



Introduction to Big Data

Ahmad Rio Adriansyah S.Si. M.Si

Instruktur

- Nama : Ahmad Rio Adriansyah
- Email :
- HP : 081573954126
- Panggilan : Rio / Arasy
 - Matematikawan,
 - Programmer,
 - Dosen,
 - Tech. Enthusiast

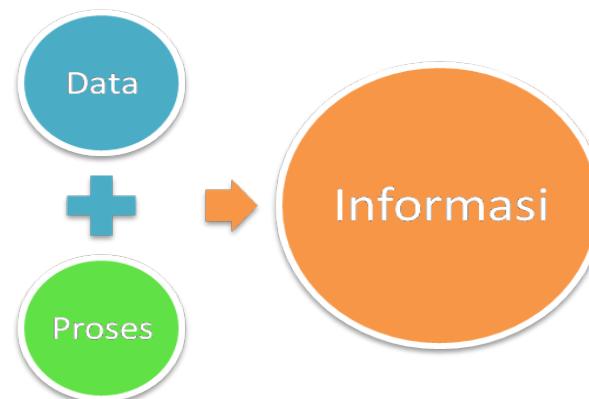


Peserta

- Nama ?
- Asal Institusi dan pekerjaan ?
- Pengalaman dengan bahasa pemrograman ?
- Pengalaman dengan big data dan data analisis ?

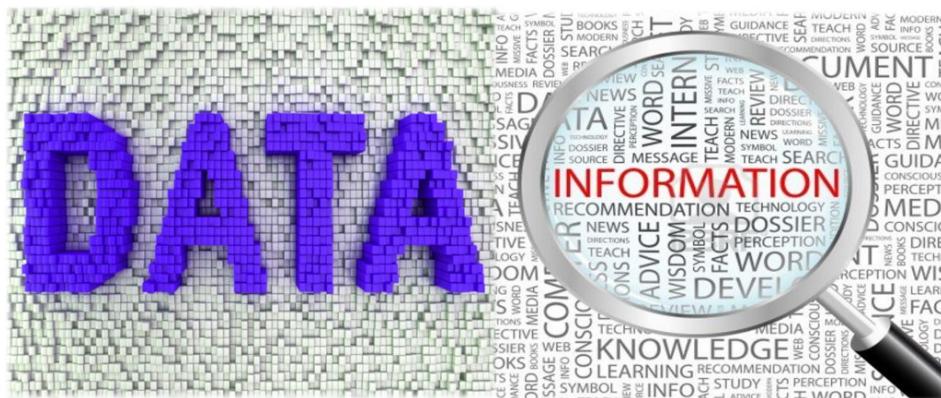
Data

- Facts and statistics collected together for reference or analysis.
- Factual information used as a basis for reasoning, discussion, or calculation.



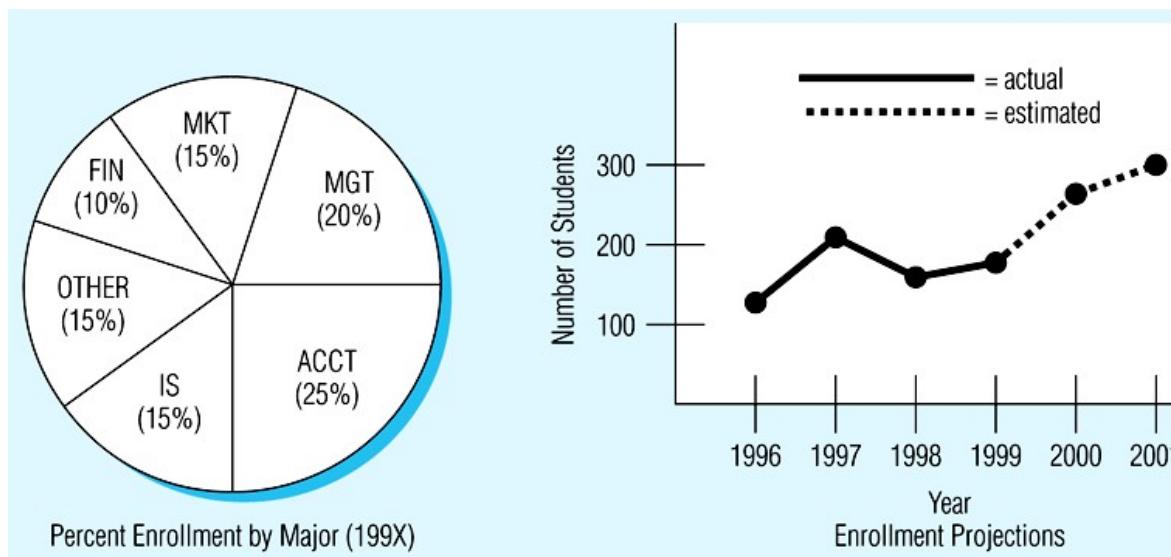
Information

- Facts provided or learned about something or someone.
- Knowledge obtained from investigation, study, or instruction.



Information

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



Data vs Information

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.

Jenis Data

- Structured

Tipe data yang dapat disimpan di database atau spreadsheet, diperlukan untuk dikelola sesuai dengan format penyimpanan standar dan ontologi, seperti : nama, alamat, telpon

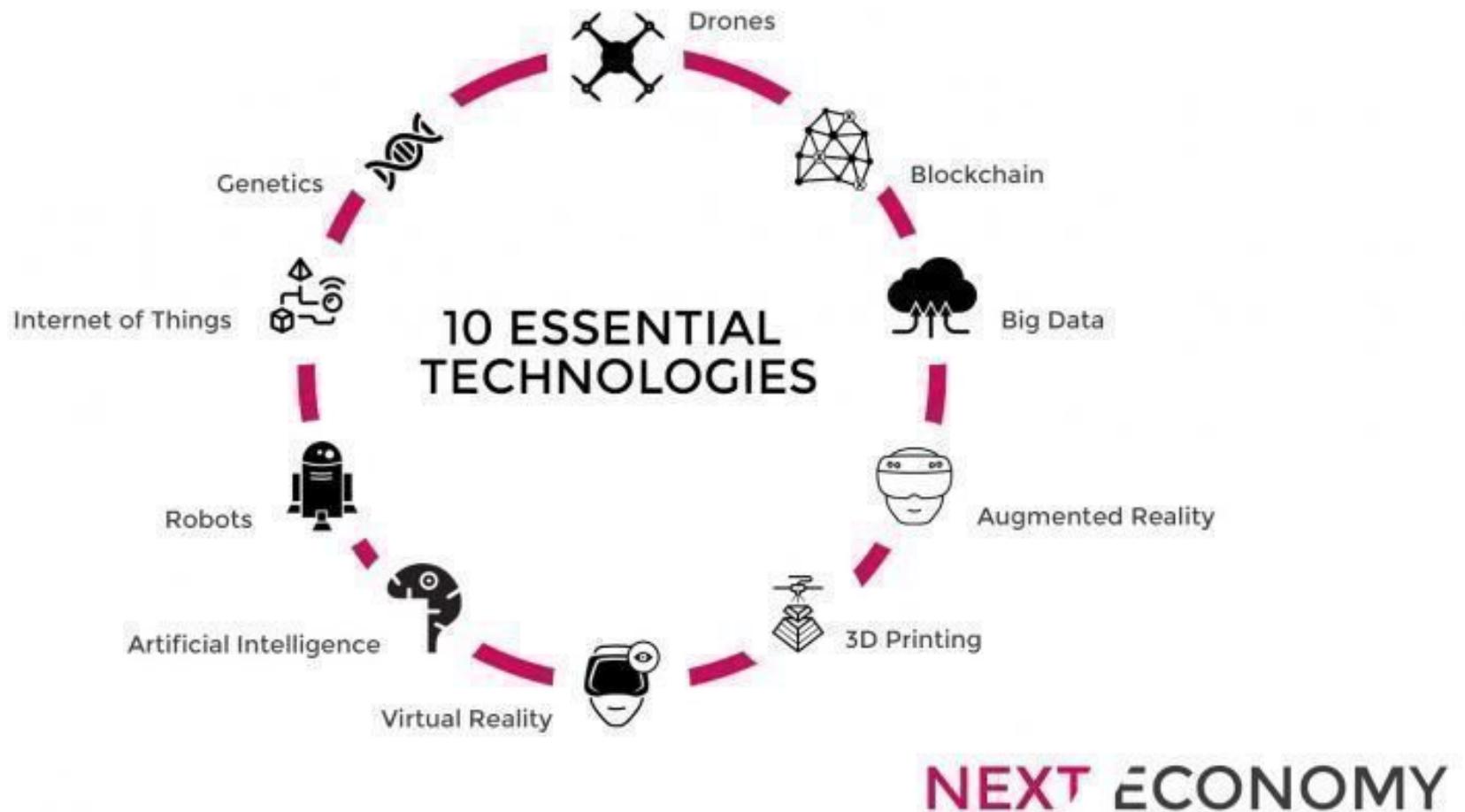
- Unstructured

text, audio, imagery, video, dan data tidak terstruktur lainnya

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 hanya berisi data berstruktur/*relational* tapi juga (majoritas) tidak berstruktur (***unstructured***).

Big Data



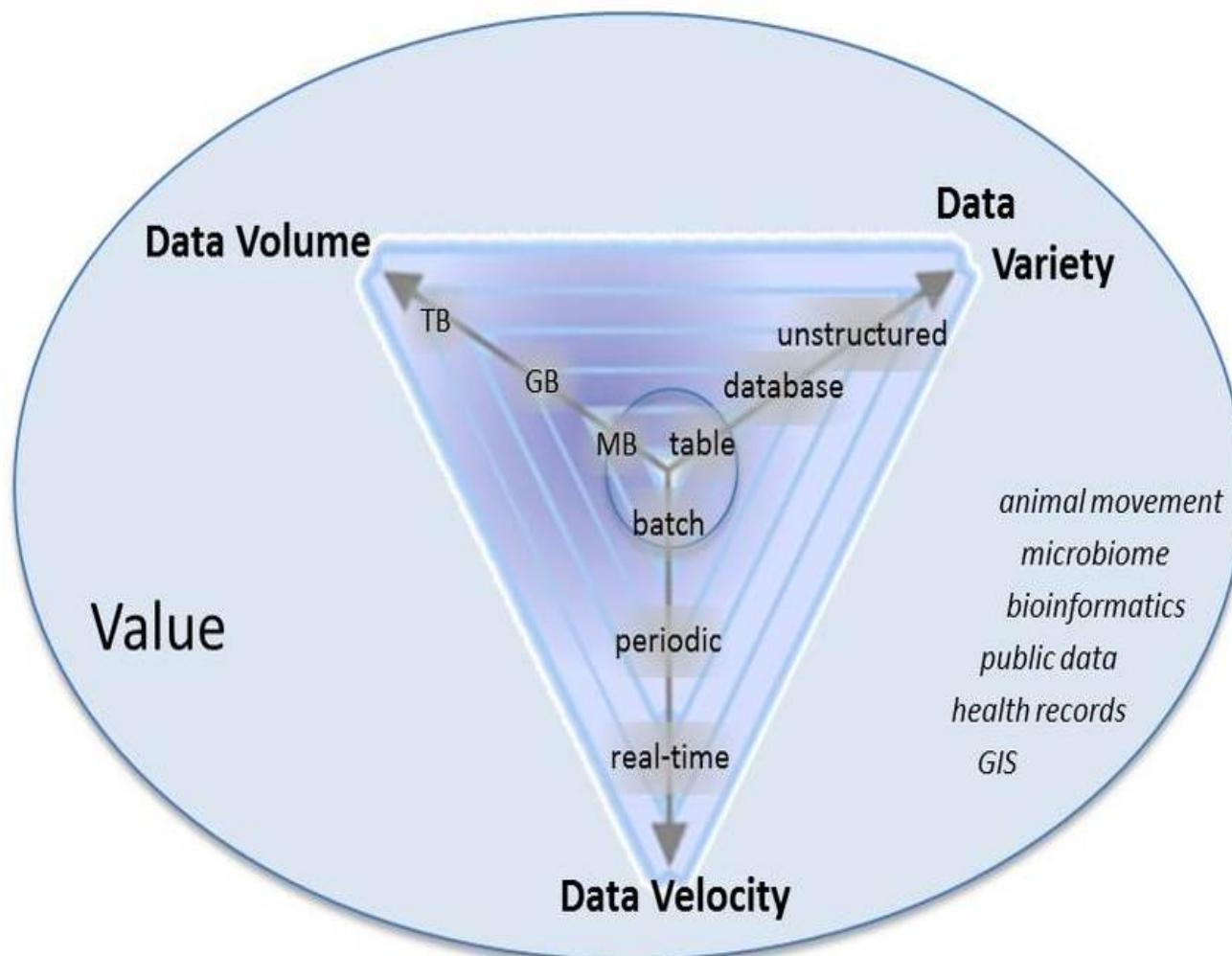
Big Data Analysis



- | | | | | | |
|---|---|--|---|--|--|
| <ul style="list-style-type: none">• Shelf-space optimization• Pricing• Product Bundling• Promotions• Up-sell / Cross-sell | <ul style="list-style-type: none">• Risk Modeling• Fraud Prediction• Customer Segmentation• Portfolio Optimization | <ul style="list-style-type: none">• Market Basket Analytics• Marketing Mix Analysis• Product Pricing• Sales Forecasting• A/B Testing | <ul style="list-style-type: none">• Clinical Trials of New Drugs• A/B Testing• Genetics Analysis• Epidemic Forecasting & Control | <ul style="list-style-type: none">• Customer Acquisition Strategies• Churn Analytics and Control• Up-Sell/ Cross-Sell• Product Bundling | <ul style="list-style-type: none">• A/B Testing• Recommendation Engine• Media Mix Analytics• Advertisements Optimization• Customer Acquisition |
|---|---|--|---|--|--|

Analytics is used by almost all industries and functional areas

Apa yang Big? (3V)

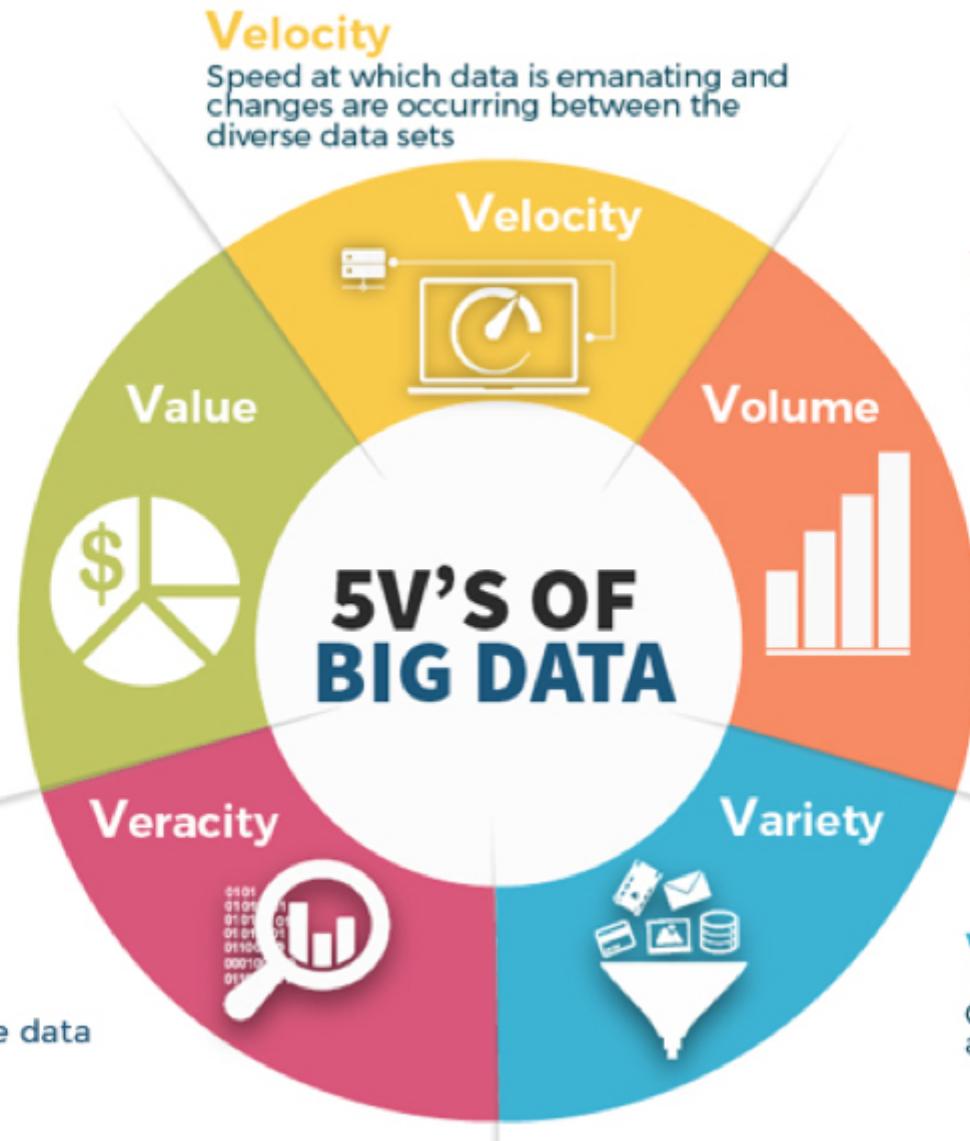


Value

Having access to big data is all well and good but that's only useful if we can turn it into a value.

Veracity

Data reliability and trust.
Verifying and validating the data



Volume

This refers to the sheer volume of data being generated every second.

Variety

Can use structured as well as unstructured data.

Sejarah Big Data

BIG DATA PHASE 1	BIG DATA PHASE 2	BIG DATA PHASE 3
Period: 1970-2000 DBMS-based, structured content: <ul style="list-style-type: none">• RDBMS & data warehousing• Extract Transfer Load• Online Analytical Processing• Dashboards & scorecards• Data mining & statistical analysis	Period: 2000-2010 Web-based, unstructured content <ul style="list-style-type: none">• Information retrieval and extraction• Opinion mining• Question answering• Web analytics and web intelligence• Social media analytics• Social network analysis• Spatial-temporal analysis	Period: 2010-present Mobile and sensor-based content <ul style="list-style-type: none">• Location-aware analysis• Person-centered analysis• Context-relevant analysis• Mobile visualization• Human-Computer-Interaction

Big Data Sekarang

? TBs of
data every
day



25+ TBs of
log data
every day

12+ TBs
of tweet data
every day

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



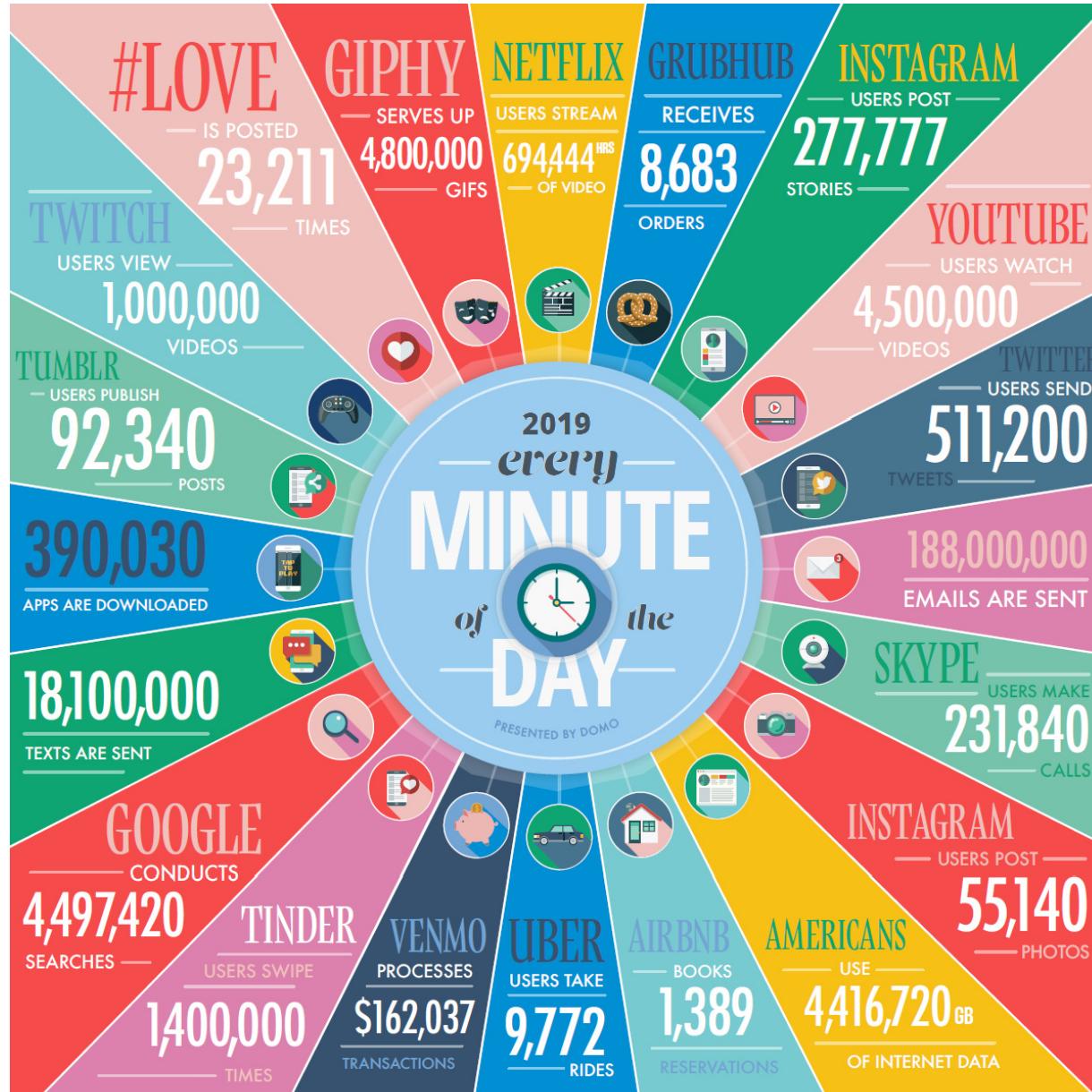
4.6
billion
camera
phones
world
wide

100s of
millions of
GPS
enabled
devices
sold
annually

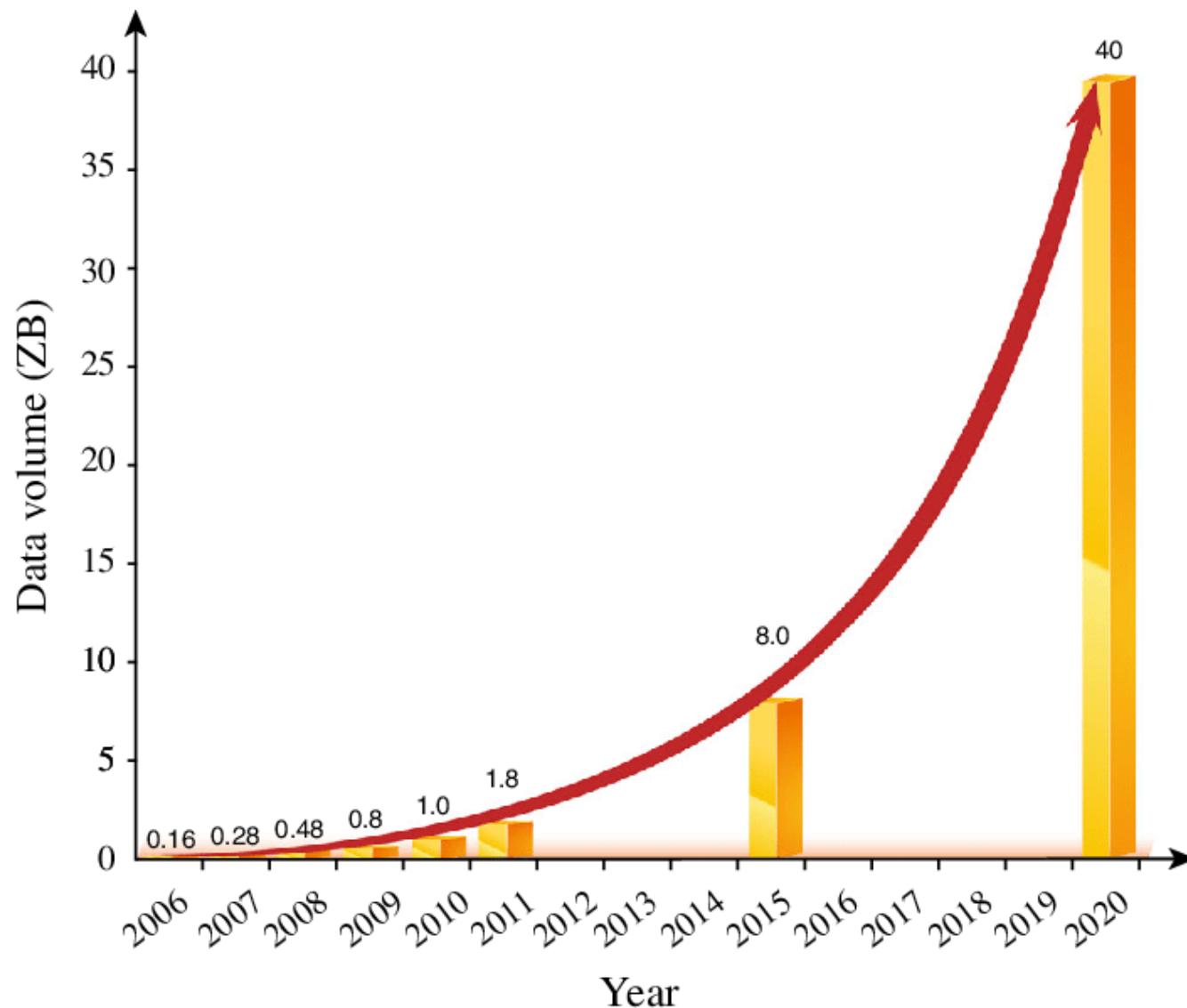
2+ billion
people
on the
Web by
end 2011



Big Data dalam Semenit



Volume Data



Rekam Jejak Data

Data Footprint of Humans



44

zettabytes

Projected volume of
global IT traffic by
2020

40

zettabytes

Volume of data
created by 2020, up
300% from 2015

2.3

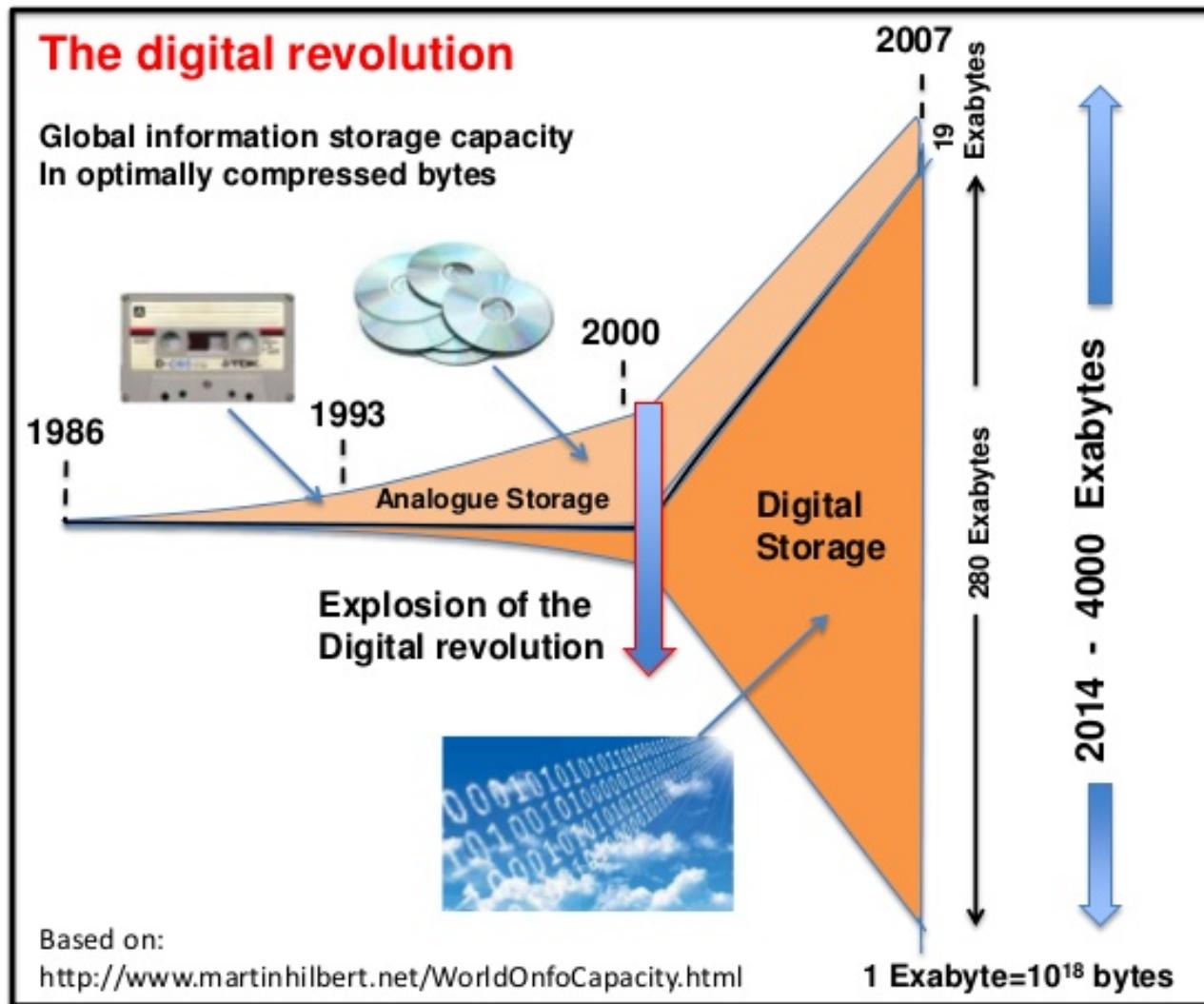
zettabytes

Volume of data that
humans produce
every day

Source: IBM, Grazziti

 FinancesOnline
REVIEWS FOR BUSINESS

Revolusi Digital



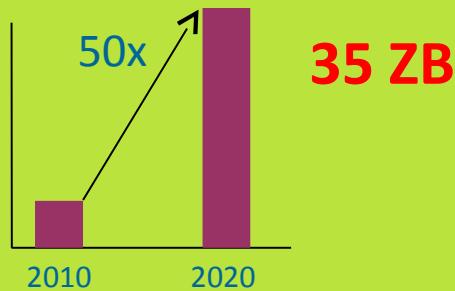
Big Data Statistics

State of Industry Statistics

- \$189.1 billion – projected worldwide revenues for big data and business analytics (BDA) solutions for 2019 (IDC)
- 12% – rate of increase for big data and business analytics from 2018 to 2019 (IDC)
- \$274.3 billion – projected worldwide revenues for big data and business analytics (BDA) solutions by 2022 (IDC)
- 13.2% – projected compound annual growth rate (CAGR) of BDA within the five-year period 2018-2022 (IDC)
- \$77 billion – projected big data market size by 2023 (Entrepreneur)
- \$103 billion – projected value of the big data analytics market by 2023 (TechJury)
- 20% – projected growth rate of the big data market for 2019 (TechJury)

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**



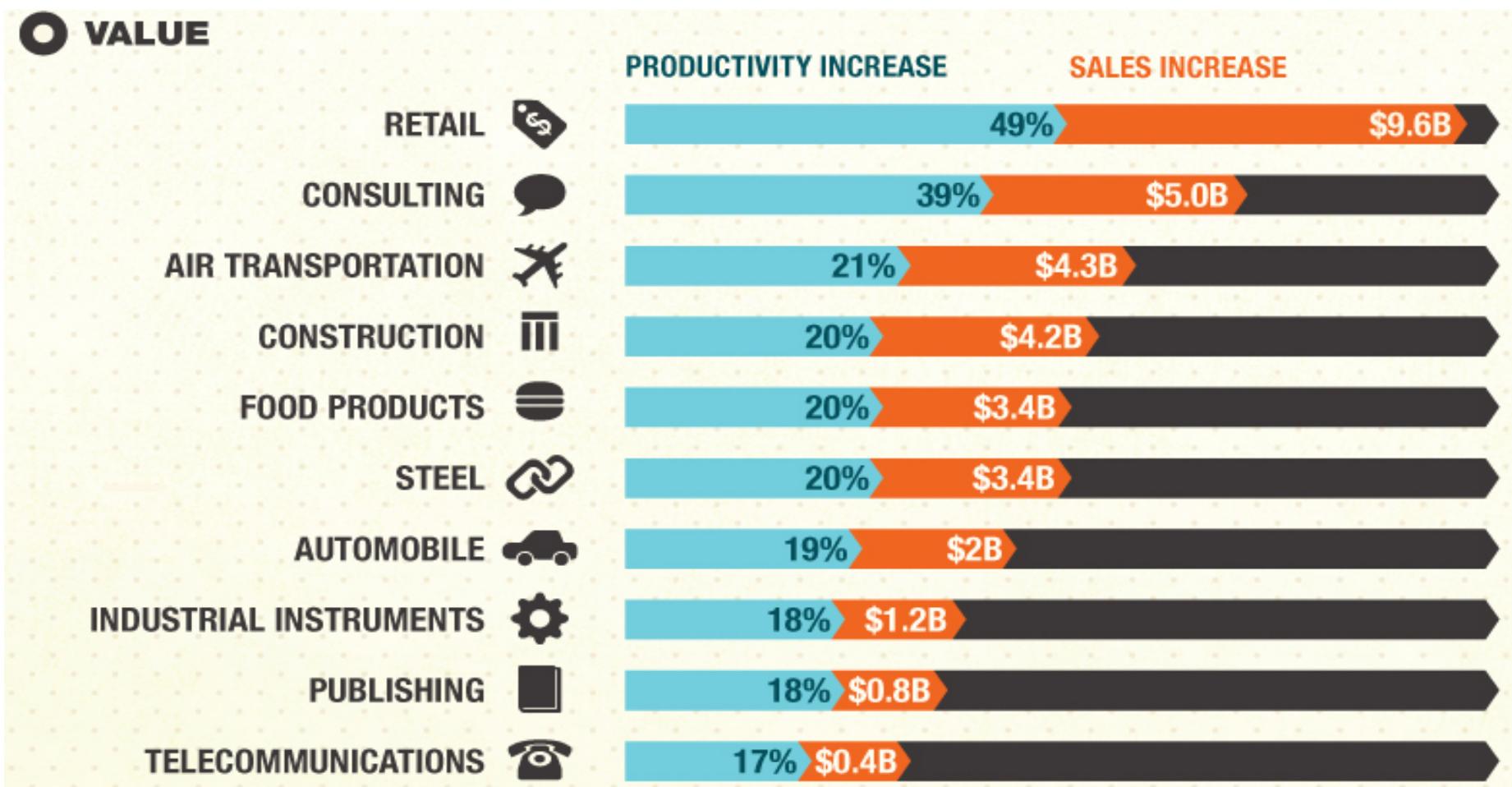
80% of the worlds data is unstructured



Establishing the **Veracity** of big data sources

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

Big Value



Big Value

Alibaba

\$120 billion
assets under
management by
Yu'E Bao¹

175 million
total Alipay
transactions in
one day²

44%
of global mobil-
wallet spending
achieved by Alipay³

Baidu

346 million
online users

130 million
users of Ping An
Good Doctor⁴

25 million
unique visitors daily
to autohome.com.cn

Tencent

889 million
we char users⁵

70 minutes
spend every day
by average
WeChat user⁶

61%
of users open
WeChat more than
ten times every day⁷

46 billion
“red packets” sent
via WeChat for the
lunar new year⁸

1 As of September 2016

2 As of August 2016

3 In 2016; see Global Payments Report 2016, Worldpay, November 8, 2016, worldpay.com

4 As of March 2017

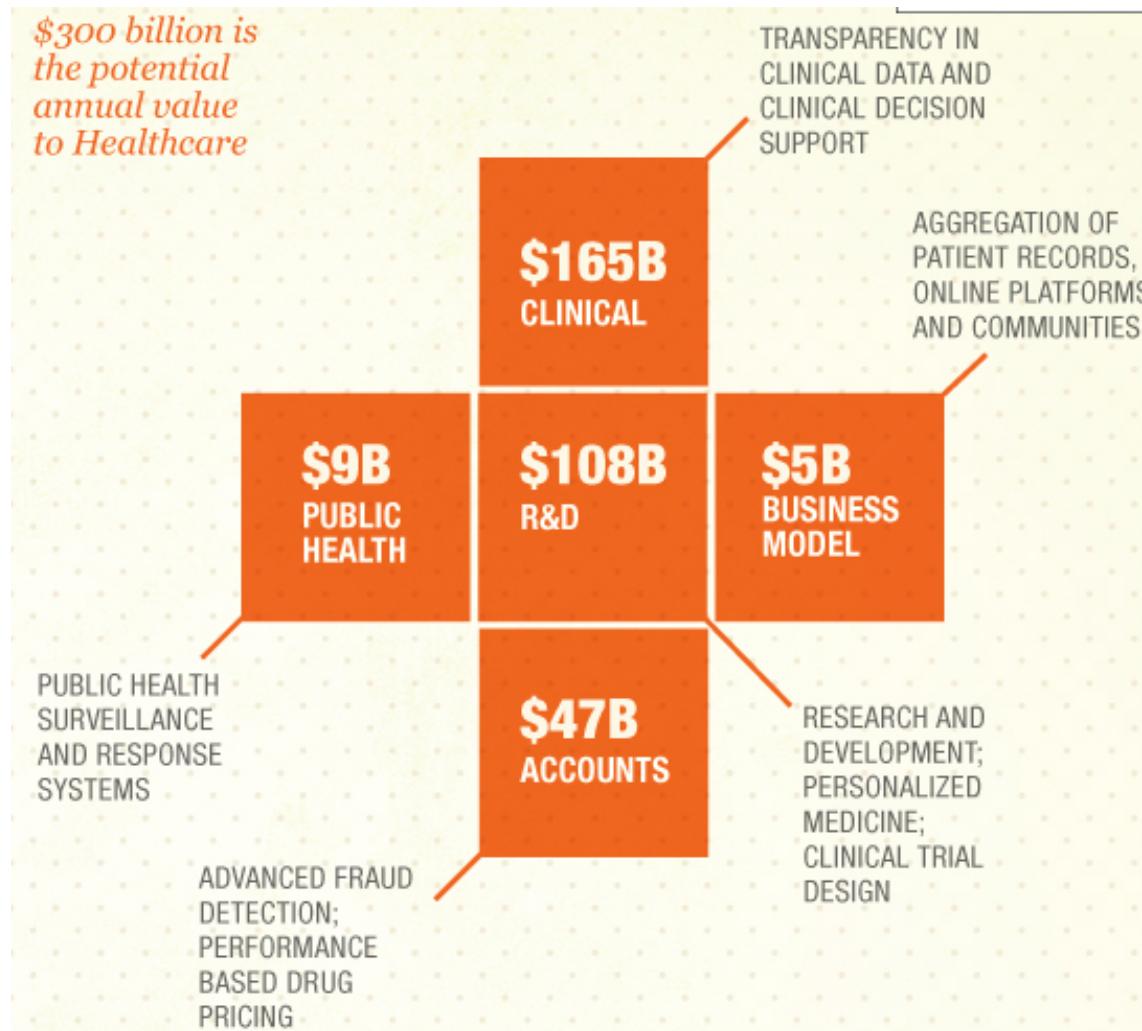
5 As of Q4 2016

6 As of March 2016

7 As of June 2016

8 For Lunar New Year falling in 2017; see “WeChat users send 46 billion digital red packets over Lunar New Year – Xinhua,” Reuters, February 6, 2017, reuters.com

Big Value



Big Value

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).

Big Data Use Case



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

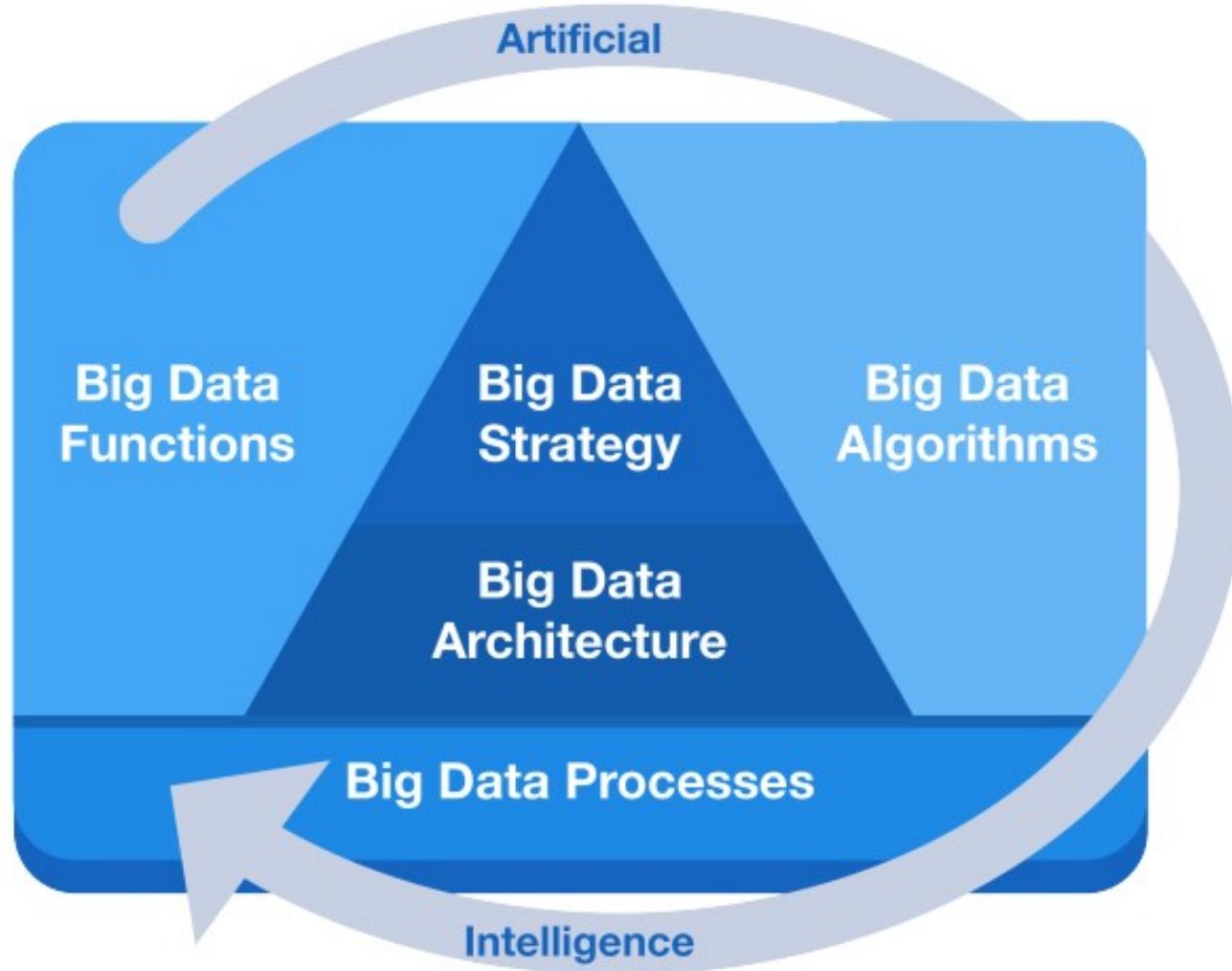
Analyze a variety of machine data for improved business results



Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency

Big Data Framework



Disclaimer

Kita akan lebih banyak menggunakan python sebagai bahasa pemrograman pengantar

Yang akan dipelajari lebih ke penggunaan toolsnya, bukan ke analisis mendalam terhadap datanya

- HADOOP
- Spark
- MongoDB
- Elastic Search