

Resiko pada data quality

Pertama, seperti yang diketahui bersama bahwa dengan diterapkannya prinsip skema data warehouse pada struktur data maka akan bertambah pula storage yang dibutuhkan untuk menunjang kelancaran proses transformasi yang akan terjadi, dari segi storage jika growth data tidak bisa diimbangi dengan penyesuaian storage maka akan dipastikan terjadinya masalah pada availability data, dari availability data tersebut maka akan terjadi masalah jika data tersebut sudah terlalu banyak dan juga demand untuk terus dilakukan update pada data, hal ini akan menjadi masalah pada data quality yang mana akan mengakibatkan kurangnya keterkinian data.

Kedua, pada redundansi data akan terjadi masalah jika diterapkannya metode datawarehouse dengan star schema, karena setiap perubahan yang terjadi pada sumber data akan mengakibatkan data yang tidak kompatibel pada star schema yang sudah diterapkan, dan masalah pada integritas data tersebut akan menimbulkan masalah yang sangat besar pada data quality jika tidak secara rutin di manage.

Mitigasi pada masalah

Mitigasi yang paling tepat untuk masalah storage data akan bisa di atasi dengan menyiapkan cloud storage untuk setiap penyimpanan yang terjadi pada datawarehouse, dengan menggunakan cloud storage setiap perubahan yang mengakibatkan ekspansi pada data dapat di mitigasi dengan cepat, sehingga ekspansi data tersebut akan dengan mudah di sesuaikan dan tidak mengakibatkan masalah yang berkelanjutan.

Mitigasi untuk integritas data adalah dengan cara menerapkan data catalog untuk setiap skema yang ada pada database, penerapan data catalog diperlukan untuk me-monitor semua perkembangan pada data dan juga menerapkan data quality rule pada data.

Data catalog yang akan diterapkan harus memiliki spesifikasi seperti :

- Data profiling

Untuk menampilkan detail dari data sebagai insight untuk menerapkan beberapa keputusan seperti menerapkan data quality atau menerapkan perubahan atau pengkinian pada data tersebut.

- Data quality integration

Data quality akan berperan sebagai penentu aturan yang akan diterapkan pada suatu table atau data, nantinya parameter yang diterapkan akan otomatis dilakukan pengecekan oleh sistem dan akan menampilkan insight berupa apakah data tersebut masih memenuhi aspek quality yang diterapkan.

- Data collaboration

Data catalog harus digunakan sebagai tempat untuk kolaborasi antar sesama owner data, sehingga owner dari data akan mendapatkan insight seputar data yang memiliki relasi dengan pekerjaannya.

- Data versioning

Data versioning akan memberi insight kepada user dari setiap table yang terdapat pada database.

- Data lineage

Data lineage akan digunakan untuk memberi insight kepada user tentang flow dari table sehingga ketika terdapat penyesuaian pada data maka user akan dengan mudah mengetahui pipeline dari data tersebut.

Data catalog yang akan diterapkan juga harus di maintain dalam 2 source berbeda, yaitu dalam bentuk sheet data dictionary (contoh google sheet), dan juga data catalog tools. Dictionary yang terdapat pada sheet akan digunakan untuk menambahkan informasi tambahan pada data seperti data descriptions dan juga flag data-data PII (Personally Identifiable Information), sementara data catalog tools akan digunakan dengan fungsi yang tertera pada 5 poin di atas.

Regulasi yang akan diterapkan

Sebagai data governance maka akan memberikan informasi ke masing-masing owner atau user data tentang data apa yang terkait dengan mereka dan juga apa saja data mereka yang perlu di jaga integritasnya, seperti memberi informasi kepada owner dari data, tentang keterkinian dari datanya yang kurang terjaga dan juga correction pada data yg mereka miliki (seperti data yang mereka miliki kurang valid), dan juga menerapkan aturan yang mengharuskan pengguna data untuk menjaga kerahasiaan data yang mereka gunakan, dengan membatasi akses mereka ke database dan juga menerapkan encryption pada data-data yang sifatnya rahasia, peraturan dan juga hak akses tersebut harus secara rutin di jaga dengan housekeeping rutin dan juga harus di atur dalam peraturan tertulis berupa SOP (Standard Operational Procedure) yang akan di sosialisasikan kepada masing-masing stakeholder atau owner data.

Kedua, setiap user yang akan meng-akses data-data rahasia harus mendapatkan persetujuan supervisi / manager langsung di atasnya, agar masing-masing user sadar akan resiko yang mereka hadapi jika mereka menyebar luaskan informasi yang bersifat rahasia.