

ABSTRACT

Monumen Pers Nasional is one of the government agencies who utilizes an IT to manage and provide an information from their national press collection to a public as a service. Based on experience and evaluation, the technology infrastructure still has limitations that could affect the availability of the data that have been stored.

Based on the literature studies, Hadoop technology especially HDFS (Hadoop Distributed File System) can be a solution because it has the ability to overcome these limitations. Moreover Hadoop can be developed as a data warehouse that fits the needs of Monumen Pers Nasional in the future. The purpose of this study is to analyze the level of HDFS Datanode availability on Hadoop features before it is being implemented in real condition.

The results showed that the level of availability on Datanode from HDFS features is affected by the replication used. The larger replication factor that is used, the higher degree of Datanode fault-tolerance that could be handled. The amount of fault-tolerant that can be guaranteed by HDFS in order to obtain 100% availability of the data stored therein is the replication factor subtracted by 1.

Keywords : Hadoop, HDFS, Datanode, fault tolerance

INTISARI

Monumen Pers Nasional sebagai salah satu instansi pemerintah, memanfaatkan teknologi informasi untuk membantu tugas dan fungsinya dalam mengelola dan penyediaan informasi dari seluruh media cetak pers nasional kepada masyarakat. Berdasarkan pengalaman dan evaluasi, infrastruktur teknologi yang digunakan masih memiliki beberapa keterbatasan sehingga ditengarai dapat mempengaruhi jaminan ketersediaan data yang disimpan.

Berdasarkan studi literatur yang telah dilakukan, teknologi Hadoop terutama fitur HDFS (*Hadoop Distributed File System*) dapat menjadi solusi karena memiliki kemampuan yang dapat mengatasi keterbatasan tersebut. Terlebih lagi Hadoop juga dapat dikembangkan sebagai *data warehouse* yang sesuai dengan kebutuhan Monumen Pers Nasional di masa depan. Penelitian ini bertujuan untuk menganalisis tingkat ketersediaan Datanode di fitur HDFS yang dimiliki oleh Hadoop sebelum diimplementasikan secara nyata.

Hasil penelitian menunjukkan bahwa tingkat ketersediaan yang dimiliki oleh Datanode pada fitur HDFS dipengaruhi oleh faktor replikasi yang digunakan. Semakin besar faktor replikasi yang digunakan berbanding lurus dengan tingkat toleransi kegagalan Datanode yang dapat ditangani. Jumlah toleransi kegagalan yang dapat dijamin oleh HDFS agar didapat tingkat ketersediaan 100% terhadap data yang disimpan didalamnya adalah faktor replikasi kurang 1.

Kata kunci – Hadoop, HDFS, Datanode, toleransi kegagalan.