

PENGENALAN TEKNOLOGI BIG DATA

Tasrif Hasanuddin

Program Studi Sistem Informasi

STMIK Profesional Makassar

tasrif.h@gmail.com

Abstrack

Istilah BIG Data telah menjadi Topik yang dominan dan sangat sering dibahas dalam dunia IT. Perkembangan volume dan jenis data yang terus meningkat secara berlipat-lipat dalam dunia maya semenjak kelahirannya adalah fakta yang tidak dapat di pungkiri. Mulai dari data berupa teks, gambar video atau foto. Google sebagai pelopor dalam implementasi Big Data seperti Google File System, Google Map Reduce dan Google Big Table, dan peluang ini tidak disiasiakn oleh Apache Software foundation dengan spesifikasi teknik google, kemudian mengembangkan Apache Hadoop, HDFS, Map Reduce Apache Hbase dalam bentuk sebuah framework yang bersifat open source, sehingga dapat dikembangkan oleh siapa saja, baik sebagai individu maupun dalam sebuah perusahaan besar.

Kata Kunci : Big Data, Google

A. PENDAHULUAN

Keberhasilan Google, Facebook dalam mengelola data tidak terstruktur, melahirkan sebuah konsep yang dikenal BIG Data. Sementara itu beberapa organisasi besar di dunia mengalami kesulitan akibat semakin massivenya data yang mereka hasilkan dari proses transaksi bisnis yang mereka lakukan, akibatnya analisis data semakin kompleks, dan pengolah data serta aplikasi yang dibuat oleh beberapa programer sudah tidak mampu lagi untuk memprosesnya.

Hal semacamnya inilah yang disebut dengan *BIG Data*, yang diartikan sebagai sekumpulan data dalam jumlah yang sangat besar, dimana yang menjadi

permasalahan bagaimana data tersebut disimpan, bagaimana melakukan pencarian, bagaimana membagi-bagikan dan akhirnya bagaimana melakan analisa terhadap data tersebut.

Sayangnya perusahaan seperti Google dan lainnya belum membuat sebuah framework yang dapat digunakan secara langsung untuk mengimplementasikan Teknologi *BIG Data* tersebut.

Apache Software foundation (AS F) kemudian menciptakan Apache Hadoop, sebuah software framework yang dikembangkan melalui proyek *open source*. Apache Hadoop utamanya terdiri dari Hadoop Distributed File System

(HDFS) dan framework Distributed Parallel Processing System yang disebut *Hadoop Map Reduce*. HDFS berfungsi sebagai penyimpanan data secara terdistribusi, sedangkan Hadoop Map Reduce berfungsi sebagai framework pemrosesan data secara parallel dan terdistribusi.

B. PEMBAHASAN

Google dan Big Data

Sejak awal, Google dibekali dengan teknologi yang disebut PageRank yang telah menjadikan *search engine* yang unik karena teknologi tersebut belum pernah dimiliki oleh search engine yang lain. Bahkan Google juga telah mampu mengelola pertumbuhan data yang membludak. Google telah memiliki teknologi *crawler* yang mampu mendownload seluruh web page yang ada di dunia maya, dan secara terus menerus mengikuti update pada jutaan web tersebut, google telah memiliki teknologi sistem penyimpanan data yang dikenal dengan nama *Google File System* (GFS), yang menyimpan data secara terdistribusi dalam ribuan computer. Kapasitas penyimpanan GFS ini bisa terus menerus diperbesar hanya dengan menambah jumlah komputer. Google memiliki teknologi pengolahan data yang disebut MapReduce yang mampu mengolah data yang disimpan

dalam GFS secara parallel dan idenpenden hingga ribuan computer, sehingga mampu mengolah data ukuran raksasa dalam waktu yang lebih cepat dibanding model konvensional. Google juga memiliki system database yang disebut BigTable, teknologi yang mampu menyajikan data berukuran raksasa yang sudah tidak memungkinkan lagi ditanganani dengan database konvensional

Google File System

Google File System (GFS) adalah file system yang diterapkan pada system terdistribusi . GFS merupakan suatu format file system yang dapat diibaratkan layaknya FAT32 ataupun NTFS. GFS digunakan untuk memformat media penyimpanan yang berupa system terdistribusi yang dari dari computer-komputer yang tergabung dalam suatu *cluster*. Cluster adalah network yang terdiri dari puluhan bahkan ribuan komputer. Berkat kemampuan GFS dapat menampung data super besar yang tidak bisa disimpan dalam HDD paling besar sekalipun.

GFS memiliki tiga unsur utama, yaitu Master Server, Chunkserver dan Client. Master servers adalah server sentral yang mengendalikan dan memonitor kondisi GFS secara keseluruhan. Chunkserver adalah unsur yang bertanggung jawab menyimpan / membaca

file dari/ke Hdd tiap node. Clients adalah aplikasi yang menggunakan GFS untuk menyimpan dan membaca file dari/ke GFS.

Google MapReduce

Map Reduce adalah model rilisan Google untuk memproses data berukuran besar secara terdistribusi dan parallel dalam cluster yang terdiri dari ribuan komputer. MapReduce dapat dibagi dalam dua proses yaitu proses Map dan proses Reduce. Kedua jenis proses ini di distribusikan ke setiap computer yang berada dalam cluster yang bersangkutan dan berjalan secara parallel tanpa saling bergantung satu dengan yang lainnya. Proses Map bertugas untuk mengumpulkan informasi dari potongan-potongan data yang terdistribusi dalam tiap computer yang membentuk cluster dimana MapReduce dijalankan. Hasilnya diserahkan kepada proses Reduce untuk diproses lebih lanjut sesuai intensitas client, Hasil akhir dari MapReduce yang akan dikirim ke client-nya.

Google Bigtable

Google menciptakan Bigtable untuk memproses data berskala besar yang harus dilakukan secara interaktif . Google Bigtable adalah system penyimpanan data terdistribusi yang ditujukan untuk mengelola data berstruktur dan didesain sebagai system yang handal untuk

memproses dalam dalam skala petabytes dan dalam ribuan mesin. Kita bisa membayangkan bahwa Bigtable memang table yang sebenarnya. Table yang berisi data dalam skala petabytes. Data tersebut disimpan secara terdistribusi dalam ribuan computer yang dapat diakses dan dikelola melalui Bigtable. Bigtable dinyatakan sebagai system yang sangat fleksibel. Bigtable mampu memproses data mulai data yang berukuran kecil seperti halnya URL. Sampai data yang berukuran besar seperti citra satelit. Bigtable bisa menangani pemrosesan data mentah yang berorientasi pada hasil akhir dan perlu diproses dalam jangka waktu tertentu, maupun pemrosesan data secara realtime yang menuntut hasil seketika seperti. Google Web Indexing , Google Analytics, Google Finance, Google Earth dan lain-lain.

Apache Hadoop

Kelahiran Apache Hadoop disebabkan Google dengan teknologi Big Datanya telah mempublikasikan konsep dan spesifikasi teknis namun sayangnya Google tidak pernah mempublikasikan teknologi Big Data ciptaanya dalam bentuk framework.

Apache Software Foundation melihat peluang besar ini pada fenomena Big Data, telah berhasil mengimplementasikan konsep dan

spesifikasi teknis dari GFS, Map Reduce dan Big Table ke dalam bentuk framework yang dikembangkan dalam bahasa Java. Kemudian pengembangan tersebut dikelola dalam bentuk proyek open source sehingga dapat diakses dan digunakan oleh semua orang entah sebagai individu maupun organisasi . Apache mengembangkan system terdistribusi yang diberi nama Hadoop yang juga dikenal sebagai Apache Hadoop.

HDFS

HDFS (Hadoop Distributed File System) sesuai dengan namanya, adalah file terdistribusi pada Hadoop. Yang dimaksud dengan system file terdistribusi adalah file system yang menyimpan data tidak dalam satu hard disk drive (HDD) atau media penyimpanan lainnya, tetapi data dipecah-pecah dan disimpan tersebar dalam suatu *cluster* yang terdiri atas sejumlah computer, bisa hanya puluhan, ratusan atau bahkan ribuan computer.

Jika disederhanakan, HDFS dapat dianggap sebagai media penyimpanan pada cluster Hadoop. HDFS didesain dan dikembangkan berdasarkan konsep dan spesifikasi teknis dari GFS, yang telah dipublikasikan oleh Google.

Hadoop MapReduce

Hadoop MapReduce dikembangkan berdasarkan konsep dan spesifikasi teknis Google MapReduce yang

diimplementasikan dengan menggunakan bahasa pemrograman Java. Olehnya itu Hadoop MapReduce memiliki programming model yang identic dan Googl MapReduce . Demikian juga dengan arsitekturnya yang ukuivalen dengan Google MapReduce . Perbedaannya adalah Google MapReduce dijalankan dalam cluster GFS, sedangkan Hadoop MapReduce dioperasikan dalam cluster Hadoop yang sudah tentunya sudah dalam format HDFS.

Apache HBase

Seperti halnya Google Map Reduce, Hadoop Map Reduce terbukti unggul dalam memproses data berkurun raksasa yang sifatnya *batch prosessing*, tetapi memiliki kelemahan dalam memproses data yang sifatnya interactive, ASF kemudian mengembangkan Apache Hbase berdasarkan konsep dan spesifikasi teknis dari Google Big Table yang diimplementasikan dalam bahaa pemrograman Java . Dengan HBase Big Data dapat diproses secara real time di dalam HDFS.

Apache HBase adalah database Hadoop , yaitu database terdistribusi yang berorientasi pada kolom, bersifat scalable , dan ditujukan untuk menyimpan Big Data. Oleh ASF Apache HBase dikembangkan dan dipublikasikan dalam proyek open source.

C. KESIMPULAN

Berdasarkan pemaparan sebelumnya, maka dapat disimpulkan beberapa hal sebagai berikut :

- a. Google sebagai peletak dasar dari teknologi BIG Data, dan mampu mengimplementasikan dalam Bentuk , Google File System, Google Map Reduce, dan Google Big Table.
- b. Kesuksesan Google dalam mengimplementasikan Big Data, diikuti oleh Apache Software Foundation yang mengembangkan dalam sebuah framework yang bersifat open source, sehingga sangat mudah untuk digunakan oleh siapa saja, baik individu, maupun organisasi besar.

DAFTAR PUSTAKA

- [1] Aapche Software Foundation
“Apache Hadoop”
[http://hadoop/apache.cgi](http://hadoop.apache.cgi)
- [2] Fay Chang, Jeffrey Dean, Sanjay Ghemawat . Bigtable. *Adistributed Storage System For Structured Data*.
<http://static.googleusercontent.com/media/research.google.com/id/archive/bigtable-osdi06.pdf>
- [3] JeffreyDeanand, Sanjay, Ghema. *Map Reduce : Simplified Data Processingon Large Clusters*.
<http://static.googleusercontent.com/media/research.google.com/id/archive/mapreduce-osdi04.pdf>
- [4] R. Kelly Rainer, C. 2011. *Introduction to Information Systems*. John Wiley & Sons (Asia) Pte Ltd.
- [5] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung . *The Google File System*.
<http://static.googleusercontent.com/media/research.google.com/id/archive/gfs-sosp2003.pdf>
- [6] Wijaya . 2015. *Teknologi Big Data: Sistem Canggih dibalik Google, Yahoo!, Facebook, IBM*. Google Play Book