

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Pemanfaatan sistem *e-learning* untuk meningkatkan fleksibilitas dan efektivitas pembelajaran di berbagai lembaga pendidikan terus berkembang seiring dengan pesatnya kemajuan teknologi informasi dan komunikasi [1]. Sistem *Learning Management System (LMS)* seperti *Moodle*, *Sakai*, *Edx*, dan *TidiaAe* menjadi *platform* penting dalam mendukung proses pengajaran [2]. Data yang dihasilkan oleh *LMS* ini meliputi aktivitas pembelajaran, penilaian, dan perilaku belajar mahasiswa yang terekam secara rinci dan lengkap pada setiap satuan waktu. Data-data ini memiliki potensi yang sangat besar untuk memberikan pengetahuan bagi institusi pendidikan, terutama dalam proses evaluasi dan peningkatan kualitas pembelajaran. Sayangnya, pengelolaan data di banyak perguruan tinggi masih dilakukan secara konvensional, yang dapat memperlambat pengambilan keputusan strategis dan mempersulit penyusunan laporan evaluasi [3].

Di sinilah pentingnya penerapan *business intelligence (BI)* yang dapat mengubah data operasional menjadi informasi yang relevan dan akurat untuk mendukung proses pengambilan keputusan di perguruan tinggi. Namun, banyak institusi pendidikan di Indonesia belum sepenuhnya memanfaatkan *BI*, terutama karena keterbatasan akses cepat terhadap data yang terstruktur dan terintegrasi dengan baik [4]. Fakta ini menunjukkan bahwa implementasi *BI* pada *LMS* di perguruan tinggi di Indonesia masih relatif kecil dan memiliki ruang besar untuk dikembangkan lebih lanjut.

Salah satu komponen kunci dalam pengembangan *BI* adalah pembangunan *data warehouse (DWH)*, yang memungkinkan pengumpulan dan pengolahan data dari berbagai sumber menjadi format multidimensi yang siap dianalisis. Proses ini tidak dapat terlepas dari proses *Extract, Transform, and Load (ETL)*,

di mana data diambil dari sumbernya, diubah ke dalam format yang diinginkan, dan kemudian dimuat ke dalam *data warehouse* [5]. Proses *ETL* inilah yang menjadi tulang punggung integrasi data, yang mendukung pembuatan analisis lebih lanjut melalui *Online Analytical Processing (OLAP)* dan penyajian informasi melalui *dashboard* atau laporan-laporan berbasis *BI*. Dalam penelitian ini, dilakukan implementasi *data warehouse e-learning* dalam melakukan transformasi data transaksional *LMS* ke bentuk data yang siap untuk dianalisis. Dalam penelitian ini, fokus utama adalah pada pengukuran efisiensi waktu eksekusi proses *ETL* sebagai salah satu parameter uji utama. Hal ini bertujuan untuk memastikan bahwa proses *ETL* dapat berjalan secara optimal dalam skenario data berskala besar, sehingga mendukung pengambilan keputusan berbasis data yang cepat [6].

Pengujian performa *ETL* sangat penting untuk memastikan bahwa proses yang diterapkan tidak hanya memenuhi kebutuhan integrasi data, tetapi juga dalam hal kecepatan waktu eksekusi [7]. Penelitian ini menitikberatkan pada pengukuran efisiensi waktu eksekusi proses *ETL* menggunakan berbagai metode ekstraksi, seperti Table Output, Switch, dan Sync After Merge. Pengujian dilakukan dengan dataset berskala besar, hingga 1 juta data, untuk mengevaluasi performansi sistem dan skalabilitas proses *ETL* yang diimplementasikan. Hasil pengujian akan digunakan untuk menilai apakah proses *ETL* yang diimplementasikan menghasilkan peningkatan performa yang signifikan dibandingkan metode konvensional serta untuk mengidentifikasi apakah terdapat penurunan performa setelah implementasi *ETL*. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam perancangan *data warehouse*, tetapi juga dalam memastikan bahwa implementasi *BI* berbasis *LMS* berjalan secara optimal dan efisien.

1.2. Rumusan Masalah

Berdasarkan latar belakang dan permasalahan yang ada, maka dirumuskan permasalahan sebagai berikut :

- Bagaimana desain dan implementasi proses ETL menggunakan Pentaho Data Integration dapat dioptimalkan untuk memproses data besar?
- Bagaimana performansi proses ETL yang diimplementasikan, terutama dalam hal efisiensi waktu eksekusi?

1.3. Tujuan dan Manfaat

Penelitian ini bertujuan untuk mengembangkan dan mengoptimalkan proses Extract, Transform, and Load (ETL) pada data warehouse e-learning dengan memanfaatkan Pentaho Data Integration. Implementasi ini difokuskan pada pengolahan data berskala besar dari Learning Management System (LMS) untuk mendukung pengambilan keputusan yang lebih efektif di lingkungan perguruan tinggi. Melalui optimalisasi proses ETL ini, diharapkan dapat meningkatkan efisiensi waktu eksekusi dalam pemrosesan data dan menghasilkan informasi yang lebih akurat untuk analisis pembelajaran.

Implementasi ini diharapkan dapat memberikan manfaat dalam meningkatkan kualitas pengelolaan data pembelajaran di institusi pendidikan, khususnya dalam konteks e-learning. Dengan adanya sistem yang teroptimasi, proses pengambilan keputusan strategis dan penyusunan laporan evaluasi dapat dilakukan dengan lebih cepat dan efisien.

Tabel 1.1. Tabel keterkaitan antara tujuan, pengujian dan kesimpulan.

No.	Tujuan	Pengujian	Kesimpulan
1	Mengimplementasikan proses ETL pada data warehouse e-learning menggunakan Pentaho Data Integration untuk memproses data besar	Pengujian implementasi proses ETL dengan berbagai metode ekstraksi (<i>Table Output, Switch, dan Sync After Merge</i>)	Proses ETL berhasil diimplementasikan dengan validasi data yang memuaskan, memverifikasi akurasi data dari sumber ke staging area hingga ke data warehouse
2	Mengukur dan menganalisis performansi proses ETL yang diimplementasikan dari segi efisiensi waktu eksekusi berdasarkan variasi ukuran data	Pengujian performa waktu eksekusi ETL dengan dataset hingga 1 juta data	Metode Sync After Merge terbukti paling optimal untuk proses ETL, menunjukkan kinerja yang stabil dan efisien terutama dalam skenario Combination Data dengan waktu eksekusi yang lebih cepat dibanding metode lain

1.4. Batasan Masalah

Dalam mengerjakan tugas akhir ini, ada beberapa batasan yang perlu ditetapkan agar penelitian bisa diselesaikan dengan baik dalam waktu satu semester. Karena keterbatasan waktu dan sumber daya, penelitian ini menggunakan data hasil simulasi dari sistem Moodle yang dipasang sendiri, bukan menggunakan data asli dari sistem pembelajaran online yang sudah ada di kampus. Hal ini dilakukan untuk menghindari kesulitan dalam mendapatkan izin penggunaan data asli yang biasanya bersifat rahasia.

Untuk penyimpanan data, penelitian ini hanya menggunakan database MySQL karena lebih mudah digunakan dan sudah familiar. Walaupun ada database lain yang mungkin lebih canggih, MySQL dipilih agar penelitian bisa lebih fokus pada proses pengolahan datanya saja.

Dalam pengujian kinerja sistem, penelitian ini membatasi pengujian hanya pada tiga cara ekstraksi data, di mana nantinya akan dipilih satu cara terbaik berdasarkan waktu prosesnya yang paling cepat. Pembatasan ini memungkinkan analisis yang lebih mendalam untuk setiap metode yang diuji. Data yang digunakan untuk pengujian dibuat melalui simulasi dengan

mengikuti struktur data Moodle, dengan jumlah data mencapai 1 juta baris. Jumlah data ini dibatasi karena mempertimbangkan kemampuan komputer yang digunakan dan waktu pengerjaan yang hanya satu semester, meskipun dalam kenyataannya sistem pembelajaran online biasanya memiliki data yang jauh lebih banyak.

1.5. Metode Penelitian

Metode yang digunakan dalam pengerjaan tugas akhir ini adalah sebagai berikut:

1. Studi Literatur

Penelitian dimulai dengan melakukan studi literatur terkait konsep dan teknologi yang digunakan dalam penelitian ini. Kajian meliputi data warehouse, proses Extract, Transform, Load (ETL), dan analisis performa menggunakan Pentaho Data Integration (PDI). Selain itu, literatur terkait Learning Management System (LMS) Moodle juga dikaji untuk memahami struktur data yang relevan dengan penelitian.

2. Pengumpulan Dataset

Data yang digunakan dalam penelitian ini diambil dari database simulasi LMS Moodle versi 4.2. Dataset mencakup data aktivitas pembelajaran seperti tugas, kuis, forum, serta data pengguna (dosen dan mahasiswa). Data ini disiapkan dalam staging area untuk diolah lebih lanjut dalam proses ETL.

3. Perancangan Data Warehouse

Tahap ini meliputi perancangan skema data warehouse menggunakan pendekatan star schema. Tabel fakta dan tabel dimensi dirancang untuk menyimpan data yang relevan, seperti aktivitas kursus, penilaian, dan aktivitas mahasiswa, sehingga mendukung analisis multidimensional.

4. Implementasi ETL

Implementasi proses ETL dilakukan menggunakan Pentaho Data Integration (PDI). Tahap ini melibatkan proses ekstraksi data dari database Moodle, transformasi data menjadi format yang sesuai dengan data warehouse, dan pemuatan data ke tabel fakta dan dimensi. Tiga metode ekstraksi data diuji, yaitu Table Output, Sync After Merge, dan Switch Case, untuk menentukan metode dengan performa terbaik.

5. Pengujian Performasi ETL

Pengujian dilakukan untuk mengukur efisiensi waktu eksekusi proses ETL dengan variasi ukuran dataset hingga 1 juta data. Skenario pengujian meliputi penambahan data baru, sinkronisasi data tanpa perubahan, pembaruan data, penghapusan data, serta kombinasi dari operasi tersebut. Analisis dilakukan untuk mengevaluasi metode yang paling optimal.

6. Analisis Hasil dan Kesimpulan

Hasil pengujian dianalisis untuk mengevaluasi performa setiap metode ekstraksi ETL dalam skenario yang berbeda. Kesimpulan diambil berdasarkan hasil pengujian, termasuk rekomendasi konfigurasi terbaik untuk implementasi ETL pada data warehouse e-learning.