

BAB 6.

INTERNET OF THINGS & BIG DATA ANALYTICS



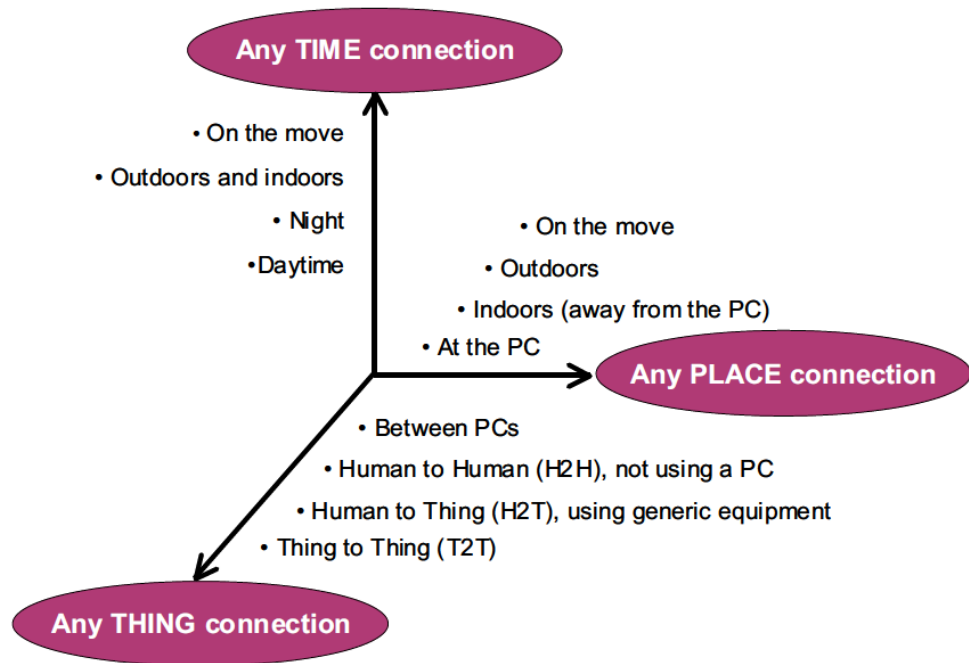
Definisi ringkas dari Internet of Things :

- Internet of Things, disebut juga internet dari obyek, mengacu pada jaringan nirkabel antar obyek.
- Dengan menanamkan transceiver seluler jarak pendek ke beragam gadget dan barang sehari-hari, memungkinkan bentuk komunikasi baru antara manusia dengan benda, dan antara benda dengan benda lainnya.
- Istilah "Internet of Things" telah datang untuk menggambarkan sejumlah teknologi dan disiplin ilmu penelitian, yang memungkinkan internet menjangkau ke dunia nyata benda-benda fisik.
- Benda benda yang memiliki identitas dan kepribadian virtual yang beroperasi di ruang cerdas, menggunakan antarmuka cerdas untuk terhubung dan berkomunikasi dalam lingkungan sosial, dan dalam konteks pengguna.

What's the Internet of Things

From any time ,any place connectivity for anyone, we will now have connectivity for anything!

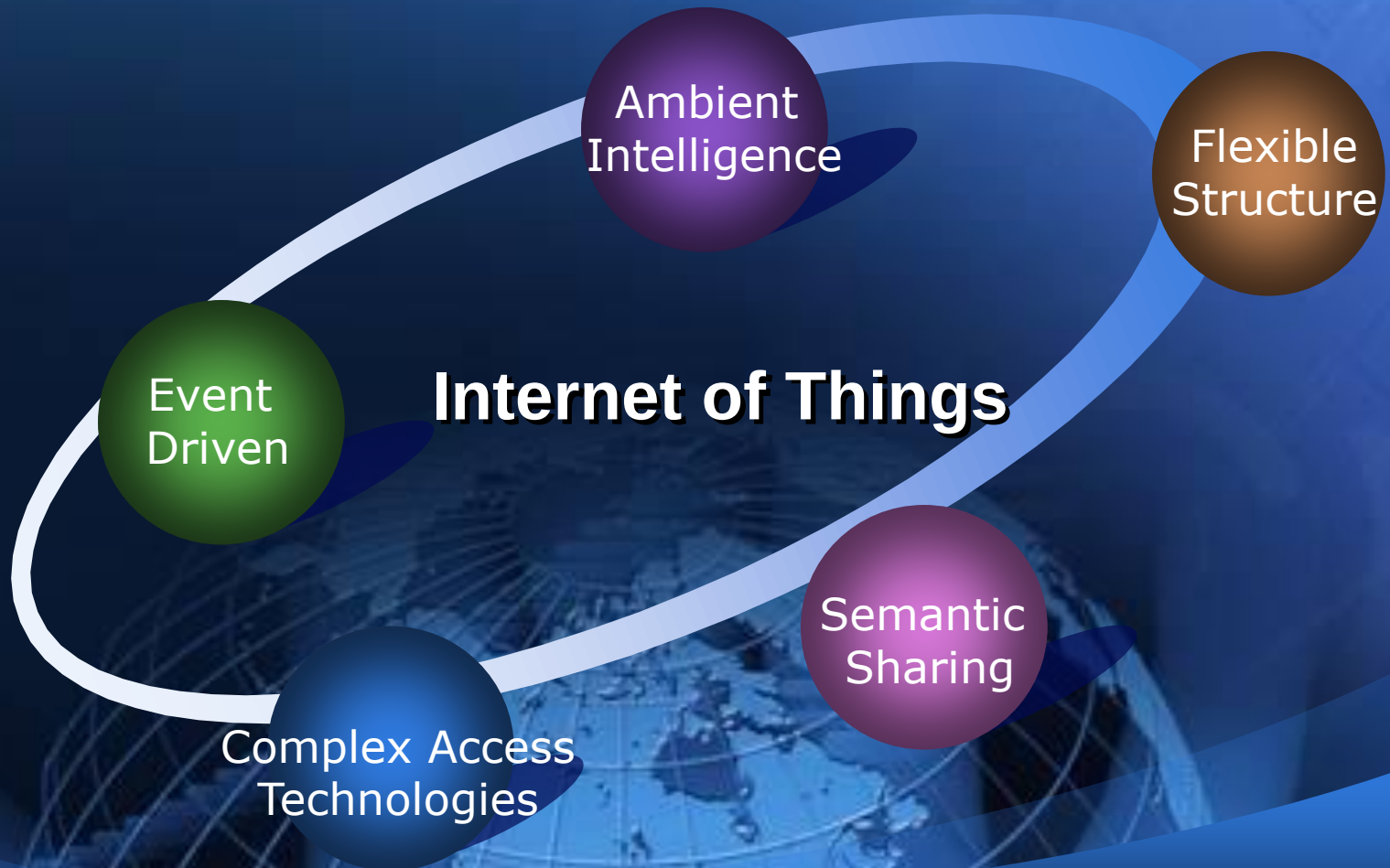
Figure 1 – A new dimension



Source: ITU adapted from Nomura Research Institute

What's the Internet of Things

Characteristics



Why Internet of Things



Dynamic control of industry and daily life



Improve the resource utilization ratio



Integrating human society and physical systems.

Why Internet of Things (ii)



Flexible configuration.



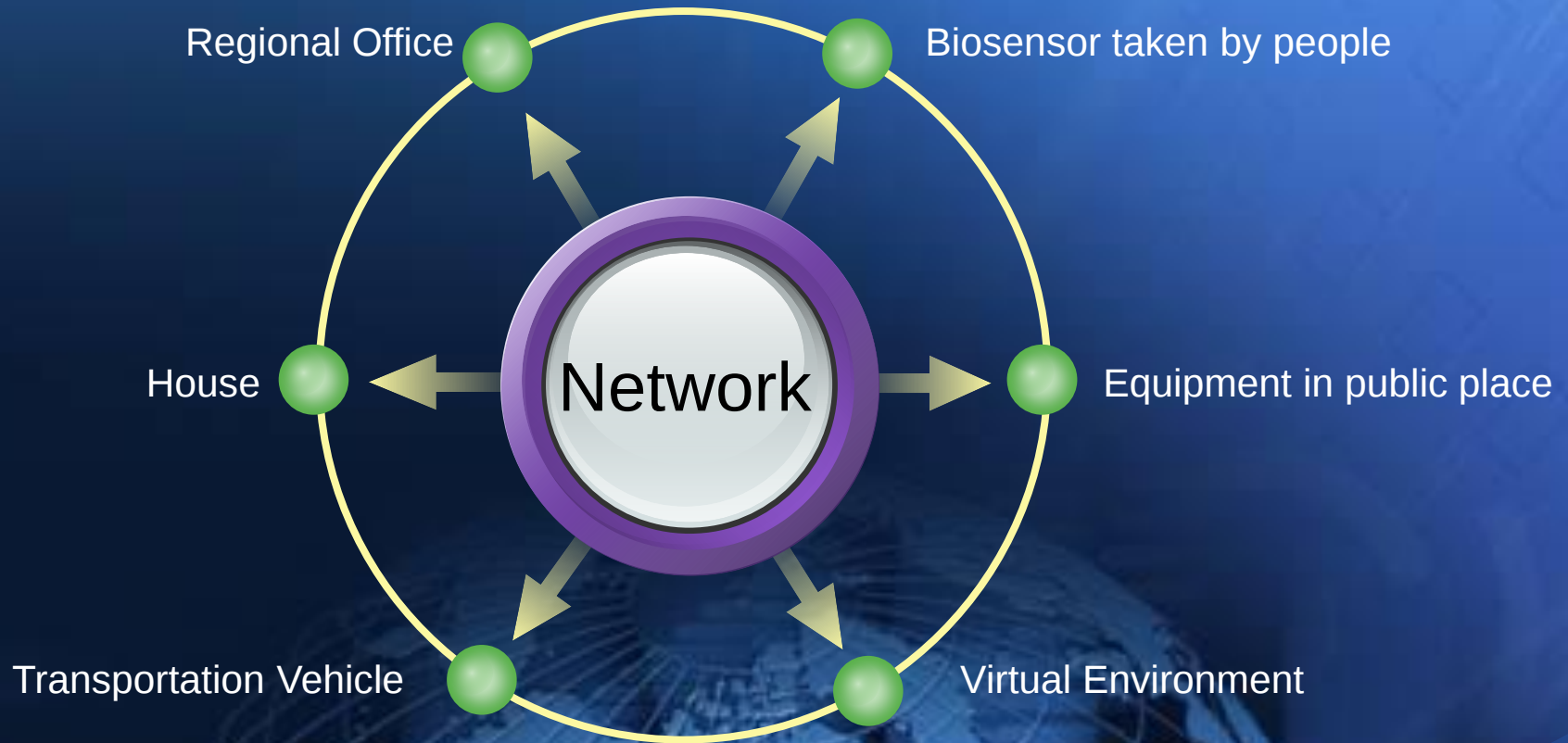
Universal transport & internetworking



Acts as technologies integrator



The application of IoT(1)



The application of IoT(2)

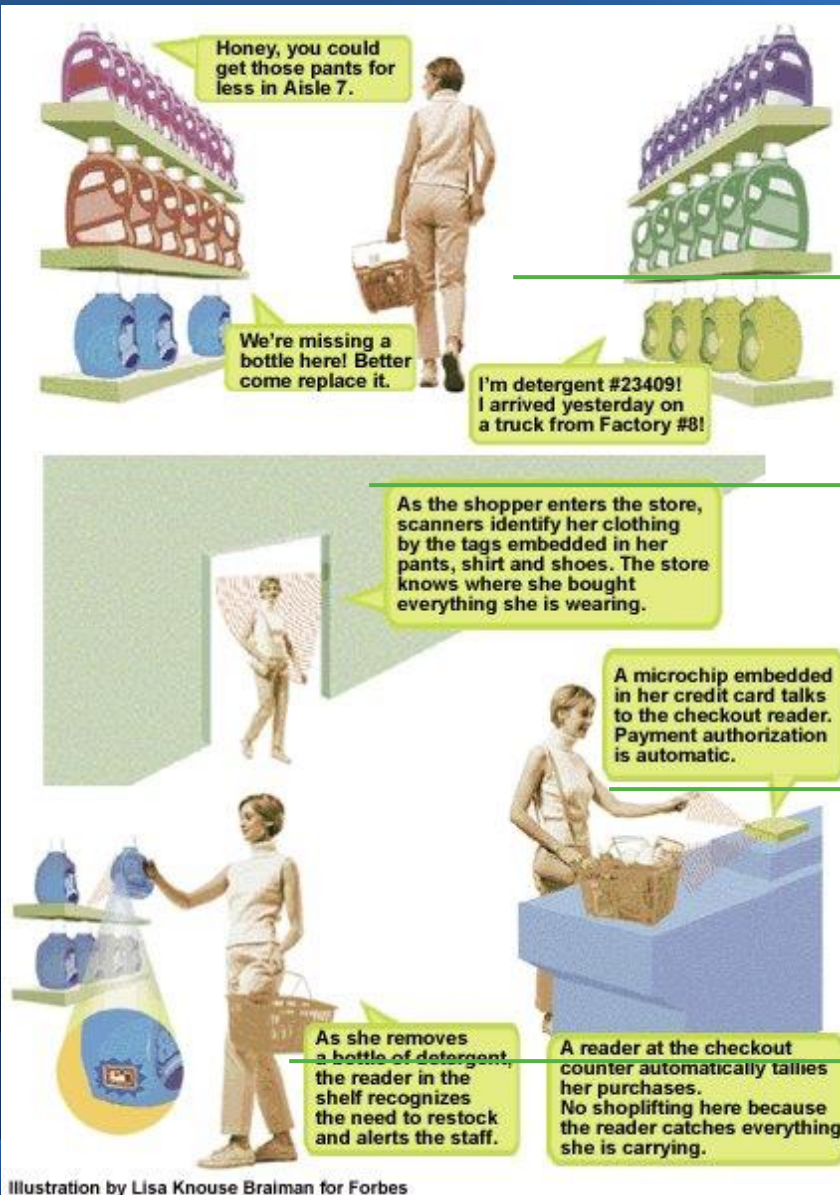
Scenario: shopping

(2) When shopping in the market, the goods will introduce themselves.

(1) When entering the doors, scanners will identify the tags on her clothing.

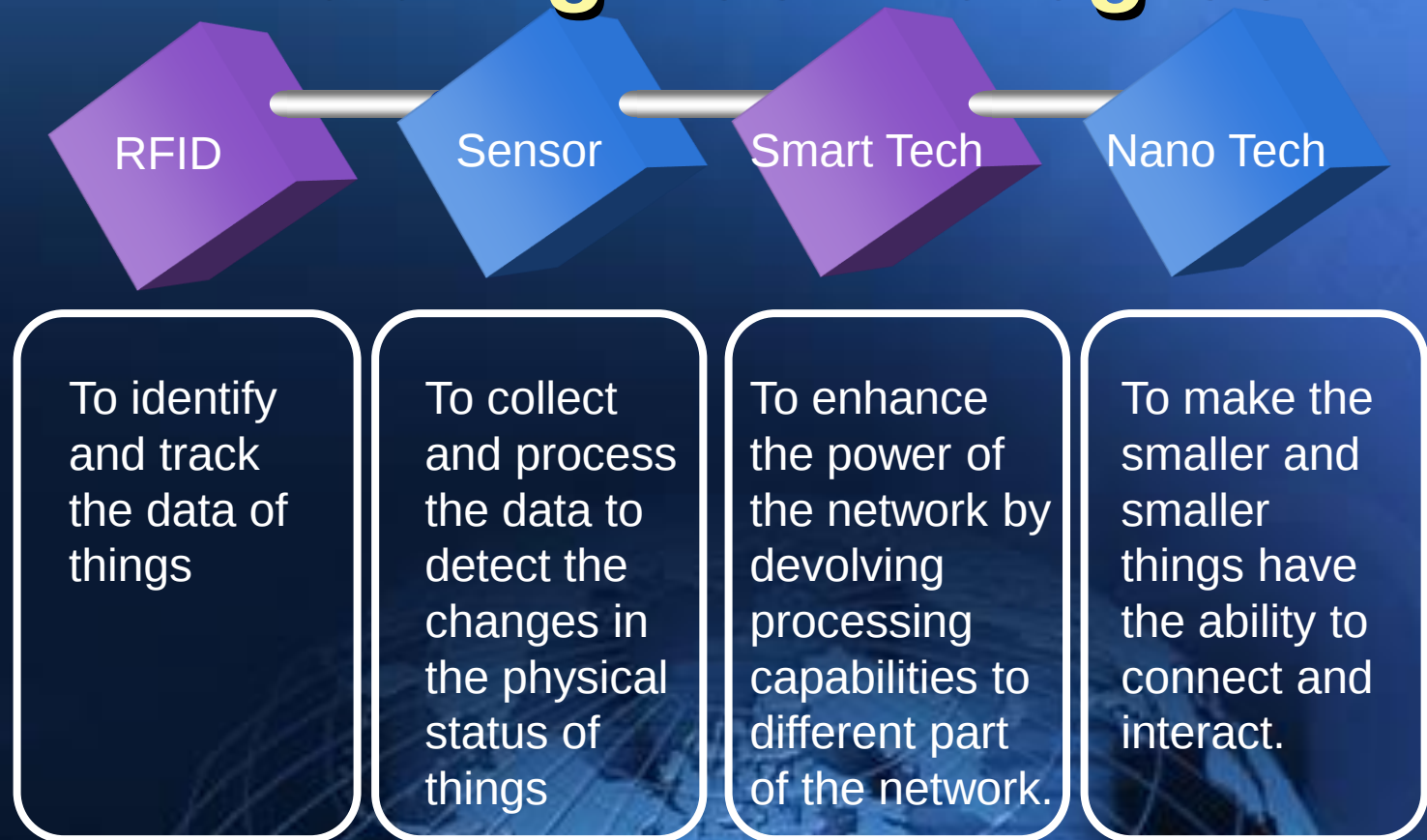
(4) When paying for the goods, the microchip of the credit card will communicate with checkout reader.

(3) When moving the goods, the reader will tell the staff to put a new one.



State of the Art of IoT

Enabling Technologies



State of the Art of IoT

Research groups

1

MIT Auto-ID Lab &
EPC Global.

Stanford University

Georgia Institute of
Technology

Cambridge Univ

2

EPFL & ETH Zurich
Information and
Communication
Systems Research
Group

Chemnitz University
of Technology
VSR Group

3

Nokia
SAP
IBM
GOOGLE
AMBIENT
Metro Group
Siemens
Sun
Cisco
GE

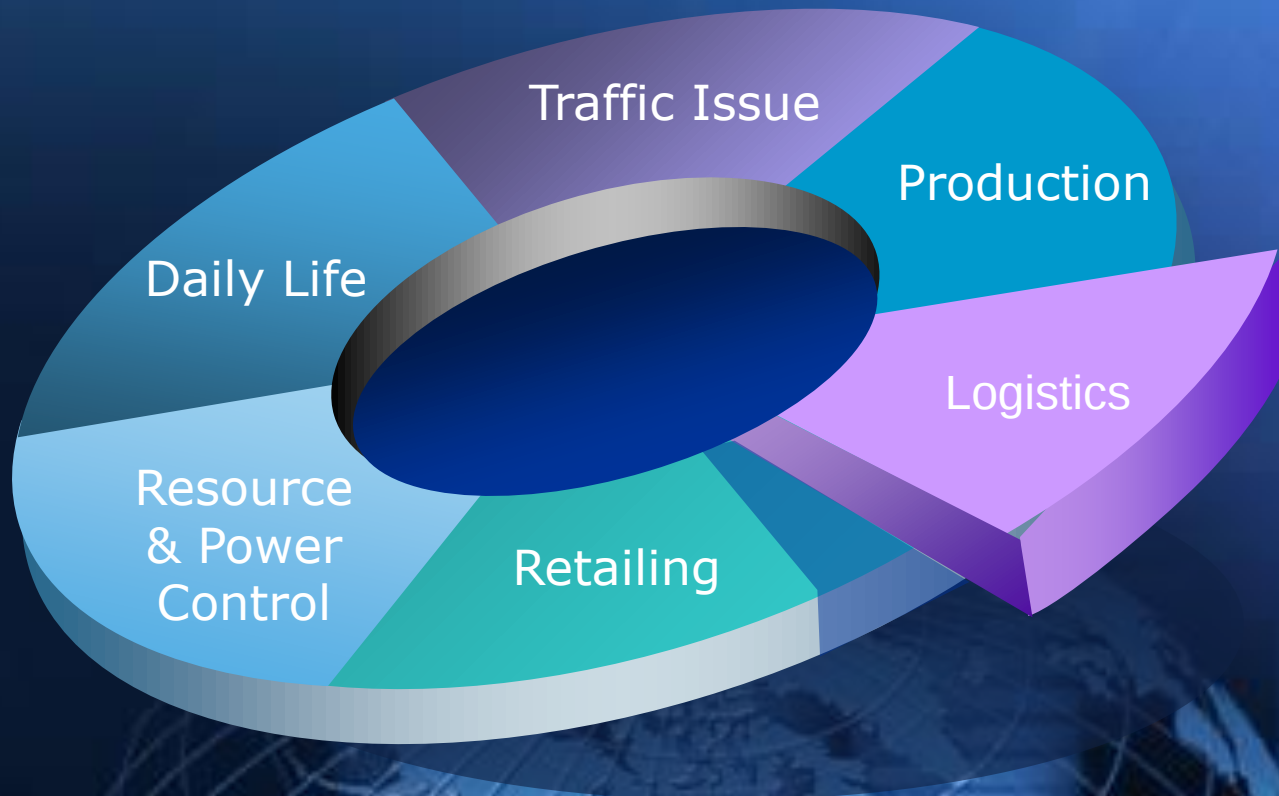
The challenge of IoT

Total challenge of IOT

1. Technological Standardization in most areas are still remain fragmented.
2. Managing and fostering rapid innovation is a challenge for governments .
3. Privacy.
4. Absence of governance.

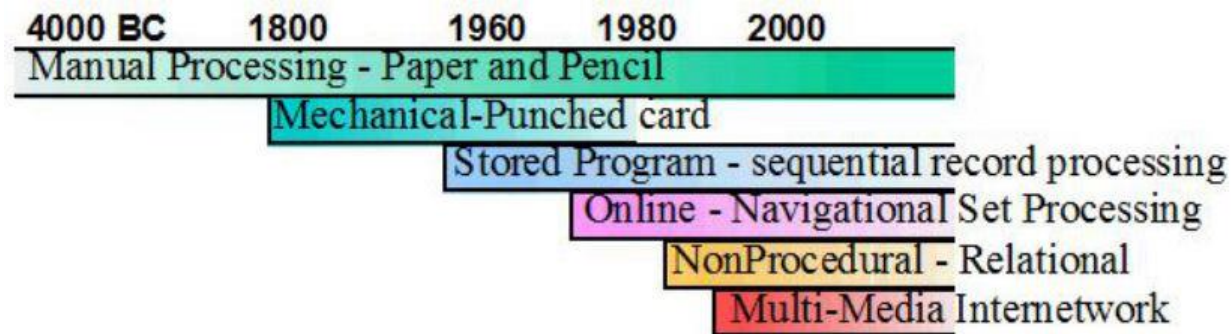


Future of IOT



BIG DATA ANALYTICS

Mahadata, lebih dikenal dengan istilah bahasa Inggris “Big Data”, adalah istilah umum untuk segala himpunan data (data set) dalam jumlah yang sangat besar, rumit dan tak terstruktur sehingga menjadikannya sukar ditangani apabila hanya menggunakan perangkat manajemen basis data biasa atau aplikasi pemroses data tradisional belaka.



Perkembangan Manajemen Data

Analisis data besar mengacu pada strategi menganalisis volume data yang besar, atau data besar. data besar ini dikumpulkan dari berbagai sumber, termasuk jaringan sosial, video, gambar digital, sensor, dan catatan transaksi penjualan. Tujuan dalam menganalisis semua data ini adalah untuk mengungkap pola dan koneksi yang mungkin tidak terlihat, dan yang mungkin memberikan wawasan berharga tentang pengguna yang menciptakannya. Melalui pemahaman ini, bisnis mungkin dapat memperoleh keunggulan atas saingan mereka dan membuat keputusan bisnis yang superior.

Ukuran Skala Data untuk Keperluan Big Data:

Bytes(8 bits)

Kilobyte (1000 bytes)

Megabyte (1 000 000 bytes)

Gigabyte (1 000 000 000 bytes)

Terabyte (1 000 000 000 000 bytes)

Petabyte (1 000 000 000 000 000 bytes)

Exabyte (1 000 000 000 000 000 000 bytes)

Zettabyte (1 000 000 000 000 000 000 000 bytes)

Yottabyte (1 000 000 000 000 000 000 000 000 bytes)

Xenottabyte (1 000 000 000 000 000 000 000 000 000 bytes)

Shilentnobyte (1 000 000 000 000 000 000 000 000 000 000 bytes)

Domegemegrottebyte (1 000 000 000 000 000 000 000 000 000 000 000 bytes)



Multiples of bytes					V • T • E
Decimal		Binary			
Value	Metric	Value	IEC	JEDEC	
1000	kB kilobyte	1024	KiB kibibyte	KB kilobyte	
1000 ²	MB megabyte	1024 ²	MiB mebibyte	MB megabyte	
1000 ³	GB gigabyte	1024 ³	GiB gibibyte	GB gigabyte	
1000 ⁴	TB terabyte	1024 ⁴	TiB tebibyte	—	
1000 ⁵	PB petabyte	1024 ⁵	PiB pebibyte	—	
1000 ⁶	EB exabyte	1024 ⁶	EiB exbibyte	—	
1000 ⁷	ZB zettabyte	1024 ⁷	ZiB zebibyte	—	
1000 ⁸	YB yottabyte	1024 ⁸	YiB yobibyte	—	
Orders of magnitude of data					

Gambar 6.10. Kelipatan dari *Byte*.

Contoh banyaknya penggunaan data:

- Google memproses 20 PB sehari (2008).
- Wayback Machine memiliki 3 PB + 100 TB / bulan (3/2009).
- Facebook memiliki 2,5 PB data pengguna + 15 TB / hari (4/2009).
- EBay memiliki 6,5 PB data pengguna + 50 TB / hari (5/2009).
- CERN's Large Hydron Collider (LHC) menghasilkan 15 PB setahun.

Jenis jenis data:

- Data Relasional (Tabel/Transaksi/Legacy Data).
- Data Teks (Web).
- Data Semi-terstruktur (XML).
- Data Grafik:
Jaringan Sosial, Semantic Web (RDF) dan lainnya.
- Data Streaming:
Kita hanya dapat memindai data satu kali.

Apa yang harus dilakukan dengan data ini?

- Agregasi dan Statistik:
 - Data warehouse dan OLAP.
- Pengindeksan, Pencarian, dan Query:
 - Pencarian berbasis kata kunci.
 - Pencocokan pola (XML / RDF).
- Penemuan Pengetahuan:
 - Data Mining.
 - Pemodelan Statistik.

Karakteristik dari Big Data meliputi:

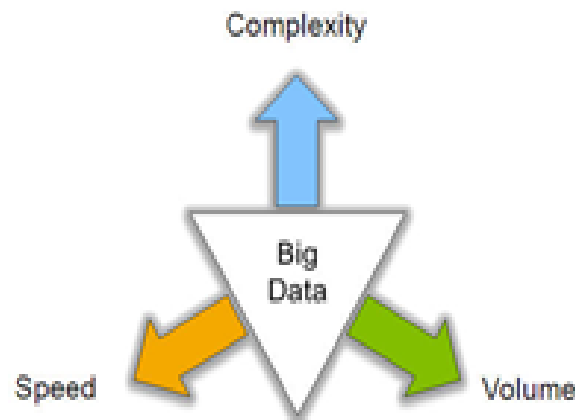
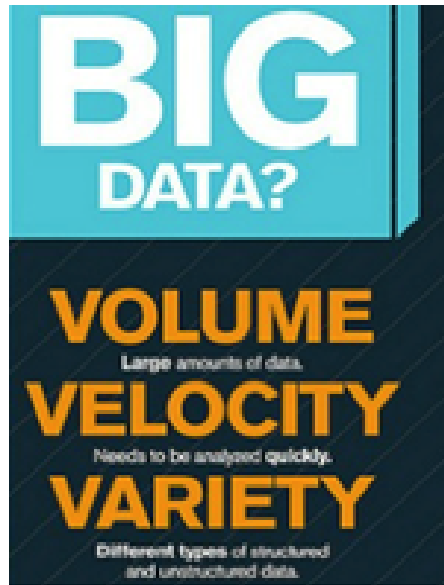
- Volume (isi/kapasitas).
- Velocity (kecepatan).
- Variety (variasi).

Dengan penjelasan sebagai berikut:

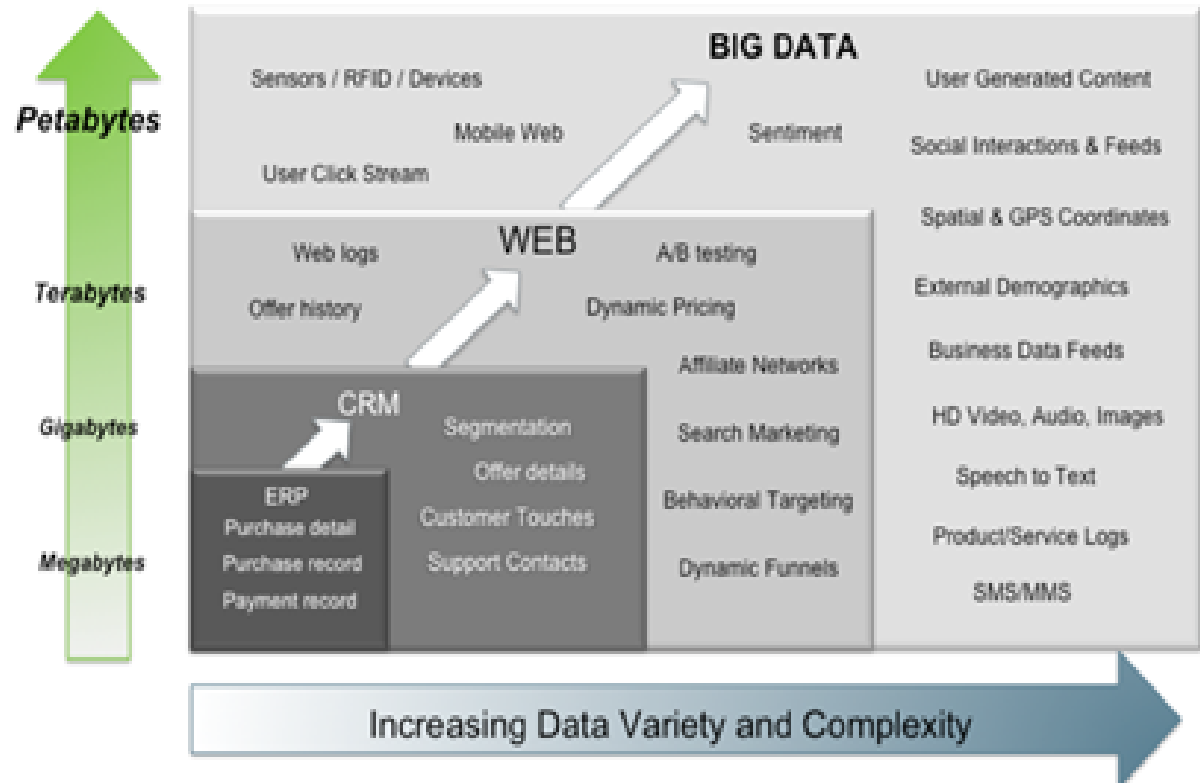
- **Volume (isi/kapasitas):**
 - 44 X naik dari tahun 2009-2020.
 - Dari 0.8 Zettabytes menjadi 35.2 Zettabytes.
 - Volume data naik secara eksponensial.
- **Velocity (kecepatan):**
 - Data mulai cepat dan perlu cepat diproses.
 - Analisis data secara online.
 - Keputusan yang terlambat akan menjadikan kehilangan kesempatan.
- **Variety (variasi):**
 - Berbagai format, jenis dan struktur.
 - Teks, numerik, gambar, audio, video, urutan, deret waktu, data media sosial, susunan multi dimensi, dan lainnya.
 - Data Statik Vs Data Streaming.
 - Aplikasi tunggal dapat menghasilkan/ mengumpulkan banyak jenis data.

Untuk mengekstrak pengetahuan, semua jenis data ini diperlukan untuk terkait secara bersamaan.

Big Data: 3V's

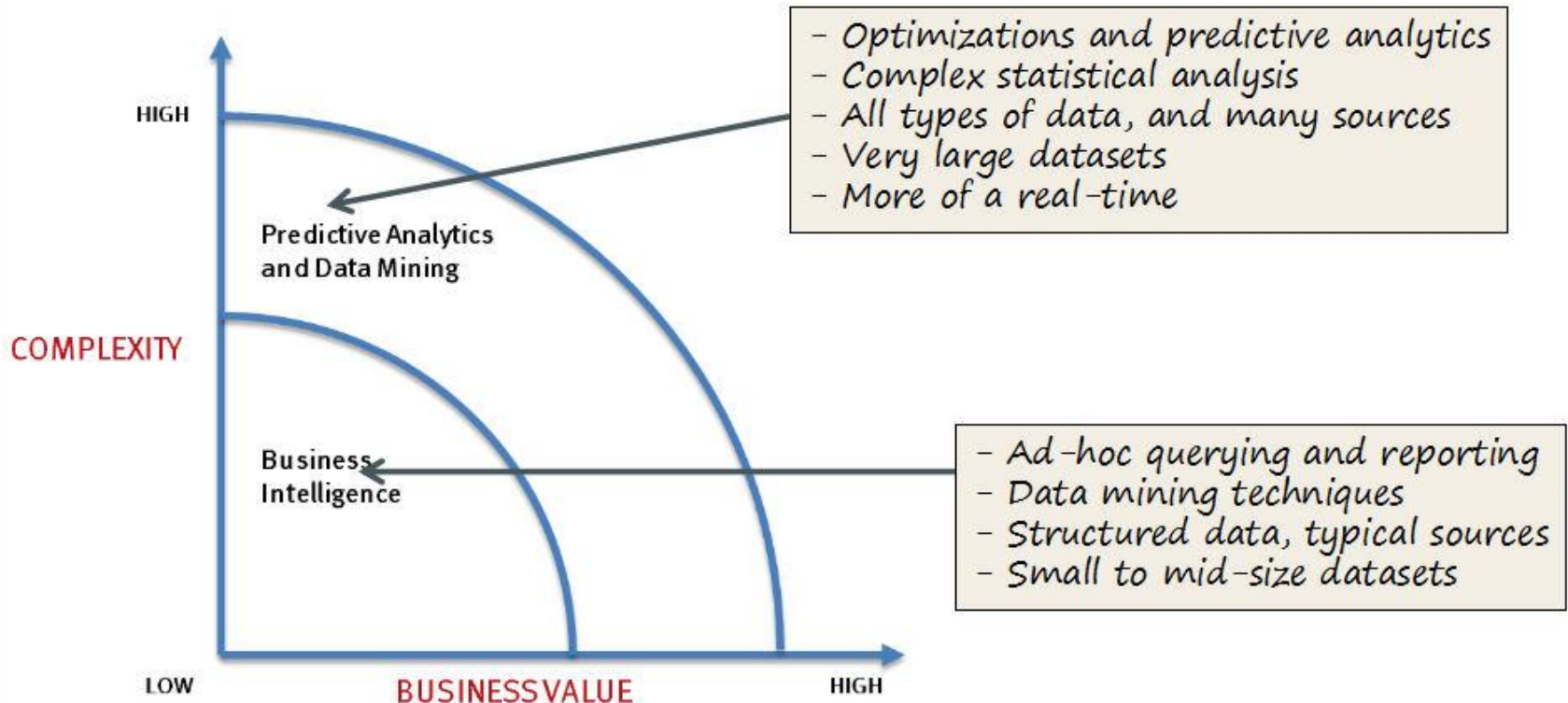


Big Data = Transactions + Interactions + Observations



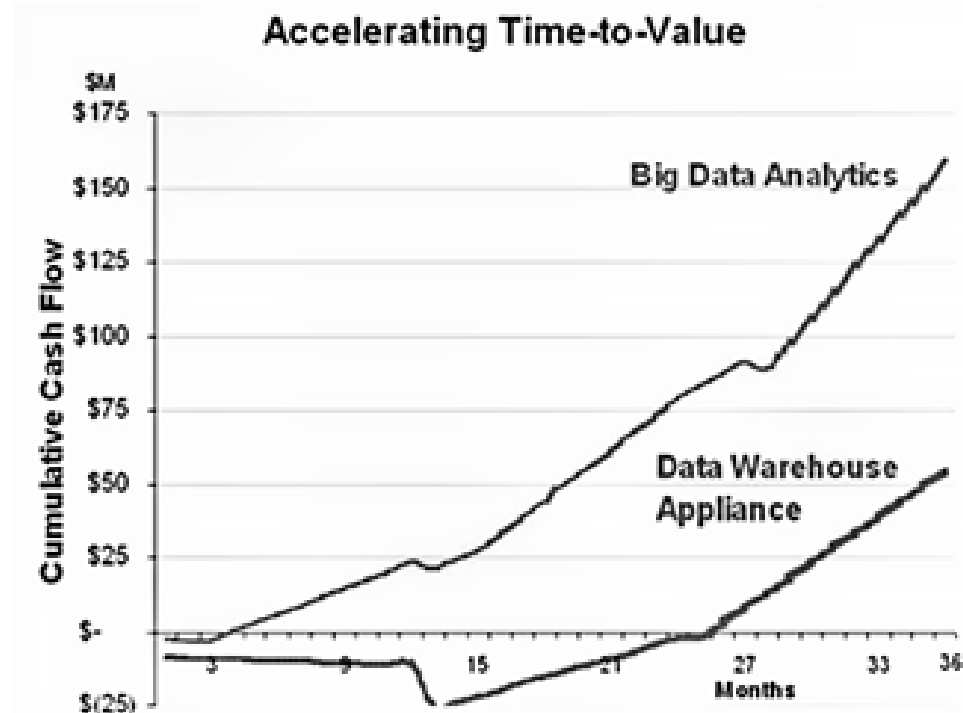
Source: Contents of above graphic created in partnership with Teradata, Inc.

Hal yang Mendorong Kemajuan Big Data



Nilai dari Analisis Big Data:

- Big Data lebih real time di alam daripada aplikasi Data Warehouse tradisional.
- Arsitektur Data Warehouse tradisional (misalnya, Exadata, Teradata) tidak sesuai untuk aplikasi data yang besar.
- Tidak berbagi apa-apa, pemrosesan paralel secara besar-besaran, merancang arsitektur yang sangat sesuai untuk aplikasi Big Data.



Perusahaan yang Menggunakan Big Data

Yahoo!, Amazon, IBM, Microsoft, Facebook, Twitter, Hewlett-Packard, LinkedIn, RECRUIT, Rakuten Japan, dan masih banyak lagi.

Contoh :

Microsoft (Windows Azure Hadoop)

Oracle (Big Data Appliance)

SAP (Hana)

EMC (GreenPlum Hadoop)

Prinsip Big Data:

Tidak membuang data apapun karena residu tersebut mungkin akan menjadi penting sejalannya waktu.

Aplikasi Proses Big Data:

Pengolahan Big Data diklarifikasi menjadi 2 jenis metode :

- Pengolahan data berbasis batch.
- Pengolahan data berbasis real-time.

Big Data Processing Software (Google):

- Google memiliki teknologi canggih yang memungkinkannya mampu mengolah dan memanfaatkan Big Data dengan tepat.
- Teknologi :
 - Google Bigtable.
 - Google MapReduce.
 - Google File System (GFS).
- Google Bigtable:

System penyimpanan data terdistribusi yang ditujukan untuk mengelola data yang terstruktur dan didesain sebagai system yang handal untuk mengelola data dalam skala petabytes dan dalam ribuan mesin (komputer).Google menggunakan Bigtable dalam lebih dari 60 produk dan proyeknya termasuk: Google web indexing, Google Analytics, Google Finance, Orkut, Personalize Search, Writely dan Google Earth.
- Google MapReduce:

Model pemrograman rilis Google yang ditujukan untuk memproses data berukuran raksasa secara terdistribusi dan paralel dalam cluster yang terdiri atas ribuan komputer.

- **Google File System (GFS):**

Salah satu jenis dari media penyimpanan data seperti halnya hard disk drive (HDD), flash disk, DVD-R dan sebagainya. Bedanya, GFS menyimpan data-nya secara terdistribusi pada komputer-komputer dalam suatu cluster.

GFS bisa menyimpan data super besar yang tidak dapat disimpan dalam suatu HDD paling besar sekalipun.

Hadoop

Apache telah merilis software open source yang dikenal dengan nama Hadoop untuk mengembangkan dan menjalankan aplikasi MapReduce.

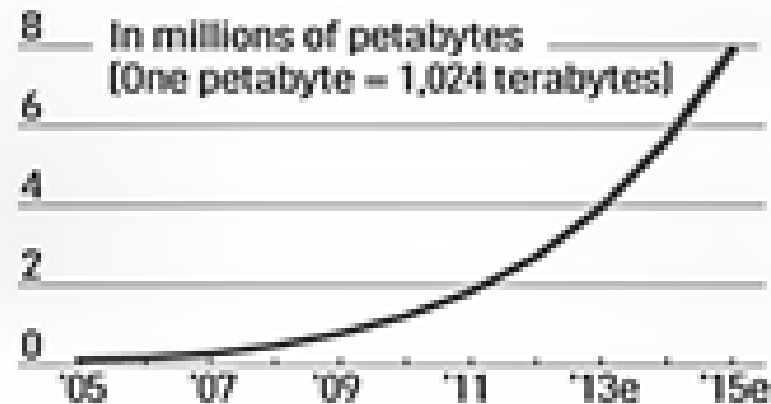
Secara garis besar Hadoop terdiri atas HDFS (Hadoop Distributed File System) dan Hadoop MapReduce. HDFS adalah versi open source-nya GFS (Google File System), dan Hadoop MapReduce adalah versi open source dari Google MapReduce.

Keunggulan:

- Sederhana.
- Fleksibel dalam Ukuran.
- Handal, Anti Gagal.

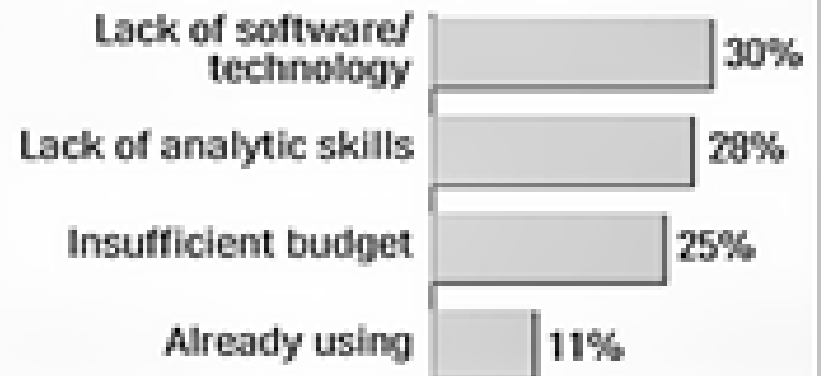
Big Data Boom

Data storage growth



Sources: IDC, DataXu

Big data challenge

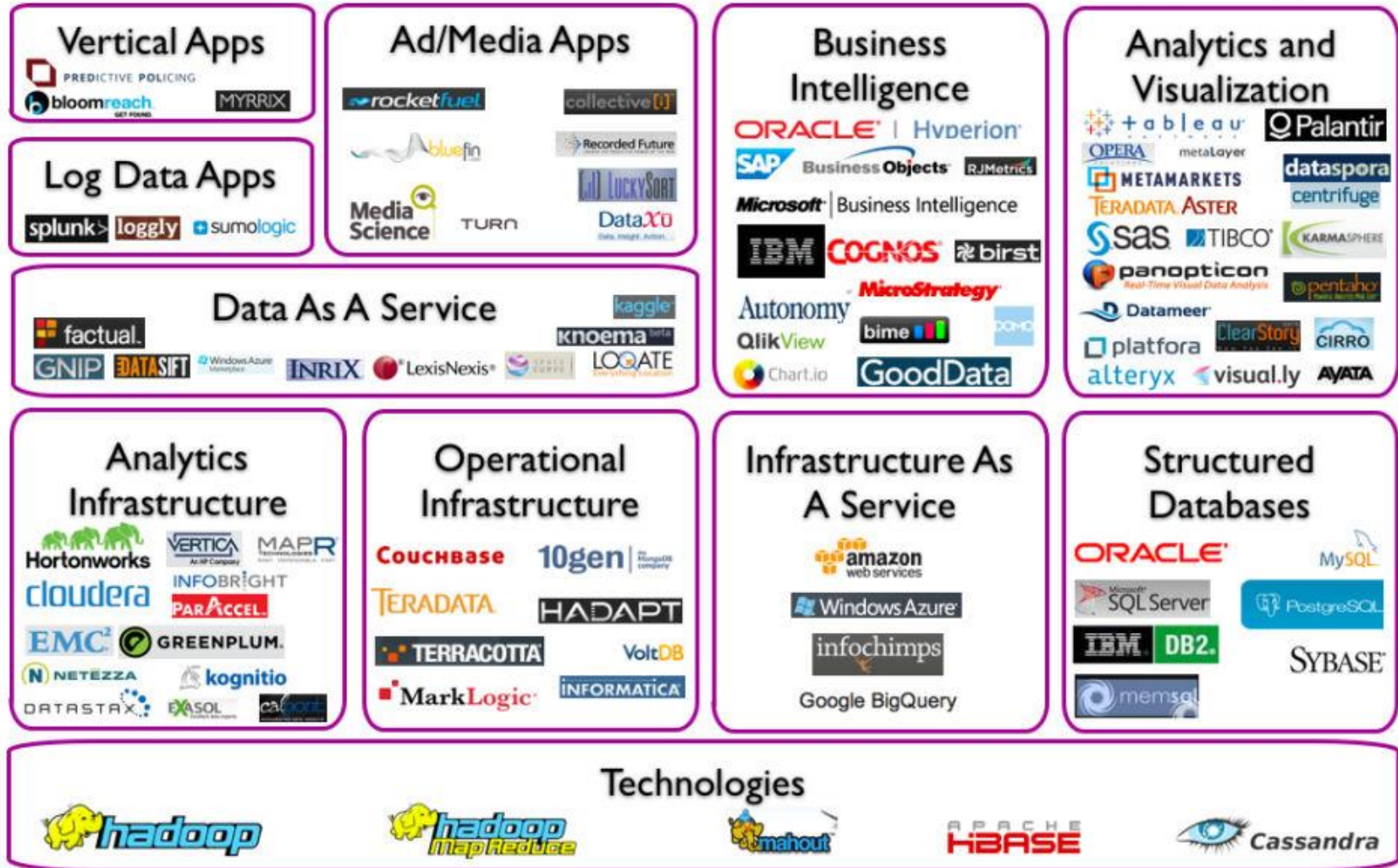


Tantangan dalam Menangani Big Data:

- Bottleneck terdapat pada teknologi:
Arsitektur baru, algoritma, teknik sangat dibutuhkan.
- Keterampilan teknis:
Pakar dalam menggunakan teknologi baru dan berurusan dengan Big Data.

Contoh beberapa teknologi yang harus dimiliki untuk menangani Big Data:

Big Data Landscape





Big Data: The Moving Parts

