

Adopsi Big Data dan Pertimbangan Perencanaan

BIG DATA – TK13025

Prasyarat Organisasi

- Data yang kedaluwarsa, tidak valid, atau teridentifikasi dengan buruk akan menghasilkan input berkualitas rendah, terlepas dari seberapa bagus solusi Big Data, akan terus menghasilkan hasil berkualitas rendah.
- Kualitas data yang ditargetkan untuk diproses oleh solusi Big Data perlu dinilai.
- *Roadmap* perlu ditentukan untuk memastikan bahwa setiap perluasan atau penambahan lingkungan yang diperlukan direncanakan untuk tetap sinkron dengan persyaratan perusahaan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Pengadaan Data

- Akuisisi solusi Big Data dapat menjadi ekonomis, karena ketersediaan platform dan peralatan open-source serta peluang untuk memanfaatkan perangkat keras komoditas.
- Namun, anggaran yang besar mungkin masih diperlukan untuk mendapatkan data eksternal.
- Sifat bisnis dapat membuat data eksternal sangat berharga.
- Semakin besar volume dan variasi data yang dapat diberikan, semakin tinggi peluang untuk menemukan wawasan tersembunyi dari pola.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Pengadaan Data

- Sumber data eksternal termasuk sumber data pemerintah dan pasar data komersial.
- Data yang disediakan pemerintah, seperti data geospasial, mungkin gratis.
- Namun, sebagian besar data yang relevan secara komersial perlu dibeli dan mungkin melibatkan kelanjutan biaya berlangganan untuk memastikan pengiriman pembaruan ke kumpulan data yang diperoleh.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Privacy

- Melakukan analitik pada kumpulan data dapat mengungkapkan informasi rahasia tentang organisasi atau individu.
- Menganalisis kumpulan data terpisah yang berisi data yang tampaknya tidak berbahaya dapat mengungkapkan informasi pribadi ketika kumpulan data dianalisis bersama.
- Hal ini dapat menyebabkan pelanggaran privasi yang disengaja atau tidak disengaja.
- Mengatasi masalah privasi ini memerlukan pemahaman tentang sifat data yang dikumpulkan dan peraturan privasi data yang relevan, serta teknik khusus untuk penandaan data dan anonimisasi.
- Misalnya, data telemetri, seperti log GPS mobil atau pembacaan data meteran pintar, yang dikumpulkan selama periode waktu yang lama dapat mengungkapkan lokasi dan perilaku seseorang.



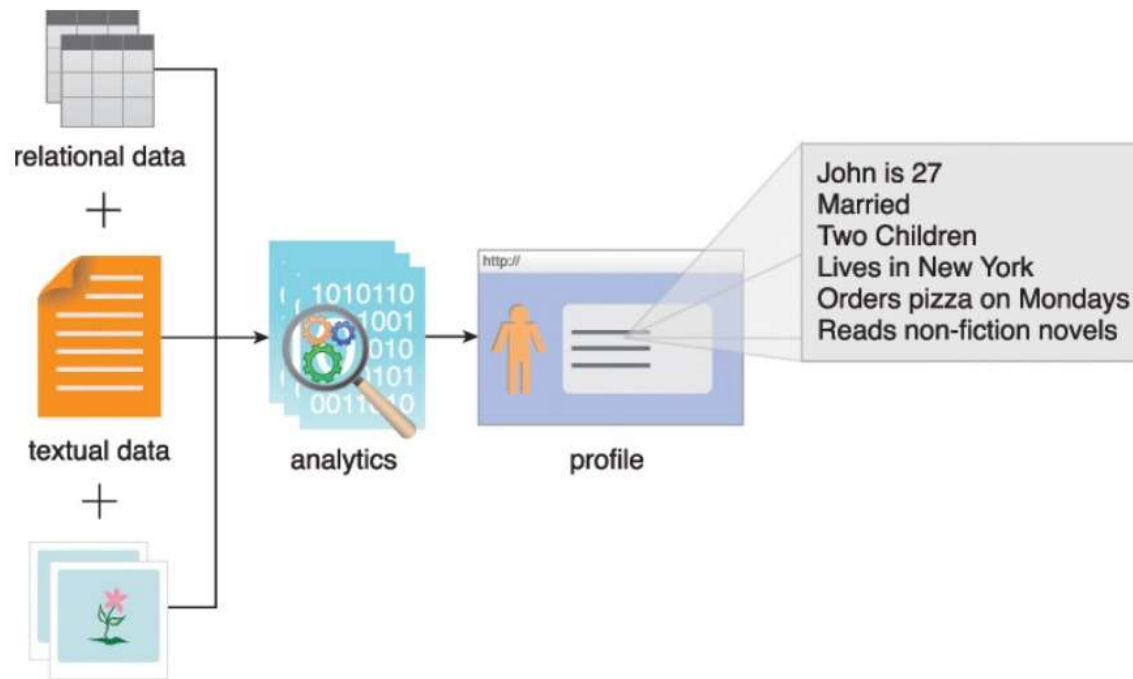
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Privacy

Informasi yang dikumpulkan dari menjalankan analitik pada file gambar, data relasional, dan data tekstual digunakan untuk membuat profil John.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
lingkup

QS
STARS
RATING SYSTEM
2016

AMBA
ACCREDITED

EFMD
EQUIS
ACCREDITED

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Keamanan

- Beberapa komponen solusi Big Data tidak memiliki solusi ketahanan pada lingkungan perusahaan tradisional dalam hal kontrol akses dan keamanan data.
- Mengamankan Big Data memastikan bahwa jaringan data dan repositori cukup diamankan melalui mekanisme otentikasi dan otorisasi.
- Keamanan Big Data selanjutnya melibatkan penetapan tingkat akses data untuk berbagai kategori pengguna.
- Misalnya, tidak seperti sistem manajemen basis data relasional tradisional, basis data NoSQL umumnya tidak menyediakan mekanisme keamanan bawaan yang kuat.
- Mereka mengandalkan API berbasis HTTP sederhana di mana data dipertukarkan dalam plaintext, membuat data rentan terhadap serangan berbasis jaringan



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Keamanan

Basis data NoSQL rentan terhadap serangan berbasis jaringan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Sumber (*Provenance*)

- *Provenance* mengacu pada informasi tentang sumber data dan bagaimana data itu diproses.
- Sumber informasi membantu menentukan keaslian dan kualitas data, dan dapat digunakan untuk tujuan audit.
- Mempertahankan sumbernya karena sejumlah besar data diperoleh, digabungkan, dan dimasukkan melalui beberapa tahap pemrosesan dapat menjadi tugas yang kompleks.
- Pada tahap yang berbeda dalam siklus hidup analitik, data akan berada dalam status yang berbeda karena faktanya mungkin sedang dikirim, diproses, atau dalam penyimpanan.
- Status-status ini sesuai dengan gagasan tentang data-dalam-gerak, data-dalam-penggunaan, dan data-diam.
- Yang penting, setiap kali Big Data berubah status, harus memicu penangkapan sumber informasi yang dicatat sebagai metadata.



Sumber (*Provenance*)

- Saat data memasuki lingkungan analitik, catatan asalnya (sumbernya) dapat diinisialisasi dengan perekaman informasi yang menangkap silsilah data.
- Pada akhirnya, tujuan menangkap sumbernya adalah untuk dapat mempertimbangkan hasil analitik yang dihasilkan dengan pengetahuan tentang asal data dan langkah atau algoritma apa yang digunakan untuk memproses data yang mengarah ke hasil.
- Sumber informasi sangat penting untuk dapat mewujudkan nilai hasil analitik.
- Sama seperti penelitian ilmiah, jika hasil tidak dapat dibenarkan dan diulang, maka tidak memiliki kredibilitas.
- Ketika sumber informasi diperoleh, hasilnya dapat lebih mudah dipercaya, dengan demikian dapat digunakan dengan percaya diri.

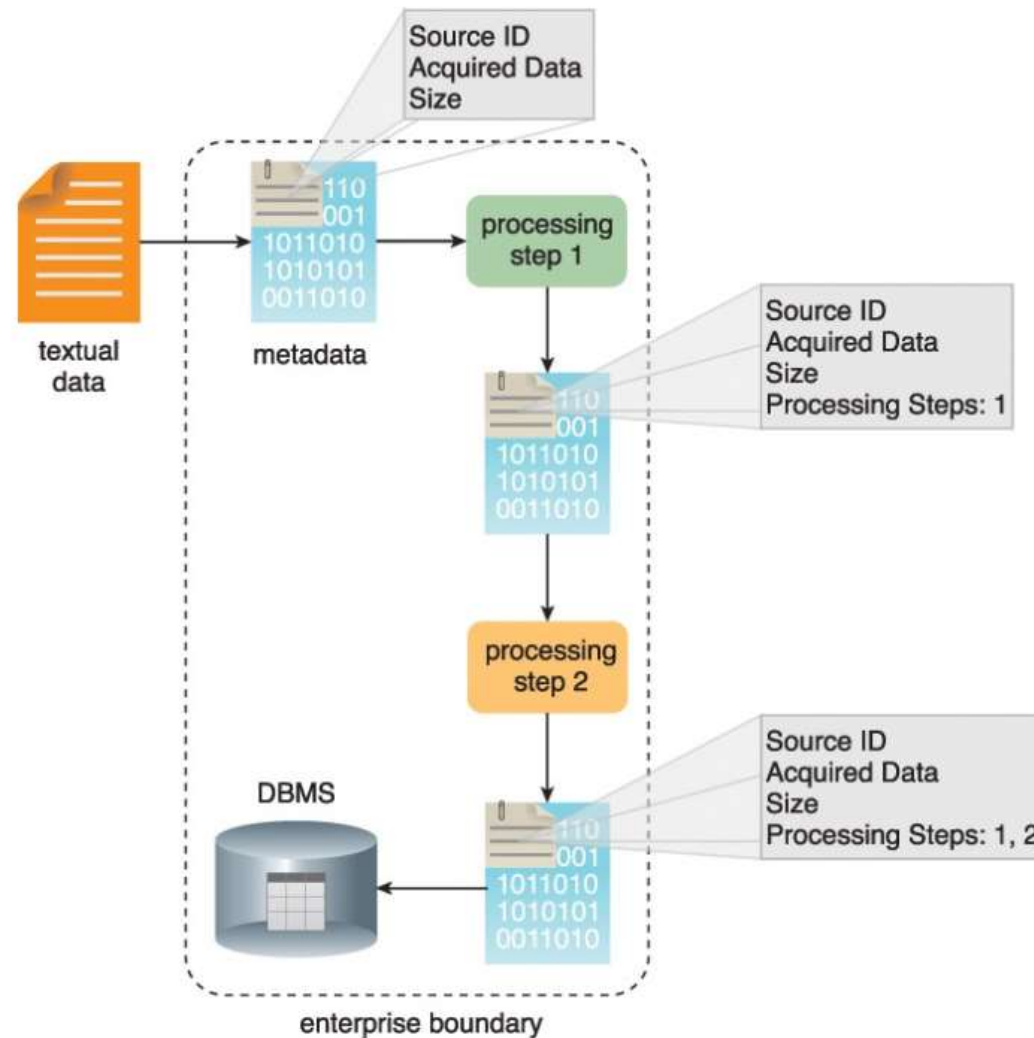


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Data perlu dianotasi dengan atribut kumpulan sumber data dan detail langkah pemrosesan saat melewati transformasi data.



Dukungan Realtime Terbatas

- Dasbor dan aplikasi lain yang memerlukan streaming data dan peringatan sering kali menuntut transmisi data *real-time* atau *near-realtime*.
- Banyak solusi dan alat Big Data *open source* berorientasi *batch*; namun, ada generasi baru alat *open source* yang mampu real-time yang memiliki dukungan untuk analisis data streaming.
- Banyak dari solusi analisis data *real-time* yang ada adalah paten.
- Pendekatan yang mencapai hasil *near-real-time* sering kali memproses data transaksional saat tiba dan menggabungkannya dengan data batch-processed yang diringkas sebelumnya.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Tantangan Performa yang Berbeda

- Karena volume data yang harus diproses oleh beberapa solusi Big Data, performa sering kali menjadi perhatian.
- Misalnya, kumpulan data besar yang digabungkan dengan algoritma penelusuran yang kompleks dapat menyebabkan waktu kueri yang lama.
- Tantangan kinerja lainnya terkait dengan bandwidth jaringan.
- Dengan meningkatnya volume data, waktu untuk mentransfer satu unit data dapat melebihi waktu pemrosesan data yang sebenarnya.

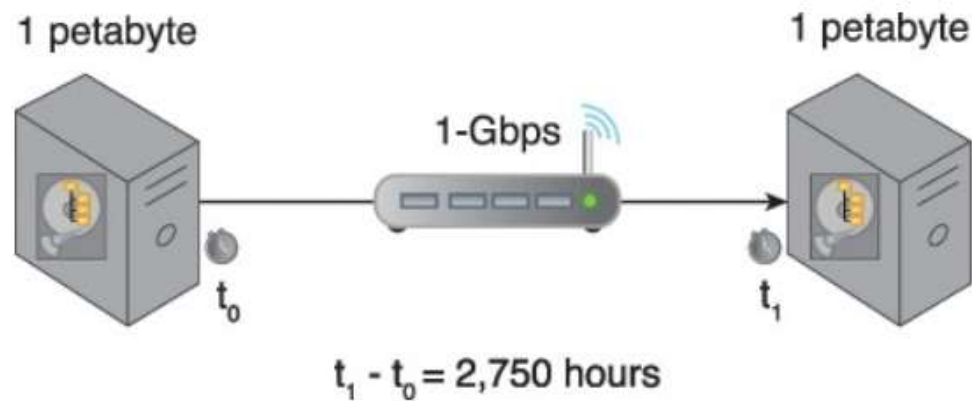


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Tantangan Performa yang Berbeda



Mentransfer 1 PB data melalui koneksi LAN 1-Gigabit memakan waktu sekitar 2.750 jam.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
tinggi

QS
STARS
RATING SYSTEM
2016

ISAS
5 STAR

IAEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Persyaratan Tata Kelola yang Berbeda

- Solusi Big Data mengakses data dan menghasilkan data, yang semuanya menjadi aset bisnis.
- Kerangka tata kelola diperlukan untuk memastikan bahwa data dan lingkungan solusi sendiri diatur, distandarisasi, dan dikembangkan secara terkendali.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Persyaratan Tata Kelola yang Berbeda

- Contoh kerangka kerja tata kelola Big Data meliputi:
 - standarisasi bagaimana data ditandai dan metadata yang digunakan untuk penandaan
 - kebijakan yang mengatur jenis data eksternal yang mungkin diperoleh
 - kebijakan mengenai pengelolaan privasi data dan anonimisasi data
 - kebijakan untuk pengarsipan sumber data dan hasil analisis
 - kebijakan yang menetapkan pedoman untuk pembersihan dan penyaringan data



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Metodologi yang Berbeda

- Metodologi akan diperlukan untuk mengontrol bagaimana data mengalir masuk dan keluar dari solusi Big Data.
- Perlu mempertimbangkan bagaimana loop umpan balik dapat dibuat untuk memungkinkan data yang diproses mengalami penyempurnaan berulang.
- Misalnya, pendekatan berulang dapat digunakan untuk memungkinkan personel bisnis memberikan umpan balik kepada personel TI secara berkala.
- Setiap siklus umpan balik memberikan peluang untuk penyempurnaan sistem dengan memodifikasi persiapan data atau langkah-langkah analisis data.



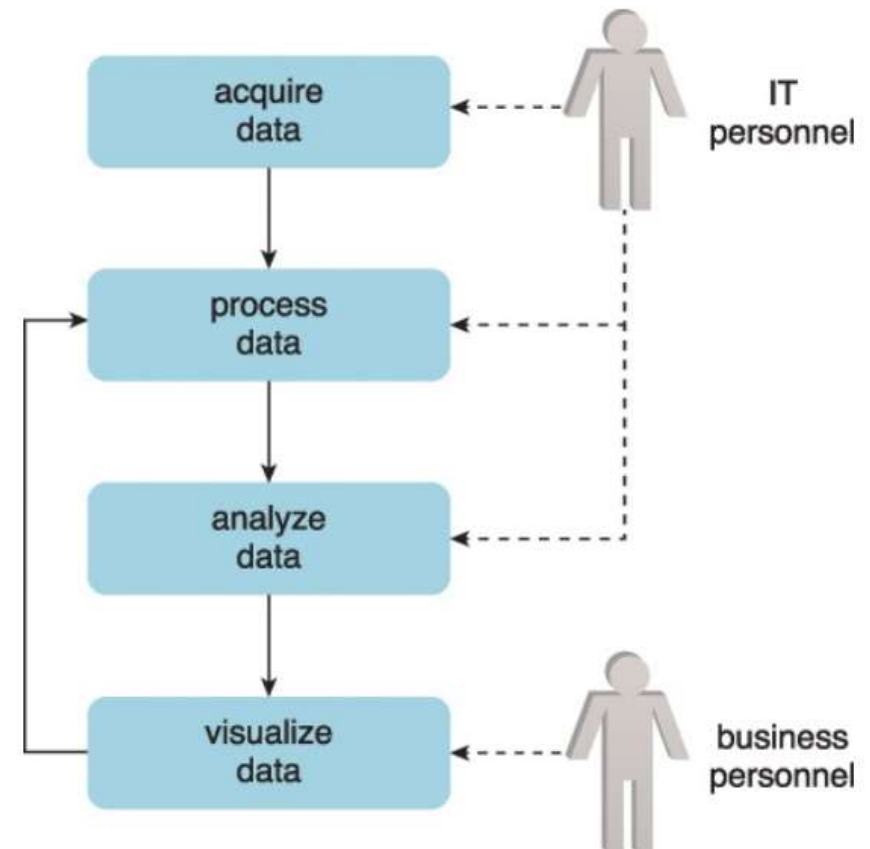
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Metodologi yang Berbeda

Setiap pengulangan dapat membantu menyempurnakan langkah pemrosesan, algoritma, dan model data untuk meningkatkan akurasi hasil dan memberikan nilai yang lebih besar.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Clouds

- Cloud menyediakan lingkungan jarak jauh yang dapat menampung infrastruktur TI untuk penyimpanan dan pemrosesan skala besar.
- Terlepas dari apakah suatu organisasi sudah mendukung cloud, adopsi lingkungan Big Data mungkin mengharuskan beberapa atau semua lingkungan ditaruh di dalam cloud.
- Misalnya, perusahaan yang menjalankan sistem CRM-nya di cloud memutuskan untuk menambahkan solusi Big Data di lingkungan cloud yang sama untuk menjalankan analitik pada data CRM-nya.
- Data ini kemudian dapat dibagikan dengan lingkungan Big Data utamanya yang berada dalam batas-batas perusahaan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Clouds

- Pembenaran umum untuk menggabungkan lingkungan cloud untuk mendukung solusi Big Data meliputi:
 - sumber daya perangkat keras internal yang tidak memadai
 - investasi modal dimuka untuk pengadaan sistem tidak tersedia
 - proyek akan diisolasi dari bisnis lainnya sehingga proses bisnis yang ada tidak terpengaruh
 - inisiatif Big Data adalah bukti konsep
 - kumpulan data yang perlu diproses sudah menjadi penduduk cloud
 - batas sumber daya komputasi dan penyimpanan yang tersedia yang digunakan oleh solusi Big Data internal telah tercapai



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Siklus Hidup Analisis Big Data

- Analisis Big Data berbeda dari analisis data tradisional terutama karena karakteristik volume, velocity (kecepatan), dan variety (variasi) dari data yang sedang diproses.
- Untuk mengatasi persyaratan berbeda untuk melakukan analisis pada Big Data, metodologi langkah demi langkah diperlukan untuk mengatur aktivitas dan tugas yang terkait dengan perolehan, pemrosesan, analisis, dan penggunaan kembali data.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
lingkup

QS
STARS
RATING SYSTEM
5 star



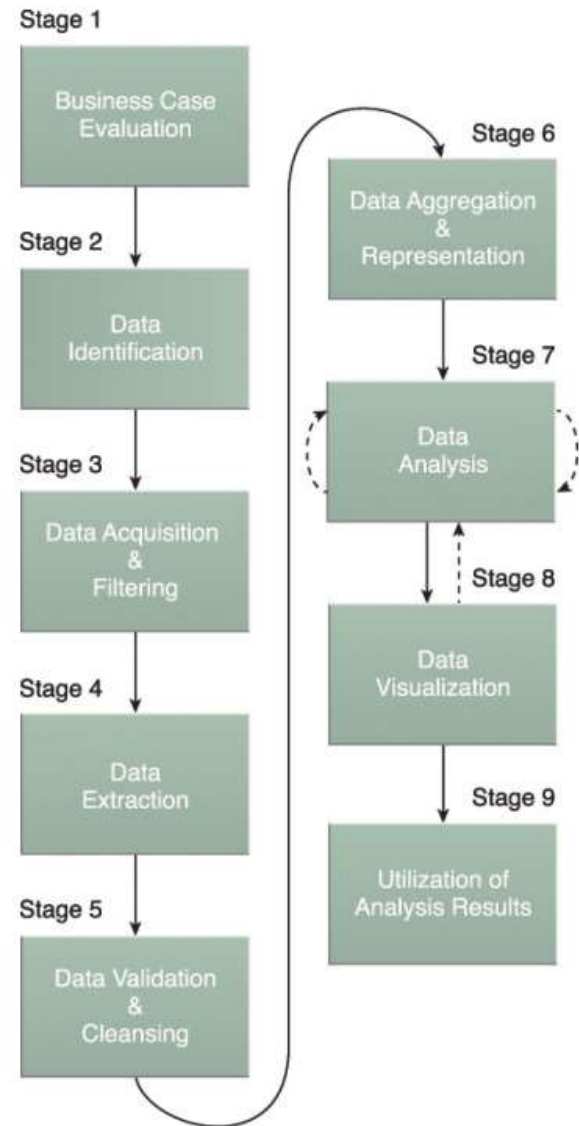
CPA
AUSTRALIA



ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Sembilan tahap siklus hidup analitik Big Data.



Evaluasi Kasus Bisnis (Business Case Evaluation)

- Setiap siklus hidup analitik Big Data harus dimulai dengan kasus bisnis yang terdefinisi dengan baik yang menyajikan pemahaman yang jelas tentang kebenaran, motivasi, dan tujuan pelaksanaan analisis.
- Tahap Evaluasi Kasus Bisnis, mengharuskan kasus bisnis dibuat, dinilai dan disetujui sebelum melanjutkan dengan tugas analisis langsung yang sebenarnya.
- Evaluasi kasus bisnis analitik Big Data membantu pengambil keputusan memahami sumber daya bisnis yang perlu digunakan dan tantangan bisnis mana yang akan ditangani oleh analisis.
- Identifikasi KPI lebih lanjut selama tahap ini dapat membantu menentukan kriteria penilaian dan panduan untuk evaluasi hasil analitik.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Evaluasi Kasus Bisnis (Business Case Evaluation)

- Jika KPI tidak tersedia, upaya harus dilakukan untuk membuat tujuan proyek analisis SMART yang spesifik, terukur, dapat dicapai, relevan, dan tepat waktu.
- Untuk memenuhi syarat sebagai masalah Big Data, masalah bisnis harus terkait langsung dengan satu atau lebih karakteristik Big Data volume, kecepatan, atau variasi.
- Perhatikan juga bahwa hasil lain dari tahap ini adalah penentuan anggaran dasar yang diperlukan untuk melaksanakan proyek analisis.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Evaluasi Kasus Bisnis (Business Case Evaluation)

- Setiap pembelian yang diperlukan, seperti peralatan, perangkat keras dan pelatihan, harus dipahami terlebih dahulu sehingga investasi yang diantisipasi dapat ditimbang terhadap manfaat yang diharapkan dari pencapaian tujuan.
- Iterasi awal dari siklus hidup analitik Big Data akan membutuhkan lebih banyak investasi di muka untuk teknologi, produk, dan pelatihan Big Data dibandingkan dengan iterasi selanjutnya di mana investasi sebelumnya ini dapat dimanfaatkan berulang kali.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Identifikasi Data (Data Identification)

- Tahap Identifikasi Data didedikasikan untuk mengidentifikasi kumpulan data yang diperlukan untuk proyek analisis dan sumbernya.
- Mengidentifikasi berbagai sumber data yang lebih luas dapat meningkatkan kemungkinan menemukan pola dan korelasi yang tersembunyi.
- Misalnya, untuk memberikan wawasan, akan bermanfaat untuk mengidentifikasi sebanyak mungkin jenis sumber data terkait, terutama jika tidak jelas apa yang harus dicari.
- Bergantung pada ruang lingkup bisnis proyek analisis dan sifat masalah bisnis yang ditangani, kumpulan data yang diperlukan dan sumbernya dapat berasal dari internal dan/atau eksternal perusahaan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Identifikasi Data (Data Identification)

- Dalam kasus kumpulan data internal, daftar kumpulan data yang tersedia dari sumber internal, seperti data mart dan sistem operasional, biasanya dikompilasi dan dicocokkan dengan spesifikasi kumpulan data yang telah ditentukan sebelumnya.
- Dalam hal kumpulan data eksternal, daftar kemungkinan penyedia data pihak ketiga, seperti pasar data dan kumpulan data yang tersedia untuk umum, dikompilasi.
- Beberapa bentuk data eksternal mungkin disematkan di dalam blog atau jenis situs web berbasis konten lainnya, dalam hal ini data tersebut mungkin perlu dikumpulkan melalui alat otomatis.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Akuisisi dan Penyaringan Data (Data Acquisition and Filtering)

- Selama tahap Akuisisi dan Penyaringan Data, data dikumpulkan dari semua sumber data yang diidentifikasi pada tahap sebelumnya.
- Data yang diperoleh kemudian difilter secara otomatis untuk menghilangkan data yang rusak atau data yang dianggap tidak memiliki nilai untuk tujuan analisis.
- Bergantung pada jenis sumber data, data mungkin datang sebagai kumpulan file, seperti data yang dibeli dari penyedia data pihak ketiga, atau mungkin memerlukan integrasi API, seperti dengan Twitter.
- Dalam banyak kasus, terutama jika menyangkut data eksternal yang tidak terstruktur, beberapa atau sebagian besar data yang diperoleh mungkin tidak relevan (noise) dan dapat dibuang sebagai bagian dari proses penyaringan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Akuisisi dan Penyaringan Data (Data Acquisition and Filtering)

- Data yang diklasifikasikan sebagai "rusak/corrupt" dapat mencakup catatan dengan nilai yang hilang atau tidak masuk akal atau tipe data yang tidak valid.
- Data yang disaring untuk satu analisis mungkin berharga untuk jenis analisis yang berbeda.
- Disarankan untuk menyimpan salinan rinci dari dataset asli sebelum melanjutkan dengan penyaringan.
- Untuk meminimalkan ruang penyimpanan yang diperlukan, salinan dapat dikompresi.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Akuisisi dan Penyaringan Data (Data Acquisition and Filtering)

- Baik data internal maupun eksternal perlu dipertahankan setelah dihasilkan atau memasuki batas perusahaan.
- Untuk analitik batch, data ini disimpan ke disk sebelum dianalisis.
- Dalam kasus analitik real-time, data dianalisis terlebih dahulu dan kemudian disimpan ke disk.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Lingguh

QS
STARS
RATING SYSTEM
5 Star



CPA
AUSTRALIA

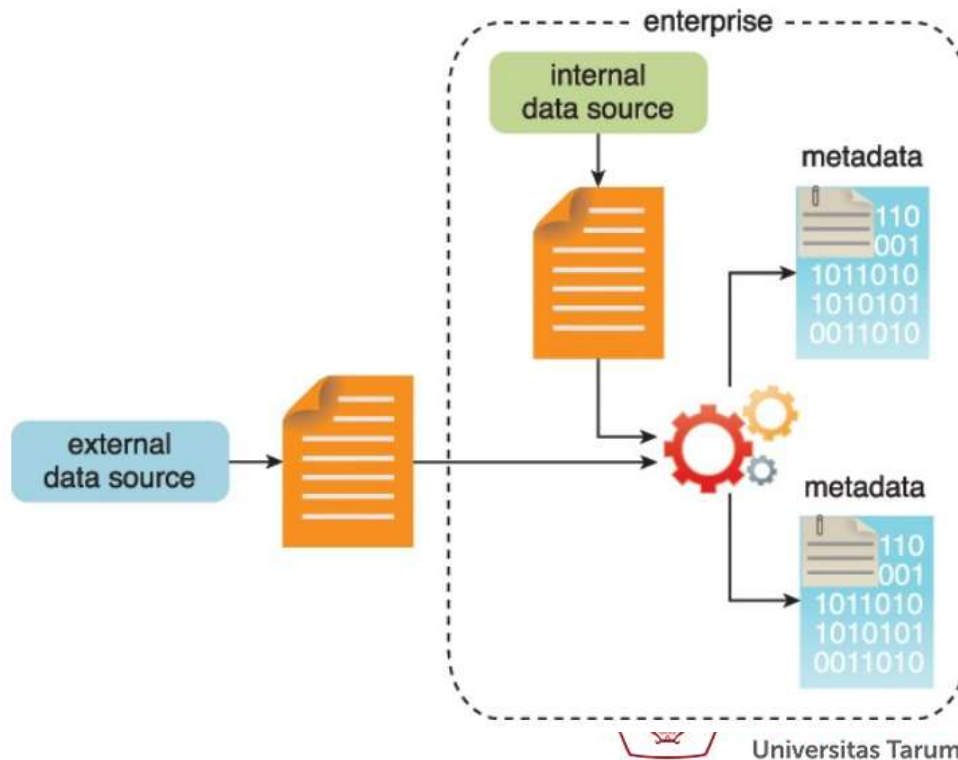


ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Akuisisi dan Penyaringan Data (Data Acquisition and Filtering)

Metadata ditambahkan ke data dari sumber internal dan eksternal.



Universitas Tarumanagara



UNTAR untuk INDONESIA

Akuisisi dan Penyaringan Data (Data Acquisition and Filtering)

- Metadata dapat ditambahkan melalui otomatisasi ke data dari sumber data internal dan eksternal untuk meningkatkan klasifikasi dan kueri.
- Contoh metadata yang ditambahkan mencakup ukuran dan struktur kumpulan data, sumber informasi, tanggal dan waktu pembuatan atau pengumpulan, dan informasi khusus bahasa.
- Sangat penting bahwa metadata dapat dibaca mesin dan diteruskan di sepanjang tahap analisis selanjutnya.
- Ini membantu menjaga asal data di seluruh siklus hidup analitik Big Data, yang membantu membangun dan menjaga akurasi dan kualitas data.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Ekstraksi Data (*Data Extraction*)

- Beberapa data yang diidentifikasi sebagai masukan untuk analisis mungkin dalam format yang tidak sesuai dengan solusi Big Data.
- Kebutuhan untuk mengatasi jenis data yang berbeda lebih mungkin dengan data dari sumber eksternal.
- Tahap siklus hidup Ekstraksi Data didedikasikan untuk mengekstraksi data yang berbeda dan mengubahnya menjadi format yang dapat digunakan oleh solusi Big Data yang mendasarinya untuk tujuan analisis data.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
linggul

QS
STARS
RATING SYSTEM
5 star



CPA
AUSTRALIA



ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Ekstraksi Data (*Data Extraction*)

- Tingkat ekstraksi dan transformasi yang diperlukan bergantung pada jenis analitik dan kemampuan solusi Big Data.
- Misalnya, mengekstrak bidang yang diperlukan dari data tekstual yang dibatasi, seperti dengan file log server web, mungkin tidak diperlukan jika solusi Big Data yang mendasarinya sudah dapat langsung memproses file tersebut.
- Mengekstrak teks untuk analitik teks, yang memerlukan pemindaian seluruh dokumen, disederhanakan jika solusi Big Data yang mendasarinya dapat langsung membaca dokumen dalam format aslinya.



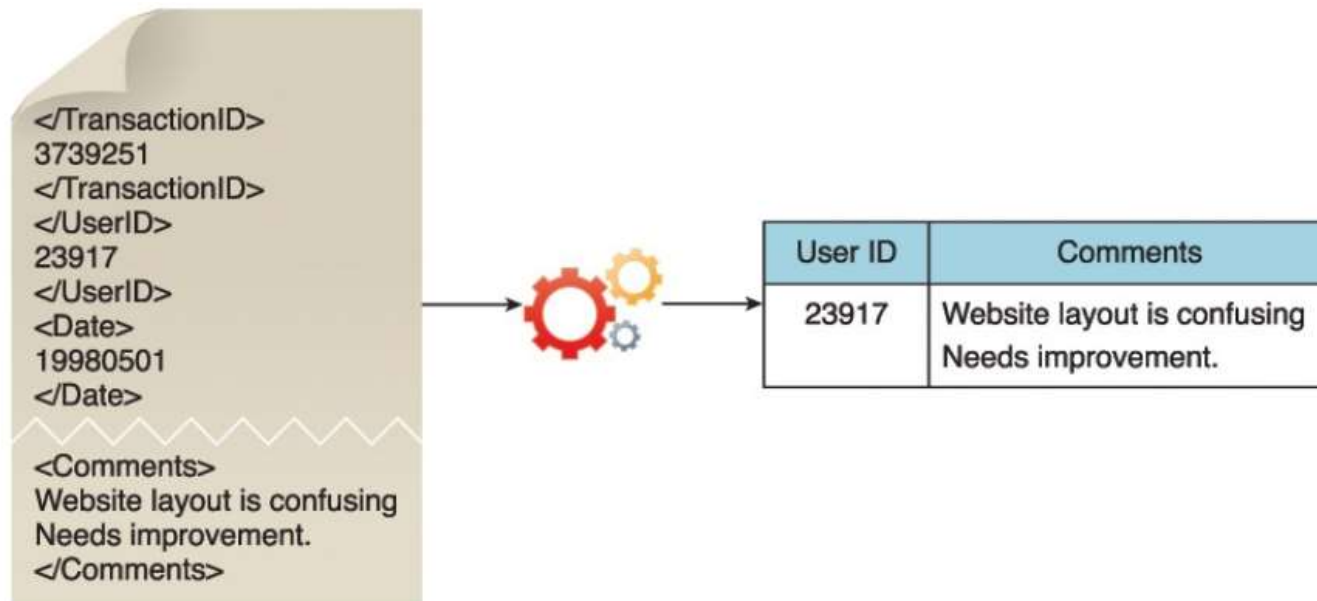
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Ekstraksi Data (*Data Extraction*)

Ilustrasi ekstraksi komentar dan ID pengguna yang disematkan di dalam dokumen XML tanpa perlu transformasi lebih lanjut.



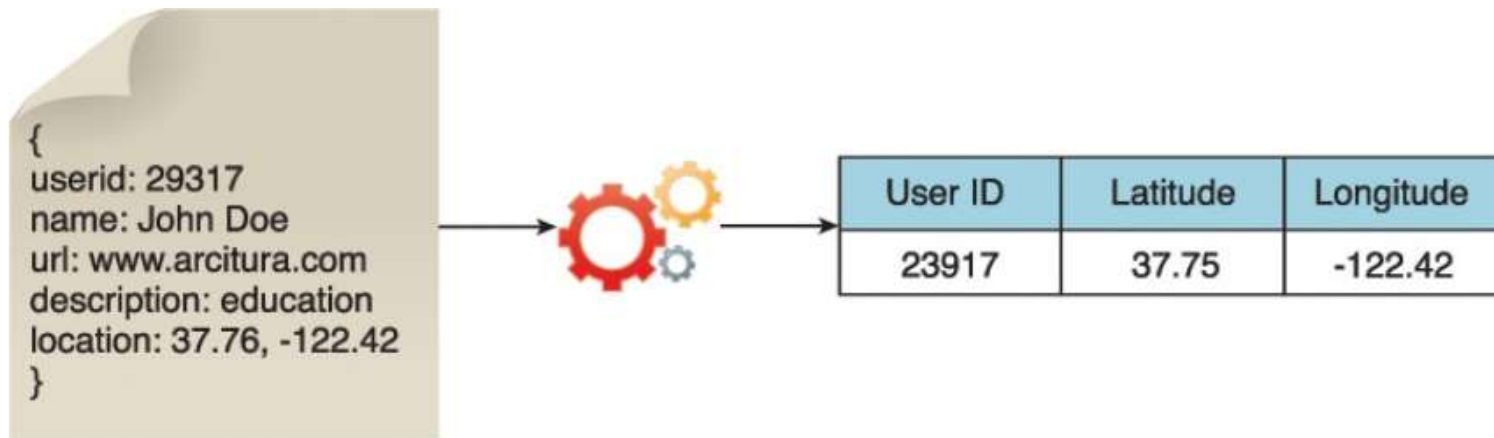
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Ekstraksi Data (*Data Extraction*)

ID pengguna dan koordinat pengguna diekstraksi dari satu JSON.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
tinggi

QS
STARS
RATING SYSTEM
2016

CSA
SIA

IAEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Validasi dan Pembersihan Data (*Data Validation and Cleansing*)

- Data yang tidak valid dapat mengacaukan dan memalsukan hasil analisis.
- Tidak seperti data perusahaan tradisional, di mana struktur data telah ditentukan sebelumnya dan data telah divalidasi sebelumnya, input data ke dalam analisis Big Data dapat tidak terstruktur tanpa indikasi validitas.
- Kompleksitasnya selanjutnya dapat mempersulit untuk sampai pada serangkaian batasan validasi yang sesuai.
- Tahap Validasi dan Pembersihan didedikasikan untuk menetapkan aturan validasi yang seringkali rumit dan menghapus data tidak valid yang diketahui.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Validasi dan Pembersihan Data (*Data Validation and Cleansing*)

- Solusi Big Data sering kali menerima data yang berlebihan di seluruh kumpulan data yang berbeda.
- Redundansi ini dapat dimanfaatkan untuk mengeksplorasi kumpulan data yang saling berhubungan untuk mengumpulkan parameter validasi dan mengisi data valid yang hilang.



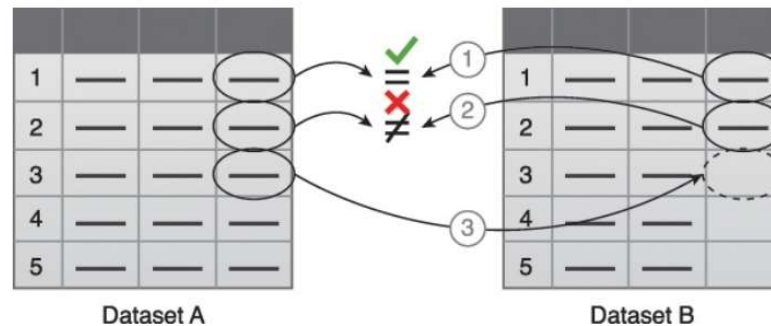
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Validasi dan Pembersihan Data (*Data Validation and Cleansing*)

- Nilai pertama di Dataset B divalidasi terhadap nilai yang sesuai di Dataset A.
- Nilai kedua di Dataset B tidak divalidasi terhadap nilai yang sesuai di Dataset A.
- Jika ada nilai yang hilang, nilai tersebut disisipkan dari Dataset A.



Validasi data dapat digunakan untuk memeriksa kumpulan data yang saling berhubungan untuk mengisi data valid yang hilang.

Validasi dan Pembersihan Data (*Data Validation and Cleansing*)

- Untuk analitik batch, validasi dan pembersihan data dapat dicapai melalui operasi ETL offline.
- Untuk analitik real-time, sistem dalam memori yang lebih kompleks diperlukan untuk memvalidasi dan membersihkan data saat tiba dari sumbernya.
- Sumber data dapat memainkan peran penting dalam menentukan keakuratan dan kualitas data yang dipertanyakan.
- Data yang tampaknya tidak valid mungkin masih berharga karena mungkin memiliki pola dan tren tersembunyi.



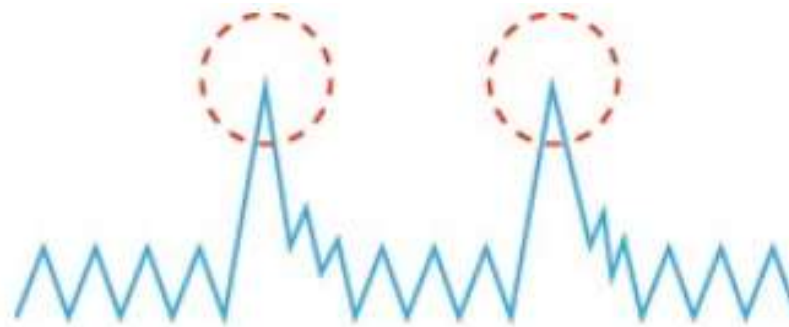
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Validasi dan Pembersihan Data (*Data Validation and Cleansing*)

- Adanya data yang tidak valid mengakibatkan lonjakan.
- Meskipun data tampak tidak normal, mungkin merupakan indikasi pola baru.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Simpul

QS
STARS
RATING SYSTEM
2020

CSA
SAS

IAEE

CPA
AUSTRALIA

ICAEW

CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Agregasi dan Representasi Data (*Data Aggregation and Representation*)

- Data dapat tersebar di beberapa kumpulan data, yang mengharuskan kumpulan data digabungkan bersama melalui bidang umum, misalnya tanggal atau ID.
- Dalam kasus lain, bidang data yang sama mungkin muncul di beberapa kumpulan data, seperti tanggal lahir.
- Metode rekonsiliasi data diperlukan atau kumpulan data yang mewakili nilai yang benar perlu ditentukan.
- Tahap Agregasi dan Representasi Data didedikasikan untuk mengintegrasikan beberapa kumpulan data bersama-sama untuk sampai pada tampilan terpadu.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Agregasi dan Representasi Data (*Data Aggregation and Representation*)

- Melakukan tahap ini bisa menjadi rumit karena perbedaan dalam:
 - Struktur Data – Meskipun format datanya mungkin sama, model datanya mungkin berbeda
 - Semantik – Nilai yang diberi label berbeda dalam dua kumpulan data yang berbeda dapat berarti hal yang sama, misalnya “nama keluarga (surname)” dan “nama belakang (last name)”.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Linggi

QS
STARS
RATING SYSTEM

CSO
KUALITAS

IAABE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Agregasi dan Representasi Data (*Data Aggregation and Representation*)

- Persyaratan analisis data di masa depan perlu dipertimbangkan selama tahap ini untuk membantu mendorong penggunaan kembali data.
- Apakah agregasi data diperlukan atau tidak, penting untuk dipahami bahwa data yang sama dapat disimpan dalam berbagai bentuk.
- Satu bentuk mungkin lebih cocok untuk jenis analisis tertentu daripada yang lain.
- Misalnya, data yang disimpan sebagai BLOB tidak akan banyak berguna jika analisis memerlukan akses ke bidang data individual.



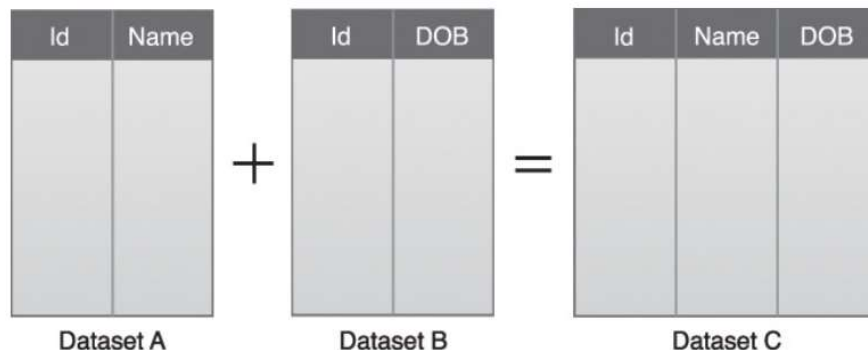
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Agregasi dan Representasi Data (*Data Aggregation and Representation*)

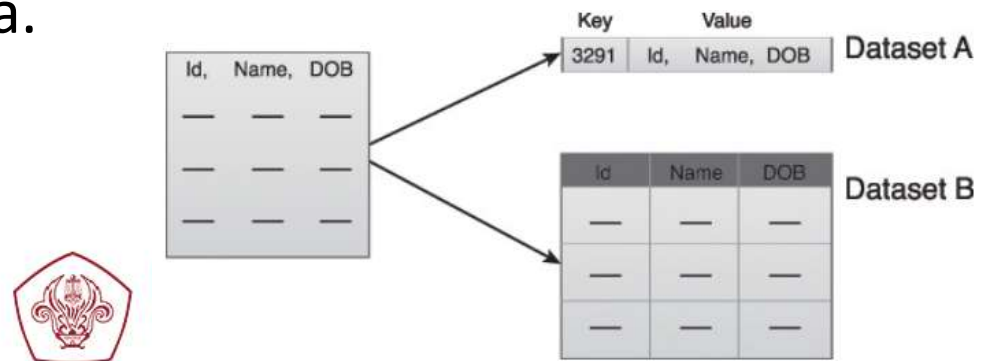
- Struktur data yang distandarisasi oleh solusi Big Data dapat bertindak sebagai penyebut umum yang dapat digunakan untuk berbagai teknik dan proyek analisis.
- Memerlukan pembuatan repositori analisis standar pusat, seperti database NoSQL.
- Contoh sederhana agregasi data di mana dua kumpulan data digabungkan bersama menggunakan bidang Id.



AR
nanagara

Agregasi dan Representasi Data (*Data Aggregation and Representation*)

- Dataset A berisi bagian data yang diinginkan, tetapi merupakan bagian dari BLOB yang tidak mudah diakses untuk query.
- Dataset B berisi bagian data yang sama yang diatur dalam penyimpanan berbasis kolom, memungkinkan setiap bidang untuk ditanyakan satu per satu.
- Dataset A dan B dapat digabungkan untuk membuat struktur data standar dengan solusi Big Data.



Analisis Data (*Data Analysis*)

- Tahap Analisis Data didedikasikan untuk melaksanakan tugas analisis yang sebenarnya, yang biasanya melibatkan satu atau lebih jenis analisis.
- Tahap ini dapat bersifat iteratif, terutama jika analisis data bersifat eksploratif, di mana analisis kasus diulang sampai pola atau korelasi yang sesuai ditemukan.
- Pendekatan analisis eksplorasi akan dijelaskan secara singkat, bersama dengan analisis konfirmatori.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis Data (*Data Analysis*)

- Bergantung pada jenis hasil analitik yang diperlukan, tahap ini bisa sederhana menanyakan kumpulan data untuk menghitung agregasi untuk perbandingan.
- Di sisi lain, ini bisa sama menantangya dengan menggabungkan penambangan data dan teknik analisis statistik yang kompleks untuk menemukan pola dan anomali atau untuk menghasilkan model statistik atau matematis untuk menggambarkan hubungan antar variabel.
- Analisis data dapat diklasifikasikan sebagai analisis konfirmasi atau analisis eksplorasi, yang terakhir terkait dengan penambangan data.



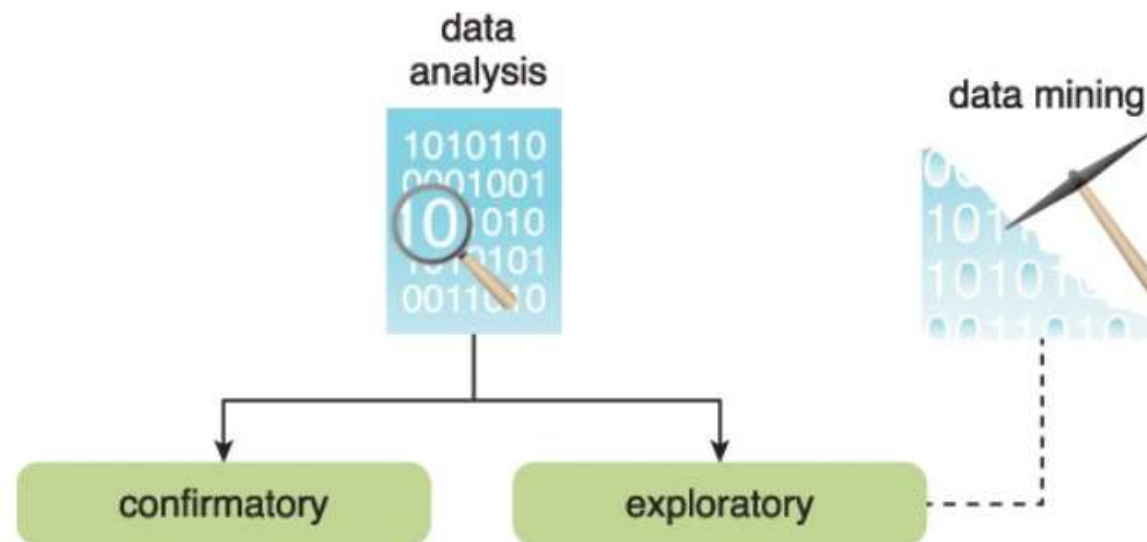
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis Data (*Data Analysis*)

- Analisis data dapat dilakukan sebagai analisis konfirmasi atau eksplorasi.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
tinggi

QS
STARS
RATING SYSTEM
2016

AMBA
AACSB
EFMD

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Analisis Data (*Data Analysis*)

- Analisis data konfirmatori adalah pendekatan deduktif di mana penyebab fenomena yang diselidiki diusulkan terlebih dahulu.
- Penyebab atau asumsi yang diajukan disebut hipotesis.
- Data tersebut kemudian dianalisis untuk membuktikan atau menyangkal hipotesis dan memberikan jawaban definitif untuk pertanyaan tertentu.
- Digunakan teknik pengambilan sampel data.
- Temuan tak terduga atau anomali biasanya diabaikan karena penyebab yang telah ditentukan diasumsikan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Analisis Data (*Data Analysis*)

- Analisis data eksplorasi adalah pendekatan induktif yang terkait erat dengan data mining.
- Tidak ada hipotesis atau asumsi yang telah ditentukan sebelumnya yang dihasilkan. Sebaliknya, data dieksplorasi melalui analisis untuk mengembangkan pemahaman tentang penyebab fenomena tersebut.
- Meskipun mungkin tidak memberikan jawaban yang pasti, metode ini memberikan arahan umum yang dapat memfasilitasi penemuan pola atau anomali.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Linggi

QS
STARS
RATING SYSTEM
5



UNTAR untuk INDONESIA

Visualisasi Data (*Data Visualization*)

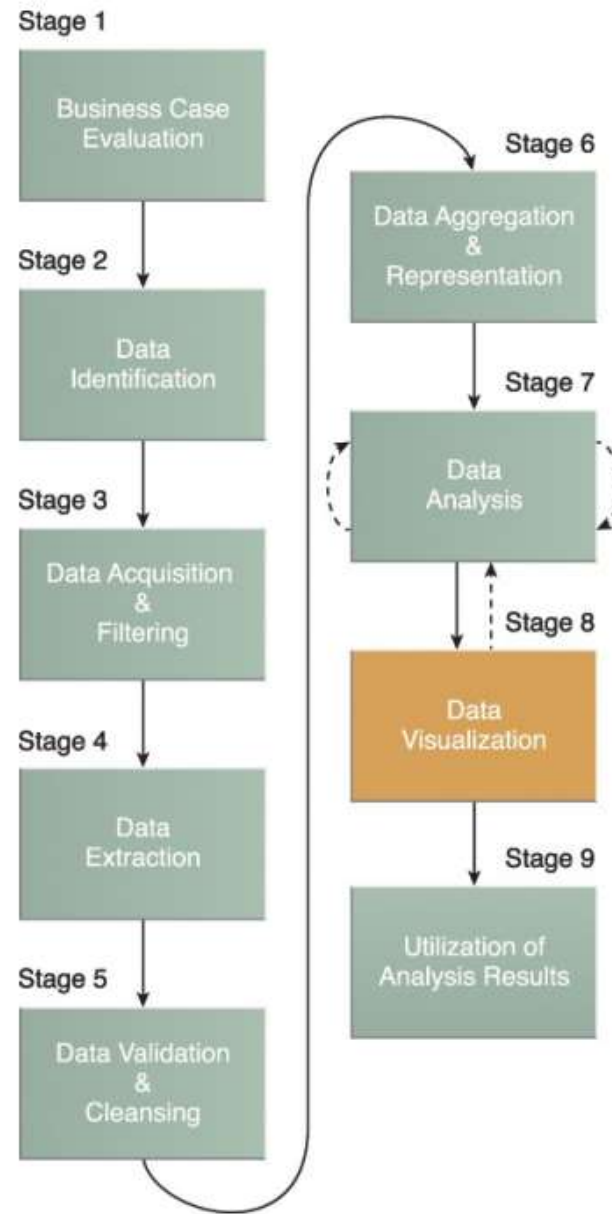
- Kemampuan untuk menganalisis data dalam jumlah besar dan menemukan wawasan yang berguna hanya memiliki sedikit nilai jika satu-satunya yang dapat menginterpretasikan hasilnya adalah para analis.
- Tahap Visualisasi Data didedikasikan untuk menggunakan teknik dan alat visualisasi data untuk mengkomunikasikan hasil analisis secara grafis untuk interpretasi yang efektif oleh pengguna bisnis.
- Pengguna bisnis harus dapat memahami hasil untuk mendapatkan nilai dari analisis dan selanjutnya memiliki kemampuan untuk memberikan umpan balik, seperti yang ditunjukkan oleh garis putus-putus yang mengarah dari tahap 8 kembali ke tahap 7.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA



Visualisasi Data (*Data Visualization*)

- Hasil menyelesaikan tahap Visualisasi Data memberi pengguna kemampuan untuk melakukan analisis visual, memungkinkan penemuan jawaban atas pertanyaan yang bahkan belum dirumuskan pengguna.
- Hasil yang sama dapat disajikan dalam beberapa cara yang berbeda, yang dapat mempengaruhi interpretasi hasil. Akibatnya, penting untuk menggunakan teknik visualisasi yang paling cocok dengan menjaga domain bisnis dalam konteks.
- Aspek lain yang perlu diingat adalah bahwa menyediakan metode pengeboran ke statistik yang relatif sederhana sangat penting, agar pengguna memahami bagaimana hasil gabungan atau gabungan dihasilkan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Pemanfaatan Hasil Analisis (*Utilization of Analysis Results*)

- Setelah hasil analisis tersedia bagi pengguna bisnis untuk mendukung pengambilan keputusan bisnis, seperti melalui dasbor, mungkin ada peluang lebih lanjut untuk memanfaatkan hasil analisis.
- Tahap Pemanfaatan Hasil Analisis ditujukan untuk menentukan bagaimana dan di mana data analisis yang diproses dapat dimanfaatkan lebih lanjut.
- Tergantung pada sifat masalah analisis yang ditangani, hasil analisis mungkin menghasilkan "model" yang merangkum wawasan dan pemahaman baru tentang sifat pola dan hubungan yang ada dalam data yang dianalisis.
- Sebuah model mungkin terlihat seperti persamaan matematika atau seperangkat aturan.
- Model dapat digunakan untuk meningkatkan logika proses bisnis dan logika sistem aplikasi, dan dapat membentuk dasar sistem atau program perangkat lunak baru.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Pemanfaatan Hasil Analisis (*Utilization of Analysis Results*)

- Area umum yang dieksplorasi selama tahap ini meliputi:
 - ✓ Input untuk Sistem Perusahaan
 - ✓ Optimasi Proses Bisnis
 - ✓ Peringatan



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Linggi

QS
STARS
RATING SYSTEM
2020



CPA
AUSTRALIA



ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Input untuk Sistem Perusahaan

- Hasil analisis data dapat dimasukkan secara otomatis atau manual langsung ke sistem perusahaan untuk meningkatkan dan mengoptimalkan perilaku dan kinerjanya.
- Misalnya, toko online dapat diumpankan ke hasil analisis terkait pelanggan yang diproses yang dapat memengaruhi cara toko tersebut menghasilkan rekomendasi produk.
- Model baru dapat digunakan untuk meningkatkan logika pemrograman dalam sistem perusahaan yang ada atau dapat membentuk dasar dari sistem baru.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Optimasi Proses Bisnis

- Pola, korelasi, dan anomali yang teridentifikasi selama analisis data digunakan untuk menyempurnakan proses bisnis.
- Contohnya adalah mengkonsolidasikan rute transportasi sebagai bagian dari proses rantai pasokan.
- Model juga dapat mengarah pada peluang untuk meningkatkan logika proses bisnis.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Peringatan

- Hasil analisis data dapat digunakan sebagai masukan untuk peringatan yang ada atau dapat menjadi dasar peringatan baru.
- Misalnya, peringatan dapat dibuat untuk memberi tahu pengguna melalui email atau teks SMS tentang suatu peristiwa yang mengharuskan mereka mengambil tindakan korektif.



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

Referensi

- Erl, T., Khattak, W. and Buhler, P., 2016. Big data fundamentals: concepts, drivers & techniques. Prentice Hall Press.
- Dietrich, D., Heller, E. and Yang, B., 2015. Data science & big data analytics: discovering, analyzing, visualizing and presenting data. Wiley.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
tinggi

QS
STARS
RATING SYSTEM
5 star



CPA
AUSTRALIA



ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA