

INTISARI

BIG DATA BENCHMARK PADA HADOOP 2, SPARK, DAN PRESTO MENGUNAKAN METODE PERBANDINGAN WAKTU RESPON QUERY

Daniel Oscar Baskoro
10/300237/PA/13187

Seiring dengan peningkatan informasi yang berbentuk digital (data digital), kebutuhan akan analisa big data menjadi hal prioritas bagi masyarakat luas khususnya pada sektor swasta dan pemerintahan. Kebutuhan tersebut menjadi prioritas karena analisa pada Big Data dapat menghasilkan suatu informasi yang dapat mempermudah para pembuat kebijakan dalam menentukan suatu kebijakan. Tingginya permintaan masyarakat dalam analisa Big data, memunculkan banyak penelitian yang menciptakan berbagai macam *framework* untuk menganalisa suatu Big Data.

Pada penelitian dan pengujian ini dilakukan suatu analisa performa beberapa *framework* yang digunakan dalam menganalisa suatu data. *framework* yang diteliti adalah Spark, Hadoop 2, dan Presto. Metode yang dipakai dalam pengujian adalah evaluasi waktu yang dibutuhkan *framework* dalam melakukan query pada parameter-parameter tertentu. Parameter dalam pengujian adalah waktu yang digunakan dalam query tingkat rendah, query tingkat menengah, query tingkat tinggi, query dengan peningkatan 2 kali jumlah *core* prosessor, dan terakhir adalah query dengan peningkatan 4 kali jumlah *core* prosessor. Penelitian ini menggunakan contoh kasus nyata dalam analisa big data, data yang diuji berada di cloud storage Amazon S3, infrastruktur analisa yang dipakai menggunakan cloud engine pada Qubole, data testing yang dianalisa adalah data Twitter yang terdiri dari 100.000 dan 3.000.000 status twitter.

Hasil akhir merupakan suatu informasi performa Spark, Hadoop, dan Presto dalam melakukan proses query. Dari informasi tersebut diketahui diantara *framework* tersebut mana yang memiliki performa terbaik sehingga dapat sebagai pertimbangan dalam pemilihan *framework* untuk menganalisa data pada berbagai kebutuhan.

Kata kunci : Big Data, Big Data *framework*, Query, High Performance Query, Cloud Storage, Cloud Engine, Komputasi, Twitter

ABSTRACT

BIG DATA BENCHMARK ON Hadoop 2, SPARK, AND PRESTO BASED COMPARISON ON QUERY RESPONSE TIME

Daniel Oscar Baskoro
10/300237/PA/13187

Along with the improvement of the information in digital form (digital data), the need for analysis of big data becomes a priority for the society, especially the private sector and government. These needs to be a priority for the Big Data analysis can provide information that can facilitate policy-makers in determining a policy. High demand in the analysis of Big Data, raises a lot of research that creates a wide variety of *framework* to analyze Big Data.

In this research and testing carried out an analysis of the performance of several *framework* that are used in analyzing the data. *framework* that testing were Spark, Hadoop 2, and Presto. The method used in testing is the evaluation of the time required to perform a high performance *framework* queries on certain parameters. The parameters in the test is the time spent in the low-Model query, Query middle Model, high-Model query, Query with an increase of 2 times the number of processors, and the last is a query with an increase of 4 times the number of processors. This research uses real case examples of big data analysis, data were tested in the Amazon S3 cloud storage, infrastructure analysis using cloud engine used in Qubole, testing the data analyzed is data of 1,000,000 and 3,000,000 Twitter twitter status.

The end result is a performance information Spark, Hadoop 2, and Presto in high performance perform queries. From the information is known among the *framework* which has the best performance that can be a consideration in the *selection* of the *framework* to analyze the data in a variety of needs.

Keywords: Big Data, Big Data Framework, Query, High Performance Query, Cloud Storage, Cloud Engine, Computing, Twitter