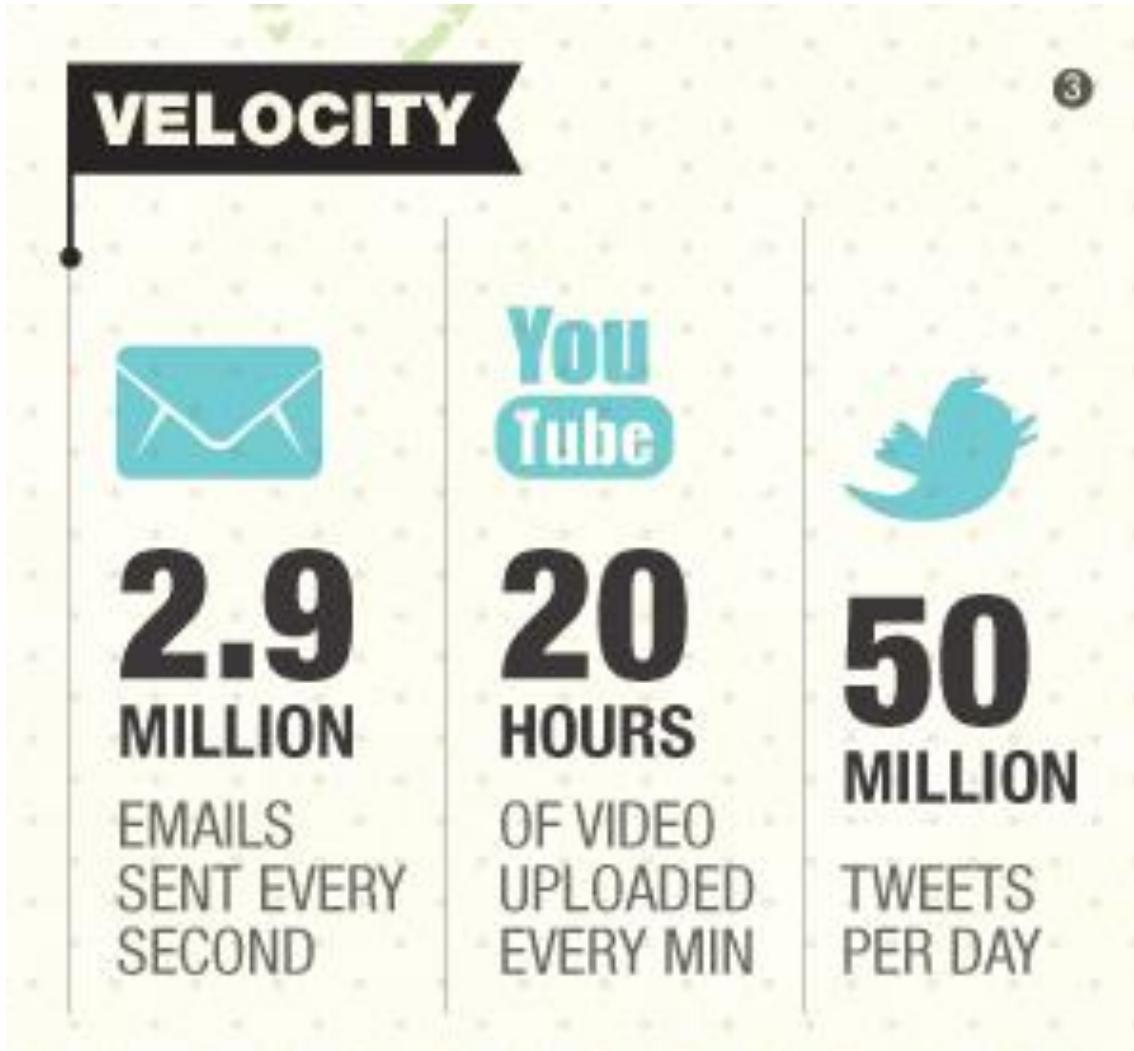
- **Volume**: Ukuran data sangat besar dari sisi jumlah yang mencapai Milyaran Terra Byte = trilyunan GB.
- **Velocity**: Kecepatan data sangat besar dari sisi kemunculan dan perubahan.
- **Variety**: Variasi jenis/tipe data sangat banyak, unstructured dan multi-structured.
- **Veracity**: Kebenaran dan keakuratan informasi yang tidak mudah dipastikan, misal salah ketik di Twitter, Berita bohong, HOAX.
- **Value**: Nilai yang dihasilkan juga sangat besar, dari sisi manfaat dalam bentuk uang maupun non uang.

Big Volume

- Volume bertambah secara eksponensial. Saat ini **2015: 4,8 Zetta Bytes = 4.800 Peta Bytes = 4,8 juta Exa Bytes = 4,8 milyar Terra Bytes = 4,8 trilyun Giga Bytes.** *)



Big Velocity



- Pertambahan (2011):
 - - Twitter 7 TB/hari
 - - Facebook 10 TB/hari
- *Sumber:*
bigdatauniversity.com

Big Variety

VARIETY



PEOPLE TO PEOPLE

NETIZENS, VIRTUAL
COMMUNITIES,
SOCIAL NETWORKS,
WEB LOGS...



PEOPLE TO MACHINE

ARCHIVES, MEDICAL
DEVICES, DIGITAL TV,
E-COMMERCE, SMART
CARDS, BANK CARDS,
COMPUTERS, MOBILES...



MACHINE TO MACHINE

SENSORS, GPS DEVICES,
BAR CODE SCANNERS,
SURVEILLANCE CAMERAS,
SCIENTIFIC RESEARCH...

Big Veracity

News / Blog / Wiki



Social Media



Mobile App



Pharmacies : Hospital & Medical System

Laboratories

Imaging Centers

Emergency Medical Services (EMS)

Hospital Information Systems

Doc-in-a-Box

Electronic Medical Records

Blood Banks

Birth & Death Records

Banking & Phone Systems



Can you hear me now?
(Heh heh heh!)



T-Mobile®

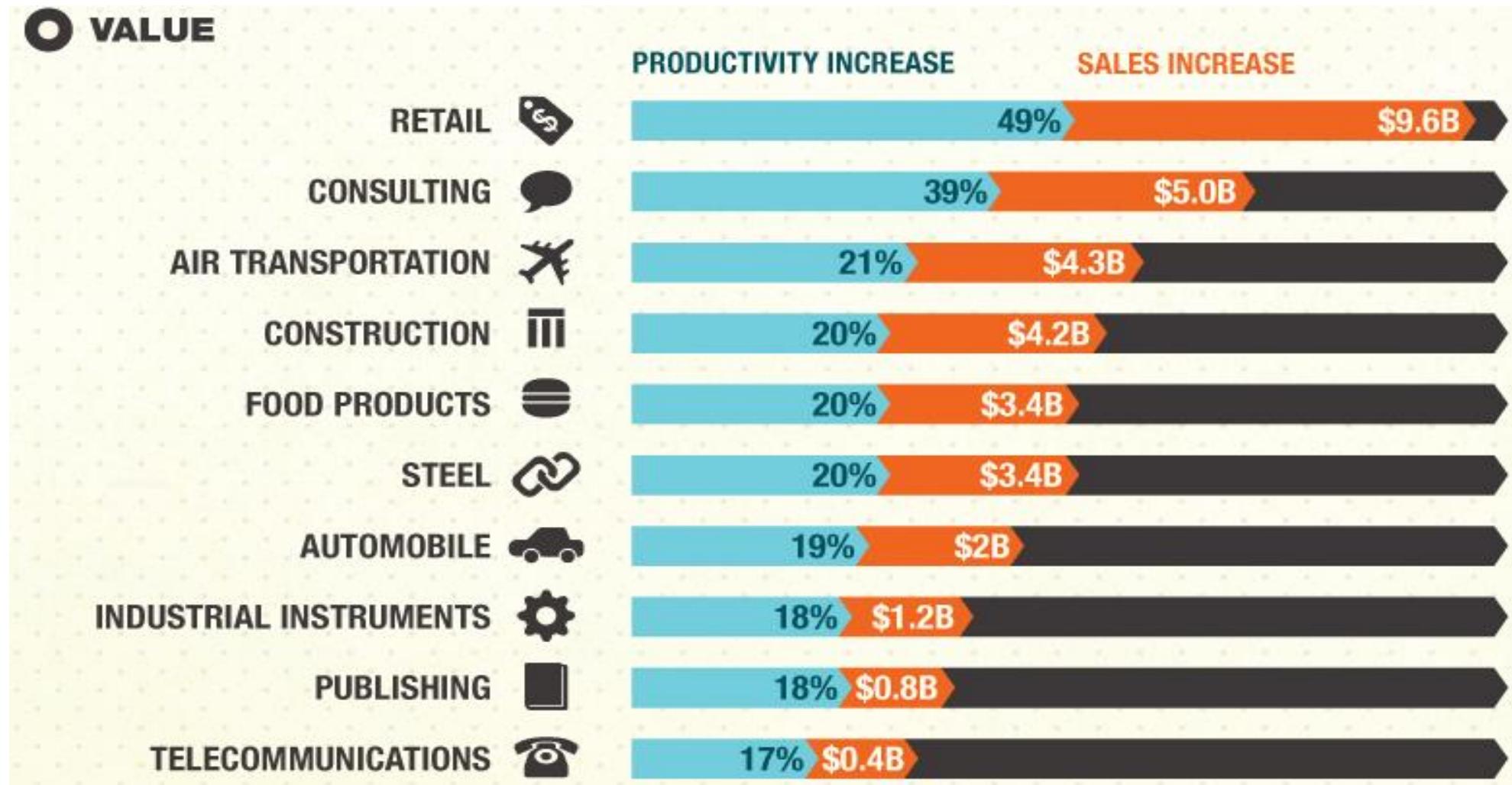
at&t



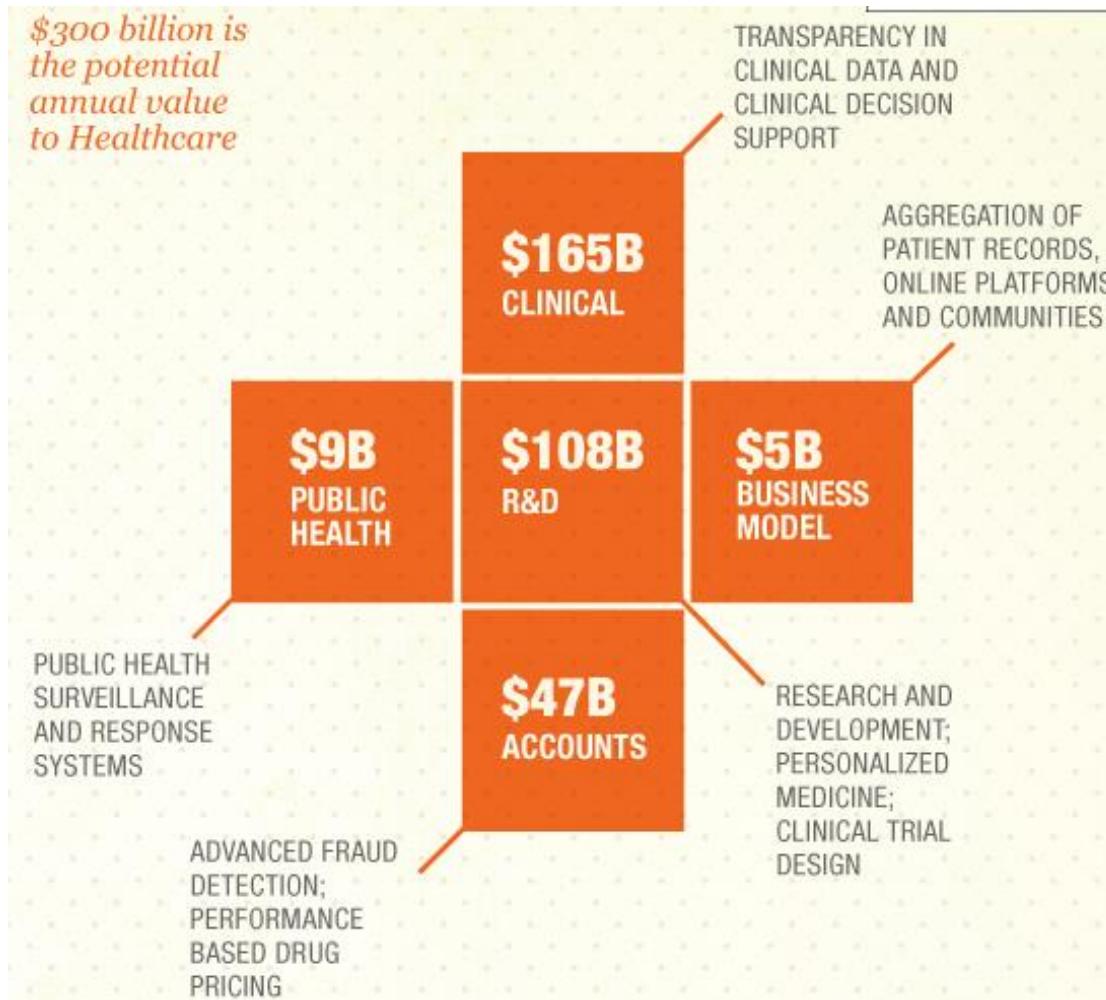
Establishing the
Veracity of big
data sources

1 in 3 business leaders don't trust the
information they use to make decisions

Big Value



Big Value (Potensi Ekonomi)



Potensi Ekonomi Big Data

- ✓ Industri kesehatan Amerika akan menghasilkan tambahan US\$ 300 billion (Rp 3.300 trilyun) per tahun, lebih besar dari APBN Negara RI.
- ✓ Pemerintah Eropa menghemat US\$ 149 billion (Rp 1.650 trilyun) per tahun.
- ✓ Kenaikan revenue Rp 1.000 trilyun bagi operator dan manfaat Rp 7.000 trilyun bagi pengguna.
- ✓ Kenaikan laba operasional retailer (USA) 60%.

Sumber: McKinsey Global Institute (2011).

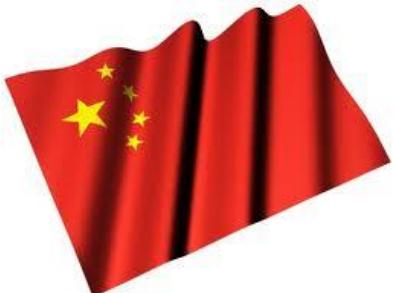
Bagaimana potensi di indonesia ?

Who is collecting all this Data ? - (1)

- Government Agencies



Department of the Treasury
Internal Revenue Service



- Big Pharmaceutical Companies



(Hey, I didn't say which government!)

Who is collecting all this data ? - (2)

- Credit Card Companies
 - What data are they getting?



Airline ticket

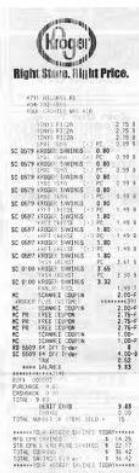


Grocery Bill

Restaurant check

1	Cab Sauv Corvina	£15.25
1	Lab Sauv Glass	£4.00
26	Net Sales Total	£73.45
1	SUDX MY D. F. FACE	£0.00
1	FISH CAKES	£4.95
2	1/2 Wings Starter	£7.90
1	MELON AND PARM HAM	£3.95
1	Calamari	£4.95
1	Garlic brd starter	£2.25
3	Meatball Starter	£17.85
1	Aub & Feta Starter	£3.95
1	Sea Bass	£12.50

Hotel Bill



Why are they collecting all this data ?

- Target Marketing
- To send you catalogs for exactly the merchandise you typically purchase.
- To suggest medications that precisely match your medical history.
- To “push” television channels to your set instead of your “pulling” them in.
- To send advertisements on those channels **just for you!**
- Targeted Information
 - To know what you need before you even know you need it based on past purchasing habits!
 - To notify you of your expiring driver’s license or credit cards or last refill on a Rx, etc.
 - To give you turn-by-turn directions to a shelter in case of emergency.

Contoh Big Data

- Walmart handles more than 1 million customer transactions ***every hour***, which is imported into databases estimated to contain more than 2.5 petabytes * of data — ***the equivalent of 167 times the information contained in all the books in the US Library of Congress.***
- FICO Credit Card Fraud Detection System protects 2.1 billion active accounts world-wide.
- The volume of business data worldwide, across all companies, ***doubles every 1.2 years***, according to estimates
- (1 Petabyte = $1000000000000000B = 1000^5 B = 10^{15} B = 1 \text{ million gigabytes}$)
- * **Think of the hard drive on your computer at home having 500 gigabytes. Now multiply that by 2,000!**

Contoh Big Data (2)

- A 2011 study predicted that roughly 1.8 zettabytes (say what? A zettabyte is equal to 1 billion terabytes. A terabyte is equal to 1024 gigabytes) of data would be generated in that year alone. That's the same amount of data that would be created ***if everyone in the U.S. posted 3 Tweets every 60 seconds for a little under 27,000 years.***
- A health care consultancy has made the data coming out of medical practices the focus of its thriving business. The company ***collects billing and diagnostic code data from 10,000 doctors on a daily, weekly and monthly basis*** to create a virtual clinical integration model. The consulting company analyzes the data to help the groups understand how well they are meeting the FTC guidelines for negotiating with health plans and whether they qualify for enhanced reimbursement based on offering a more cost-effective standard of care.

Contoh Big Data (3)

- Global position satellite technology now allows trucking firms to track their trucks - and the merchandise inside them. Practically anything you can attach an **RFID** tag to can be tracked. How a company uses that information – to re-route trucks to create efficient routes, **alert customers to deliveries**, and forecast and price services – depends on the ability to manage and analyze data effectively.

The 5 Key Big Data Use Cases



Big Data Exploration

Find, visualize, understand all big data to improve decision making



Enhanced 360° View of the Customer

Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information sources



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis

Analyze a variety of machine data for improved business results



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency

How Can You Avoid Big Data ?

- Pay cash for everything!
- Never go online!
- Don't use a telephone!
- Don't use credit cards!
- Don't fill any prescriptions!
- Never leave your house!

Peluang Kerja Big Data

- 2011: Di Amerika saja butuh 190.000 tenaga ahli baru di bidang Big Data.
- 2018: Diproyeksikan Amerika butuh 490.000 tenaga ahli baru di bidang Big Data.
- Lowongan kerja di seluruh dunia untuk Big Data pada 2015 sekitar 4,4 juta orang.

Sumber: McKinsey Global Institute (2011).

Peluang Bisnis/Kerja di Indonesia

- Peluang besar perusahaan dan pemerintahan butuh TI untuk Big Data, dengan pilihan hemat memilih produk Open Source seperti Linux, berbagai produk Apache, pemrograman Java, dll. Sehingga kebutuhan **SDM Open Source, Cloud, dan Big Data meningkat**.
- **Kebutuhan Cloud Computing akan terus bertambah besar** (sumber: IBM) akibat kebutuhan Big Data, karena lebih murah dan mudah dalam perawatan, ditambah lagi semakin banyak jumlah pengguna smartphone dan tablet (di atas 250 juta).

Mulai dari mana ?

- Big Data, Cloud Computing, dan teknologi terbaru, termasuk produk-produk Open Source belum banyak diajarkan di lembaga pendidikan formal, padahal dunia industri dan pemerintah memakainya.
- Mulai meningkatkan ilmu dan skill Big Data, misal dengan bergabung ke komunitas IdBigData, Klub Studi Linux, KPLI (Kelompok Pengguna Linux Indonesia), Komunitas BlankOn, dan lain-lain.
- Free E-learning: Bigdatauniversity.com.

Bidang yang perlu di kuasai mahasiswa

- Sistem operasi **Linux** dengan berbagai software untuk jaringan, virtualisasi, clustering, cloud.
- Aplikasi untuk **Cloud**: web Apache, database MySQL, pemrograman PHP, Java, Framework PHP/Java, dan paket-paket sistem informasi.
- Aplikasi untuk **Big Data**: Apache Hadoop, YARN, Cassandra, MongoDB, Hive, Pig, Python, R, dll.
- Mobile Application : Android , Mobile Framework Programming, GeoLocation, JavaScript, NodeJs, Web Service
- Linux untuk **Internet of Things**: smartphone, tablet, multimedia, kontrol, sensor, robot, dll.

Quiz !!

- Jelaskan 5 karakteristik Big Data (5V)
- Sebutkan Contoh Big Data disekitar anda (yang anda gunakan sehari2) !!
(3 contoh)
- Apakah teknologi Big Data dapat mengatasi berita HOAX ? Beri opini anda