

6.1 Framework Kualitas Data

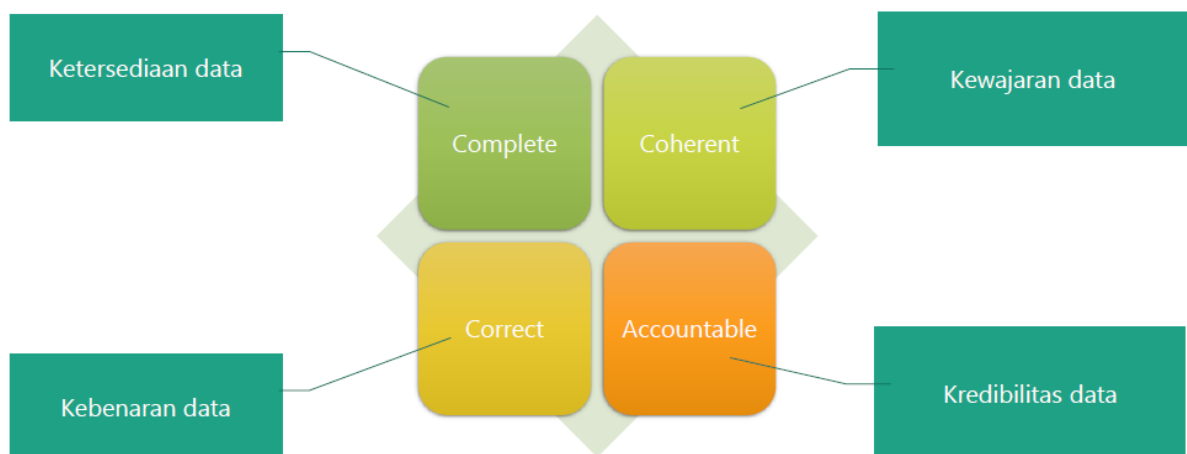
Kualitas data merupakan kondisi data yang sesuai dengan kebutuhan penggunaannya.

Ada banyak definisi tentang kualitas data. Dua yang mendominasi adalah:

- ✓ Data berkualitas, jika datanya sesuai dengan tujuan penggunaan yang dimaksudkan
- ✓ Data berkualitas, jika data merepresentasikan konstruksi dunia nyata dengan benar

Kualitas data ditentukan oleh berbagai faktor seperti keakuratan, kelengkapan, konsistensi, atau ketepatan waktu. Kualitas tersebut diperlukan untuk memenuhi kebutuhan organisasi dalam hal operasional, perencanaan dan pengambilan keputusan.

Berikut dimensi dari Kualitas data



Framework Kualitas Data

A. Complete (Ketersediaan Data)

Kelengkapan data menunjukkan terdapat informasi yang cukup untuk membuat kesimpulan. Kelengkapan dapat diukur dengan menentukan apakah setiap entri data merupakan entri data "lengkap" atau tidak. Semua bidang entri data yang tersedia harus lengkap, dan kumpulan catatan data tidak boleh kehilangan informasi terkait.

Misalnya, metrik kualitas sederhana yang dapat Anda gunakan adalah jumlah nilai kosong dalam kumpulan data: dalam konteks inventaris / pergudangan, itu berarti

bahwa setiap baris item mengacu pada produk dan masing-masing harus memiliki pengenalan produk. Hingga pengenalan produk tersebut diisi, item baris tersebut tidak valid.

✓ *Apakah saya memiliki semua data yang saya butuhkan untuk menjawab pertanyaan saya?*

- Tersedianya semua variabel yang dibutuhkan
- Tersedianya semua unit observasi yang dibutuhkan
- Tersedianya semua nilai dari variabel pada unit observasi

✓ *Apakah sebenarnya Anda memiliki semua record untuk pertanyaan yang diajukan?*

- Pastikan jumlah data yang dimiliki sesuai dengan data di system lainnya. Untuk data yang jumlahnya sangat besar, terutama karena sekarang kita berada di era Big Data, pengecekan data dapat dilakukan menggunakan teknik sampling.

Record-Level Completeness

Ketika menemukan ketidak lengkapan data pada level baris, pilihan aksinya adalah:

1. Perbaiki data yang hilang.
2. Hapus baris yang tidak lengkap.
3. Memberikan tanda untuk baris yang tidak lengkap dan menginformasikan adanya ketidaklengkapan data ketika menjawab kebutuhan atas suatu informasi.
4. Tidak melakukan apa-apa.

B. Coherent (Kewajaran Data)

Juga dikenal sebagai validasi data, integritas mengacu pada struktural data untuk memastikan bahwa data sesuai dengan prosedur. Ini berarti tidak ada kesalahan data yang tidak diinginkan, dan ini sesuai dengan peruntukannya (misalnya, tanggal, bulan, dan tahun). Semuanya berawal pada tingkat kesalahan transformasi data. Kita dapat melacak berapa banyak operasi transformasi data yang relatif gagal secara keseluruhan - atau dengan kata lain, seberapa sering proses pengambilan data yang disimpan dalam satu format dan mengonversinya ke format lain tidak berhasil dilakukan.

Apakah data yang dimiliki konsisten dan sesuai dengan logika?

✓ **Referential integrity**

Referential integrity adalah sebuah cara untuk menjaga konsistensi data antara tabel yang saling berhubungan. Cara ini diterapkan untuk data yang disimpan dalam bentuk tabular.

Contoh:

Terdapat pengecekan ke dalam data mata kuliah tawar program studi sedemikian sehingga hanya mata kuliah yang ditawarkan saja yang dapat diambil oleh mahasiswa ketika masa registrasi semester baru.

✓ **Value Integrity**

Apakah nilai yang disimpan konsisten secara internal?

Contoh:

Kuota yang tersedia untuk suatu mata kuliah X adalah 40 mahasiswa. Jika sudah ada 30 mahasiswa yang melakukan registrasi dan mengambil matakuliahX, maka kuota yang tersisa seharusnya adalah 10 mahasiswa.

Informasi yang terdapat pada data registrasi, data matakuliah (yang mencakup kuota mahasiswa) harus dipastikan konsistensinya agar jumlah pendaftar tidak melebihi kuota yang tersedia.

Evaluasi Koherensi

- Tentukan level koherensi yang dibutuhkan.
- Apakah validasi pada referential integrity atau value integrity, atau mungkin diperlukan validasi integritas data yang lebih kompleks?
- Tentukan seberapa lengkap validasi yang dibutuhkan, dan apa saja batasan performansi dan waktunya.
- Apakah perlu melakukan validasi untuk setiap baris atau setiap relasi antar table yang ada?
- Ataupun dapat dilakukan *statistical sampling* untuk mengevaluasi sebagian baris-baris yang representatif?

C. Correct (Kebenaran Data)

✓ *Apakah data Anda sudah benar?*

mengevaluasi kebenaran data

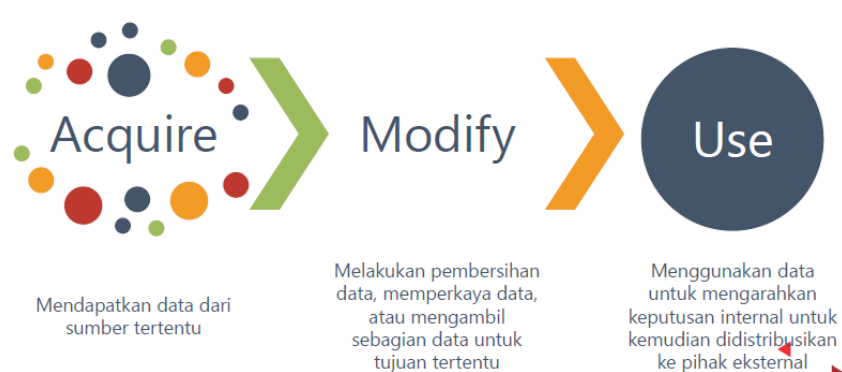
- Rincilah data apa saja yang dapat di verifikasi dengan mudah. Misal Anda memiliki data penjualan dari sebuah supermarket, apakah data yang perlu Anda analisis seharusnya adalah data penjualan yang terjadi pada tahun 2018-2019?

- Tentukan hal apa saja yang penting untuk divalidasi
- Pahami, seberapa banyak data yang harus benar?
- Putuskan apa yang akan dilakukan pada data yang tidak benar Apakah mungkin untuk memperbaiki datanya?
- Bisakah proses dilanjutkan tanpa menggunakan data yang tidak benar tersebut?
- Beri tanda untuk data yang tidak benar, dan nyatakan fakta tersebut dalam proses analisis.
- Pilihan lainnya, pisahkan data yang tidak benar, dan lakukan analisis secara terpisah.

D. Accountability (Kredibilitas Data)

Konsep "akuntabilitas" telah muncul sebagai tema dominan dalam privasi global dan undang-undang perlindungan data, kebijakan, dan praktik organisasi dan dianggap mendasar untuk manajemen privatisasi. Ada tren yang muncul yang menunjukkan bahwa prinsip akuntabilitas mengharuskan organisasi mengambil pendekatan proaktif dan terstruktur untuk manajemen privatisasi melalui penerapan langkah-langkah perlindungan data dan privasi yang sesuai dan dapat dibuktikan.

- ✓ Siapa yang bertanggung jawab terhadap data Anda?



Catat Semua Sumber Data

- Bagaimana cara mengakses datanya?
- Tentukan pihak mana yang memiliki dan mengelola setiap data yang didapatkan
- Tentukan siapa saja yang dapat dihubungi jika ada pertanyaan terkait data dan apa saja tanggung jawab atau kesepakatan layanan dari datanya: Apa jaminannya terkait kualitas, akurasi, format data, dan ketepatan waktu penyediaan datanya.

Simpan semuanya

- Simpan data asli (data mentah) yang didapatkan.
- Data asli dapat digunakan untuk mengecek modifikasi apa yang sudah dilakukan

Lakukan Audit Mandiri

- Periksa data untuk mengkonfirmasi apakah salinan data yang dimiliki sesuai dengan versi yang berada di sumber data.
- Konfirmasi bahwa setiap modifikasi atau peningkatan kualitas data tidak memunculkan kesalahan pada data.
- Pastikan setiap data tidak mengalami modifikasi tanpa rencana atau berubah untuk tujuan tertentu yang tidak sesuai (dipalsukan).

Telusuri Akses terhadap Data

- Pahami siapa saja yang dapat mengakses data dan bagaimana caranya
- Untuk data yang tersimpan dalam sistem pengelolaan data yang mapan, pengaksesan data dapat dikendalikan dan dimonitor dengan baik hingga pada level apa saja data yang dapat diakses oleh setiap pengguna.

Kesimpulan

- Memperhatikan kualitas data sebelum mulai bekerja dengan data tidak hanya menghemat waktu, melainkan juga memberikan pemahaman yang lebih baik mengenai apa saja kemampuan dan keterbatasan Anda dalam menganalisis data.
- Pastikan Anda membuat dan mendokumentasikan keputusan dan aksi yang dilakukan terkait perhatian Anda terhadap kualitas data.

Analisis terhadap kualitas data dan penyiapan data adalah langkah awal yang penting untuk mendapatkan *insight* dari data yang dimiliki.

6.2 Manfaat data berkualitas

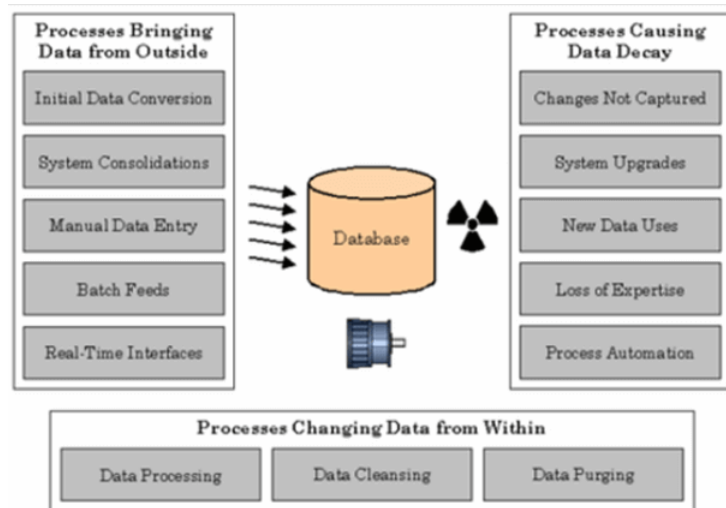
Pentingnya kualitas data sangat jelas: orang hanya dapat membuat keputusan berdasarkan data yang benar jika data yang mereka gunakan benar. Tanpa kualitas data yang memadai, data praktis tidak berguna dan berbahaya.

Misalkan data di bidang: pemasaran. Diperoleh memiliki daftar yang dari 10.000 email, nama, nomor telepon, bisnis, dan alamat di dalamnya. Kemudian, bayangkan 20% dari daftar itu tidak akurat (yang sesuai dengan data bagan dari Ringlead di atas). Artinya, 20% dari daftar tersebut memiliki email, nama, nomor telepon, dll yang salah. Bagaimana cara menerjemahkannya ke dalam angka?

Misal jika Anda menjalankan kampanye iklan di Facebook yang menargetkan nama-nama dalam daftar ini, biayanya akan menjadi 20% lebih tinggi dari yang seharusnya - karena entri nama palsu itu. Jika Anda mengirim surat fisik, hingga 20% dari surat Anda bahkan tidak akan sampai ke penerima. Dengan panggilan telepon, perwakilan penjualan Anda akan membuang-buang waktu mereka untuk nomor yang salah atau nomor yang tidak dapat diambil. Dengan email, bukan masalah besar, tetapi jika melibatkan tarif akan terdistorsi berdasarkan daftar "kesalahan" data menyebabkan biaya bertambah, berkontribusi pada masalah biaya.

Konsekuensi dari kontrol kualitas data yang buruk dapat mempengaruhi setiap aspek pada sebuah organisasi, termasuk di antaranya :

- Biaya kampanye/iklan pemasaran
- Akurasi dalam memahami pelanggan
- Efektifitas kerja yang mengubah prospek menjadi penjualan
- Akurasi membuat keputusan bisnis



Sumber data berkualitas rendah