

# Pengenalan Big Data

BIG DATA – TK13025

# Penilaian

- Penilaian: Tugas 40%, UTS 30%, UAS 30%
- **Tugas TI: (analisis Big Data 40%, quis 30%, praktikum 30%)**
- **Tugas SI: (analisis Big Data 40%, quis 30%, Tugas Mandiri 30%)**
- 16 Pertemuan: 14 kuliah, 1 UTS, 1 UAS
- Absensi mengikuti aturan universitas



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Definisi

- Big Data adalah bidang yang didedikasikan untuk analisis, pemrosesan, dan penyimpanan kumpulan data besar yang sering kali berasal dari sumber yang berbeda.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
Linggi

QS  
STARS  
RATING SYSTEM  
2016



CPA  
AUSTRALIA



ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Definisi

- “Big data adalah aset informasi bervolume tinggi (high-**volume**), berkecepatan tinggi (high-**velocity**), dan beragam (high-**variety**) yang menuntut bentuk pemrosesan informasi yang hemat biaya dan inovatif untuk meningkatkan wawasan dan pengambilan keputusan.” – Gartner
- “Sementara perusahaan berjuang untuk mengkonsolidasikan sistem dan meruntuhkan database yang berlebihan untuk memungkinkan konsistensi operasional, analitis, dan kolaboratif yang lebih besar, perubahan kondisi ekonomi telah membuat pekerjaan ini lebih sulit. E-commerce, khususnya, telah meledakkan tantangan manajemen data dalam tiga dimensi: **volume**, **kecepatan**, dan **variasi**. Pada tahun 2001/02, organisasi TI banyak menyusun berbagai pendekatan yang mereka miliki untuk menangani masing-masing.” – Doug Laney



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Apa yang membuat Big Data dibutuhkan?



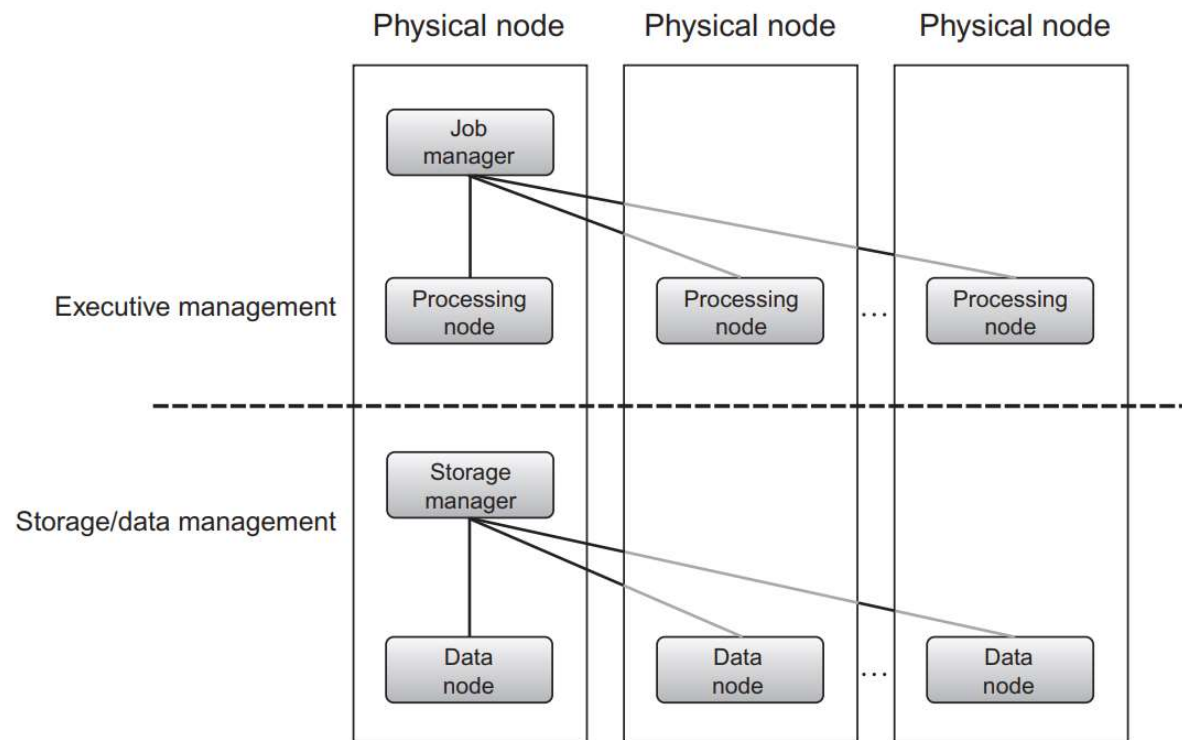
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

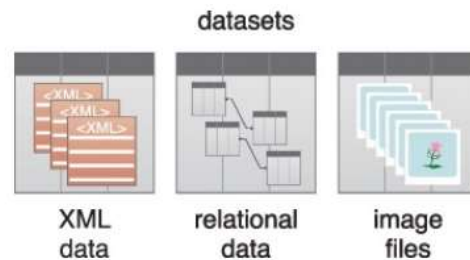
# Sumber Daya Komputasi Utama untuk Big Data

- Kemampuan pemrosesan: CPU, prosesor, atau node
- Memory
- Storage
- Jaringan



# Konsep dan Terminologi: Datasets

- Kumpulan atau kelompok data terkait umumnya disebut sebagai kumpulan data.
- Setiap kelompok atau anggota dataset (datum) berbagi set atribut atau properti yang sama dengan yang lain dikumpulan data yang sama.
- Beberapa contoh kumpulan data adalah:
  - ✓ tweet disimpan dalam file
  - ✓ kumpulan file gambar dalam direktori
  - ✓ ekstrak baris dari tabel database yang disimpan dalam file berformat CSV
  - ✓ pengamatan cuaca historis yang disimpan sebagai file XML
- Datasets dapat ditemukan dalam berbagai format.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
tinggi

QS  
STARS  
RATING SYSTEM

CSA  
SAR

IAEE

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**



## Konsep dan Terminologi: *Data Analysis*

- *Data analysis* adalah proses pemeriksaan data untuk menemukan fakta, hubungan, pola, wawasan dan/atau tren.
- Tujuan keseluruhan dari analisis data adalah untuk mendukung pengambilan keputusan yang lebih baik.
- Contoh analisis data sederhana adalah analisis data penjualan es krim untuk mengetahui hubungan jumlah es krim yang terjual dengan suhu harian.
- Hasil analisis semacam itu akan mendukung keputusan yang terkait dengan berapa banyak es krim di toko harus disediakan sehubungan dengan informasi ramalan cuaca.
- Melakukan analisis data membantu membangun pola dan hubungan di antara data yang dianalisis



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**



# Konsep dan Terminologi: *Data Analytics*



- *Data analytics* adalah istilah yang lebih luas yang mencakup analisis data.
- *Data analytics* adalah disiplin yang mencakup pengelolaan siklus hidup data yang lengkap, yang meliputi pengumpulan, pembersihan, pengorganisasian, penyimpanan, analisis, dan pengaturan data.
- Istilah ini mencakup pengembangan metode analisis, teknik ilmiah, dan otomatisasi peralatan.
- Di lingkungan Big Data, *data analytics* telah mengembangkan metode yang memungkinkan data analisis terjadi melalui penggunaan teknologi terdistribusi yang sangat skalabel dan kerangka kerja yang mampu menganalisis data dalam jumlah besar dari berbagai sumber



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
Linggi

QS  
STARS  
RATING SYSTEM

ISAS  
LIVABLE

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Konsep dan Terminologi: *Data Analytics*

- Siklus hidup analitik Big Data umumnya melibatkan identifikasi, pengadaan, persiapan, dan menganalisis sejumlah besar data mentah dan tidak terstruktur untuk mengekstrak informasi bermakna yang dapat berfungsi sebagai masukan untuk mengidentifikasi pola, memperkaya data perusahaan yang ada dan melakukan pencarian skala besar.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
tinggi

QS STARS  
RATING SYSTEM  
5 star



CPA  
AUSTRALIA



ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Konsep dan Terminologi: *Data Analytics*

- Berbagai jenis organisasi menggunakan alat dan teknik analisis data dengan cara yang berbeda.
- Contoh:
  - ✓ Dalam lingkungan berorientasi bisnis, hasil analisis data dapat menurunkan biaya operasional dan memfasilitasi pengambilan keputusan strategis
  - ✓ Dalam domain ilmiah, analitik data dapat membantu mengidentifikasi penyebab suatu fenomena untuk meningkatkan akurasi prediksi.
  - ✓ Dalam lingkungan berbasis layanan seperti organisasi sektor publik, analisis data dapat membantu memperkuat fokus dalam memberikan layanan berkualitas tinggi dengan menurunkan biaya.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Konsep dan Terminologi: *Data Analytics*

- Data analytics memungkinkan pengambilan keputusan berdasarkan data dengan dukungan ilmiah sehingga keputusan dapat didasarkan pada data faktual dan bukan hanya pada pengalaman atau intuisi masa lalu.
- Ada empat kategori umum analitik yang dibedakan berdasarkan hasil yang diperoleh:
  - ✓ descriptive analytics
  - ✓ diagnostic analytics
  - ✓ predictive analytics
  - ✓ prescriptive analytics



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Konsep dan Terminologi: *Data Analytics*

- Jenis analitik yang berbeda memanfaatkan teknik dan algoritma analisis yang berbeda.
- Ini menyiratkan bahwa mungkin ada berbagai data, penyimpanan, dan persyaratan pemrosesan untuk memfasilitasi penyampaian berbagai jenis hasil analitik.



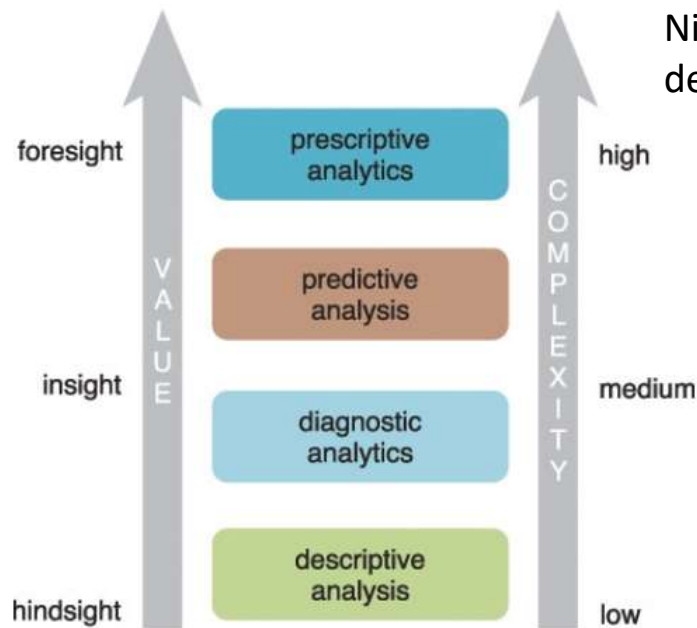
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Konsep dan Terminologi: *Data Analytics*

- Gambar di bawah ini menggambarkan kenyataan bahwa generasi hasil analitik bernilai tinggi meningkatkan kompleksitas dan biaya analitik lingkungan.



Nilai (*value*) dan kompleksitas (*complexity*) meningkat dari analitik deskriptif menjadi preskriptif.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Descriptive Analytics

- Analisis deskriptif dilakukan untuk menjawab pertanyaan tentang peristiwa yang telah terjadi.
- Bentuk analitik ini mengontekstualisasikan data untuk menghasilkan informasi.
- Contoh pertanyaan dapat mencakup:
  - ✓ Berapa volume penjualan selama 12 bulan terakhir?
  - ✓ Berapa jumlah “telepon darurat” yang diterima yang dikategorikan berdasarkan tingkat keparahan dan lokasi geografis?
  - ✓ Berapa komisi bulanan yang diperoleh setiap agen penjualan?
- Diperkirakan bahwa 80% dari hasil analitik yang dihasilkan bersifat deskriptif.
- Dari segi nilai, analitik deskriptif memberikan nilai yang paling rendah dan membutuhkan keahlian yang relatif mendasar.



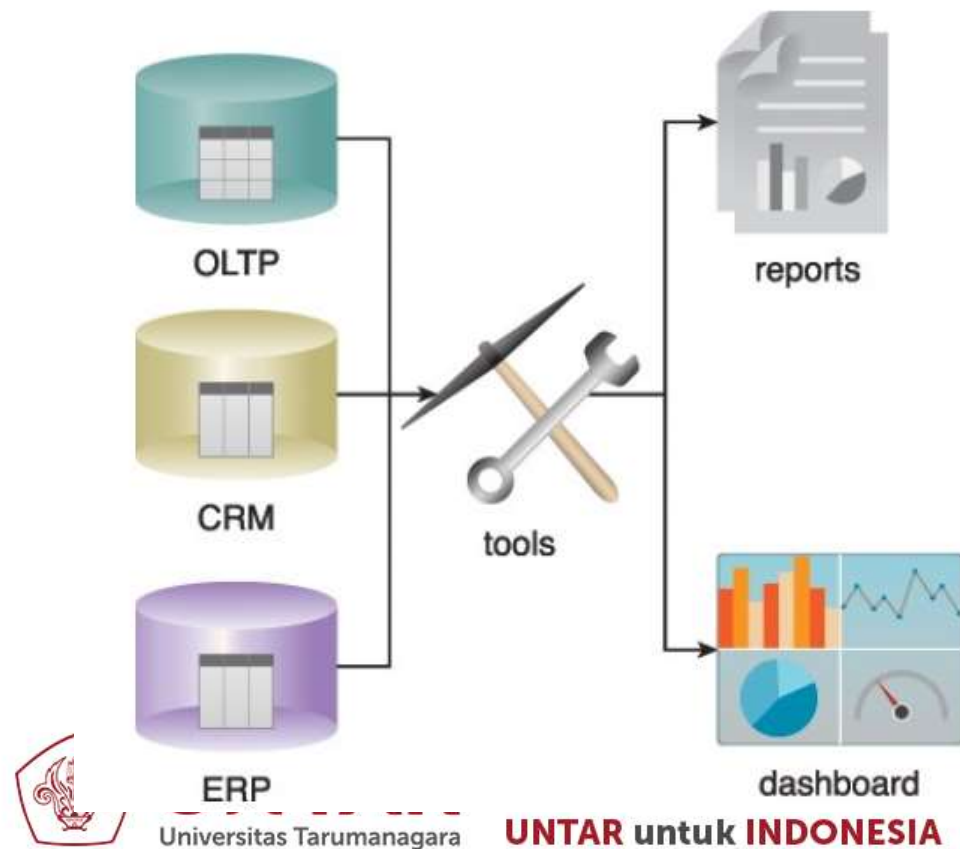
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Descriptive Analytics

- Sistem operasional, gambar kiri, ditanyakan melalui analitik deskriptif alat untuk menghasilkan laporan atau dasbor, gambar kanan.
- Laporan umumnya bersifat statis dan menampilkan data historis yang disajikan dalam bentuk grid data atau grafik.
- Kueri dieksekusi pada data operasional toko dari dalam suatu perusahaan, misalnya sistem *Customer Relationship Management (CRM)* atau sistem *Enterprise Resource Planning (ERP)*.
- *Online transaction processing (OLTP)*.





# Analisis Diagnostik (Diagnostic Analytics)

- Analisis diagnostik bertujuan untuk menentukan penyebab dari suatu fenomena yang terjadi di masa lalu menggunakan pertanyaan yang berfokus pada alasan di balik peristiwa tersebut.
- Tujuannya adalah untuk menentukan informasi apa yang terkait dengan fenomena untuk memungkinkan menjawab pertanyaan yang berusaha untuk menentukan mengapa sesuatu telah terjadi.
- Pertanyaan tersebut meliputi:
  - ✓ Mengapa penjualan Q2 kurang dari penjualan Q1?
  - ✓ Mengapa ada lebih banyak “telepon darurat” yang berasal dari wilayah Timur dari pada dari wilayah barat?
  - ✓ Mengapa ada peningkatan tingkat penerimaan kembali pasien selama tiga bulan terakhir?



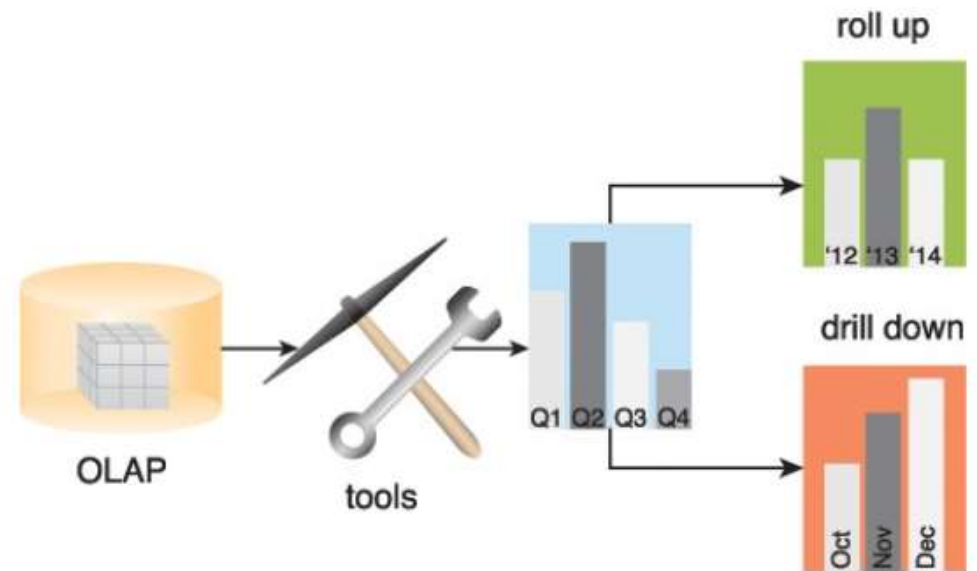
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Analisis Diagnostik (Diagnostic Analytics)

- Analitik diagnostik memberikan nilai lebih dari pada analitik deskriptif tetapi membutuhkan lebih banyak keahlian tingkat lanjut.
- Analitik diagnostik biasanya memerlukan pengumpulan data dari beberapa sumber dan menyimpannya dalam struktur yang cocok untuk melakukan *drill-down* dan *roll-up* analisis



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Analisis Diagnostik (Diagnostic Analytics)

- Analisis drill-down berarti mendapatkan data rinci tentang ringkasan informasi yang menekankan pada hal tertentu.
- Analisis roll-up untuk mendapatkan ringkasan.
- Hasil analitik diagnostik dilihat melalui interaktif alat visualisasi yang memungkinkan pengguna untuk mengidentifikasi tren dan pola.
- Kueri yang dieksekusi lebih kompleks dibandingkan dengan analitik deskriptif dan dilakukan pada data multidimensi yang disimpan dalam sistem pemrosesan analitik.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Predictive Analytics

- Analisis prediktif dilakukan dalam upaya untuk menentukan hasil dari suatu peristiwa yang mungkin terjadi di masa depan.
- Dengan analitik prediktif, informasi ditingkatkan dengan makna untuk menghasilkan pengetahuan yang menyampaikan bagaimana informasi itu terkait.
- Kekuatan dan besarnya asosiasi membentuk dasar model yang digunakan untuk menghasilkan masa depan prediksi berdasarkan peristiwa masa lalu.
- Penting untuk dipahami bahwa model yang digunakan untuk analitik prediktif memiliki ketergantungan implisit pada kondisi di mana masa lalu peristiwa terjadi.
- Jika kondisi yang mendasari ini berubah, maka model yang membuat prediksi perlu diperbarui.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Predictive Analytics

- Pertanyaan biasanya dirumuskan dengan menggunakan *what-if*, seperti berikut:
  - ✓ Berapa peluang seorang nasabah akan gagal membayar pinjaman jika mereka melewatkan pembayaran bulanan?
  - ✓ Berapa tingkat kelangsungan hidup pasien jika Obat B diberikan sebagai pengganti Obat A?
  - ✓ Jika seorang pelanggan telah membeli Produk A dan B, berapa peluang mereka akan membeli Produk C?



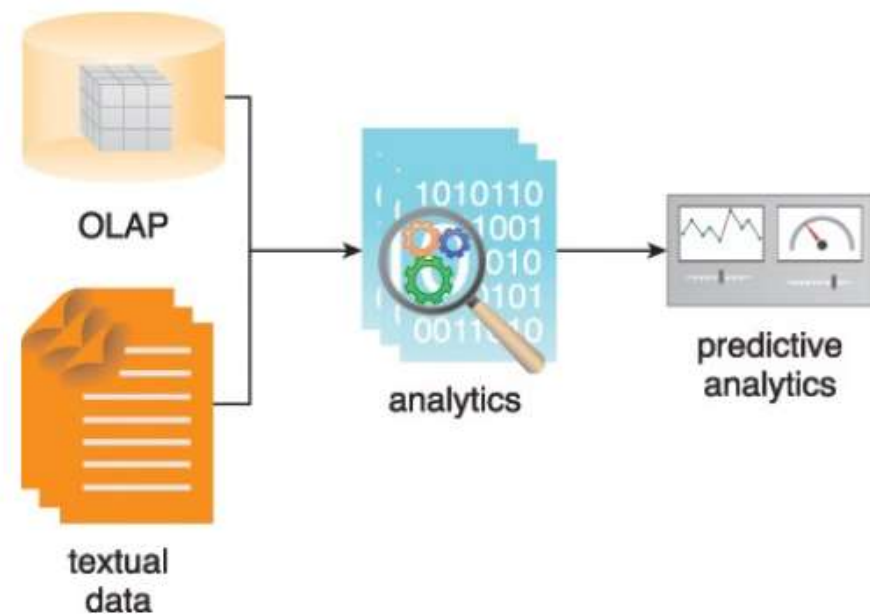
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Predictive Analytics

- Analitik prediktif mencoba memprediksi hasil peristiwa, dan prediksi dibuat berdasarkan pada pola, tren, dan pengecualian yang ditemukan dalam data historis dan terkini.
- Hal ini dapat menyebabkan identifikasi resiko dan peluang.
- Analisis semacam ini melibatkan penggunaan kumpulan data besar yang terdiri dari data internal dan eksternal dan berbagai teknik analisis data.
- Hal ini memberikan nilai yang lebih besar dan membutuhkan lebih banyak keahlian lanjutan dari pada analitik deskriptif dan diagnostik.
- Alat analitik prediktif dapat menyediakan antarmuka *front-end* yang ramah pengguna.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Prescriptive Analytics

- Analitik preskriptif dibangun berdasarkan hasil analitik prediktif dengan menentukan tindakan yang harus diambil.
- Fokusnya tidak hanya pada pilihan mana yang terbaik untuk diikuti, tapi kenapa.
- Dengan kata lain, analitik preskriptif memberikan hasil yang dapat dinalar karena mereka menanamkan unsur pemahaman situasional.
- Dengan demikian, analitik semacam ini dapat digunakan untuk mendapatkan keuntungan atau mengurangi risiko.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Prescriptive Analytics

- Analitik preskriptif memberikan nilai lebih dari pada jenis analitik lainnya dan karenanya memerlukan keahlian paling canggih, serta perangkat lunak dan peralatan khusus.
- Pendekatan bergeser dari penjelasan ke nasihat dan dapat mencakup simulasi berbagai skenario.
- Contoh pertanyaan dapat mencakup:
  - Di antara tiga obat, mana yang memberikan hasil terbaik?
  - Kapan waktu terbaik untuk memperdagangkan saham tertentu?



**UNTAR**  
Universitas Tarumanagara

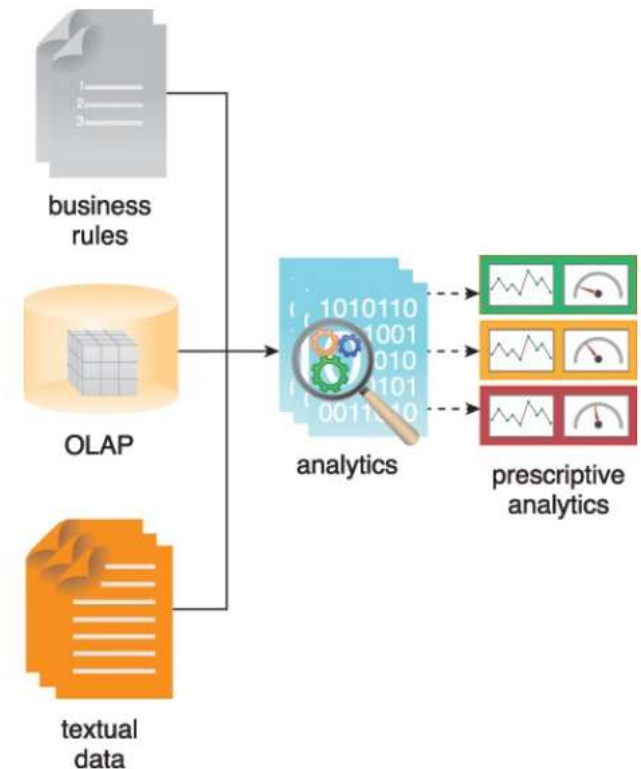


**UNTAR untuk INDONESIA**



# Prescriptive Analytics

- Analisis semacam ini menggabungkan data internal dengan data eksternal.
- Data internal termasuk data penjualan saat ini dan data historis, informasi pelanggan, data produk, dan aturan bisnis.
- Data eksternal dapat mencakup data media sosial, prakiraan cuaca, dan data demografi yang dihasilkan pemerintah.
- Analitik preskriptif melibatkan penggunaan aturan bisnis dan sejumlah besar data internal dan eksternal untuk mensimulasikan hasil dan meresepkan tindakan yang terbaik.
- Analitik preskriptif melibatkan penggunaan aturan bisnis dan internal dan/atau data eksternal untuk melakukan analisis mendalam.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Business Intelligence (BI)

- BI memungkinkan organisasi untuk mendapatkan wawasan tentang kinerja perusahaan dengan menganalisis data yang dihasilkan oleh proses bisnis dan sistem informasinya.
- Hasil analisis dapat digunakan oleh manajemen untuk mengarahkan bisnis dalam upaya memperbaiki masalah yang terdeteksi atau meningkatkan kinerja organisasi.
- BI menerapkan analitik ke sejumlah besar data di seluruh perusahaan, yang biasanya telah dikonsolidasikan ke dalam gudang data perusahaan untuk menjalankan kueri analitik.



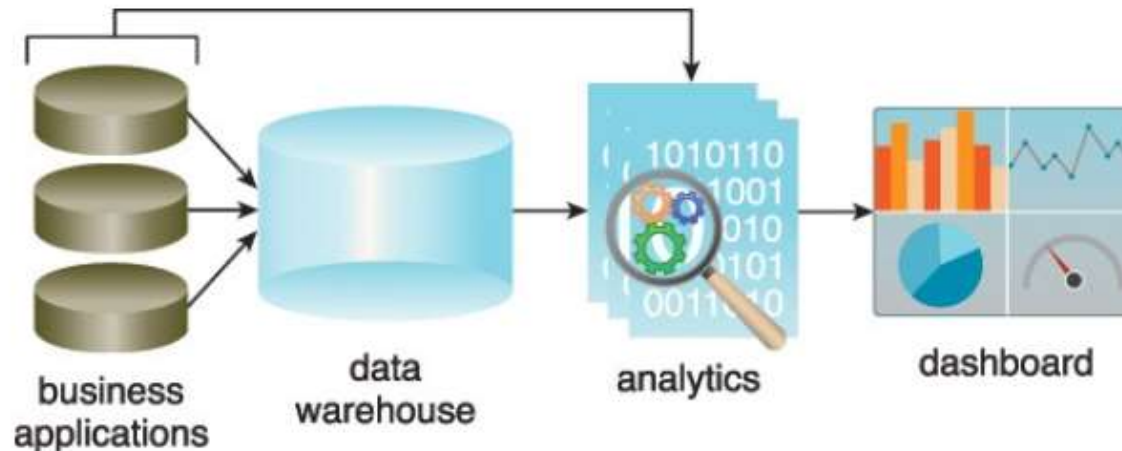
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Business Intelligence (BI)

- Keluaran BI dapat ditampilkan ke dasbor yang memungkinkan manajer mengakses dan menganalisis hasilnya dan berpotensi menyaring kueri analitik untuk mengeksplorasi data lebih lanjut.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Key Performance Indicators (KPI)

- KPI adalah metrik yang dapat digunakan untuk mengukur kesuksesan dalam konteks bisnis tertentu.
- KPI terkait dengan tujuan dan sasaran strategis perusahaan secara keseluruhan.
- KPI sering digunakan untuk mengidentifikasi masalah kinerja bisnis dan menunjukkan kepatuhan terhadap peraturan.
- Oleh karena itu, KPI bertindak sebagai titik referensi yang dapat diukur untuk mengukur aspek tertentu dari suatu kinerja keseluruhan bisnis.
- Dasbor menggabungkan tampilan beberapa KPI dan membandingkan pengukuran aktual dengan nilai ambang batas yang menentukan rentang nilai yang dapat diterima dari KPI.



KPI dashboard



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
linggul

QS  
STARS  
RATING SYSTEM  
5 stars



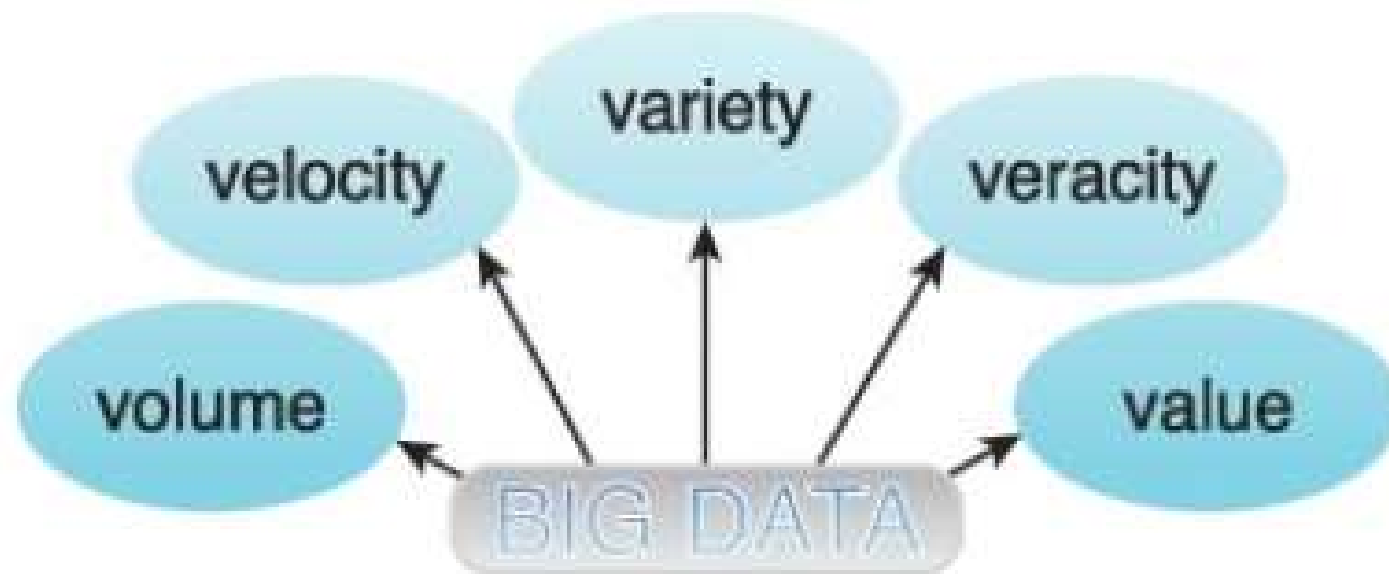
CPA  
AUSTRALIA



ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Karakteristik Big Data



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
tinggi

QS  
STARS  
RATING SYSTEM  
2016

AMBA  
AACSB  
EQUIS

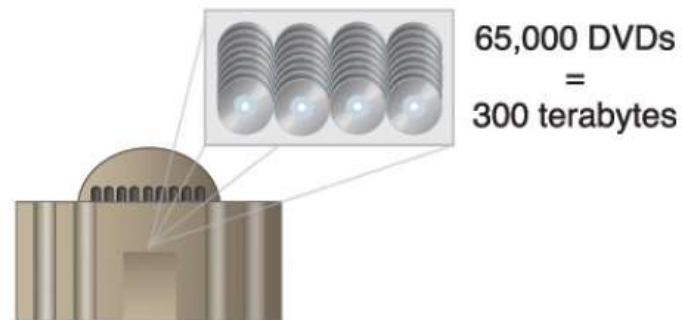
CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Karakteristik Big Data: Volume

- Volume data yang tinggi memaksakan permintaan penyimpanan dan pemrosesan data yang berbeda, serta proses persiapan, kurasi, dan pengelolaan data tambahan.



**UNTAR untuk INDONESIA**

# Karakteristik Big Data: Volume

- Sumber data yang menghasilkan volume data yang tinggi dapat mencakup:
  - ✓ transaksi online, seperti point-of-sale dan perbankan
  - ✓ eksperimen ilmiah dan penelitian
  - ✓ sensor, seperti sensor GPS, RFID, pengukur pintar, dan telematika
  - ✓ media sosial, seperti Facebook dan Twitter



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Karakteristik Big Data: Velocity

- Di lingkungan Big Data, data dapat tiba dengan kecepatan tinggi, dan kumpulan data yang sangat besar dapat terakumulasi dalam waktu yang sangat singkat.
- Dari sudut pandang perusahaan, kecepatan data diterjemahkan ke dalam jumlah waktu yang dibutuhkan untuk data diproses sekali itu memasuki perimeter perusahaan.
- Mengatasi arus masuk data yang cepat perusahaan harus untuk merancang solusi pemrosesan data yang sangat elastis dan tersedia dan kemampuan penyimpanan data yang sesuai.



**UNTAR**  
Universitas Tarumanagara

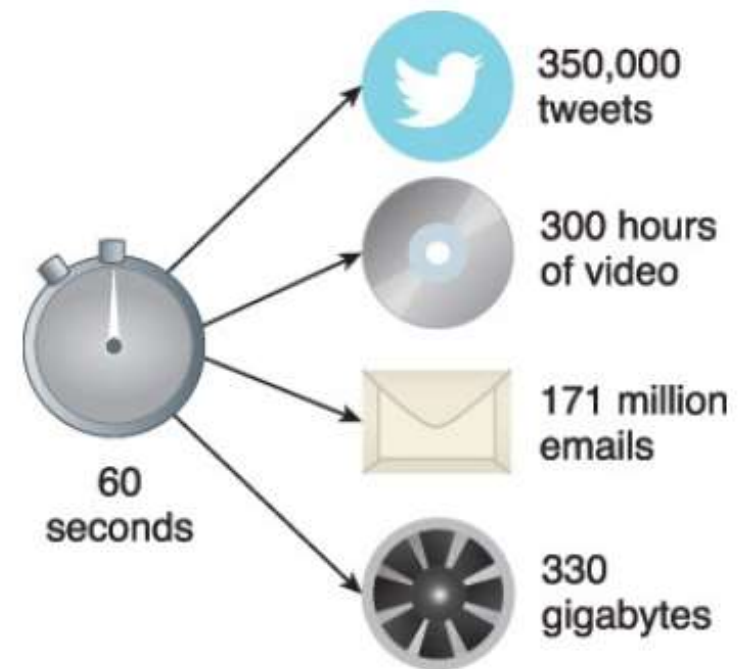


**UNTAR untuk INDONESIA**



# Karakteristik Big Data: Velocity

- Contoh kumpulan data Big Data berkecepatan tinggi yang dihasilkan setiap menit termasuk tweet, video, email, dan GB yang dihasilkan dari mesin jet.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Karakteristik Big Data: Variety

- Keragaman data mengacu pada berbagai format dan jenis data yang perlu didukung oleh solusi Big Data.
- Keragaman data membawa tantangan bagi perusahaan dalam hal data integrasi, transformasi, pemrosesan, dan penyimpanan.
- Contoh kumpulan data Big Data yang sangat beragam termasuk data terstruktur, tekstual, gambar, video, audio, XML, JSON, data sensor dan metadata.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN PT

A  
tinggi

QS  
STARS  
RATING SYSTEM  
2020

5  
4  
3  
2  
1

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Karakteristik Big Data: Veracity

- Veracity mengacu pada kualitas data.
- Data yang masuk ke lingkungan Big Data perlu dinilai kualitasnya, yang dapat menyebabkan kegiatan pemrosesan data mengatasi masalah data yang tidak valid dan menghilangkan *noise*.
- Sehubungan dengan kebenaran, data dapat menjadi bagian dari sinyal atau noise dari kumpulan data.
- *Noise* adalah data yang tidak dapat diubah menjadi informasi dan karenanya tidak memiliki nilai, sedangkan *signals* memiliki nilai dan mengarah pada informasi yang bermakna.
- Data dengan rasio signal-to-noise tinggi memiliki lebih banyak kebenaran daripada data dengan rasio yang lebih rendah.
- Data yang didapat secara terkontrol, misalnya melalui pendaftaran pelanggan online, biasanya berisi lebih sedikit noise daripada data yang diperoleh melalui sumber yang tidak terkontrol, seperti posting blog.
- Jadi rasio signal-to-noise data tergantung pada sumber data dan jenisnya.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Karakteristik Big Data: Value

- *Value* didefinisikan sebagai kegunaan data untuk suatu perusahaan.
- Sifat nilainya adalah secara intuitif terkait dengan karakteristik kejujuran karena semakin tinggi fidelitas data, semakin banyak nilai yang dimiliki untuk bisnis.
- *Value* tergantung pada berapa lama pemrosesan data berlangsung karena hasil analitik memiliki masa simpan; misalnya, penawaran saham tertunda 20 menit memiliki sedikit atau tidak ada nilai untuk melakukan perdagangan dibandingkan dengan penawaran yang berumur 20 milidetik.
- Seperti yang ditunjukkan, nilai dan waktu berbanding terbalik.
- Semakin lama waktu yang dibutuhkan untuk data berubah menjadi informasi yang berarti, semakin sedikit nilainya untuk bisnis.



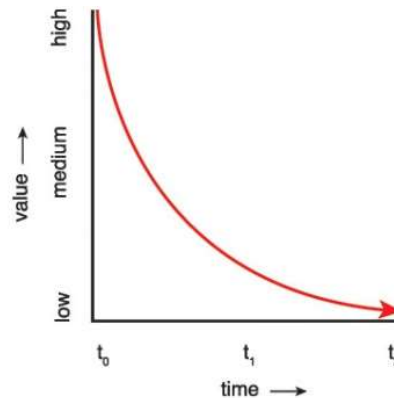
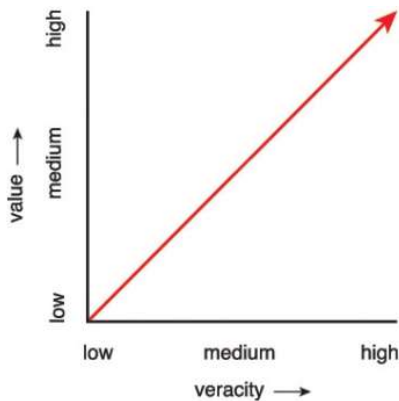
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Karakteristik Big Data: Value

- Data yang memiliki veracity yang tinggi dan dapat dianalisis dengan cepat memiliki nilai lebih untuk bisnis.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
tinggi

QS  
STARS  
RATING SYSTEM  
5 star

AMBA  
AACSB  
EFMD

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Karakteristik Big Data: Value

- Terlepas dari *veracity* dan waktu, nilai juga dipengaruhi oleh siklus hidup berikut yang terkait kekhawatiran:
  - ✓ Seberapa baik data telah disimpan?
  - ✓ Apakah atribut berharga dari data dihapus selama pembersihan data?
  - ✓ Apakah jenis pertanyaan yang tepat diajukan selama analisis data?
  - ✓ Apakah hasil analisis dikomunikasikan secara akurat kepada pihak pembuat keputusan yang tepat?



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Berbagai Jenis Data

- Data yang diproses oleh solusi Big Data dapat dibuat oleh manusia atau dibuat oleh mesin, meskipun pada akhirnya merupakan tanggung jawab mesin untuk menghasilkan hasil analitik.
- Data yang dihasilkan manusia adalah hasil interaksi manusia dengan sistem, seperti layanan online dan perangkat digital.
- Data yang dihasilkan mesin dihasilkan oleh program perangkat lunak dan perangkat keras direspon terhadap peristiwa dunia nyata.
- Misalnya, file log menangkap keputusan otorisasi dibuat oleh layanan keamanan, dan sistem point-of-sale menghasilkan transaksi terhadap persediaan untuk mencerminkan barang yang dibeli oleh pelanggan.
- Dari perspektif perangkat keras, dan contoh data yang dihasilkan mesin adalah informasi yang disampaikan dari banyak sensor di ponsel yang mungkin melaporkan informasi, termasuk posisi dan menara seluler kekuatan sinyal.

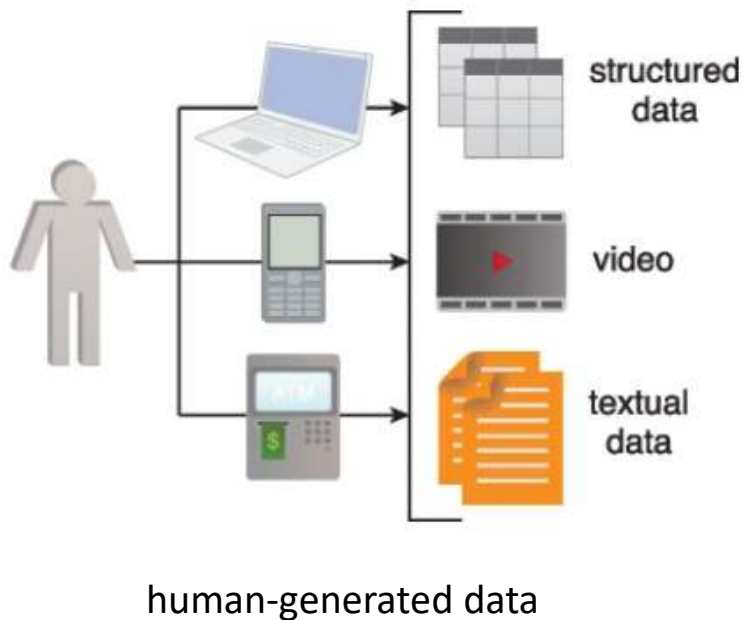


**UNTAR**  
Universitas Tarumanagara

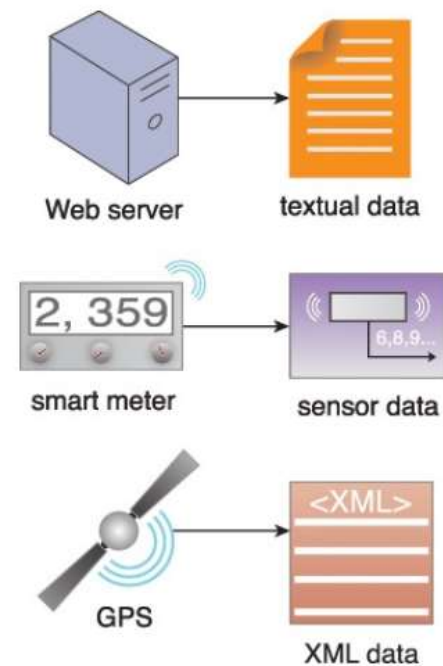


**UNTAR untuk INDONESIA**

# Berbagai Jenis Data



Contoh data yang dihasilkan manusia termasuk media sosial, posting blog, email, berbagi foto, dan pemesanan.



machine-generated data

Contoh data yang dihasilkan mesin termasuk log web, data sensor, data telemetri, data smart meter dan data penggunaan alat.



# Berbagai Jenis Data

- Data yang dibuat oleh manusia dan yang dibuat oleh mesin dapat berasal dari berbagai sumber dan direpresentasikan dalam berbagai format atau jenis.
- Jenis data utama adalah:
  - ✓ structured data
  - ✓ unstructured data
  - ✓ semi-structured data



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
Linggi

QS  
STARS  
RATING SYSTEM  
2020

CSO  
KEMEN  
KEMEN  
KEMEN

IAEE

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Structured Data

- Data terstruktur (*structured data*) sesuai dengan model atau skema data dan sering disimpan dalam bentuk tabel.
- Data terstruktur (*structured data*) digunakan untuk menangkap hubungan antara entitas yang berbeda dan paling sering disimpan dalam database relasional.
- Data terstruktur (*structured data*) sering dibuat oleh perusahaan aplikasi dan sistem informasi seperti ERP dan sistem CRM.
- Karena alat melimpah dan database yang secara asli mendukung data terstruktur, jarang memerlukan pertimbangan dalam hal pengolahan atau penyimpanan.
- Contoh dari jenis data ini meliputi: transaksi perbankan, faktur, dan catatan pelanggan.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Unstructured Data

- Data yang tidak sesuai dengan model data atau skema data dikenal sebagai data tidak terstruktur (*unstructured data*).
- Diperkirakan bahwa data tidak terstruktur membentuk 80% dari data yang diberikan perusahaan.
- Data tidak terstruktur memiliki tingkat pertumbuhan yang lebih cepat daripada data terstruktur.
- Bentuk data ini adalah tekstual atau biner (*binary files*) dan sering disampaikan melalui file yang mandiri dan non-relasional.
- Sebuah file teks mungkin berisi konten berbagai tweet atau posting blog.
- File biner sering menjadi media file yang berisi data gambar, audio atau video.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Unstructured Data

- Secara teknis, file teks dan biner memiliki: struktur ditentukan oleh format file itu sendiri, tetapi aspek ini diabaikan, dan gagasan tentang menjadi tidak terstruktur dalam kaitannya dengan format data yang terkandung dalam file itu sendiri.
- Logika khusus biasanya diperlukan untuk memproses dan menyimpan data yang tidak terstruktur.
- Misalnya, untuk memutar file video, penting tersedia codec (decoder-coder).
- Data tidak terstruktur tidak dapat langsung diproses atau di-query menggunakan SQL.
- Jika diperlukan untuk disimpan dalam database relasional, data disimpan dalam tabel sebagai Binary Large Object (BLOB).
- Not-only SQL (NoSQL) adalah database non-relasional yang dapat digunakan untuk menyimpan data tidak terstruktur bersama data terstruktur.



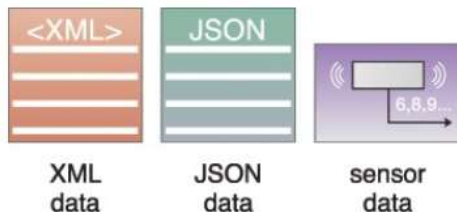
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Semi-structured Data

- Data semi-terstruktur (*semi-structured data*) memiliki tingkat struktur dan konsistensi yang ditentukan, tetapi tidak relasional.
- Sebaliknya, data semi-terstruktur bersifat hierarkis atau berbasis grafik.
- Data semacam ini biasanya disimpan dalam file yang berisi teks.
- XML dan file JSON adalah bentuk umum dari data semi terstruktur.
- Karena sifat tekstual ini data dan kesesuaiannya dengan beberapa tingkat struktur, lebih mudah diproses daripada data tidak terstruktur.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
tinggi

QS  
STARS  
RATING SYSTEM

5  
stars

IAEE

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Semi-structured Data

- Contoh sumber umum dari data semi-terstruktur meliputi: electronic data interchange (EDI) files, spreadsheets, RSS feeds and sensor data.
- Data semi-terstruktur sering memiliki persyaratan pra-pemrosesan dan penyimpanan khusus, terutama jika format dasarnya tidak berbasis teks.
- Contoh pra-pemrosesan data semi-terstruktur adalah validasi dari file XML untuk memastikan bahwa file sesuai dengan definisi skemanya.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
tinggi

QS  
STARS  
RATING SYSTEM



CPA  
AUSTRALIA



ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

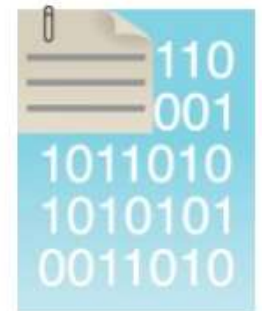
# Metadata

- Metadata memberikan informasi tentang karakteristik dan struktur kumpulan data.
- Jenis data ini sebagian besar dihasilkan oleh mesin dan dapat ditambahkan ke data.
- Pelacakan metadata sangat penting untuk pemrosesan, penyimpanan, dan analisis Big Data karena menyediakan informasi tentang silsilah data dan asalnya selama pemrosesan.
- Contoh metadata meliputi:
  - ✓ tag XML memberikan penulis dan tanggal pembuatan dokumen.
  - ✓ atribut yang menyediakan ukuran file dan resolusi foto digital.



**UNTAR**  
Universitas Tarumanagara

**UNTAR untuk INDONESIA**



RED  
KANTAS

# Latar Belakang Studi Kasus

- Ensure to Insure (ETI) adalah perusahaan asuransi terkemuka yang menyediakan berbagai paket asuransi di sektor kesehatan, bangunan, kelautan dan penerbangan kepada 25 juta pelanggannya yang tersebar di seluruh dunia.
- Lingkungan TI ETI terdiri dari kombinasi platform client-server dan mainframe yang mendukung pelaksanaan sejumlah sistem, termasuk kutipan kebijakan, administrasi kebijakan, manajemen klaim, penilaian risiko, manajemen dokumen, penagihan, enterprise resource planning (ERP) dan customer relationship management (CRM).



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**



# Hambatan Bisnis

- Selama beberapa dekade terakhir, profitabilitas perusahaan telah menurun.
- Temuan komite mengungkapkan bahwa alasan utama di balik memburuknya posisi keuangan perusahaan adalah meningkatnya jumlah klaim penipuan dan pembayaran terkait yang dilakukan terhadap mereka.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
Linggi

QS STARS  
RATING SYSTEM  
5★



CPA  
AUSTRALIA



ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Contoh Studi Kasus

- ETI memilih Big Data untuk implementasi tujuan strategisnya, tetapi ETI tidak memiliki keterampilan Big Data internal dan perlu memilih antara menyewa konsultan Big Data atau mengirim tim TI-nya ke pelatihan Big Data.
- Perusahaan memutuskan tim TI senior yang dikirim ke pelatihan untuk mengantisipasi solusi jangka panjang yang hemat biaya di mana anggota tim yang terlatih akan menjadi sumber daya Big Data internal permanen yang dapat dikonsultasikan kapan saja dan juga dapat melatih tim junior anggota untuk lebih meningkatkan keahlian Big Data internal.
- Setelah menerima pelatihan Big Data, anggota tim yang terlatih menekankan perlunya istilah kosakata umum sehingga seluruh tim berada di halaman yang sama ketika berbicara tentang Big Data.
- Mereka mengadopsi *example-driven approach*.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Contoh Studi Kasus

- Saat membahas kumpulan data, beberapa kumpulan data terkait yang ditunjukkan oleh anggota tim mencakup klaim, kebijakan, kutipan, data profil pelanggan, dan data sensus.
- Meskipun konsep *data analysis* dan *data analytics* cepat dipahami, beberapa anggota tim yang tidak memiliki banyak eksposur bisnis mengalami kesulitan memahami BI dan penetapan KPI yang sesuai.
- Salah satu anggota tim TI yang terlatih menjelaskan BI dengan menggunakan proses pembuatan laporan bulanan untuk mengevaluasi kinerja bulan sebelumnya sebagai contoh.
- Proses ini melibatkan mengimpor data dari sistem operasional ke EDW (*Enterprise Data Warehouse*) dan menghasilkan KPI seperti: polis yang dijual dan klaim yang diajukan, diproses, diterima, dan ditolak yang ditampilkan di dasbor dan kartu skor yang berbeda.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Contoh Studi Kasus

- Dalam hal analitik, ETI menggunakan analitik deskriptif dan diagnostik.
- Analisis deskriptif meliputi query sistem administrasi polis untuk menentukan jumlah polis yang terjual setiap hari, query sistem manajemen klaim untuk mengetahui berapa banyak klaim yang diajukan setiap hari dan query sistem penagihan untuk mengetahui berapa banyak pelanggan yang terlambat membayar premi.
- Diagnostic analytics dilakukan sebagai bagian dari berbagai aktivitas BI, seperti melakukan query untuk menjawab pertanyaan seperti mengapa target penjualan bulan lalu tidak tercapai.
- Ini termasuk melakukan operasi penelusuran untuk mengelompokkan penjualan menurut jenis dan lokasi sehingga dapat ditentukan lokasi mana yang berkinerja buruk untuk jenis kebijakan tertentu.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
Simpul

QS  
STARS  
RATING SYSTEM

ISAS  
ISAS

IAEE

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Contoh Studi Kasus: Mengidentifikasi Karakteristik Data

- Volume

- Sejumlah besar data transaksi dihasilkan sebagai hasil dari pemrosesan klaim, penjualan polis baru, dan perubahan polis yang ada.
- Volume data tidak terstruktur yang besar, baik di dalam maupun di luar perusahaan, mungkin terbukti membantu dalam mencapai tujuan ETI.
- Data ini meliputi catatan kesehatan, dokumen yang diserahkan oleh pelanggan pada saat mengajukan aplikasi asuransi, jadwal properti, data media sosial dan data cuaca.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Contoh Studi Kasus: Mengidentifikasi Karakteristik Data

- **Velocity**

- Data tergolong low velocity, seperti data pengajuan klaim dan data polis baru yang diterbitkan.
- Data log server web dan kutipan asuransi adalah data berkecepatan tinggi.
- Data media sosial dan data cuaca tiba dengan cepat.
- Manajemen bencana dan deteksi klaim penipuan, data perlu diproses dengan cukup cepat untuk meminimalkan kerugian.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
Linggi

QS  
STARS  
RATING SYSTEM



CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Contoh Studi Kasus: Mengidentifikasi Karakteristik Data

- Variety
  - Kumpulan data yang mencakup catatan kesehatan, data kebijakan, data klaim, data kutipan, data media sosial, catatan agen pusat panggilan, catatan penyesuaian klaim, foto insiden, laporan cuaca, sensus data, log server web, dan email.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
Simpul

QS STARS  
RATING SYSTEM  
5 Star

AMBA  
AACSB  
EFMD EQUIS

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Contoh Studi Kasus: Mengidentifikasi Karakteristik Data

- Veracity
  - Sampel data yang diambil dari sistem operasional dan EDW menunjukkan tanda-tanda kebenaran yang tinggi (*high veracity*).
  - Validasi data dilakukan pada beberapa tahap termasuk validasi pada saat entri data, validasi di berbagai titik saat aplikasi memproses data, seperti validasi input tingkat fungsi, dan validasi yang dilakukan oleh database saat data disimpan.
  - Data yang diperoleh dari sosial media dan data cuaca memerlukan proses validasi dan pembersihan data.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
tinggi

QS  
STARS  
RATING SYSTEM  
5 stars

ISAS  
ISAS

IAEE

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**



# Contoh Studi Kasus: Mengidentifikasi Karakteristik Data

- Value
  - Perlu menentukan nilai maksimum dari kumpulan data yang tersedia dengan memastikan kumpulan data disimpan dalam bentuk aslinya dan sesuai dengan jenis analitik yang tepat.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
linggul

QS  
STARS  
RATING SYSTEM  
5 star



CPA  
AUSTRALIA



ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

## Contoh Studi Kasus: Mengidentifikasi Jenis Data

- Data terstruktur (*structured data*): data polis, data klaim, data profil pelanggan dan data penawaran.
- Data tidak terstruktur (*unstructured data*): data media sosial, dokumen aplikasi asuransi, panggilan catatan agen pusat, catatan penyesuai klaim, dan foto-foto insiden.
- Data semi terstruktur (*semi-structured data*): catatan kesehatan, data profil pelanggan, laporan cuaca, data sensus, log server web, dan email.



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
Simpul

QS  
STARS  
RATING SYSTEM  
5 Star

ISAS  
ISAS

IAEE

CPA  
AUSTRALIA

ICAEW  
CHARTERED  
ACCOUNTANTS

**UNTAR untuk INDONESIA**

# Tugas Analisis Big Data Satu Semester

- Masing – masing mahasiswa memilih satu kota untuk analisis harga pangan di Indonesia
- Mahasiswa diperbolehkan memilih kota yang sama tetapi berbeda variabel.
- Mahasiswa diperbolehkan memilih variabel yang sama tetapi berbeda kota.
- Mahasiswa **tidak diperbolehkan memilih kota dan variabel yang sama.**
- Dataset <https://www.bi.go.id/hargapangan/>
- Donwload data harian dari Januari 2020 – Maret 2023.
- Dataset dikumpulkan pada pertemuan ke 7.
- Analisis data akan dilakukan setelah UTS.



**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Pilih variabel (komoditas)

- Pilih 5 variabel (bebas memilih), misalnya: beras, minyak goreng, telur, cabai rawit, bawang merah, bawang putih dll.
- Susun ulang dataset menjadi seperti tampilan berikut:



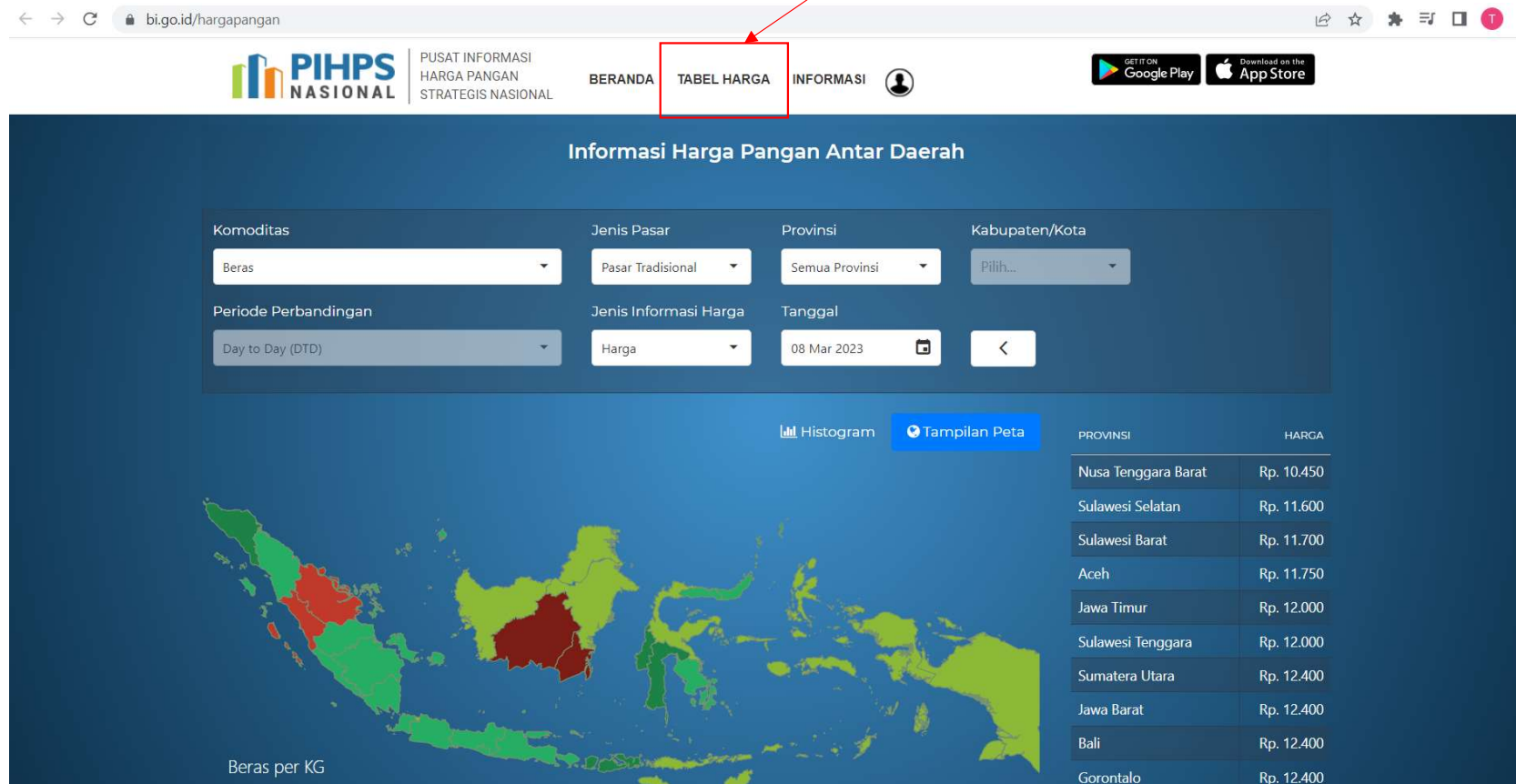
**UNTAR**  
Universitas Tarumanagara



**UNTAR untuk INDONESIA**

# Dataset Harga Pangan

Pilih Tabel Harga



Universitas Tarumanagara

UNTAR untuk INDONESIA



# Dataset Harga Pangan

Pilih komoditas dan daerah

← → ↺ bi.go.id/hargapangan/TabellHarga/PasarTradisionalDaerah

**PIHPS NASIONAL** PUSAT INFORMASI HARGA PANGAN STRATEGIS NASIONAL BERANDA TABEL HARGA INFORMASI

GET IT ON Google Play Download on the App Store

## Tabel Harga Berdasarkan Daerah

**Komoditas**

- ☐ Telur Ayam Ras Segar
- ☒ Bawang Merah
- ☐ Bawang Merah Ukura...
- ☒ Bawang Putih
- ☐ Bawang Putih Ukuran...
- ☒ Cabai Merah

**Provinsi**

- ☐ Bengkulu
- ☐ DI Yogyakarta
- ☒ DKI Jakarta
- ☐ Dum
- ☐ Dummy
- ☐ Gorontalo

**Kabupaten/Kota**

- ☐ Kota Jakarta Pusat

**Perkembangan Harga Pangan**

Periode : 28 Feb 2023 - 7 Mar 2023  
Provinsi : Semua Provinsi  
Kabupaten/Kota : Semua Kabupaten/Kota  
Pasar : Semua Pasar  
Tipe Laporan : Laporan Harian

NO	KOMODITAS (RP)	28/ 02/ 2023	01/ 03/ 2023	02/ 03/ 2023	03/ 03/ 2023	06/ 03/ 2023	07/ 03/ 2023
Q	Q	Q	Q	Q	Q	Q	Q
I	Beras	12,650	12,650	12,600	12,550	12,550	12,550
1	Beras Kualitas Bawah I	11,400	11,400	11,400	11,350	11,350	11,350
2	Beras Kualitas Bawah II	11,150	11,100	11,100	11,000	11,000	11,000
3	Beras Kualitas Medium I	12,850	12,850	12,850	12,700	12,700	12,700
4	Beras Kualitas Medium II	12,650	12,650	12,650	12,500	12,500	12,500
5	Beras Kualitas Super I	14,050	14,000	14,000	14,000	14,000	14,000
6	Beras Kualitas Super II	13,400	13,350	13,350	13,350	13,400	13,350

# Dataset Harga Pangan

Atur tanggal dan download data

bi.go.id/hargapangan/TabelHarga/PasarTradisionalDaerah

☐ Maluku Utara  
☐ Nusa Tenggara Barat

Kabupaten/Kota  
☐ Kota Jakarta Pusat

Tipe Laporan  
Laporan Harian

Tanggal Mulai  
1/2/2023

Tanggal Selesai  
28/2/2023

[Lihat Laporan](#)

[Download](#)

2	Beras Kualitas Bawari II	11,150	11,100	11,100	11,000	11,000	11,050
3	Beras Kualitas Medium I	12,850	12,850	12,850	12,700	12,700	12,700
4	Beras Kualitas Medium II	12,650	12,650	12,650	12,500	12,500	12,500
5	Beras Kualitas Super I	14,050	14,000	14,000	14,000	14,000	14,000
6	Beras Kualitas Super II	13,400	13,350	13,350	13,350	13,400	13,350
II	<b>Daging Ayam</b>	<b>32,250</b>	<b>32,200</b>	<b>32,200</b>	<b>32,100</b>	<b>32,050</b>	<b>32,000</b>
1	Daging Ayam Ras Segar	32,250	32,200	32,200	32,100	32,050	32,000
III	<b>Daging Sapi</b>	<b>129,100</b>	<b>129,100</b>	<b>129,100</b>	<b>129,050</b>	<b>129,100</b>	<b>129,100</b>
1	Daging Sapi Kualitas 1	133,100	133,100	133,100	133,050	133,100	133,050
2	Daging Sapi Kualitas 2	124,800	124,750	124,800	124,700	124,800	124,800
IV	<b>Telur Ayam</b>	<b>27,350</b>	<b>27,300</b>	<b>27,250</b>	<b>27,150</b>	<b>27,250</b>	<b>27,400</b>
1	Telur Ayam Ras Segar	27,350	27,300	27,250	27,150	27,250	27,400
V	<b>Bawang Merah</b>	<b>36,400</b>	<b>36,200</b>	<b>36,100</b>	<b>36,200</b>	<b>36,450</b>	<b>36,450</b>
1	Bawang Merah Ukuran Sedang	36,400	36,200	36,100	36,200	36,450	36,450
VI	<b>Bawang Putih</b>	<b>30,350</b>	<b>30,400</b>	<b>30,450</b>	<b>30,550</b>	<b>31,150</b>	<b>31,450</b>
1	Bawang Putih Ukuran Sedang	30,350	30,400	30,450	30,550	31,150	31,450
VII	<b>Cabai Merah</b>	<b>42,400</b>	<b>42,300</b>	<b>42,050</b>	<b>41,600</b>	<b>42,200</b>	<b>42,450</b>
1	Cabai Merah Besar	40,600	40,400	40,100	39,700	40,250	40,650
2	Cabai Merah Kecil	42,000	42,000	42,350	42,200	42,050	41,950



# Contoh Data

Tanggal	Beras	Bawang Merah
02/ 01/ 2019	12,650	15000
03/ 01/ 2019	12,700	15200
04/ 01/ 2019	12,700	15200
07/ 01/ 2019	12,950	15200
08/ 01/ 2019	12,950	15200
09/ 01/ 2019	12,950	15200
10/ 01/ 2019	13,000	15200
11/ 01/ 2019	13,000	15000
14/ 01/ 2019	13,000	15000
15/ 01/ 2019	13,000	15000
16/ 01/ 2019	13,000	15000
17/ 01/ 2019	12,950	16000
18/ 01/ 2019	12,950	16000
21/ 01/ 2019	12,950	16000
22/ 01/ 2019	12,950	16000
23/ 01/ 2019	12,950	16000
24/ 01/ 2019	12,950	15900
25/ 01/ 2019	12,950	15900
28/ 01/ 2019	12,950	15900
29/ 01/ 2019	12,950	15900
30/ 01/ 2019	12,950	15900
31/ 01/ 2019	12,950	15900
01/ 02/ 2019	12,950	15900
04/ 02/ 2019	12,950	15900
06/ 02/ 2019	12,950	15900
07/ 02/ 2019	12,950	15900

**AR**

managara



**UNTAR untuk INDONESIA**



# Koordinasi dengan Asisten Dosen

- Setiap kelompok wajib melaporkan 'kota yang dipilih' kepada asisten dosen.
- Asisten dosen akan membantu memeriksa jika adanya kesamaan



**UNTAR**  
Universitas Tarumanagara

Terakreditasi  
BAN-PT

A  
linggul

QS  
STARS  
RATING SYSTEM  
5 star



**UNTAR untuk INDONESIA**