BAB I PENDAHULUAN

I.1 Latar Belakang

Big Data perlu sebuah infrastruktur untuk menjalankan sistemnya beserta salah satu *framework* pengolahan *big data* yaitu Hadoop. Hadoop adalah sebuah *software framework* untuk komputasi yang *reliable*, *scalable*, *parallel*, dan komputasi terdistribusi (Ghazi & Gangodkar, 2015). Hadoop memiliki fungsi untuk memproses, mengatur serta menganalisis dari berbagai macam tipe data misalnya *structured*, *unstructured* dan *semi-structured* (Yadav et al., 2019). *Volume, velocity* dan *variety* adalah karakteristik umum yang dimiliki *big data* dan biasa disebut dengan 3 *V's of data*. Namun hal-hal tersebut masih kurang lengkap untuk dapat menjelaskan mengenai *big data* pada saat ini. Ada beberapa faktor lain seperti *veracity*, *validity*, *value*, *variability*, *venue*, *vocabulary*, and *vagueness* yang dapat menjelaskan *big data* secara lebih lengkap (Yadav et al., 2019). Dengan perkembangan teknologi informasi saat ini, muncul berbagai jenis faktor V lainnya, salah satu faktor V yang perlu diperhatikan saat ini pada *big data*, yaitu *vulnerabilities* (Bhathal & Singh, 2019).

Vulnerabilities pada infrastruktur Hadoop tidak terlalu diperhatikan pada awalnya. Namun ketika Hadoop memiliki *issue* mengenai keamanan yang tidak menjadi prioritas dalam menghadapi distribusi data dan proses data secara paralel maka dari itu masalah *vulnerabilities* semakin diperhatikan (Abouelmehdi et al., 2017). Pada awalnya, dalam perkembangan *big data*, *developer* tidak mempertimbangkan isu yang muncul dari segi keamanan dan privasi. Namun, pada saat ini, masalah utama *big data* yang paling menantang yaitu mengenai keamanan dan privasi. Data adalah aset yang paling penting, masalah keamanan data merupakan masalah yang besar. Mengatasi keamanan *big data* adalah tugas yang sangat sulit. Keamanan pada big data mencakup dalam melindungi data dari akses yang tidak sah (*unauthorized access*), modifikasi, manipulasi, penghancuran dan memastikan kerahasiaan (*confidentiality*), ketersediaan, dan integritas pada data (Yadav et al., 2019).

Hadoop memang memiliki masalah keamanan yang melekat karena menggunakan arsitektur terdistribusi, Hal ini tidak mungkin terjadi pada instalasi Hadoop yang dikelola dengan aman. Instalasi Hadoop yang memiliki peran pengguna yang jelas dan berbagai tingkat otentikasi dan enkripsi untuk *data confidentiality* tidak akan

membiarkan akses yang kepada pengguna yang tidak berwenang. Pada awalnya, Hadoop dimaksudkan untuk memproses sejumlah besar *data web* di domain publik, dan karenanya keamanan bukanlah fokus pengembangan. Itu sebabnya Hadoop tidak memiliki model keamanan yang baik dan hanya menyediakan otentikasi dasar untuk HDFS, yang sangat tidak baik, karena sangat mudah untuk meniru pengguna lain (Perwej, 2021).

Dengan sistem yang terdistribusi, Hadoop memiliki data yang tersebar di sejumlah besar *host* dan disimpan secara lokal. Terjadi sejumlah besar komunikasi data antara *host-host* ini, oleh karena itu data menjadi rentan pada saat berpindah (*data-in-transit*) maupun pada saat diam dan disimpan di penyimpanan lokal (*data at rest*). Hadoop dimulai sebagai penyimpanan data untuk mengumpulkan data penggunaan web serta bentuk lain dari data besar yang tidak rahasia (*non confidential data*). Hal tersebut menyebabkan Hadoop tidak memiliki ketentuan bawaan untuk mengenkripsi data. Saat ini, keadaan berubah dan Hadoop semakin banyak digunakan untuk menyimpan data rahasia di dunia bisnis. Hal ini telah menciptakan kebutuhan akan data untuk dienkripsi pada saat berpindah (*data-in-transit encryption*) dan saat istirahat (*data at rest encryption*) (Perwej, 2021).

Maka dari itu untuk mendapatkan informasi mengenai kerentanan sistem pada Hadoop yang dapat menjadi acuan untuk membangun sistem Hadoop yang memiliki keamanan data dan privasi yang lebih baik, perlu dilakukan *vulnerability assessment* menggunakan *tool* GSM Trial pada infrastruktur Hadoop sehingga akan menghasilkan informasi mengenai *vulnerability* yang ada. Dari hasil yang ditemukan, dapat dilakukan *hardening* dengan metode *data encryption, encryption* pada saat berpindah (*data-in-transit*) dan pada saat data disimpan (*data at rest*) dapat menambah aspek *confidentiality* untuk menjaga kerahasiaan data, sehingga tidak ada kebocoran data penting dan hanya orang yang berwenang yang dapat mengakses data tersebut. Hal ini tentu meningkatkan keamanan data pada infrastruktur Hadoop serta dapat membantu mengurangi kerentanan yang ditemukan pada infrastruktur Hadoop.

I.2 Rumusan Masalah

Isi berdasarkan latar belakang diatas, permasalahan yang dihadapi dirumuskan sebagai berikut:

1. Bagaimana menemukan kerentanan pada Infrastruktur Hadoop?
2. Bagaimana mengendalikan kerentanan yang ditemukan pada Infrastruktur Hadoop dengan menggunakan fungsi enkripsi?

I.3 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui *vulnerability assessment tools* yang digunakan untuk mencari kerentanan dalam HDFS Infrastruktur Hadoop.
2. Mengetahui implementasi fungsi enkripsi berupa *confidentiality* untuk mengatasi kerentanan yang ada pada infrastruktur Hadoop.

I.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Secara teoritis, dapat menambah ilmu pengetahuan terkait Big data, Hadoop, *confidentiality*, *vulnerability assessment*, konsep *Hardening* menggunakan metode *encryption*.
2. Secara praktis, penelitian ini dapat menambah pengetahuan mengenai pemahaman cara penggunaan *tool vulnerability assessment* yang dapat digunakan untuk mencari kerentanan pada konfigurasi infrastruktur Hadoop dan cara implementasi *hardening* dengan metode *encryption* untuk *confidentiality data*.

I.5 Batasan Penelitian

Batasan penelitian ini adalah sebagai berikut:

1. Penelitian ini membahas kerentanan pada HDFS dan tidak pada *core service* lain.
2. Penelitian ini tidak membahas mengenai *hardening* pada aspek *authentication*.
3. Penelitian ini dilakukan untuk mengimplementasikan fitur keamanan pada aspek *confidentiality*, tapi sejauh mana *confidentiality* nya kuat tidak dibahas.
4. Penelitian ini tidak membahas implementasi *tool vulnerability assessment* lain selain GSM Trial.
5. Penelitian ini tidak membahas implementasi *tool hardening* lain selain eCryptfs.

I.6 Sistematika Penulisan

Berikut sistematika penulisan yang digunakan untuk menyusun tugas akhir ini. Pada Bab I menjelaskan tentang latar belakang yang berisi masalah dan pemecahan masalah yang mengacu pada studi literatur. Kemudian menjelaskan tujuan dari penelitian yang dilakukan. Setelah itu terdapat manfaat yang bisa didapatkan dari penelitian. Kemudian melakukan perumusan masalah dan membuat batasan masalah dalam penelitian yang dilakukan.

Pada Bab II berisi uraian mengenai penelitian terdahulu dan teori dasar yang digunakan untuk mendukung penelitian ini. Penelitian terdahulu yang dicantumkan dalam Bab II yaitu penelitian yang berkaitan dengan topik penelitian tugas akhir ini. Kemudian ada sejumlah teori dasar yang juga berkaitan dengan topik penelitian dan dapat mendukung teknik pengerjaan penelitian ini.

Pada Bab III berisi uraian mengenai metode penelitian serta penjelasan langkah-langkah secara rinci yang digunakan untuk menyelesaikan permasalahan dengan model konseptual serta sistematika pemecahan masalah.

Pada Bab IV berisi uraian mengenai rincian perancangan sistem yang akan dilakukan dan implementasi eksperimen penelitian dalam melakukan proses *vulnerability scanning* menggunakan *tool* GSM dan *hardening* menggunakan metode *data encryption.*

Pada Bab V berisi uraian mengenai analisis hasil eksperimen penelitian yang dilakukan sesuai dengan alur pengerjaan, serta analisis dari data hasil eksperimen penelitian.

Pada Bab VI berisi uraian mengenai kesimpulan dari hasil penelitian dalam perancangan, pengujian sistem, dan analisis yang dilakukan, serta saran yang bermanfaat untuk penelitian selanjutnya.