

Komputasi Awan

Oleh:

Tim Penyusun Komputasi Awan
PENS

Materi Ajar

1. Pengenalan terhadap Komputasi Awan
2. Layanan Komputasi Awan
3. Kerangka dan Teknologi Komputasi Awan
4. Integrasi dan Standarisasi Komputasi Awan
5. Keamanan Komputasi Awan
6. Aplikasi Awan pada AWS dan GCP
7. Aplikasi Awan dengan Openstack
- 8. Big Data dan Analytics pada Komputasi Awan**
9. Performa, Keandalan dan Ketersediaan Komputasi Awan
10. Migrasi dan Manajemen Komputasi Awan
11. SKKNI Komputasi Awan
12. Prospek Komputasi Awan

Big Data dan Analytic pada Komputasi Awan

Tujuan

- Memperkenalkan Aplikasi Big Data dan Data Data Analytic
- Memperkenalkan Komputasi awan dalam aplikasi Big Data
- Memperkenalkan contoh-contoh cloud untuk bigdata
- Memperkenalkan contoh komputasi awan dan penggunaannya pada Data Analytics

Definisi Big Data

- Big data pertama kali disebut dalam sebuah artikel ilmiah berjudul Application controlled demand paging for out-of-core visualization. Yang ditulis oleh Michael Cox dan David Ellsworth tahun 1997.
- Persoalan yang muncul mengenai Big data dinyatakan dalam rumusan berikut: Visualization provides an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of big data

Definisi BigData

Wikipedia

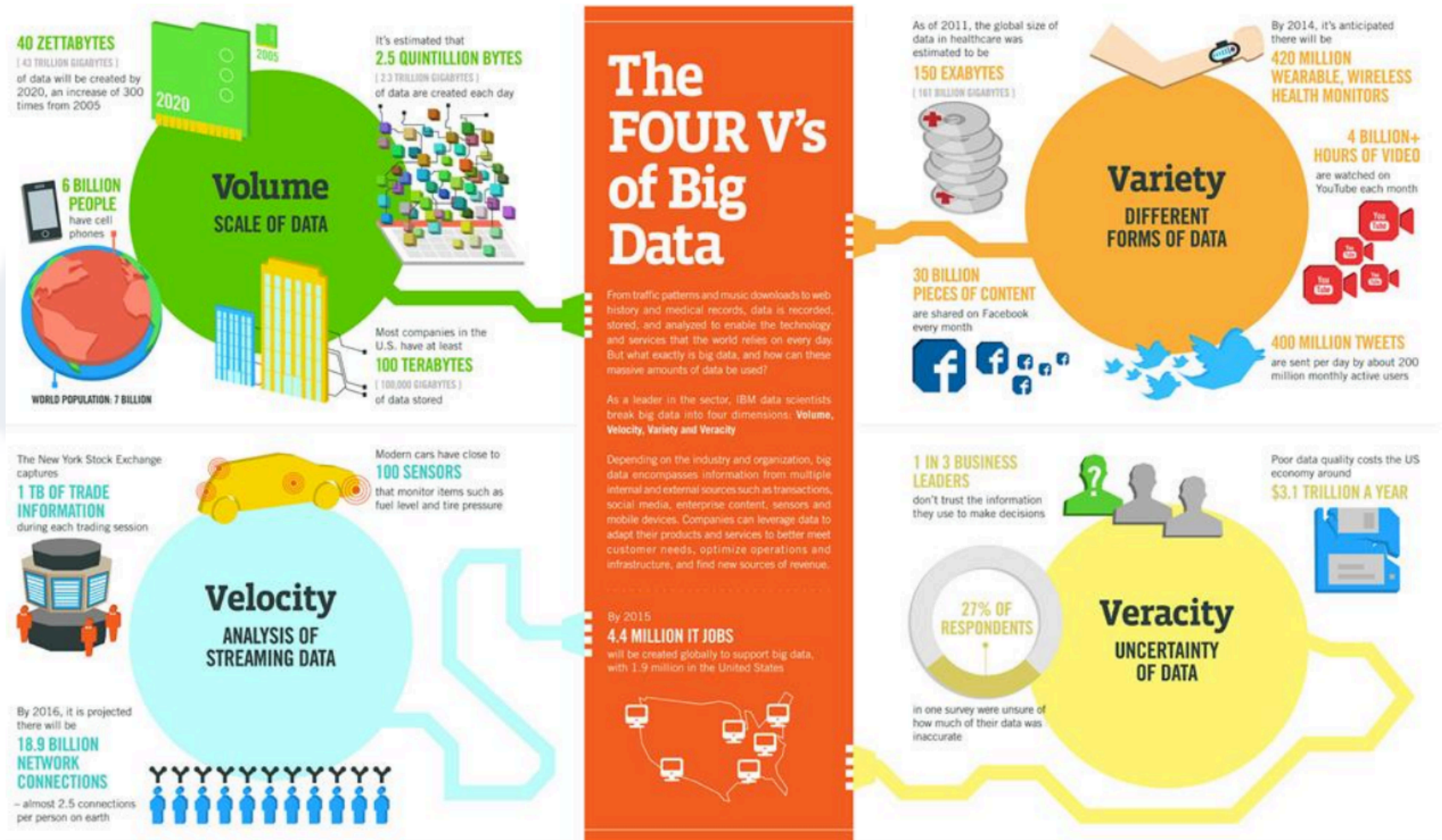
Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Challenges include capture, storage, analysis, data curation, search, sharing, transfer, visualization, querying, updating and information privacy

Kumpulan data yang begitu besar dan kompleks sehingga perangkat lunak aplikasi pemrosesan data tradisional tidak memadai untuk memproses data tersebut

Faktor Munculnya Big Data

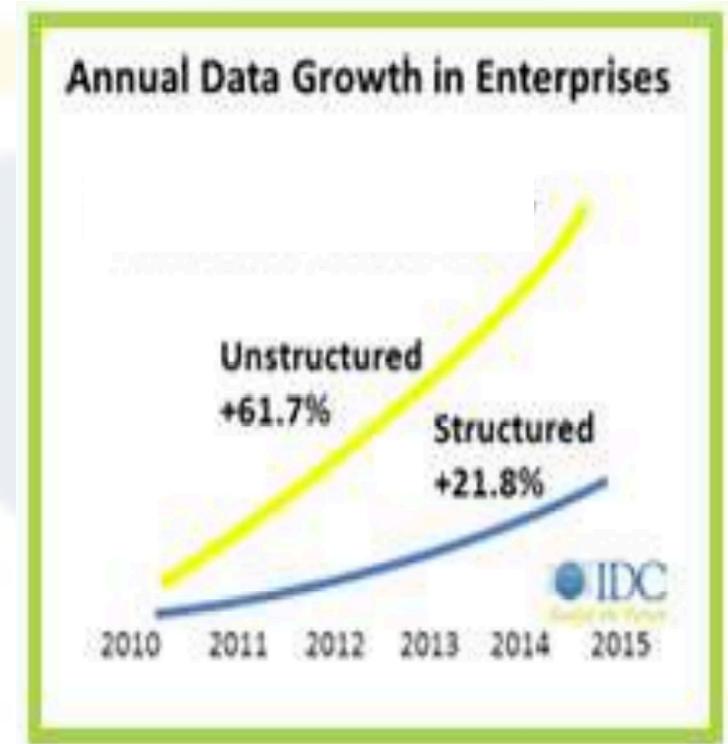
- Perkembangan technology
- Pemakaian Internet
- Perilaku Orang
- Digitalisasi terhadap segala aspek
- Kompetisi Bisnis

Empat V Big Data



Volume

- 2.7 Zetabytes di tahun 2012
- Lebih 90% adalah unstructured data
- Ketika dataset menjadi begitu besar perlu inovasi untuk mengumpulkan, menyimpan, menganalisa supaya berguna



Velocity

- Kebutuhan Analisa data secara real-time untuk pengambilan keputusan
- Tidak adanya jeda waktu antara saat data dibuat atau data yang ditangkap, dan ketika itu juga dapat diakses dan dianalisa karna kebutuhan bisnis.
- Sulit menggunakan sistem tradisional untuk menangkap, menyimpan dan menganalisis pada kondisi tersebut, sehingga perlu sistem baru.

Variety

- keberagaman data : Structure, semi structure dan unstructure data format
- Structure : sistem basis data, sistem datawarehouse
- Semi-Strcutured : data Logs, tweets, sensor, blog
- Unstructured : Freeform text, reports, feedback, dll
- Perlu cara baru dalam data processing

Veracity

- Veracity berhubungan dengan keandalan data.
- System harus tetap bisa menganalisa data yang mungkin ada kesalahan di dalamnya, misal Twitter dengan hashtag, singkatan, kesalahan ketik, dll.

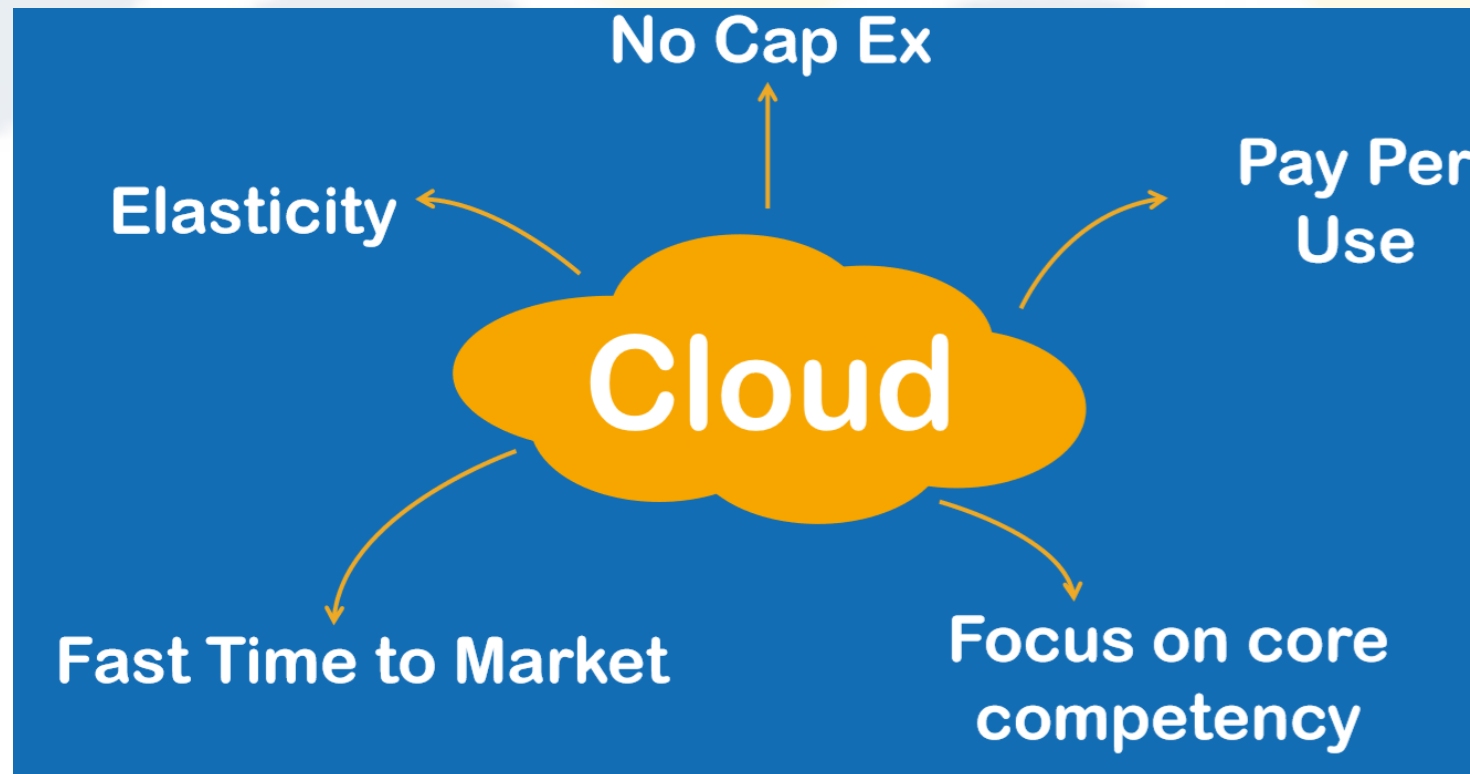
Big Data Tool

- Perlu tool baru yang dibangun untuk data besar
 - Skalabel
 - Parallel computing
 - Bisa Menggunakan hardware komoditas tidak butuh hardware khusus dan digunakan sesuai kebutuhan
 - Butuh skill khusus untuk maintenance



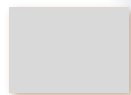
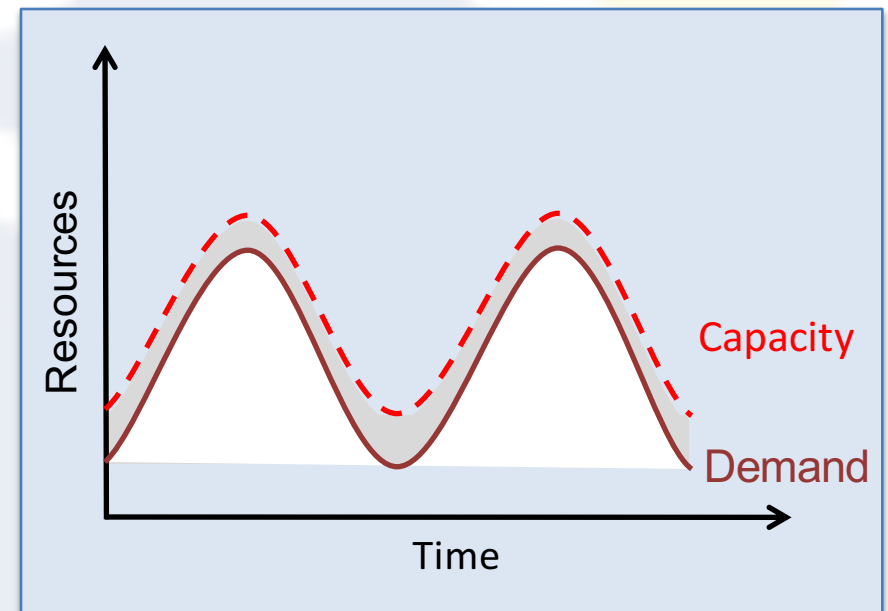
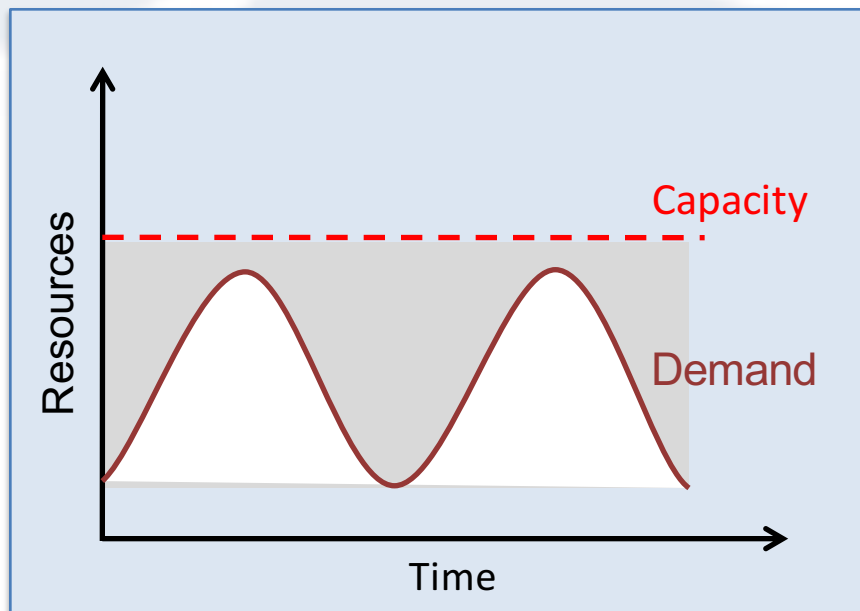
Cloud dan Big Data

- Cloud adalah Best Friend Big Data
- Karakteristik cloud sesuai dengan kebutuhan dari big data



Cloud adalah Best Friend Big Data

- Jika kita butuh “*collect, store, organize, analyze dan sharing*” tapi dengan resource terbatas
- Cloud merupakan pilihan dengan melakukan optimasi secara tepat resource yang dibutuhkan sesuai kebutuhan
- Tantangan Big Data adalah elasticity, data bisa menjadi sangat besar proses bisa menjadi sangat besar.
- Memakai sesuai kebutuhan.



Unused resources

Cloud adalah Best Friend Big Data

Modalitas dari cloud

- Hosted Applications and services
- Pay-as-you-go model
- Scalability, fault-tolerance, elasticity, and self-manageability
- Very large data repositories
- Complex analysis
- Distributed and parallel data processing

Data Analytic dan Cloud Computing

- Skalabel untuk data besar :
 - Scan 100 TB pada 1 node @ 50 MB/detik = 23 hari
 - Scan pada 1000-node cluster = 33 menit
- ➔ Divide-And-Conquer (data partitioning)

Implementasi data Analytics di Cloud

Penelitian tentang data analytic dan cloud computing

- Asynchronous Views over Cloud Data - Yahoo! Research (SIGMOD-International Conference on Management of data tahun 2009)
- DataPath: Data-centric Analytic Engine – Alin Dobra dan Christopher M. Jermaine) (SIGMOD-International Conference on Management of data tahun 2010)

Implementasi MapReduce dengan cloud:

- MapReduce Online (UC Berkeley)
- HadoopDB (Yale)
- Multi-way Join in MapReduce (Afrati&Ullman: EDBT'2010)
- Hadoop++ (Dittrich et al.: VLDB'2010)

Referensi

- Encyclopedia of Cloud Computing, Wiley 2016
- https://www.researchgate.net/publication/221212756_The_DataPath_system_A_data-centric_analytic_processing_engine_for_large_data_warehouses
- <https://www.slideshare.net/AmazonWebServices/big-data-and-the-cloud-a-best-friend-story>

TERIMA KASIH