# BAB I PENDAHULUAN

## 1.1 Latar Belakang Masalah

*Extract, Transform, Loading* atau biasa disingkat dengan *ETL*, sebagaimana kita tahu merupakan inti dari pemrosesan data *warehouse* atau saat ini digunakan juga pada pengolahan *Big Data*. *ETL* menjadi solusi pengolahan data yang sangat popular dan banyak diimplementasikan di banyak perusahaan.

Dalam melakukan analisis data, tahap transformasi data pada *ETL* merupakan suatu proses yang penting untuk mengubah data menjadi format yang lebih tepat dan mudah diproses. Namun, proses transformasi data juga dapat mengungkapkan informasi sensitif yang tidak diinginkan, seperti data pribadi atau informasi rahasia lainnya. Hal ini dapat menjadi masalah yang serius, terutama jika data yang terungkap tersebut jatuh ke tangan yang salah.

Untuk mengatasi masalah tersebut, teknik *data masking* dapat digunakan untuk menyembunyikan informasi sensitif yang tidak diinginkan dalam proses transformasi data. Teknik *masking* data menggunakan nilai acak atau penggantian karakter untuk mengaburkan data sensitif tanpa mengubah informasi penting lainnya. Dengan demikian, teknik *masking* data dapat membantu menjaga kerahasiaan data saat proses transformasi dilakukan.

Namun, teknik *masking* data juga dapat mempengaruhi analisis data karena informasi sensitif yang disembunyikan mungkin mempengaruhi hasil akhir analisis. Oleh karena itu, teknik *masking* data harus dilakukan secara hati-hati dan dipilih dengan baik sehingga dapat menjaga kerahasiaan data tanpa mempengaruhi hasil analisis yang dihasilkan.

Berdasarkan latar belakang masalah di atas, penelitian "Mengamankan Data Sensitif Proses Transform *ETL* Menggunakan *Data masking*" bertujuan untuk memperkenalkan teknik *masking* yang efektif dan aman dalam proses transformasi data serta memastikan bahwa hasil analisis yang dihasilkan tetap akurat dan dapat dipercaya.

## 1.2 Rumusan Masalah

Pada dasarnya, *data masking* memiliki dua buah tipe cara pemrosesannya, yaitu *static data masking* dan *dynamic data masking.* Biasanya proses *data masking* dilakukan untuk data yang sudah tersimpan di tabel pada suatu *database*. Hal ini kurang efisien, karena bagaimanapun data sebelum masuk ke dalam *database*, akan tetap disimpan di *file sistem* (*staging*) dan bisa dibuka sembarang oleh mereka yang memiliki akses. Hal ini sangat berbahaya, apabila terdapat data-data sensitif atau data pribadi seseorang. Sehingga dari beberapa point tersebut, dapat dirumuskan masalahnya, sebagai berikut:

1. Bagaimana teknik *data masking* dapat digunakan untuk menjaga kerahasiaan data sensitif pada saat proses transformasi data?
2. Bagaimana efektivitas teknik *data masking* dalam menjaga kerahasiaan data sensitif pada saat proses transformasi data dilakukan?
3. Bagaimana cara mengukur dan mengevaluasi hasil proses *data masking* dalam menjaga kerahasiaan data-data?

## 1.3 Batasan Masalah

Dalam penelitian ini, memberikan batasan terhadap masalah dalam implementasi penggunaan teknik *data masking* dalam perlindungan data sensitif. *Data masking* adalah metode yang digunakan untuk mengamankan data sensitif dengan menggantinya dengan data palsu atau terenkripsi, sehingga mengurangi risiko pelanggaran keamanan dan privasi. Namun, dalam mengimplementasikan teknik ini, ada beberapa aspek yang perlu dipertimbangkan dan dibatasi.

Dengan membatasi masalah ini secara jelas, penelitian ini akan memberikan pemahaman yang lebih mendalam tentang efektivitas dan tantangan yang terkait dengan penggunaan *data masking* dalam konteks keamanan data. Sehingga, batasan masalah dari penelitian ini adalah

1. Penelitian berfokus pada desain workflow untuk mengimplementasikan *data masking* menggunakan HGrid247 DE.
2. Penelitian berfokus pada proses transformasi data, menggunakan fungsi *data masking* dan menjalankannya di atas framework *Apache Spark*.
3. Penelitian tidak menghasilkan package atau library untuk *Apache Spark*.
4. Penelitian tidak dibatasi hanya untuk pengolahan data *Stream Proccesing*

## 1.4 Tujuan

Tujuan utama dari penelitian ini adalah untuk mengidentifikasi dan menerapkan teknik *data masking* yang dapat digunakan untuk menjaga kerahasiaan data sensitif pada saat proses transformasi data dilakukan. Diharapkan teknik *masking* data ini dapat membantu pengguna *database* dan manajer *data warehouse* dalam mengelola data sensitif dengan aman dan efektif.

Selain itu, tujuan penelitian adalah untuk memberikan rekomendasi alat, juga memberikan alternatif dalam memilih teknik *data masking* yang tepat untuk menjaga kerahasiaan data-data sensitif pada saat proses transformasi data. Diharapkan rekomendasi ini dapat membantu pengguna dalam mengambil keputusan yang tepat, terutama implementasi *data masking* yang paling sesuai untuk digunakan dalam konteks penggunaan data mereka.

Dengan mencapai tujuan-tujuan tersebut, diharapkan penelitian ini dapat memberikan kontribusi yang signifikan dalam pengembangan teknik *data masking* yang dapat digunakan untuk menjaga kerahasiaan data sensitif pada proses transformasi data.

### 1.5 Manfaat

Adapun beberapa manfaat yang akan didapat dari penelitian "Mengamankan Data Sensitif Proses Transform *ETL* Menggunakan *Data masking*" yaitu, dengan menggunakan teknik *data masking* yang tepat, pengguna dapat memastikan bahwa data sensitif tidak terungkap. Dengan demikian, data sensitif dapat dijaga kerahasiaannya dan tidak digunakan oleh pihak yang tidak berwenang.

Meningkatkan efisiensi pengelolaan data, dalam konteks penggunaan data, efisiensi adalah hal yang sangat penting. Dengan menggunakan teknik *data masking*, pengguna *database* dapat mengelola data sensitif dengan aman dan efektif, sehingga menghemat waktu dan sumber daya yang diperlukan untuk melindungi data sensitif.

Memberikan solusi yang inovatif, dalam hal ini, teknik *data masking* yang diusulkan dapat memberikan solusi yang inovatif untuk masalah keamanan data

sensitif dalam konteks transformasi data.

Dengan demikian, penelitian ini memiliki manfaat yang penting bagi pengguna dalam pengelolaan data sensitif secara aman dan efektif. Selain itu, penelitian ini juga dapat memberikan kontribusi yang signifikan dalam pengembangan teknologi keamanan data yang lebih canggih dan inovatif.