



# Data Warehouse

---

Sirojul Munir S.SI, M.KOM – Semester Genap TA 20182



---

## PENGANTAR BIG DATA – DATA WAREHOUSE

# Jenis Data

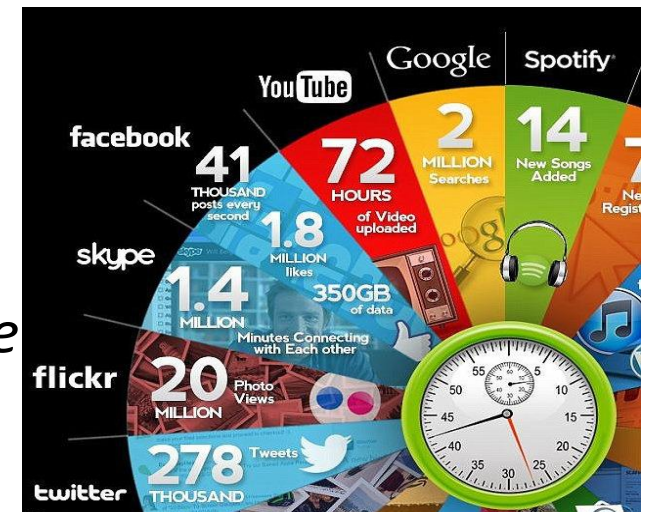
- **Structured data**

- 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-learning , 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 (mayoritas) tidak berstruktur (*unstructured*).

# Where Is This “Big Data” Coming From ?

**12+ TBs**  
of tweet data  
every day

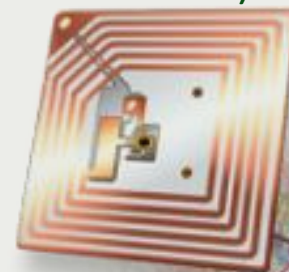


? TBs of  
data every  
day



**25+ TBs** of  
log data  
every day

**30 billion** RFID tags  
today  
(1.3B in 2005)



**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



# 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.

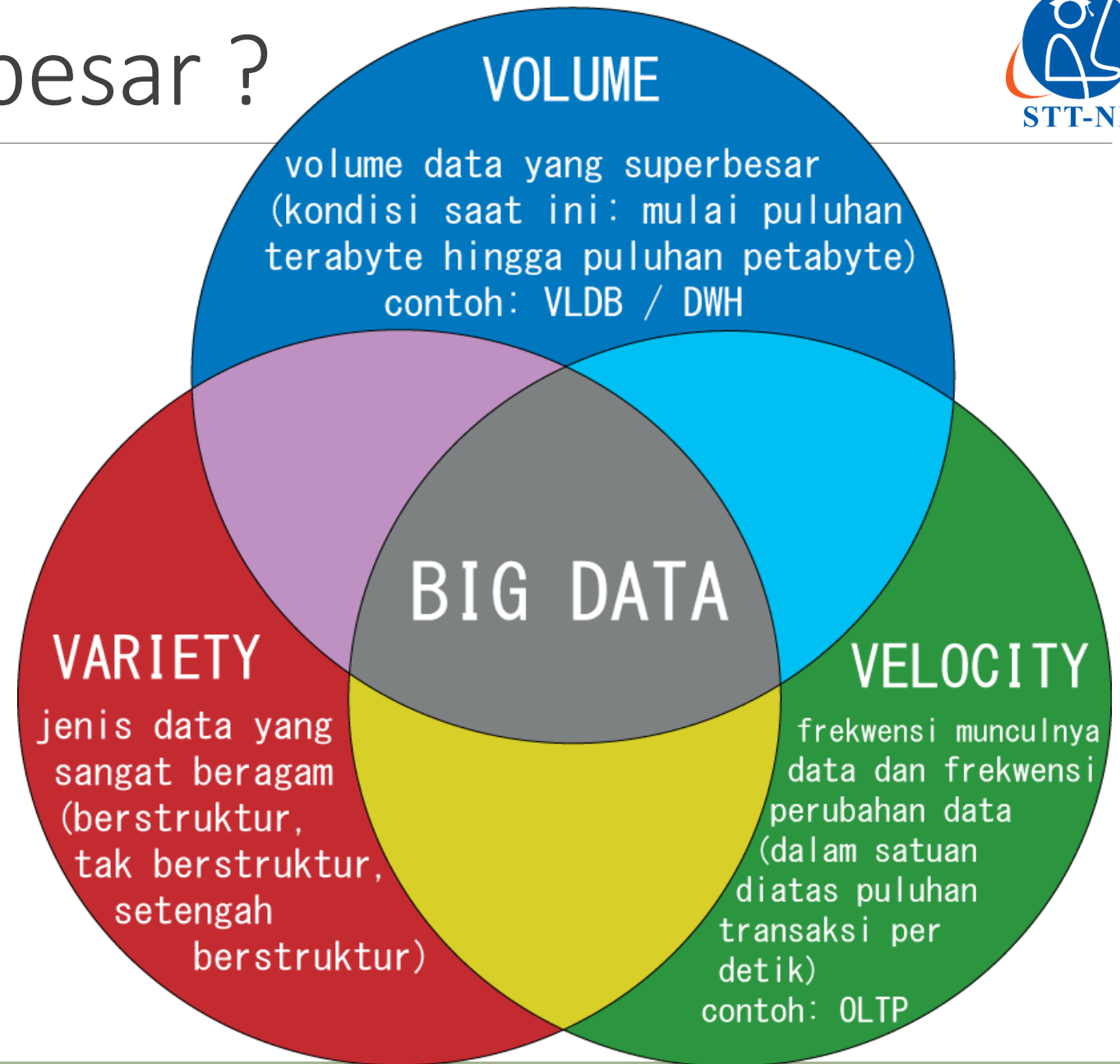
# Sumber Big Data (2)

---

- ✓ 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



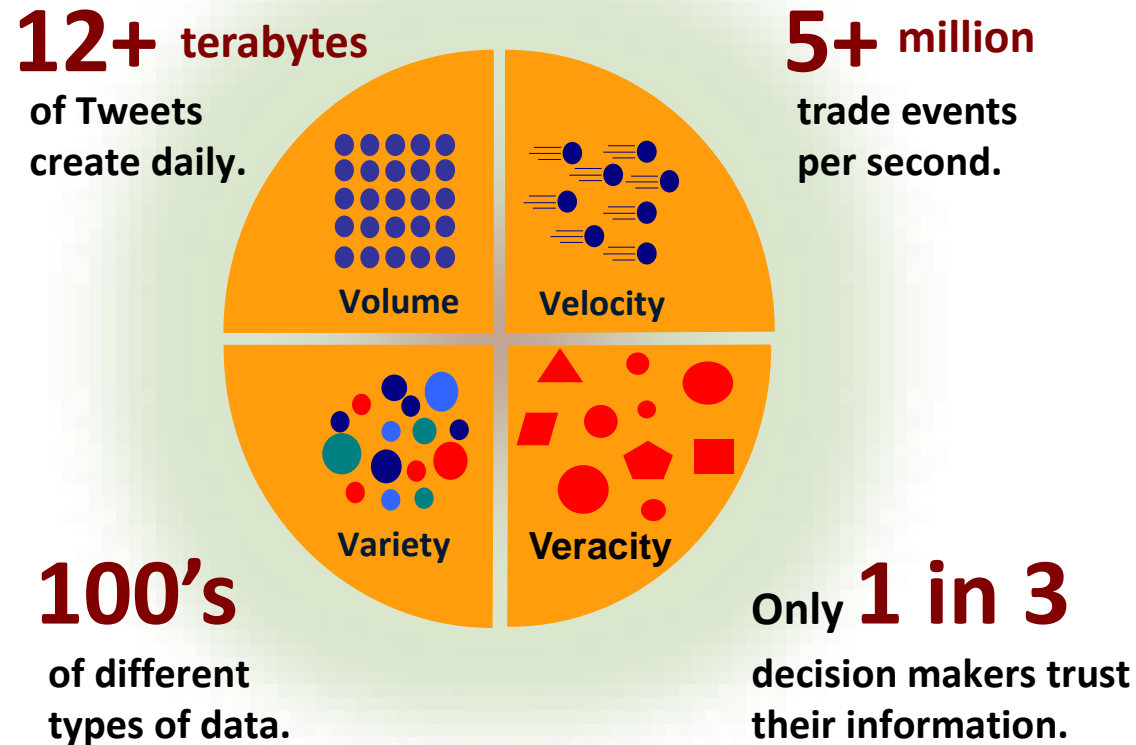
# Apa yang sangat besar ?



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

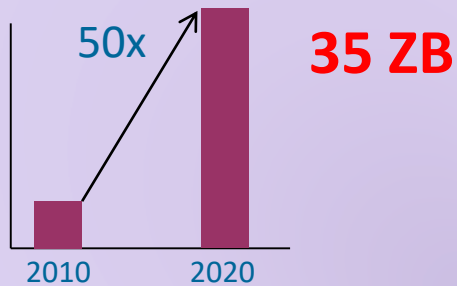


## With Big Data, We've Moved into a New Era of Analytics



# 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

# Kebutuhan Analisa Data



**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.

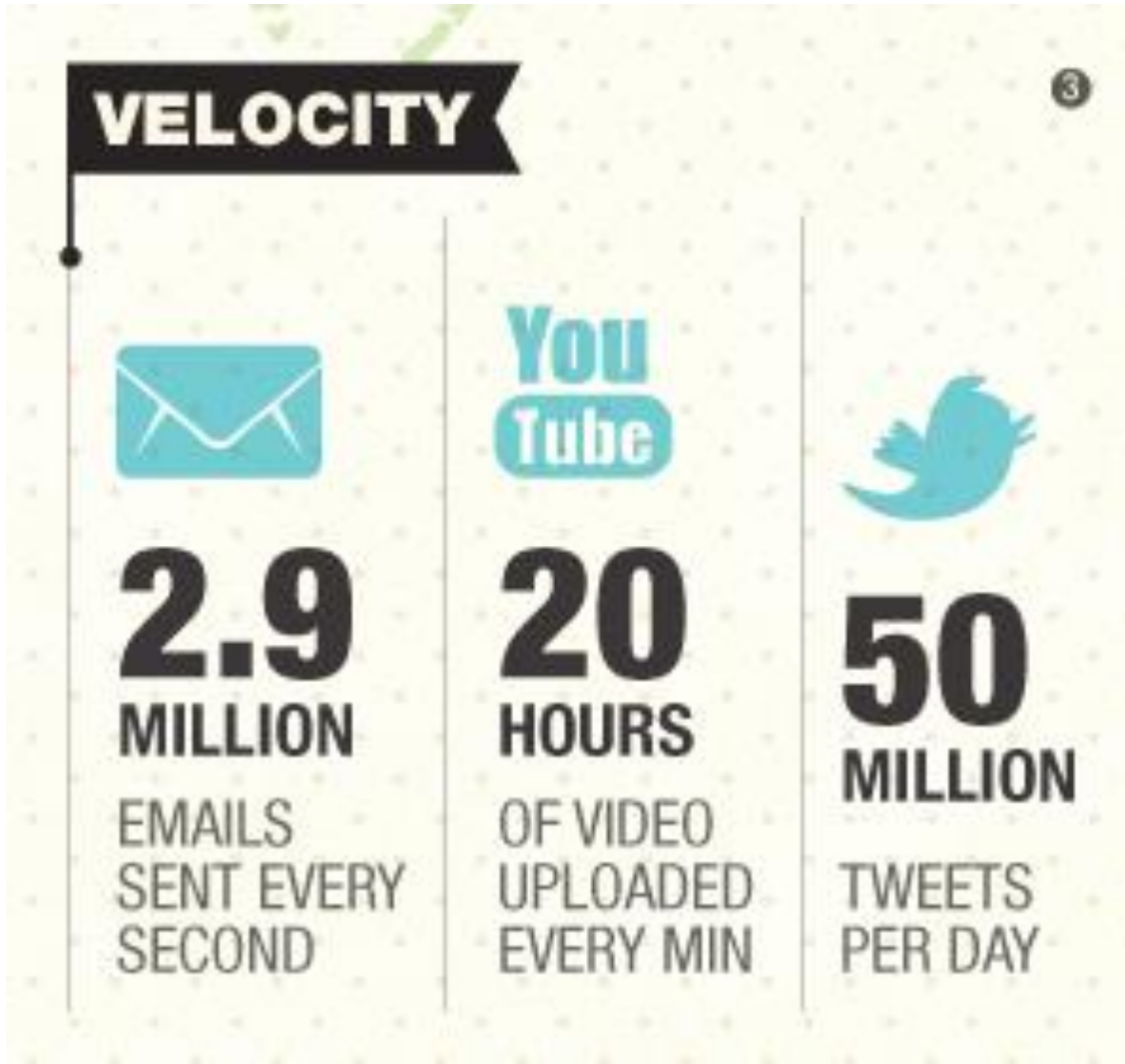
# Big Volume

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



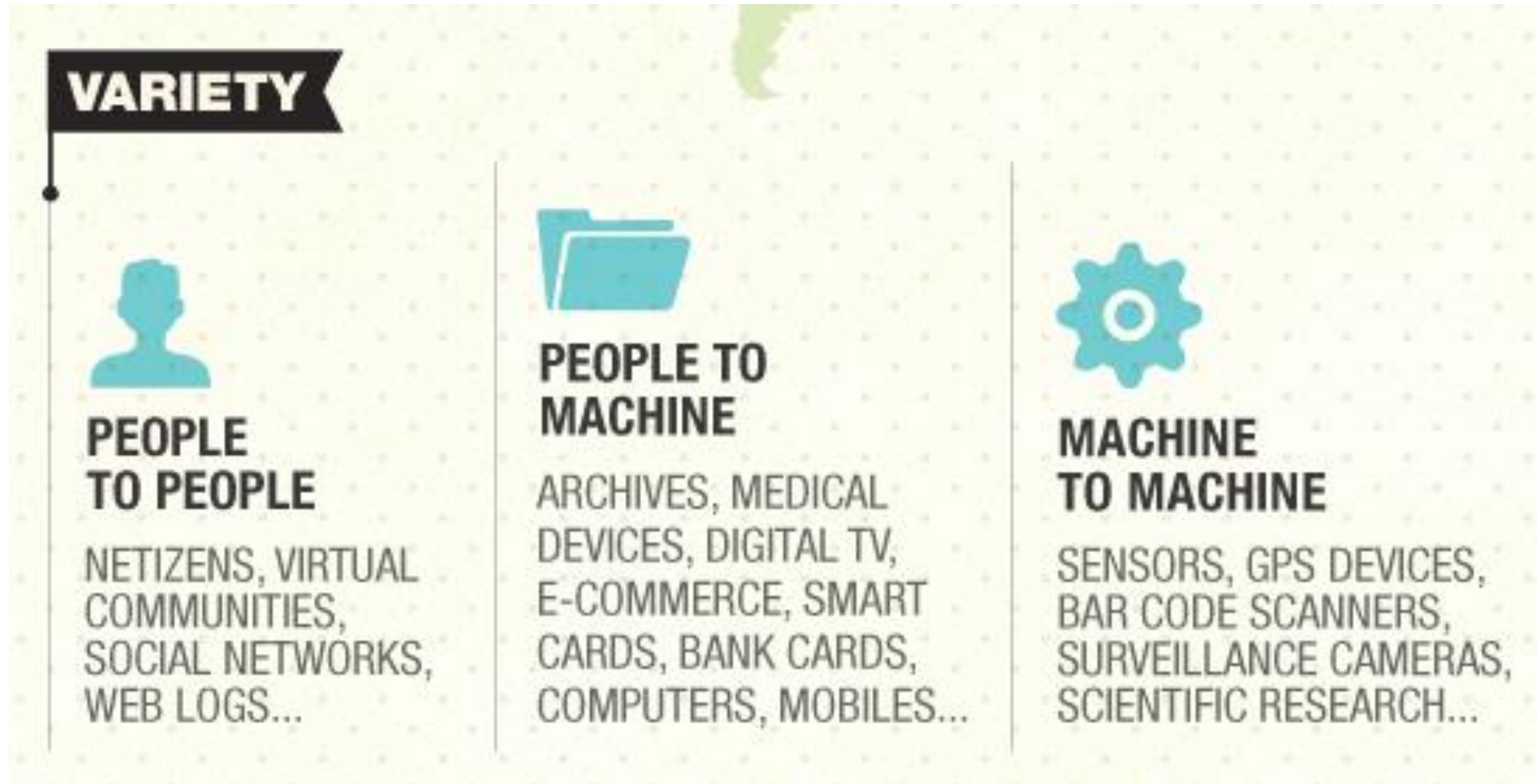
Pertambahan (2011):

- Twitter 7 TB/hari
- Facebook 10 TB/hari

*Sumber: [bigdatauniversity.com](http://bigdatauniversity.com)*



# Big Variety



# 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!)



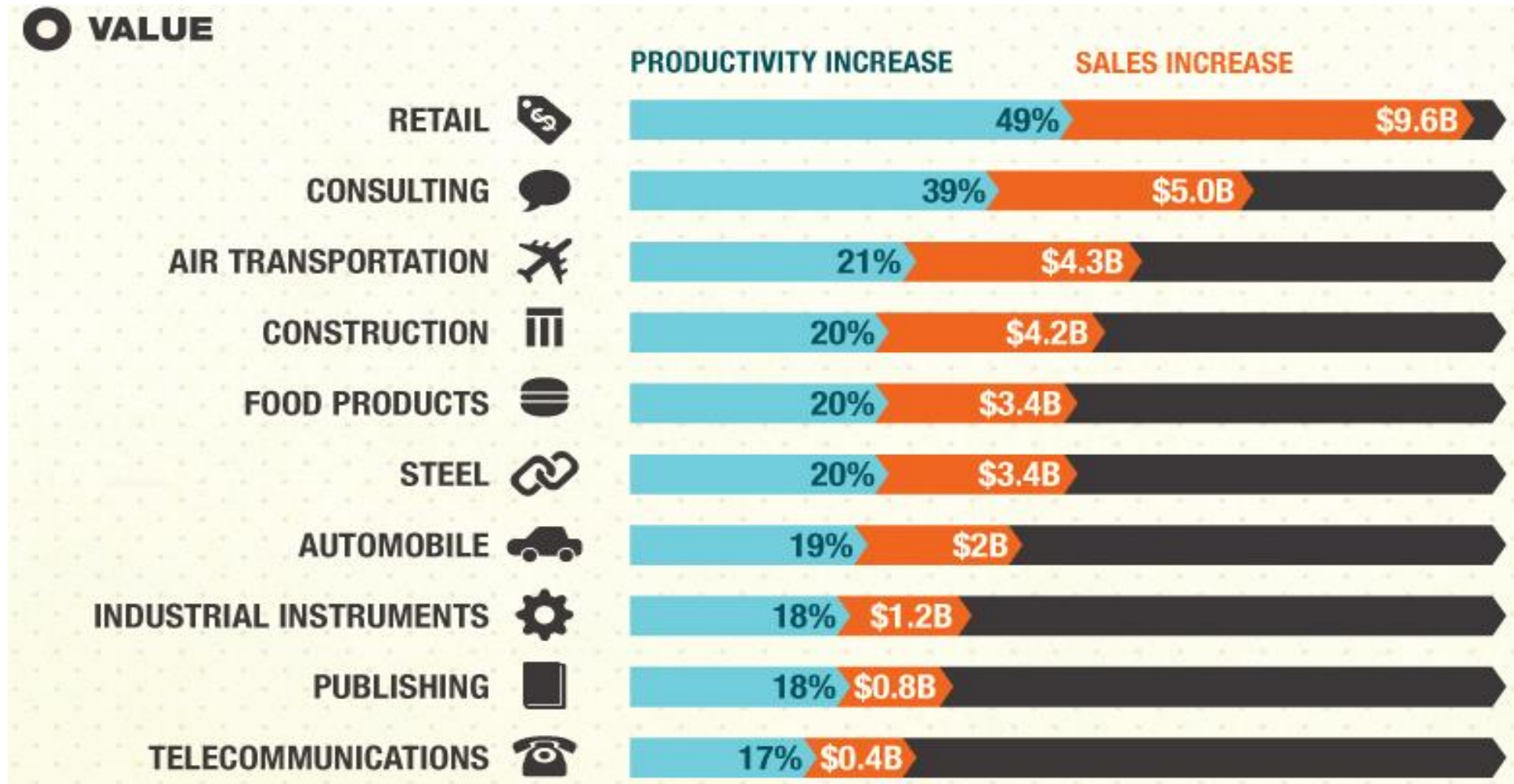
at&t



Establishing the  
**Veracity** of big  
data sources

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

# Big Value



# Quiz

- ✓ Pada aplikasi e-commerce seperti bukalapak / Lazada / tokopedia , sebutkan jenis data:
  - terstruktur
  - tidak terstruktur
  
- ✓ Pada system informasi yang ada di kampus (perguruan tinggi) sebutkan jenis data:
  - terstruktur
  - tidak terstruktur