BAB II TINJAUAN PUSTAKA

Bab tinjauan pustaka ini bertujuan untuk menyajikan gambaran luas tentang penelitian mengenai *data masking* dan penelitian lain yang berkaitan dengan *data masking* dan juga *ETL*. Fokus utama tinjauan pustaka ini adalah untuk mengidentifikasi dan menganalisis berbagai metode dan pendekatan yang digunakan dalam pelaksanaan *data masking*, serta mengeksplorasi manfaat, keterbatasan, dan tantangan yang terkait dengan penerapan teknik ini.

Tinjauan pustaka ini akan memberikan dasar yang kokoh untuk memahami penggunaan *data masking* dan membantu mengidentifikasi celah penelitian yang masih ada, yang akan menjadi dasar untuk pengembangan penelitian ini. Beberapa penelitian yang telah ada sebelumnya yang berkaitan dengan *Data masking*, *ETL* proses, dan juga enkripsi terangkum dalam Tabel 2.1.

(Archana, R. A., Hegadi, R. S., & Manjunath, T. N., 2018) dalam karyanya yang berjudul *A study on Big Data privacy protection models using data masking methods*, menyarankan pemanfaatan *data masking* untuk menjaga privasi data, terutama pada organisasi yang menggunakan *Big Data*.

Penggunaan beberapa tipe dalam *data masking* digunakan, agar data memiliki format yang sama dengan data aslinya. *The masking type will be decided based on the data such as substitution, replacement, multiplier, randomizer and shuffling*. Dimana, data hasil proses *masking*, formatnya akan tetap sama, terutama dari jumlah karakternya, hanya saja berbeda makna dan isinya.

(Yanti et al., 2018), Implementasi Algoritma Data Encryption Standard Pada Penyandian Record *Database* dalam penelitiannya mengatakan bahwa

record-record yang ada di *database* umumnya masih ditampilkan dalam bentuk teks sebagai informasi bagi pengguna. Sehingga mempermudah kriptanalis untuk mengakses serta memberi peluang untuk melakukan pembocoran, mendistribusikan maupun memodifikasi *record database* tersebut. Dalam penelitiannya, Yanti menggunakan algoritma enkripsi *DES* terutama untuk mengamankan data.

Penelitian berikutnya yaitu (Hidayat & Faizin, 2019), Perbandingan Kriptografi Menggunakan *Algoritma Data Encryption Standard (DES)* dan *Algoritma Rivest Shamir Adleman (RSA)* Untuk Keamanan Data. Dalam risetnya dikatakan, dalam pertukaran informasi yang semakin berkembang pesat, teknologi informasi mendorong kinerja bidang keamanan data.

Dalam penelitiannya lebih menitik beratkan pembahasan pada perbandingan algoritma *Data Encryption Standar (DES)* dengan algoritma *Rivest Shamir Adleman (RSA).*

Pada kesimpulannya, didapatkan hasil dari algoritma *RSA* lebih cepat secara performa, namun lebih baik algoritma *DES* dalam mengamankan data. Dimana disebutkan algoritma *DES* dalam hal keamanannya mengandalkan perhitungan biner, sedangkan *RSA* mengandalkan perhitungan pemfaktoran.

Berikutnya yaitu penelitian oleh (Amrulloh & Ujianto, 2019), Kriptografi Simetris Menggunakan Algoritma *Vigenere Cipher*. Penelitian ini menyebutkan bahwa keamanan informasi itu sangat penting, salah satunya dengan memanfaatkan algoritma kriptografi.

Dikatakan, kriptografi adalah salah satu cara untuk mencegah kebocoran data yang bersifat rahasia. Penelitian ini menggunakan algoritma *Vigenere Cipher*,

dimana teks alfabet di enkripsi menggunakan serangkaian *Caesar Cipher* yang berbeda berdasarkan huruf dari kata kunci dan merupakan bentuk substitusi *polyalphabetic* yang sederhana.

(Prasser et al., 2019), Privacy-enhancing *ETL*-processes for biomedical data. Penelitian ini fokus pada pengamanan informasi dibidang kesehatan, khususnya data pasien. Dikatakan, perlindungan privasi perlu pertimbangan yang cermat, ketika data akan disimpan atau akan digunakan kembali.

Penelitian ini bertujuan untuk menjembatani lingkungan *ETL* yang umumnya tidak mendukung anonimisasi, dengan alat anonimisasi yang umumnya juga tidak mudah untuk diintegrasikan dengan *ETL*.

Berikutnya yaitu penelitian yang berjudul Extract Transform Load (*ETL*) Process in Distributed *Database* Academic Data *Warehouse* (Yulianto, 2019). Penelitian ini melibatkan distributed *database* yang digunakan sebagai solusi pemrosesan data akademis.

ETL Proses juga dikatakan sebagai komponen utama dalam pengembangan *data warehouse* dan perlu perhatian khusus dari manajer *data warehouse*. Hasil dari proses transformasi pada *ETL* juga perlu diuji dan dibuktikan kebenarannya, agar informasi data dari sumber ke tujuan sesuai yang diinginkan.

(Yesin & Vilihura, 2019), dalam penelitiannya yang berjudul "*Some approach to data masking as means to counteract the inference threat"*, mengatakan bahwa *database* saat ini menjadi alat yang diperlukan dihampir semua bidang, yang dapat diandalkan sebagai informasi untuk pengambilan keputusan.

Sehingga *database* menjadi salah satu target pencurian akan data, dimana

biasanya didalam *database* tersimpan informasi penting, seperti informasi operasional perusahaan, data pribadi karyawan, informasi keuangan, informasi pelanggan, dan lainnya.

Dan penelitian berikutnya yaitu, *Database Security And Study Of Data Encryption Methods in Cloud Storage* , (Shcherbinina, Ye., Martseniuk, B., & Filonenko, A, 2020). Penelitian ini mengatakan bahwa, pentingnya melakukan enkripsi *database*, dan keamanan akan data adalah tugas yang penting saat ini.

Banyak kerugian yang dihasilkan dari kebocoran data, seperti kehilangan banyak uang. Terutama sekali, bagi organisasi yang menyimpan datanya pada layanan *cloud*. Disebutkan, dahulu penyedia layanan cloud masih bisa diandalkan untuk merahasiakan data-data yang dititipkan oleh konsumennya. Namun belakangan, dalam beberapa kasus penyedia layanan cloud tidak dapat memenuhi komitmen tersebut, terutama ketika menanggapi permintaan informasi oleh pemerintah.

Dengan memanfaatkan teknik kriptografi algoritma *symmetric encryption*, memungkinkan untuk melakukan enkripsi data pelanggan pada layanan *cloud*. *Symmetric Encryption* ini memiliki kecepatan dan efisiensi komputasi yang cukup baik, terutama dalam menangani enkripsi data yang besar di *cloud-storage*.

Tabel 2.1 Tabel Referensi

Nama Tahun	Permasalahan	Akibat	Data	Jumlah Data	Metode	Jenis Penelitian	Teknologi	Solusi	Hasil
(Archana et al., 2018)	Data besar yang kurang perlindungan	Dapat diambil oleh orang yang tidak berwenang	Raw Big Data Set	Tidak disebutkan	Data Masking	Kualitatif, Kuantitatif	Informatica Big Data Quality, Hadoop	Fleksibilitas dalam implementasi data masking	Melakukan data masking di big data menjadi lebih mudah. mempertahankan privasi data dan memberikan hasil yang serupa dengan data asli.
(Yanti et al., 2018)	Record database masih ditampilkan dalam bentuk teks	Mempermudah kriptanalis memproses serta membuka peluan kebocoran data	Record Database Mahasiswa	Beberapa baris	kriptografi	Kualitatif, Kuantitatif	Algoritma DES	Penyandian record database	Tidak mudah melihat dan memahami isi suatu tabel, dikarenakan datanya dienkripsi
(Hidayat & Faizin, 2019)	Pertukaran data tanpa pengamanan data	Penyalahgunaan oleh orang yang tidak bertanggung jawab	Plain text	Tidak disebutkan	Kriptografi	Kualitatif, Kuantitatif	Algoritma DES dan RSA	membandingkan algoritma DES dan RSA untuk mengenkripsi data	Enkripsi data menggunakan algoritma DES lebih baik, terutama dalam hal pengamanan data
(Amrulloh & Ujianto, 2019)	Akses informasi penting oleh sembarang penerima	Bocornya kerahasiaan data	Data form dari web	Data tidak disebutkan	Kriptografi	Kualitatif, Kuantitatif	Algoritma Vigenere Cipher	Mengimplementasikan algoritma Vigenere Cipher untuk mengenkripsi data diaplikasi	Data tidak mudah di akses dan disalahgunakan, karena terenkripsi
(Prasser et al., 2019)	Kurangnya keamanan pada data medis	Data sensitif pasien tidak terlindungi	Data pasien	Tidak disebutkan	ETL	Kualitatif, Kuantitatif	Algoritma Cell Suppression	Plugin Pentaho	Data terlindungi dari berbagai ancaman
(Yulianto, 2019)	Performa yang lambat dalam proses ETL di datawarehouse	Kebutuhan akan data lambat dipenuhi	Data akademis	497 baris	ETL	Kualitatif, kuantitatif	Pentaho Data Integration, Distributed Database	mengimplementasikan distributed database untuk pengolahan data akademis	Pengolahan data menjadi lebih singkat
(Yesin & Vilihura, 2019)	Serangan untuk mencuri data pada data warehouse	Pencurian dan kebocoran data	Row database	18000 baris data	Data Masking	Kualitatif, kuantitatif	Oracle Database, Algoritma AES128	Implementasi berbagai metode data masking, untuk menyembunyikan data	Data terjamin keamanannya
(Shcherbinina et al., 2020)	Kurangnya perlindungan akan keamanan data	Kerugian yang cukup besar karena data mudah dicuri peretas	Plain text	Data tidak disebutkan	Kriptografi	Kualitatif, Kuantitatif	Symmetric Algorithm	Mengimplementasikan proses enkripsi dan deskripsi pada Cloud sistem	Data di cloud lebih terkontrol, terutama dalam keamanannya
(Thahara & Siregar, 2021)	Tidak menjaga keamanan computer dengan baik	Pencurian data, kerusakan oleh virus	Plaintext	Data tidak disebutkan	Kriptografi	Kualitatif, Kuantitatif	Algoritma DES	Penggunaan Algoritma DES	Keamanan data dan jaringan yang terjaga
(Gomes et al., 2021)	Penggunakan teknik keamanan enkripsi di database	Mempengaruhi kinerja sistem pada operasi baca tulis di basis data	Data TPC.H (transaction processing and database benchmark)	10 GB	Kriptografi	Kualitatif, Kuantitatif	PGP dan AES	Manajemen Enkripsi database	Penggunaan engkrpsi AES
(Mahanan et al., 2021)	Pelanggaran privasi data	Data privasi tidak terjamin	Data Kesehatan pasien	Data tidak disebutkan	Kriptografi	Kualitatif, kuantitatif	k-anonymity algorithm	Pemanfaatan algroitma k-anonymity	Implementasi algrotritma yang digunakan memiliki performa yang baik terutama

Nama Tahun	Permasalahan	Akibat	Data	Jumlah Data	Metode	Jenis Penelitian	Teknologi	Solusi	Hasil
									dalam melakukan enkripsi data
(Pamungkas & Zaney, 2021)	Penggunaan Akses Data secara Tidak Sah	Merugikan Pemilik Data	Data Login	Data tidak disebutkan	Kriptografi	Kualitatif, Kuantitatif	Hashing SHA1 dan Algoritma Asimetris RSA	Melakukan enkripsi data	Data terenkripsi
(Ghann et al., 2022)	Mudahnya data dengan informasi yang sensitif di publikasikan	Pelanggaran privasi	Tidak disebutkan	Tidak disebutkan	Kriptografi	Kualitatif, kuantitatif	Bit Code Sensitif Algorithm (BSCA)	Mengimplementasikan algoritma BSCA untuk menjaga keamanan data yang sesnsitif	Efisien, efektif dan lebih aman dalam menjaga keamanan data
(Prajapati & Mary, 2022)	Untuk membangun proses ETL membutuhkan biaya yang mahal	Proses ETL dianggap sebagai masalah, sehingga jarang digunakan	Tidak disebutkan	Data tidak disebutkan	ETL	Kualitatif, kuantitatif	Data Warehouse	Mengimplementasikan berbagai metode terbaik yang ada di ETL	semua ketergantungan pada kualitas data dan pemeriksaan data yang diperlukan akan berkurang
Mengamankan Data Sensitif Proses Transform ETL Menggunakan Data masking, 2023	Akses RAW data dan data di Database Oleh Orang diluar organisasi	Kebocoran data, penyalahgunaan data, privasi tidak terjaga	Kaggle	100 baris, 3 kolom	Scrambling, Substitution, Shuffling, Date Aging, Variance Applies, Masking Out, Nullifying	Kualitatif, Kuantitatif	Hadoop, HDFS, Spark, Python, Database	Menyediakan Database terpisah yang mana datanya sudah di*masking*	Pengaburan data

Usulan penelitian, 2023, dengan menitik beratkan pada efisiensi proses *data masking*, diusulkan untuk melakukan penelitian Teknik *data masking*, terutama pada proses transformasi di *ETL*. Hal ini diusulkan, agar penyimpanan data yang sudah di*masking* lebih sederhana, dibandingkan dengan Teknik *data masking* yang dilakukan di level *database*, terutama tabel.

Teknik *data masking* di *database*, akan membuat tabel baru, khusus untuk menyimpan data hasil proses *masking*, sehingga bisa dikatakan akan ada dua tabel berbeda, antara tabel yang menyimpan data asli, dengan tabel yang menyimpan data hasil *masking*. Data dan kebutuhan storage menjadi tidak ringkas, terutama jika data yang ada berukuran cukup besar.

Begitu juga untuk *raw data* yang tetap tersimpan di file sistem, data tersebut akan aman dengan format yang sudah di-*masking*. Karena data hasil proses transormasi di *ETL*, selain di load (masukan) ke dalam *database*, biasanya akan ditinggalkan di file sistem. Formula atau logika penggunaan dari implementasi *data masking* hanya diketahui oleh beberapa orang saja, diantaranya yaitu mereka yang ada di tim development dan juga pemilik aplikasi. Sehingga jauh lebih sedikit, dibandingkan dengan data yang terekspos tanpa di*masking*, dalam arti semua yang memiliki akses ke dalam sistem operasi dan dapat mengakses file sistemnya, dapat membaca semua data apa adanya.