

Pengantar Big Data



Big Data STF1724

Rini Nuraini Sukmana, M.T
0020087901 - 08882024236
rini.nuraini@usbypkp.ac.id



Outline

- Pengantar
- Gambaran Umum Big Data
- 5 V (Karakteristik pada Big Data)
- Ekosistem Big Data
- Contoh Big Data Analytics

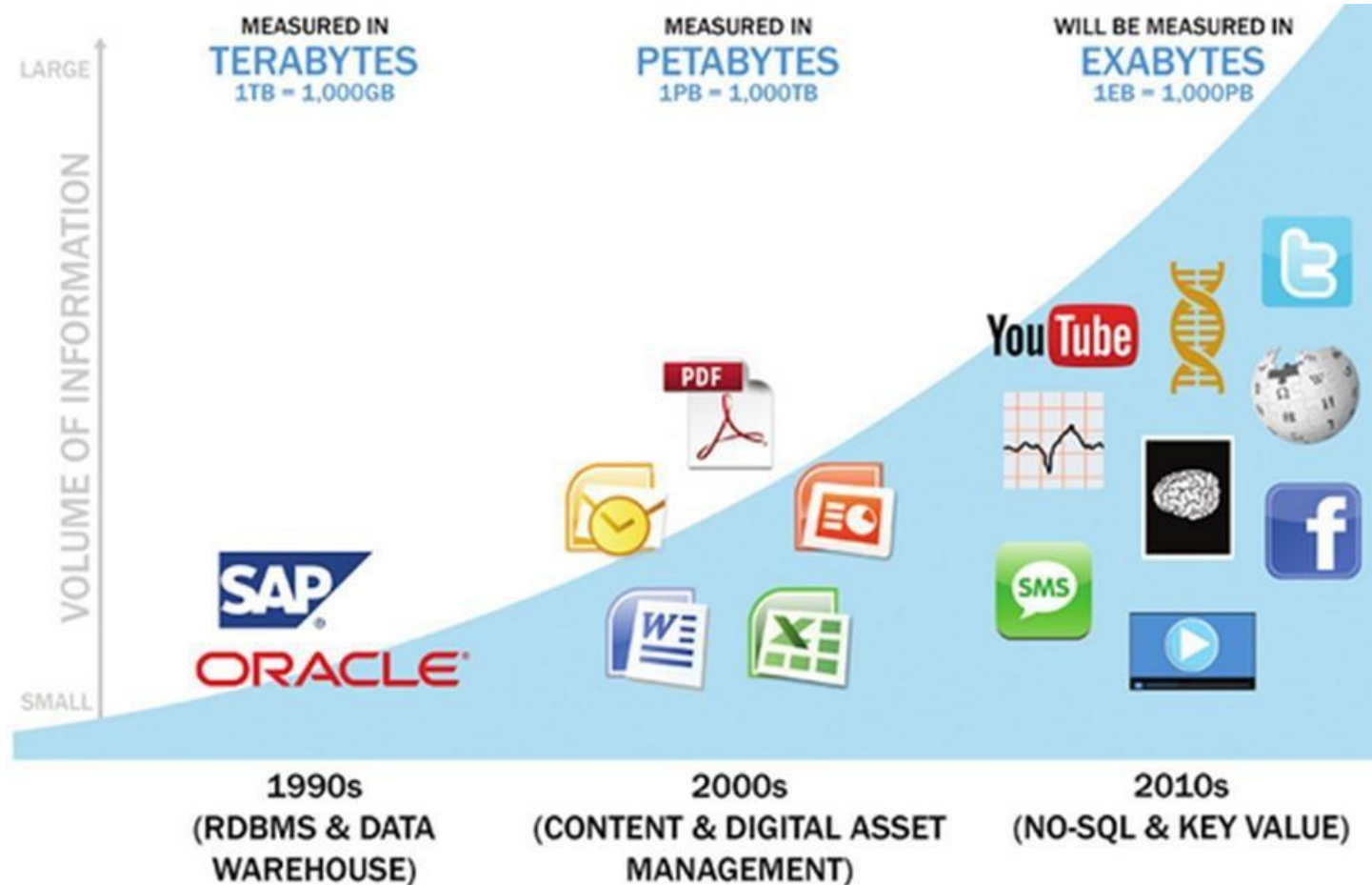


Pengantar

- Big Data “meledak” di awal dekade abad 21
- Perusahaan online dan startup, misalnya Google, Amazon, eBay, LinkedIn, Facebook, dibangun dari data yang besar
- Data yang sangat besar dapat memengaruhi banyak hal, di berbagai bidang: medis, industri, ekonomi dan perbankan, dll



Perkembangan Data





Apa itu Big Data?

- Big Data → mirip dengan 'small data', namun jauh lebih besar ukurannya
 - banyak perdebatan tentang apa itu Big Data
- Dengan data sangat besar perlu pendekatan berbeda
 - Teknik, alat bantu, arsitektur
- Tujuan: menyelesaikan masalah baru atau masalah lama dengan cara lebih baik
- Big Data memberikan nilai dari penyimpanan dan pemrosesan dari kuantitas sangat besar yang tidak bisa dianalisis dengan teknik komputasi tradisional



Apa itu Big Data?

- Walmart menangani > 1 juta transaksi dari *customer* tiap jam
- Facebook menangani 40 miliar foto dari *user base*
- Decoding genome manusia awalnya memerlukan pemrosesan selama 10 tahun → sekarang 1 minggu



1 **NEW**
DEFINITION
IS ADDED ON
urban

1,600+
READS ON
Scribd.

13,000+ HOURS
MUSIC
STREAMING ON
PANDORA

12,000+
NEW ADS
POSTED ON
craigslist

370,000+ MINUTES
VOICE CALLS ON

98,000+
TWEETS

320+
NEW
twitter
ACCOUNTS

100+
NEW
Linked in
ACCOUNTS

1 associatedcontent
NEW
ARTICLE IS
PUBLISHED

6,600+
NEW
PICTURES ARE
UPLOADED ON
flickr

50+
WORDPRESS
DOWNLOADS

695,000+
facebook
STATUS
UPDATES

125+
PLUGIN
DOWNLOADS

79,364
WALL
POSTS

510,040
COMMENTS

1,700+
Firefox
DOWNLOADS

694,445
SEARCH
QUERIES

168 MILLION
EMAILS
ARE SENT

60+
NEW
BLOGS

1,500+
BLOG
POSTS

70+
DOMAINS
REGISTERED

600+
NEW
VIDEOS

100+
Answers.com

40+
YAHOO! Answers

QUESTIONS
ASKED ON THE
INTERNET...

25+ HOURS
TOTAL
DURATION



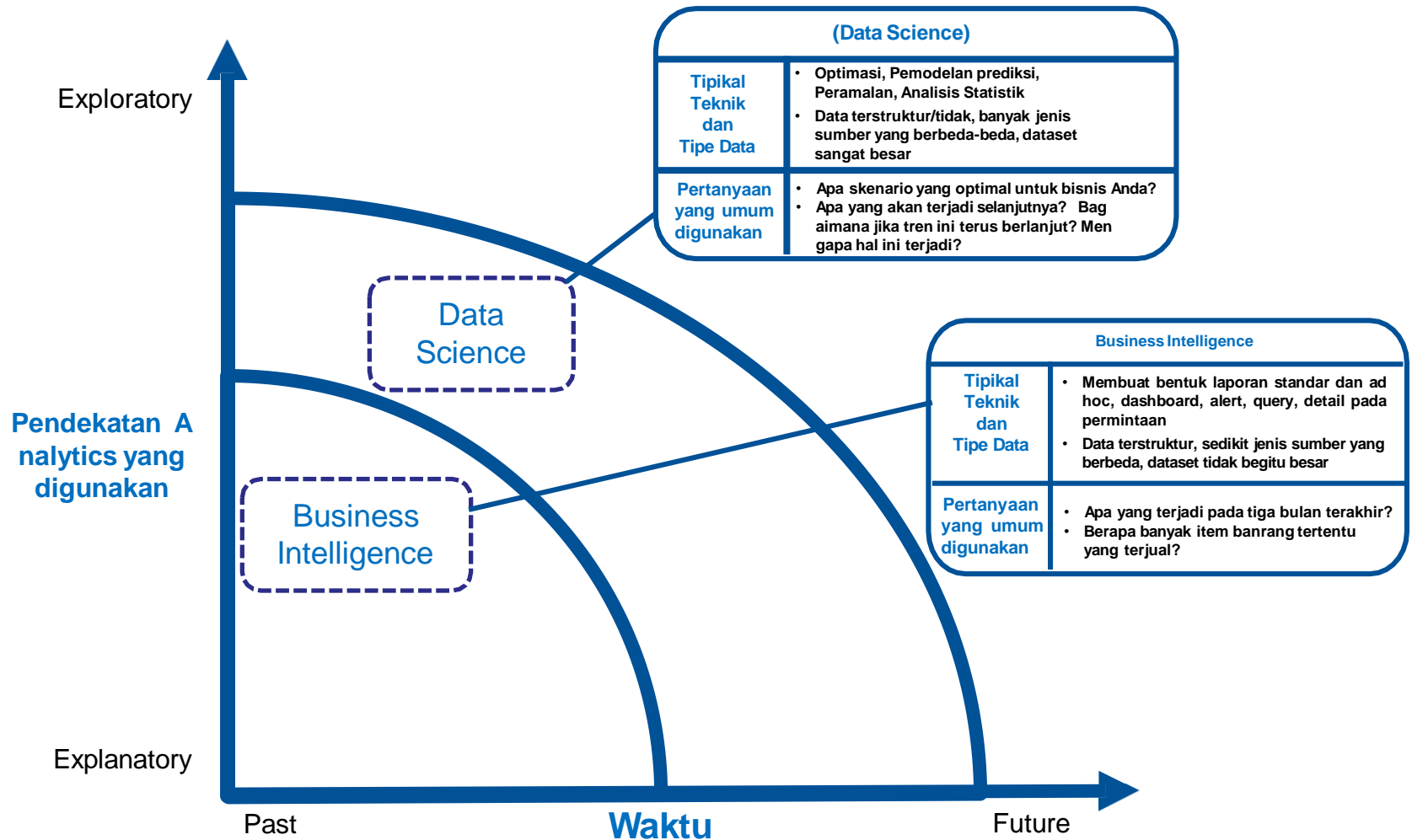
Mengapa belajar Analisis Big Data?



- **Banyak yang menulis tentang Big Data** dan kebutuhan untuk analisis yang canggih dalam industri, akademisi, dan pemerintah, maupun lainnya.
- **Ketersediaan sumber data baru** dan munculnya **peluang analitis** yang lebih **kompleks** telah menciptakan kebutuhan untuk memikirkan kembali arsitektur data yang ada untuk memungkinkan analisis yang dapat dengan optimal memanfaatkan Big Data
- Bab ini menjelaskan beberapa konsep utama Big Data, mengapa analisis canggih diperlukan, perbedaan **Data Science** vs **Business Intelligence** (BI), dan apa peran baru yang diperlukan untuk ekosistem Big Data.



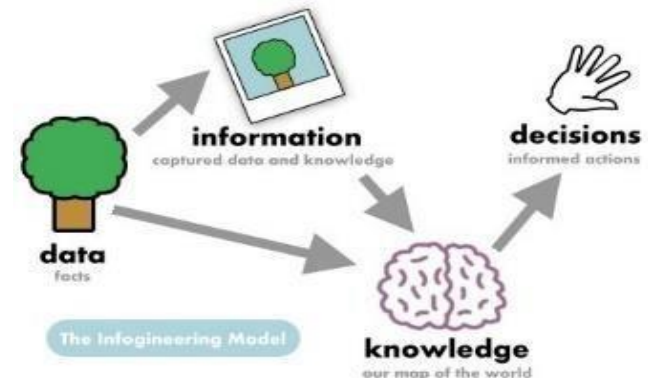
Data Science vs Business Intelligence





Gambaran Umum Big Data

- **Big Data vs Big Information vs Big Knowledge:**
 - **Data** (*Facts, a description of the World*)
 - **Information** (*Captured Data and Knowledge*): Merekam atau mengambil **Data** dan **Knowledge** pada satu waktu tertentu (*at a single point*). Sedangkan **Data** dan **Knowledge** dapat terus berubah dan bertambah dari waktu ke waktu.
 - **Knowledge** (*Our personal map/model of the world*): apa yang kita ketahui (***not the real world itself***) Anda saat ini tidak dapat menyimpan pengetahuan dalam diri anda dalam apa pun selain otak, dan untuk membangun pengetahuan perlu informasi dan data.





Gambaran Umum Big Data



- **McKinsey Global (2011)**

- Big Data dapat didefinisikan dengan data yang memiliki skala (volume), distribusi (velocity), keragaman (variety) yang sangat besar, dan atau abadi, sehingga membutuhkan penggunaan arsitektur teknis dan metode analitis yang inovatif untuk mendapatkan wawasan yang dapat memberikan nilai bisnis baru (informasi yang bermakna).

- **Hurwitz, et al. (2013)**

- Big data merupakan istilah untuk sekumpulan data yang begitu besar atau kompleks dimana tidak bisa ditangani lagi dengan sistem teknologi komputer konvensional

- **Kapan suatu data dapat dikatakan sebagai “Big Data”?**



5 V Pada Big Data

**3V
(Awal)**

- Volume
- Velocity
- Variety

**5 V
(Sekarang)**

- Veracity
- Value



5 V Pada Big Data:

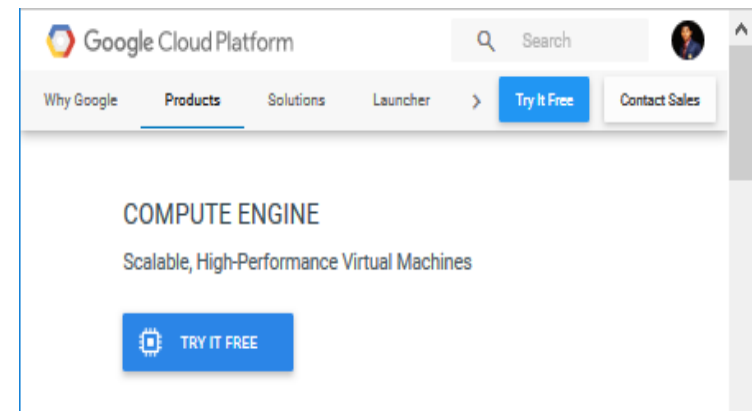
Volume (1)



- Facebook menghasilkan 10TB data baru setiap hari, Twitter 7TB
- Sebuah Boeing 737 menghasilkan 240 terabyte data penerbangan selama penerbangan dari satu wilayah bagian AS ke wilayah yang lain
- Microsoft kini memiliki satu juta server, kurang dari Google, tetapi lebih dari Amazon, Ballmer (2013).
- **Catatan:** Kita dapat menggunakan semua data ini untuk memberitahu kita sesuatu, jika kita mengetahui pertanyaan yang tepat untuk bertanya.



Bentuk infrastruktur data center



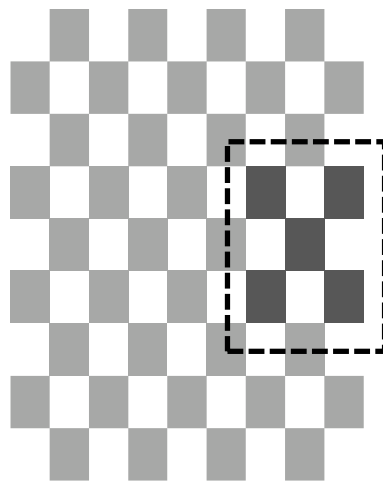
Google Cloud Platform



5 V Pada Big Data:

Volume (2)

Traditional Approach

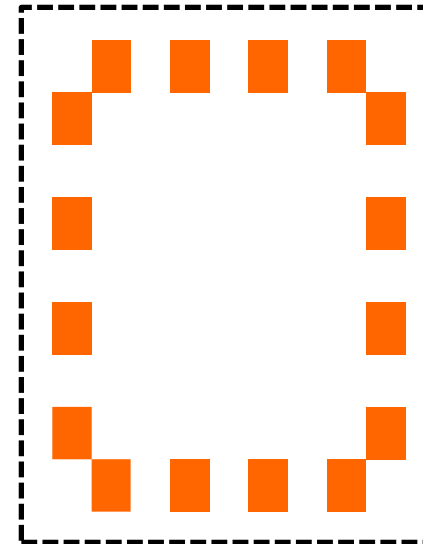


Analyzed information

All available information

Analyze small subsets of data

Big Data Approach



All available information analyzed

Analyze *all* data



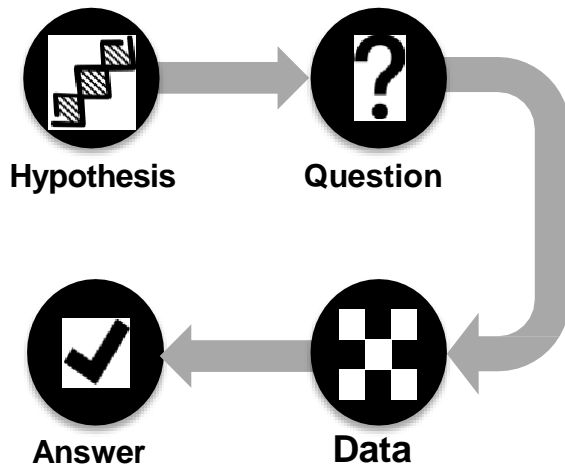
5 V Pada Big Data:

Velocity



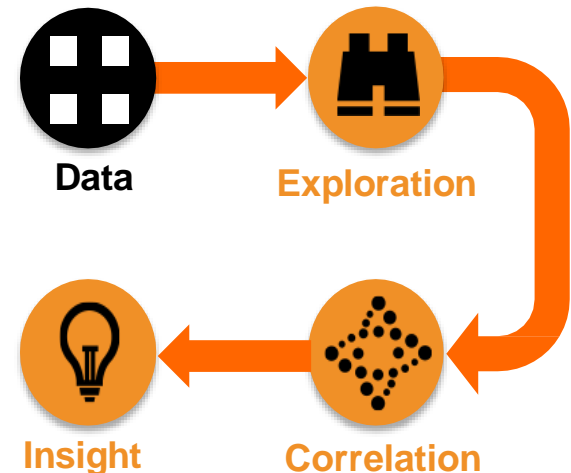
- Kecepatan data yang masuk (per jam, per detik, etc). *Clickstreams* (web log) dan transfer data *asynchronous* yang dapat menangkap apa saja yang dilakukan oleh jutaan atau lebih pengguna yang lakukan saat ini.

Traditional Approach



Start with hypothesis and test against selected data

Big Data Approach



Explore *all* data and identify correlations



5 V Pada Big Data:

Variety



- Kumpulan dari berbagai macam data, baik data yang terstruktur, semi terstruktur maupun data tidak terstruktur (bisa dipastikan lebih mendominasi).
- Tampilan data semakin komprehensif (lengkap dan menyeluruh).



5 V Pada Big Data:

Veracity



- Ketidakpastian akan data.
- Business process rawan akan kesalahan, tergantung datanya
- Bagaimana suatu data dapat dipercaya mengingat keandalan sumbernya
- Bagaimana mengelola, mengolah data mana yang benar dan mana yang salah



5 V Pada Big Data:

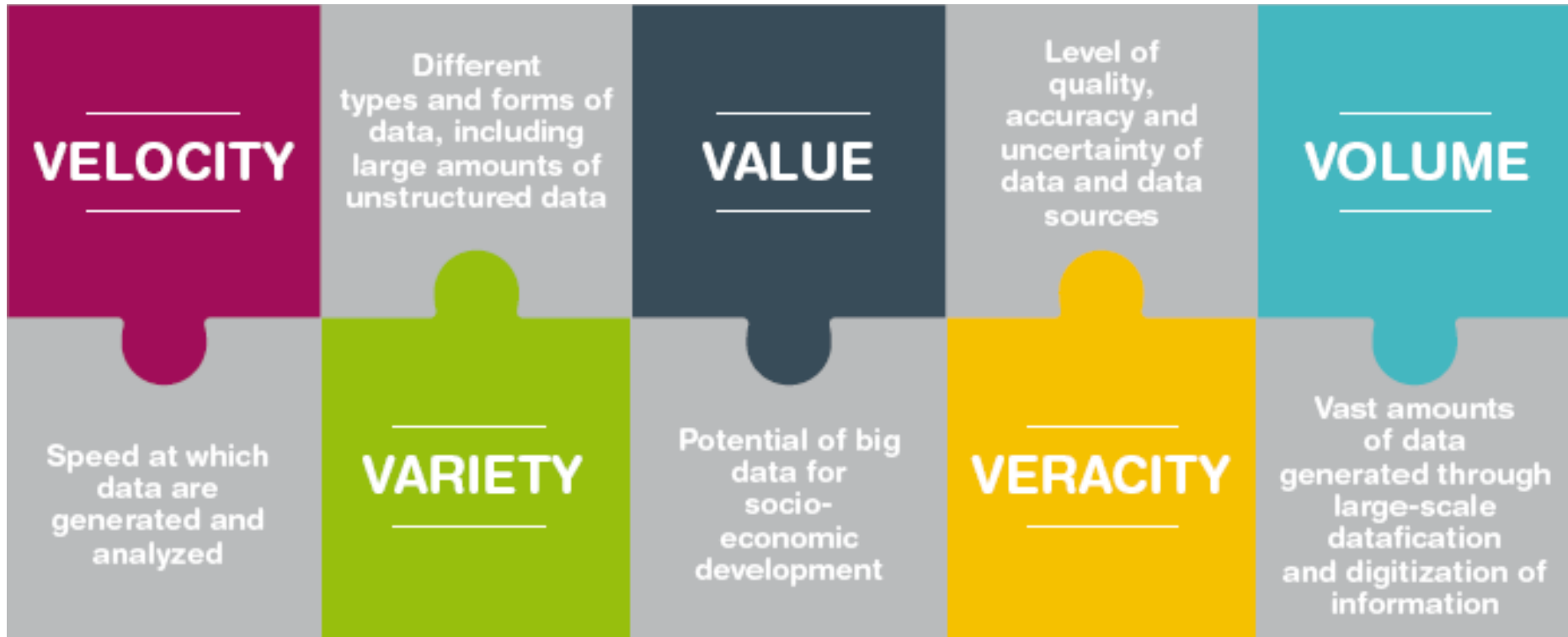
Value



- Data yang besar seharusnya berdampak (secara moneter) terhadap suatu perusahaan yang menggunakan komputasi Big Data
- Akan sia-sia bila memiliki data yang sangat besar tapi tidak tahu bagaimana cara mengolah dan menganalisisnya, hanya akan buang-buang *resource*



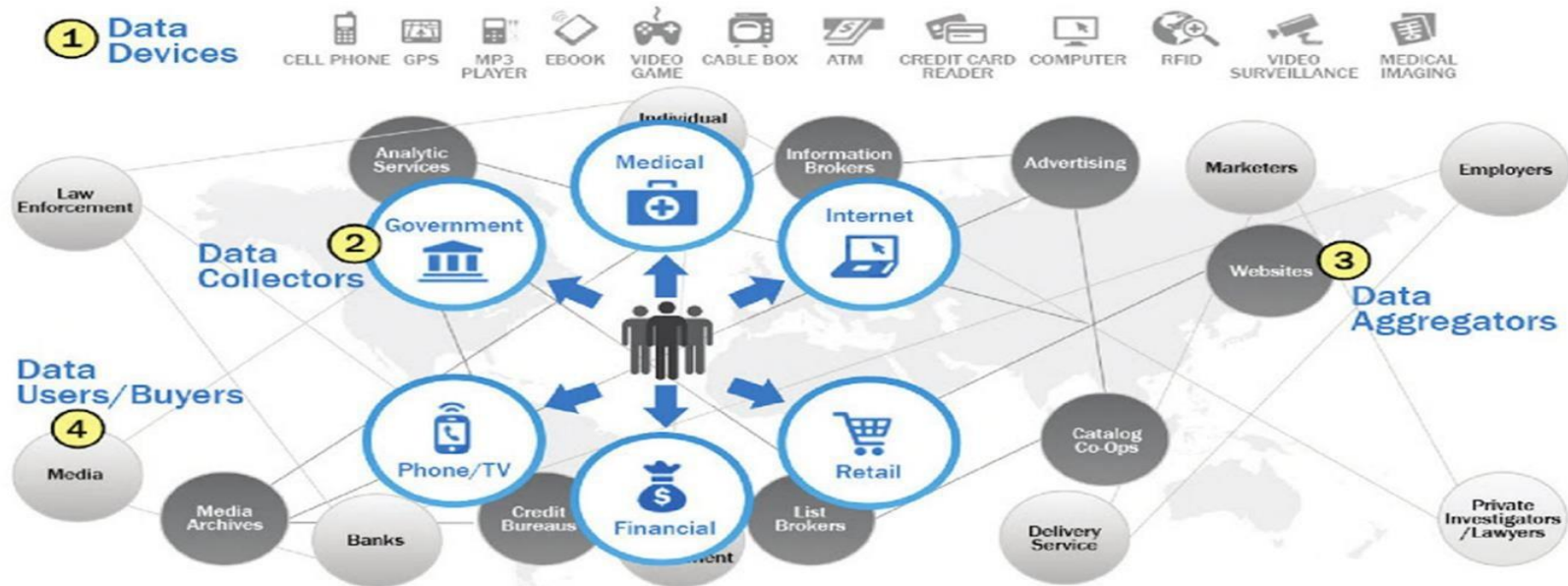
5 V Secara Singkat





Ekosistem Pada Big Data

Gambaran dari Ekosistem Big Data



1. Data Devices
2. Data Collectors
3. Data Aggregators: kompilasi informasi dari database dengan tujuan untuk mempersiapkan dataset gabungan untuk pengolahan data.
4. Data Users/ Buyers



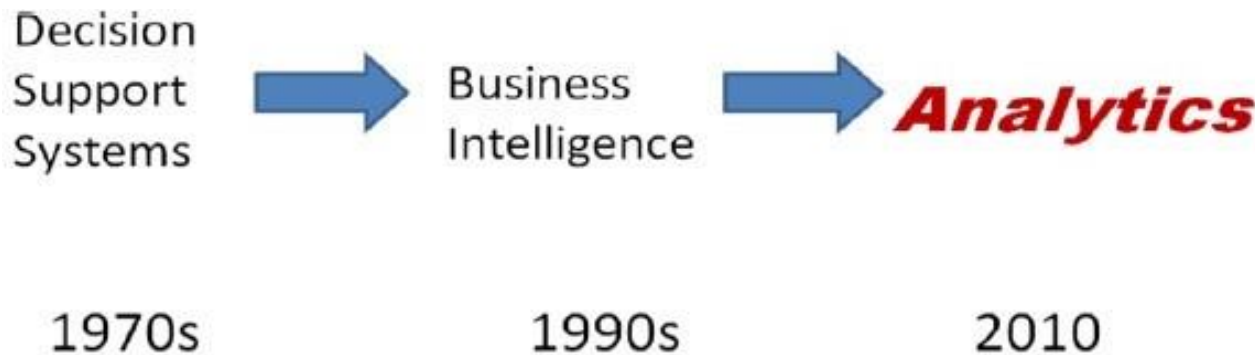
Ekosistem Big Data

- Bidang Pekerjaan baru Big Data Analytics:
 - **Deep Analytical Talent/Data scientists** (Memiliki bakat analitis yang mendalam/Ilmuwan Data):
 - orang-orang dengan latar belakang yang kuat dalam algoritma-algoritma sistem cerdas, atau matematika terapan, atau ekonomi, atau ilmu pengetahuan lainnya melalui inferensi data dan eksplorasi.
 - **Data Savvy Professionals** (para profesional data Cerdas):
 - Mereka tahu bagaimana untuk berpikir tentang data, bagaimana mengajukan jenis pertanyaan (goal) yang tepat sesuai dengan kebutuhan lembaga/perusahaan/lainnya dan mampu memahami dan mengklarifikasi jawaban (hasil analisis) yang mereka terima.
 - **Technology and data enablers:**
 - Mampu memberikan dukungan integrasi antara data dengan teknologi yang sesuai, dan yang paling berkembang saat ini.



Big Data Analytics

- Apa yang dimaksud dengan Analytics? Sebuah titik awal untuk memahami Analytics adalah Cara untuk mengeksplorasi/menyelidiki/ memahami secara mendalam suatu objek sampai ke akar-akarnya
- Hasil analytics biasa tidak menyebabkan banyak kebingungan, karena konteksnya biasanya membuat makna yang jelas



Gambar: Dari DSS berkembang menjadi BI kemudian menjadi Analytics.



Big Data Analytics

- BI dapat dilihat sebagai istilah umum untuk semua aplikasi yang mendukung DSS, dan bagaimana hal itu ditafsirkan dalam industri dan semakin meluas sampai di kalangan akademisi.
- BI berevolusi dari DSS, dan orang dapat berargumentasi bahwa Analytics berevolusi dari BI (setidaknya dalam hal peristilahan). Dengan demikian, Analytics merupakan istilah umum untuk aplikasi analisis data.
- Big Data Analytics: Alat dan teknik analisis yang akan sangat membantu dalam memahami big data dengan syarat algoritma yang menjadi bagian dari alat-alat ini harus mampu bekerja dengan jumlah besar pada kondisi real-time dan pada data yang berbeda-beda.



Contoh Big Data Analytics

- Contoh perusahaan yang menggunakan analisis Big Data

Starbucks (Memperkenalkan Produk Coffee Baru). Pagi itu kopi itu mulai dipasarkan, pihak **Starbucks memantau** melalui **blog, Twitter, dan kelompok forum diskusi kopi** lainnya untuk menilai reaksi pelanggan. Pada pertengahan-pagi, Starbucks menemukan **hasil dari analisis Big Data** bahwa meskipun orang menyukai rasa kopi tersebut, tetapi mereka berpikir bahwa harga kopi tersebut terlalu mahal. Maka dengan segera pihak Starbucks menurunkan harga, dan menjelang akhir hari semua komentar negatif telah menghilang. Bagaimana jika menggunakan **analisis tradisional**?

- Contoh tersebut menggambarkan penggunaan sumber data yang berbeda dari Big Data dan berbagai jenis analisis yang dapat dilakukan dengan **respon sangat cepat oleh pihak Starbucks**.



Sumber



Presentasi Putra pandu Adikara & Imam Cholisoddin, Universitas Brawijaya