

BAB I PENDAHULUAN

Bab Pendahuluan ini memuat latar belakang dari metadata, perumusan masalah, tujuan penelitian, batasan-batasan, serta potensi manfaat dari penelitian mengenai metadata.

I.1 Latar Belakang

Data merupakan suatu aset penting yang memiliki nilai tinggi bagi semua perusahaan maupun organisasi terutama pada skala *enterprise*. Pada skala ini, jumlah data yang dihasilkan dan digunakan sangat dan harus memiliki penyimpanan suatu basis data yang dapat diakses secara cepat, dan berkapasitas besar. Setiap individual yang ada pada *enterprise* memiliki pengetahuan yang berbeda satu sama lain mengenai data yang terdapat pada perusahaannya, dan tidak akan ada satupun individual yang mengetahui semua mengenai datanya (DAMA International, 2017). Menurut Stefanov (2006) konteks informasi pada data dibutuhkan agar individual mengetahui untuk apa data itu, seperti *goals* dari perusahaan, dan metrik yang digunakan untuk mengevaluasi datanya.

Data yang sangat banyak menjadi tantangan bagi *enterprise* akan sulitnya mengontrol data-datanya. Data yang tidak terkontrol akan membuat perusahaan kesulitan dalam mengatur data, memelihara data, bahkan menggunakan data secara efektif. Pemakaian data yang tidak tepat dapat menimbulkan redudansi data yang dapat memenuhi *storage*, juga kualitas data yang kurang baik. Untuk mencegah hal tersebut diharuskan adanya suatu standar dan informasi dari data-data tersebut dengan menggunakan metadata. Selain itu metadata berfungsi agar suatu perusahaan dapat mengetahui mengenai datanya, representasi dari apa data tersebut, darimana data tersebut, bagaimana data tersebut berpindah melalui sistem, siapa yang dapat mengaksesnya, dan apa arti dari data tersebut agar dapat berkualitas tinggi (DAMA International, 2017). Fungsi utama dari metadata sendiri yaitu untuk menyediakan informasi dari suatu data yang tersedia agar data tersebut dapat ditemukan menggunakan kriteria yang relevan.

Elemen pada metadata dapat digunakan untuk menjelaskan suatu atribut seperti pada koleksi museum, informasi yang terkait pada hak, lokasi *repository* juga

cukup penting untuk dokumentasi pada museum (Chen & Ke, 2013). Terutama pada suatu *file system*, metadata digunakan untuk menjelaskan organisasi dan struktur dari *file system* tersebut yang termasuk atribut *file*, direktori *file* tersebut, dan juga informasi dari *physical space* dari *file* yang ada. Pemrosesan pada metadata selain dilakukannya *maintenance* untuk *namespace*, juga diperlukan *maintenance* untuk atribut dari *file* tersebut dan juga suatu lokasi dari blok *file*. Meskipun jumlah metadata tergolong sedikit, namun operasi dari metadata ini dapat bekerja sebanyak 60% dari *workloads* pada suatu *file*. Oleh karena itu efisiensi pemrosesan metadata akan sangat berpengaruh pada kinerja suatu *file system* (Xiong, 2011). Untuk mengefisienkan suatu pemrosesan metadata perlu dilakukannya distribusi dan juga integrasi metadata pada *file system* yang akan digunakan.

Metadata sangat diperlukan untuk kelestarian, interpretasi, penggunaan ulang suatu data, dan juga suatu proses untuk mendokumentasikan data pada transisi untuk pengarsipan koleksi data. Maka dari itu menggunakan metadata sebagai pengintegrasian data di seluruh fase pencatatan dan siklus hidup pengarsipan serta untuk mengelola hubungan antara objek data, tahapan penelitian, dan teknologi. Metadata menjadi suatu pusat transisi dari agregasi data yang tidak teratur, termasuk pada *long tail* pada penelitian, dan juga data baru yang terbuat secara aktif ketika publikasi dan studi, menjadi suatu arsitektur koleksi. Transisi Integrasi metadata memerlukan suatu strategi yang jelas agar asal data (*data provenance*) tidak menjadi buram dan agar tidak mustahil (Kulasekaran, Trelogan, & Esteva, 2014).

Pada suatu data statistik (Dayal, 2008) yang mencakup lebih dari 13 *file systems* dari lima situs *supercomputing* dan dua *file systems* dari departmen server *file*, banyaknya jumlah *file* dan juga direktori dari setiap *file system* adalah beberapa juta sampai sepuluh juta *file*. Proses yang berjalan pada setiap inti CPU menjadi masalah akses I/O secara independen yang menjadikan banyaknya akses yang bersamaan pada *file systems*. Maka dari itu suatu *file system* yang memiliki klaster dengan skala besar perlu menangani puluhan ribu metadata dari *file* per detik. Hal ini mengindikasikan bahwa pemrosesan metadata secara terpusat kurang praktis dan diperlukan prosesi metadata secara terdesentralisasi atau

terdistribusi untuk suatu *file system* untuk memanfaatkan kelompok server metadata. Partisi suatu *namespace* dan distribusi metadata merupakan suatu kunci masalah pada prosesi metadata terdistribusi yang mengukur hasil kinerja prosesi metadata. (Xiong, 2011). Banyak dari sistem yang ada menggunakan partisi secara statis yang berarti administrator secara manual melakukan partisi pada *namespace* dan juga mendistribusikan metadatanya. Salah satu *policy* suatu distribusi metadata yaitu S2PC-MP yang dapat menjadikan metadata konsisten untuk operasi distribusinya.

Penelitian ini merupakan pengembangan dari penelitian sebelumnya yang dilakukan oleh (Khairunisak, 2021) mengenai teknik penilaian manajemen metadata dan assesment model. Pengembangan yang dilakukan yaitu pengaplikasian tools metadata menggunakan *open source tools* dari *assesment model* yang berfungsi untuk standarisasi dari data-data yang memiliki perbedaan dalam satu basis data agar tidak ada replikasi atau redudansi, menciptakan data yang terstruktur, dan agar *lineage* dari data-data tersebut jelas terdefinisi. Hal ini bertujuan untuk memudahkan akses data-datanya bagi perusahaan. Dengan mengimplementasikan integrasi metadata dan distribusi metadata yang dilakukan secara benar, perusahaan juga dapat mengetahui hubungan antar data dan dapat meminimalkan risiko akan kebocoran suatu data yang merupakan aset penting. Untuk itu metadata diperlukan untuk memastikan bahwa perusahaan dapat mengidentifikasi data yang rahasia dan sensitif dan juga dapat mengatur siklus hidup data. Hal ini bertujuan agar meminimalkan munculnya risiko pada data perusahaan (DAMA International, 2017). Penelitian ini juga akan diaplikasikan pada suatu aplikasi manajemen data berbasis web yang bertemakan *Data Governance*. Aplikasi ini memiliki dua fitur utama yaitu Data Quality Management dan juga Master Data Management yang kemudian akan ditambahkan modul mengenai integrasi dan distribusi metadata.

I.2 Perumusan Masalah

Berdasarkan uraian masalah yang telah dijelaskan pada latar belakang, terdapat perumusan masalah yang akan dikaji dari penelitian ini yaitu:

1. Bagaimana cara menambahkan modul integrasi metadata pada aplikasi manajemen data dengan open source tools?
2. Bagaimana cara menambahkan modul distribusi metadata pada aplikasi manajemen data dengan open source tools?

I.3 Tujuan Penelitian

Terdapat beberapa tujuan dari penelitian ini yang ingin dicapai, yaitu:

1. Menambahkan modul integrasi metadata pada aplikasi manajemen data dengan open source tools
2. Menambahkan modul distribusi metadata pada aplikasi manajemen data dengan open source tools

I.4 Batasan Penelitian

Adapun suatu batasan dari penelitian ini yaitu sebagai berikut:

1. Pengembangan aplikasi dilakukan untuk mengimplementasikan modul integrasi dan distribusi metadata
2. Modul yang akan diimplementasikan berupa suatu *business rule* berbentuk *metadata* dari table database, dengan kerangka kerja *website* menggunakan *Laravel*.
3. Basis data yang akan diimplementasikan menggunakan satu jenis RDBMS yaitu MySQL
4. Pengimplementasian modul dengan open source tools menggunakan Pentaho Data Integration
5. Metadata yang digunakan berupa metadata *table* dari suatu *database*
6. Logging yang dilakukan untuk dashboard hanya untuk satu aktivitas melakukan *generate data*
7. Jumlah pengguna pada aplikasi hanya satu pengguna.

I.5 Manfaat Penelitian

Manfaat dari penelitian ini terdiri dari manfaat secara teknis dan secara keilmuan. Manfaat secara teknis yaitu dapat mengimplementasikan metadata pada aplikasi suatu perusahaan agar dijadikan suatu rule untuk standarisasi struktur tabel dengan integrasi pada aplikasi, dan distribusi dari metadata yang

tersedia pada perusahaan yang dapat menambah keakuratan identifikasi data, konsistensi suatu data, serta penggunaan data yang ada pada perusahaan dapat terpantau secara jelas dengan suatu dashboard yang memberikan gambaran umum mengenai metadata yang dapat memberikan suatu *insight* agar dapat memberikan keputusan bagi perusahaan. Sedangkan manfaat keilmuan yang diharapkan yaitu dapat mengetahui fungsi dari metadata untuk dijadikan suatu rule dengan mengimplementasikan integrasi modul metadata pada aplikasi dan mendistribusikannya pada pengguna.

I.6 Sistematika Penulisan

Sistematika penulisan laporan ini terbagi menjadi beberapa bab dari pokok pembahasan, secara umum dapat dijabarkan sebagai berikut:

Bab I Pendahuluan

Pada bab ini berisi uraian mengenai latar belakang implementasi metadata pada aplikasi dengan perumusan masalah, tujuan penelitian, hingga manfaat penelitian

Bab II Tinjauan Pustaka

Bab ini berisi literatur yang relevan dengan permasalahan yang diteliti dan dibahas pula hasil-hasil penelitian terdahulu. Minimal terdapat lebih dari satu metodologi/metode/kerangka kerja yang disertakan pada bab ini untuk menyelesaikan permasalahan atau meminimalkan gap antara kondisi eksisting dengan target. Pada akhir bab, analisis pemilihan metodologi/metode/kerangka kerja harus dijelaskan untuk menentukan metodologi/metode/kerangka kerja yang akan digunakan di penelitian ini.

Bab III Metodologi Penelitian. Metodologi penelitian merupakan strategi dan langkah-langkah (*plan of attack*) yang akan dilakukan di penelitian dalam rangka menjawab rumusan masalah yang disusun sebelumnya. Penyusunan metodologi penelitian harus dilakukan secara kritis apakah metode atau teknik yang dipilih memang tepat sesuai tujuan penelitian. Pada bab ini dijelaskan langkah-langkah

penelitian secara rinci meliputi: tahap merumuskan masalah penelitian, merumuskan hipotesis, mengembangkan model penelitian, mengidentifikasi dan melakukan operasionalisasi variabel penelitian, menyusun kuesioner penelitian, merancang pengumpulan dan pengolahan data, melakukan uji instrumen, merancang analisis pengolahan data.

Bab V Hasil dan Evaluasi

Pada bab ini, disajikan hasil rancangan, temuan, analisis dan pengolahan data. Selain itu bab ini juga berisi tentang validasi atau verifikasi hasil dari penelitian, sehingga hasil tersebut apakah telah benar-benar menyelesaikan masalah atau menurunkan gap antara kondisi eksisting dan target yang akan dicapai. Analisis sensitivitas juga dapat digunakan di bab ini untuk lebih mengetahui hasil penelitian dapat diterapkan baik secara khusus di konteks penelitian maupun secara umum di konteks serupa (misal perusahaan di sektor serupa). Selain itu metode-metode evaluasi yang lain dapat diterapkan untuk memvalidasi hasil TA sesuai dengan kebutuhan.

Secara keseluruhan bab ini membahas secara mendetail mengenai hasil dari penelitian dan refleksinya terhadap tujuan penelitian. Untuk penelitian yang berfokus pada merancang sistem informasi/ aplikasi maka penamaan bab ini mengikuti tahapan penerapan SDLC yang digunakan dalam penelitian.

Bab VI Kesimpulan dan Saran

Pada bab ini dijelaskan kesimpulan dari cara untuk mengintegrasikan dan mendistribusikan metadata dengan jawaban dari pertanyaan penelitian yang disajikan di pendahuluan. Saran penelitian dikemukakan pada bab ini untuk penelitian selanjutnya.