

BAB I PENDAHULUAN

1. PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi telah mendorong digitalisasi pada berbagai sektor, seperti *e-commerce*, *smart city*, *Internet of Things* (IoT), dan sistem berbasis komputasi awan. Dalam mendukung kebutuhan berbagai data secara cepat dan real-time, sistem komputasi awan terdistribusi menjadi solusi utama dalam pengelolaan data berskala besar yang tersebar lintas lokasi dan node. Namun, sistem ini masih menghadapi tantangan dalam melakukan sinkronisasi data secara efisien, terutama dalam menjaga konsistensi antar node, menurunkan latensi, serta menghemat bandwidth komunikasi (Ali, dkk, 2020).

Permasalahan infrastruktur jaringan di Indonesia menjadi hambatan utama dalam penerapan sistem komputasi awan terdistribusi. Stabilitas internet yang belum merata serta biaya bandwidth yang cukup tinggi menyebabkan proses sinkronisasi data antar node menjadi tidak efisien. Ketidakefisienan ini berdampak langsung pada keterlambatan pembaruan informasi yang sangat dibutuhkan dalam sistem layanan real-time berbasis *cloud* (The World Bank, 2021).

Blockchain hadir sebagai solusi teknologi *ledger* terdistribusi yang mampu menjamin integritas dan keamanan data. Teknologi ini memungkinkan pencatatan data yang tidak dapat diubah dan transparan antar *node* dalam jaringan. Namun demikian, *blockchain* memiliki keterbatasan, seperti kapasitas penyimpanan yang terbatas dan latensi validasi yang relatif tinggi. Hal ini menjadi kendala ketika sistem *blockchain* diterapkan pada proses sinkronisasi metadata dalam skala besar (Benet, 2014).

Teknologi *blockchain* dapat dikombinasikan dengan *InterPlanetary File System* (IPFS) untuk mengatasi keterbatasan sistem dalam hal penyimpanan data terdistribusi. *IPFS* memungkinkan penyimpanan data secara efisien dengan sistem terdistribusi. Setiap data yang disimpan akan menghasilkan *Content Identifier* (CID) yang bersifat unik dan permanen. *CID* ini memungkinkan proses sinkronisasi

yang selektif, yaitu hanya data yang berubah yang perlu didistribusikan ulang, sehingga menghemat bandwidth antar node (Ali, dkk, 2020).

Proses pengelompokan data dilakukan melalui metode klasterisasi berbasis teks menggunakan pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma *K-Means*. Metode ini menghasilkan metadata klaster yang lebih terstruktur, ringan, dan mudah disinkronkan dalam sistem terdistribusi. Kombinasi ini menjadikan proses sinkronisasi data tidak hanya efisien, tetapi juga konsisten dan toleran terhadap gangguan koneksi jaringan (Zhang, dkk, 2021).

Penelitian ini bertujuan untuk mengembangkan *Clustered Data Synchronization Framework* (CDSF), yaitu *framework* sinkronisasi data klaster yang mengintegrasikan metode klasterisasi, penyimpanan metadata melalui *InterPlanetary File System* (IPFS), serta pencatatan metadata pada jaringan blockchain menggunakan algoritma konsensus *Proof-of-Authority* (PoA). *Framework CDSF* dikembangkan untuk mendukung proses sinkronisasi data klaster secara efisien, selektif, dan toleran terhadap gangguan dalam lingkungan komputasi awan terdistribusi.

1.2. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan dan mengimplementasikan algoritma konsensus *Proof-of-Authority* (PoA) *blockchain* pada sistem sinkronisasi data klaster yang terdistribusi.

Tujuan utama dari penelitian ini adalah untuk membangun sistem *Clustered Data Synchronization Framework* (CDSF) yang mampu menjalankan proses sinkronisasi data klaster secara efisien, konsisten, dan toleran terhadap kegagalan dalam lingkungan komputasi awan terdistribusi. Penelitian ini mengintegrasikan 3 komponen utama dalam satu alur arsitektur sistem yang saling mendukung, yaitu:

- a. *Clustering Engine*, yang bertugas melakukan pengelompokan data teks menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma *K-Means* untuk menghasilkan *metadata* klaster yang terstruktur.

- b. *Data Storage*, yaitu penyimpanan *metadata* secara terdistribusi menggunakan *InterPlanetary File System* (IPFS) untuk menyimpan hasil klusterisasi dan menghasilkan *Content Identifier* (CID) yang unik.
- c. *Blockchain Engine*, yang menggunakan algoritma konsensus *PoA* dengan protokol *QBFT* untuk mencatat *CID metadata* ke dalam *smart contract*, menyebarkan informasi *metadata* tersebut ke seluruh *node* dalam jaringan, serta menjaga konsistensi data melalui mekanisme toleransi terhadap kegagalan *node validator*.

1.3. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem sinkronisasi *metadata* yang efisien pada lingkungan komputasi awan terdistribusi. Adapun manfaat utama dari penelitian ini antara lain:

1. Menghasilkan *framework* sinkronisasi data kluster (CDSF) yang menggabungkan proses klusterisasi teks, penyimpanan terdistribusi, dan pencatatan *metadata* melalui *blockchain*.
2. Menyediakan mekanisme sinkronisasi selektif berbasis *CID* untuk mengurangi duplikasi dan konsumsi *bandwidth* antar *node*.
3. Meningkatkan efisiensi dan akurasi dalam pengelompokan data berbasis teks menggunakan metode *TF-IDF* dan *K-Means*.
4. Menjamin integritas dan konsistensi *metadata* antar *node* dengan pencatatan pada *blockchain PoA* yang tahan terhadap kegagalan *node*.
5. Menjadi rujukan bagi pengembangan sistem federasi data, *edge computing*, atau integrasi data dalam sistem *cloud* yang membutuhkan efisiensi dan ketahanan tinggi.