

TESIS
MENGAMANKAN DATA SENSITIF PROSES TRANSFORM
ETL MENGGUNAKAN DATA MASKING



ARIEF DOLANTS

22/2008/0087/TSD/14

PROGRAM STUDI TEKNOLOGI INFORMASI
PROGRAM MAGISTER
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA
2024

TESIS
MENGAMANKAN DATA SENSITIF PROSES TRANSFORM
ETL MENGGUNAKAN DATA MASKING

Diajukan sebagai salah satu syarat untuk menyelesaikan studi



Disusun Oleh

ARIEF DOLANTS

22/2008/0087/TSD/14

PROGRAM STUDI TEKNOLOGI INFORMASI
PROGRAM MAGISTER
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA

2024

**HALAMAN PERSETUJUAN
UJIAN TESIS**

Judul : Mengamankan Data Sensitif Proses Transform ETL
Menggunakan Data Masking
Nama : Arief Dolants
NIM : 22/2008/0087/TSD/14
Program Studi : Teknologi Informasi
Program : Magister
Semester : Genap
Tahun Akademik : 2023/2024

Telah diperiksa dan disetujui untuk diujikan di hadapan Dewan Penguji Tesis

Yogyakarta, 14 April 2023

Dosen Pembimbing,



Dr. Widyastuti Andriyani, S.Kom., M.Kom

0217038201

HALAMAN PENGESAHAN


TESIS

**MENGAMANKAN DATA SENSITIF PROSES TRANSFORM *ETL*
MENGUNAKAN *DATA MASKING***

Telah dipertahankan di depan Dewan Penguji Tesis dan dinyatakan diterima
untuk memenuhi sebagian persyaratan guna memperoleh Gelar
Magister Komputer
Program Studi Teknologi Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta

Yogyakarta, 4 Desember 2023

| Dewan Penguji | NIDN | Tandatangan |
|--|------------|--|
| 1. Dr. Widyastuti Andriyani, S.Kom., M.Kom (Ketua) | 0217038201 |  |
| 2. Dr. Bambang P.D.P, S.E., Ak., S.Kom.,MMSI (Sekretaris) | 0525087201 |  |
| 3. Dr. Domy Kristomo, S.T., M.Eng (Penguji) | 0530078302 |  |

Mengetahui
Ketua Program Studi Teknologi Informasi

Dr. Widyastuti Andriyani, S.Kom., M.Kom
NPP 121176

PERNYATAAN KEASLIAN TESIS

Dengan ini saya menyatakan bahwa naskah tesis ini belum pernah diajukan untuk memperoleh gelar Magister Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 9 Maret 2024



Arief Dolants
22/2008/0087/TSD/14

HALAMAN PERSEMBAHAN

Alhamdulillah. Puji dan syukur kehadiran Allah ﷻ, Tuhan semesta alam, yang telah memberikan rahmat, hidayah serta karunia-Nya, sehingga penulis dapat menyelesaikan tesis ini dengan baik. Shalawat dan salam semoga senantiasa tercurahkan kepada Nabi Muhammad ﷺ.

Dengan penuh rasa syukur, tesis ini saya persembahkan untuk :

- Kedua orang tua, yang selama ini telah mendidik dengan baik, menyayangi, serta selalu berkorban untuk kami, putra dan putri tercinta mereka.
- Istri tercinta, yang tak pernah lelah untuk selalu memberikan semangat dan dukungannya.
- Teman-teman, serta rekan kerja di Solusi247 yang memberikan semangat dan selalu berbagi ilmu.
- Bapak dan Ibu Dosen, yang telah sabar mendidik dan semangat dalam mengajar dan memberikan ilmunya.
- Dosen Pembimbing, yang selalu menyediakan waktunya untuk mengarahkan kami dalam menyelesaikan tesis ini.
- Rekan-rekan di kampus Universitas Teknologi Digital Indonesia, saya haturkan terima kasih untuk bantuan dan kerjasama kalian sampai dengan saat ini.
- Pihak lain, keluarga, saudara, dan juga teman-teman lain yang tak dapat saya sebutkan satu persatu, saya persembahkan tesis ini untuk kalian semua.

HALAMAN MOTO

*Bukanlah (hakikat) ilmu yang hanya sekedar dihafal, namun
ilmu yang sebenarnya adalah yang bermanfaat*

-Al Imam Asy-Syafi'I rahimahullah-

KATA PENGANTAR

Alhamdulillah, puji syukur penulis panjatkan kepada Allah SWT atas segala limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul "**Mengamankan Data Sensitif Proses Transform ETL Menggunakan Data Masking**". Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister pada Program Studi **Teknologi Informasi, Universitas Teknologi Digital Indonesia**. Penelitian ini bertujuan untuk mengkaji dan menerapkan teknik data masking dalam proses ETL (Extract, Transform, Load) guna melindungi data sensitif dari ancaman yang mungkin terjadi selama proses transformasi data.

Dalam penyusunan tesis ini, penulis banyak menerima bimbingan, dukungan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan penghargaan dan terima kasih yang sebesar-besarnya kepada:

1. **Dr. Widyastuti Andriyani, S.Kom., M.Kom**, selaku pembimbing utama yang dengan sabar telah memberikan arahan, bimbingan, dan masukan yang sangat berharga selama proses penyusunan tesis ini.
2. **Dr. Bambang P.D.P, S.E., Ak., S.Kom., MMSI**, selaku penguji yang telah memberikan saran dan tanggapan yang bermanfaat.
3. Para dosen di **Fakultas Teknologi Informasi, Universitas Teknologi Digital Indonesia**, yang telah memberikan ilmu pengetahuan dan wawasan yang mendalam selama masa perkuliahan.
4. Teman-teman mahasiswa dan semua pihak yang telah memberikan dukungan moral dan material selama penyusunan tesis ini.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak guna penyempurnaan penelitian ini di masa mendatang.

Akhir kata, penulis berharap semoga tesis ini dapat memberikan kontribusi yang berarti bagi perkembangan ilmu pengetahuan, khususnya dalam bidang keamanan data dan proses ETL, serta dapat bermanfaat bagi pihak-pihak yang membutuhkan.

Yogyakarta, 9 Maret 2024
Penulis,

Arief Dolants

DAFTAR ISI

| | |
|--|------|
| HALAMAN PERSETUJUAN TESIS | iii |
| HALAMAN PENGESAHAN TESIS | iv |
| PERNYATAAN KEASLIAN TESIS | v |
| HALAMAN PERSEMBAHAN | vi |
| HALAMAN MOTO | vii |
| KATA PENGANTAR | viii |
| DAFTAR ISI..... | ix |
| DAFTAR GAMBAR | xi |
| DAFTAR TABEL..... | xii |
| INTISARI..... | xiii |
| ABSTRACT..... | xiv |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang Masalah | 1 |
| 1.2 Rumusan Masalah | 2 |
| 1.3 Batasan Masalah..... | 3 |
| 1.4 Tujuan..... | 3 |
| 1.5 Manfaat..... | 4 |
| BAB II TINJAUAN PUSTAKA..... | 6 |
| BAB III LANDASAN TEORI..... | 13 |
| 3.1 Transformasi Data | 13 |
| 3.2 Data Privasi | 16 |
| 3.3 <i>ETL – Extract Transform Loading</i> | 18 |
| 3.4 <i>Data masking</i> | 21 |
| BAB IV METODOLOGI PENELITIAN | 25 |
| 4.1 Studi Literasi | 25 |

| | |
|--|-----------|
| 4.2 Bahan atau Data..... | 26 |
| 4.3 Alat atau Perangkat..... | 27 |
| 4.4 Pengumpulan data | 28 |
| 4.5 Analisis dan perancangan sistem..... | 30 |
| 4.5.1. Analisis Masalah..... | 30 |
| 4.5.2. Arsitektur Sistem | 31 |
| 4.6 HGrid247 DE - Data Engineering | 34 |
| BAB V HASIL DAN PEMBAHASAN..... | 36 |
| 5.1 Extract, Transform, Loading | 36 |
| 5.1.1. Extract Data | 36 |
| 5.1.2. Proses Transform | 39 |
| 5.1.3. Loading Data..... | 43 |
| 5.2 <i>Data masking</i> | 45 |
| 5.3 Pengujian | 48 |
| 5.4 Confusion Matrix | 52 |
| BAB VI KESIMPULAN DAN SARAN | 56 |
| 6.1 Kesimpulan..... | 56 |
| 6.2 Saran | 58 |
| DAFTAR PUSTAKA..... | 61 |

DAFTAR GAMBAR

| | | |
|-------------|---|----|
| Gambar 3.1 | Gambaran umum proses ETL..... | 19 |
| Gambar 4.1 | Halaman Kaggle untuk dataset siswa pendaftar kampus IIT ... | 29 |
| Gambar 4.2 | Data transaksi bank yang ada di Kaggle..... | 30 |
| Gambar 4.3 | Gambaran proses implementasi data masking | 31 |
| Gambar 4.4 | Arsitektur sistem proses ETL implementasi data masking | 32 |
| Gambar 4.5 | Proses Pembuatan Workflow di HGrid247 DE..... | 34 |
| Gambar 5.1 | Isi dari data transaksi bank | 36 |
| Gambar 5.2 | Data siswa pendaftar kampus IIT dalam bentuk csv | 37 |
| Gambar 5.3 | Isi dari data transaksi bank | 38 |
| Gambar 5.4 | Sumber data disimpan di folder lokal file system | 38 |
| Gambar 5.5 | Spark-shell dari Apache Spark | 39 |
| Gambar 5.6 | Design workflow proses ETL untuk data transaksi bank..... | 40 |
| Gambar 5.7 | Data transaksi bank kolom-kolom dipisahkan oleh delimiter .. | 41 |
| Gambar 5.8 | Inisialisasi membagi perkolom untuk data transaksi bank..... | 41 |
| Gambar 5.9 | Logic yang digunakan pada filtering data | 42 |
| Gambar 5.10 | Fungsi join reference dan replace text..... | 43 |
| Gambar 5.11 | Data hasil proses masking yang sudah tersimpan di database . | 44 |
| Gambar 5.12 | Data hasil masking diquery lewat CLI MySQL Client | 44 |
| Gambar 5.13 | Simulasi penggunaan data reference | 45 |
| Gambar 5.14 | Reference file untuk mengaburkan nilai Transaction Details .. | 46 |
| Gambar 5.15 | Proses mengubah Date format String menjadi Timestamp | 47 |
| Gambar 5.16 | Design workflow pemberian nilai tambahan..... | 48 |
| Gambar 5.17 | Tampilan ketika perintah submit job dijalankan | 48 |
| Gambar 5.18 | Output hasil proses data masking data transaksi bank | 50 |
| Gambar 5.19 | Hasil data masking dan hasil query pengembalian data..... | 51 |
| Gambar 5.20 | Grafik hasil pengujian responden..... | 53 |
| Gambar 5.21 | Salah satu pertanyaan yang diajukan kepada responden | 53 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1 Tabel Referensi | 10 |
| Tabel 5.1 Transformasi perubahan data transaksi bank | 43 |
| Tabel 5.2 Perbandingan antara data sumber, dan data hasil proses transformasi | 49 |
| Tabel 5.3 Tabel confusion matrix untuk data masking transaksi bank | 54 |

INTISARI

Extract, Transform, Loading atau *ETL*, melibatkan beberapa proses, diantaranya ekstraksi data dari berbagai sumber data, kemudian proses transformasi untuk menjadikan data agar format dan isinya sesuai dengan yang diinginkan, dan menyimpannya ke dalam *database* dengan proses loading. Proses transformasi data pada *ETL* merupakan proses penting dalam analisis data, namun terkadang tanpa disadari dapat mengungkapkan informasi sensitif yang tidak diinginkan, terutama hasil proses transformasi data di *staging* yang tidak terlindungi. Penelitian ini memperkenalkan teknik *data masking* yang dapat digunakan pada proses transformasi data untuk menjaga kerahasiaan data sebelum data tersebut disimpan ke dalam *database*. Teknik *data masking* yang digunakan meliputi penggunaan nilai acak, substitusi dan juga penggantian karakter. Penelitian ini menunjukkan bahwa teknik *data masking* yang diusulkan dapat secara efektif melindungi data sensitif, terutama hasil proses transformasi dan juga sebelum data di load ke dalam *database*, dan meminimalkan pembuatan tabel duplikasi atau penyimpanan baru khusus untuk menyimpan data yang sudah dimasking.

Kata Kunci : Transformasi data, Data Privasi, *ETL*, *Data masking*

ABSTRACT

Extract, Transform, Load or ETL involves several processes, including data extraction from various data sources, followed by transformation processes to adjust the data format and content to the desired state, and then storing it in a database through the loading process. The data transformation process in *ETL* is a crucial step in data analysis, but sometimes, unintentionally, it can expose unwanted sensitive information, especially in the transformed data's unprotected staging phase. This study introduces data masking techniques that can be applied during the data transformation process to preserve data confidentiality before it's stored in the database. The employed data masking techniques encompass the use of random values, substitution, and character replacement. This research demonstrates that the proposed data masking techniques can effectively safeguard sensitive data, particularly in the results of the transformation process and prior to loading the data into the database. This approach also minimizes the need for creating duplicate tables or new storage specifically for holding the masked data.

Kata Kunci : Data Transformation, Data Privacy, ETL, Data Masking