

SISTEM PENGELOLAAN KEARSIPAN GOOGLE MELALUI BIG DATA

GOOGLE'S RECORDS AND ARCHIVES ADMINISTRATION SYSTEM THROUGH BIG DATA

Aldha Shafrieldha Shihab¹, Anugerah Pagiyon Nurfajar²

^{1,2}KSM Eka Prasetya Universitas Indonesia

Email: aldha.shafrieldha@ui.ac.id¹, anugerah.pagiyon@ui.ac.id²

Abstract

This article examines the use of Big Data technology for the records and archives administration system in the contemporary era. Big Data can be defined as a collection of data that is extremely large and complex which cannot be dealt with the traditional data processing software. This technology makes it possible to store large amounts of data and can be accessed quickly. By implication, Big Data served sufficiently as a digital records and archives administration system. One of companies in the world that uses Big Data as a digital system is Google. Inc. This phenomenon is quite new in the contemporary era, hence it is quite interesting to see "how Google manages its records and archives administration system through Big Data". In line with the previous question, this article aims to find out how Google manages its records and archives administration system through Big Data. This article uses a qualitative research method through the literature review. The data used in this article is from the secondary data sourced from books and journal articles. One of the important findings in this article is that Google uses a distributed storage system to manage structured data called Bigtable. Bigtable makes the other Google services operate, so it can be stated that Bigtable is a fundamental system in Google. Although there are quite a few Google services which use Bigtable as the base, there are three main services that everyone use the most, namely Google Analytics, Google Earth, and Personalize Search. Through this Big Data technology, it can be seen that the archive management can use not only the conventional methods, but also the latest technology.

Keywords: Google, Big Data, Digital Archives, Bigtable, Machine Learning Google

Abstrak

Artikel ini mengkaji mengenai pemanfaatan teknologi Big Data dalam sistem kearsipan di era kontemporer. *Big Data* sendiri merupakan kumpulan data yang begitu besar dan kompleks sehingga tidak memungkinkan lagi untuk dikelola dengan perangkat lunak tradisional. Teknologi ini memungkinkan untuk menyimpan data dalam jumlah besar dan dapat diakses secara cepat. Implikasinya, Big Data sangat cocok untuk dimanfaatkan sebagai sistem kearsipan digital. Salah satu perusahaan di dunia yang memanfaatkan Big Data sebagai sistem digital adalah Google. Inc. Fenomena ini terbilang cukup baru di dunia kontemporer, maka dari itu cukup menarik untuk melihat "*bagaimana Google mengelola sistem kearsipannya melalui Big Data*". Linear dengan pertanyaan sebelumnya, artikel ini bertujuan untuk mengetahui cara-cara Google dalam mengelola sistem kearsipannya melalui Big Data. Artikel ini menggunakan metode penelitian

kualitatif melalui studi pustaka. Data yang digunakan berupa data sekunder bersumber dari buku maupun artikel jurnal. Salah satu temuan penting dalam tulisan ini adalah Google menggunakan sistem penyimpanan terdistribusi untuk mengelola data terstruktur yang disebut *Bigtable*. Melalui *Bigtable* ini layanan-layanan Google lainnya beroperasi, sehingga dapat dikatakan jika *Bigtable* merupakan sistem fundamental dalam Google. Meskipun cukup banyak layanan Google yang menggunakan *Bigtable* sebagai dasar, tetapi terdapat tiga layanan utama yang tentu digunakan semua orang. Layanan Google dengan basis fundamental *Bigtable* tersebut antara lain Google Analytics, Google Earth dan Personalize Search. Melalui teknologi *Big Data* ini kemudian dapat dilihat jika pengelolaan arsip tidak hanya melalui metode konvensional, tetapi bisa menggunakan teknologi mutakhir.

Kata Kunci: Google, *Big Data*, Arsip Digital, *Bigtable*, Pembelajaran Mesin Google

PENDAHULUAN

Sejak ditemukannya internet oleh Leonard Kleinrock dan tim ARPANET pada tahun 1969, teknologi basis data dunia maya terus berkembang. Rangkaian perkembangan dunia maya terus berkembang semenjak pertama kalinya ditemukan, hingga lahirnya konsep WWW (*World Wide Web*). WWW dicetuskan oleh sekelompok ilmuwan yang bekerja di lab fisika CERN sebuah lembaga riset nuklir Eropa. Mereka mengusulkan sebuah sistem manajemen informasi yang terdesentralisasi melalui sistem koneksi *hypertext* dan pencarian dengan *keyword*. Hingga pada akhirnya sistem manajemen informasi tersebut menjadi familiar bagi masyarakat dunia dan mulai dikomersilkan melalui pendirian *Thefacebook.com* oleh Mark Zuckerberg pada 2003.

Melalui sistem yang baru ini aliran data menjadi semakin besar dan menghimpun banyak akses informasi. Akses informasi serta himpunan data tersebut bisa mencapai miliaran *kilo bytes*, sehingga teknologi bernama *Big Data* pun mulai diperkenalkan. Menurut Hurwitz (2013) *Big Data* merupakan istilah untuk sekumpulan

data yang begitu besar atau kompleks dimana tidak bisa ditangani lagi dengan sistem teknologi komputer konvensional. Menurut McKinsey (2011, dalam Cholissodin dan Riyandani) *Big Data* adalah data yang memiliki skala (*Volume*), distribusi (*Velocity*) dan keragaman (*Variety*) yang sangat besar. Konsekuensinya, dalam pengelolaan *Big Data* membutuhkan penggunaan arsitektur teknikal dan metode analitik yang inovatif untuk mendapatkan wawasan yang dapat memberikan nilai bisnis baru (Informasi yang bermakna). Menurut Cholissodin dan Riyandani (2016) *Big Data* memiliki beberapa elemen penting yaitu data, informasi, dan pengetahuan.

Saat ini internet hampir tidak mungkin dikecualikan dari kehidupan manusia modern. Di Indonesia saja, berdasarkan survei yang diselenggarakan oleh APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) pada tahun 2018 sekitar 171,17 juta penduduk Indonesia telah mampu mengakses internet. Hal ini berarti sekitar 64,8% dari total penduduk Indonesia telah menggunakan internet, padahal pada

tahun sebelumnya yaitu 2017 terdapat sekitar 54,68% penduduk Indonesia menggunakan internet. Dalam survei yang sama mayoritas pengguna internet di Indonesia atau sekitar 45,3% sering mengunjungi situs film atau video, dan sekitar 50,7% menggunakan Facebook sebagai media sosial yang paling banyak dikunjungi. Ini berarti telah terhimpun miliaran data dalam sistem *Big Data*, tidak hanya di Indonesia, bahkan di dunia. Tahun 2019 lalu, Internet Archive salah satu lembaga yang mengarsipkan situs web dan objek didalamnya telah menyimpan sekitar 273 miliar halaman web dengan lebih dari 361 juta situs web. Seluruhnya disimpan dalam penyimpanan berkapasitas lima belas *peta byte* atau jika dikonversi menjadi 15.360 *terabyte* atau 15.728.640 *gigabyte*.

Teknologi *Big Data* tidak hanya menghimpun data-data untuk informasi semata. Dalam perkembangannya *Big Data* juga bisa digunakan untuk keperluan bisnis dan arsip digital. Arsip sendiri memiliki beberapa pengertian, berdasarkan UU No. 43 tahun 2009 arsip didefinisikan sebagai rekaman kegiatan atau peristiwa dalam bentuk media sesuai dengan perkembangan teknologi informasi dan komunikasi yang dibuat dan diterima lembaga negara, pemerintahan daerah, lembaga pendidikan, perusahaan, organisasi politik, organisasi kemasyarakatan, dan perorangan dalam pelaksanaan kehidupan bermasyarakat, berbangsa dan bernegara. Pengertian arsip yang tercantum dalam ISO 15489-1 (*Records Management – Part 1: General*) adalah informasi yang diciptakan, diterima dan disimpan sebagai bukti dan informasi oleh organisasi atau seseorang dengan

tujuan memenuhi kewajiban hukumnya atau digunakan sebagai transaksi bisnis. Pengertian lain yang tercantum dalam *Electronics Record: A Workbook for Archive* yaitu menyebutkan bahwa arsip merupakan informasi terekam yang dibuat atau diterima dalam suatu proses yang terdiri dari memulai, melaksanakan dan menyelesaikan aktivitas dari suatu institusi atau perorangan, dalam definisi ini arsip mengandung konten, konteks dan struktur yang memadai sebagai bukti adanya aktivitas tersebut. Melalui beberapa definisi diatas, dapat dipersingkat bahwa arsip merupakan sebuah dokumen baik dalam bentuk digital maupun konvensional yang berisi rekaman aktivitas atau kejadian ketika suatu institusi atau perorangan melakukan sesuatu.

Arsip dan *Big Data* di era kontemporer ini semakin erat kaitannya dan hampir tidak dapat dipisahkan. Keduanya saling berkelit kelindan satu sama lain, arsip memanfaatkan teknologi *Big Data* sebagai tempat penyimpanan barunya yang hampir tidak memiliki batas. Sedangkan *Big Data* semakin berkembang karena banyak terisi oleh arsip-arsip digital dari berbagai penjuru dunia. Uniknya, kedua ‘alat digital’ ini dapat bersatu dalam sebuah perusahaan bernama Google. Inc (dalam artikel ini nantinya akan konsisten disebut Google). Google merupakan salah satu perusahaan di dunia yang bergerak dalam teknologi *Big Data* serta pelopor perpaduan antara teknologi *Big Data* – kearsipan, dan dikomersialisasi. Komersialisasi tersebut menghasilkan sebuah bisnis baru yaitu menjual data serta informasi yang terdapat dalam *server* mereka. Dalam artikel ini akan lebih fokus

membahas mengenai *Big Data* dan pemanfaatan dalam kearsipan dalam konteks kontemporer. Sebagai contoh kasus, penulis menggunakan sistem kearsipan Google yang telah *menggunakan Big Data* dalam operasionalisasinya. Artikel ini memulai pembahasan dari awal pendirian Google yang menjadi salah satu perusahaan pertama di bidang *Big Data*. Kemudian menjabarkan perkembangan-perkembangan pengelolaan arsip Google melalui teknologi *Big Data*.

Maka dalam artikel ini dirumuskan sebuah pertanyaan penelitian yaitu “Bagaimana cara Google mengelola arsip digitalnya melalui teknologi *Big Data*?”

Penelitian ini bertujuan untuk mengetahui cara-cara Google mengelola kearsipannya melalui *Big Data*. *Big Data* tersebut nantinya dapat digunakan dengan banyak hal seperti bidang usaha baru dan sebagainya. Melalui *Big Data* tersebut Google mampu menjadi perusahaan multinasional dan mendominasi sektor digital.

METODE PENELITIAN

Artikel ini berfokus pada pengelolaan kearsipan Google melalui teknologi *Big Data*. Maka dari itu penulisan ini menggunakan metode studi pustaka dengan analisis deskriptif. Pengumpulan data dilakukan melalui studi pustaka melalui data sekunder dari berbagai sumber jurnal maupun buku. Analisis yang diberikan berupa penjabaran mengenai Google dan pengelolaan kearsipan Google melalui *Big Data*.

Adapun kerangka konseptual yang digunakan dalam penelitian ini adalah:

a. Arsip

Undang-undang No 43 tahun 2009 tentang Kearsipan, dinyatakan bahwa arsip adalah rekaman kegiatan atau peristiwa dalam bentuk media sesuai dengan perkembangan teknologi informasi dan komunikasi yang dibuat dan diterima lembaga negara, pemerintahan daerah, lembaga pendidikan, perusahaan, organisasi politik, organisasi kemasyarakatan, dan perorangan dalam pelaksanaan kehidupan bermasyarakat, berbangsa dan bernegara. Dalam Undang-Undang yang sama disebutkan bahwa arsip dapat dibedakan menjadi dua golongan menurut fungsinya, yaitu arsip statis dan arsip dinamis. Arsip statis adalah arsip yang dihasilkan oleh pencipta arsip karena memiliki nilai guna kesejarahan, telah habis retensinya, dan berketerangan dipermanenkan yang telah diverifikasi baik secara langsung maupun tidak langsung oleh Arsip Nasional Republik Indonesia (ANRI) dan atau lembaga kearsipan lain. Arsip Dinamis adalah arsip yang dipergunakan secara langsung dalam kegiatan pencipta arsip dan disimpan selama jangka waktu tertentu, yang terbagi menjadi tiga, yaitu:

1. Arsip vital, adalah arsip yang keberadaannya merupakan persyaratan dasar bagi kelangsungan operasional pencipta arsip, tidak dapat diperbaharui, dan tidak bisa tergantikan apabila rusak atau hilang.

2. Arsip aktif, adalah arsip yang frekuensi penggunaannya tinggi dan atau terus menerus.
3. Arsip inaktif, adalah arsip yang frekuensi penggunaannya telah menurun.

Arsip diolah menjadi informasi, pengolah tersebut adalah pengolahan yang disesuaikan dengan kebutuhan dari informasi yang akan diambil. Perkembangan ilmu pengetahuan dan teknologi modern berdampak pada kemajuan bidang kearsipan yang semakin maju, kemajuan ini membawa kemudahan untuk melaksanakan tugas-tugas kearsipan, apalagi jika memiliki volume arsip yang sangat besar, maka teknologi akan sangat membantu mempercepat dan mengefisienkan proses pengelolaan arsip. Salah satu pengaruh terbesar dari penggunaan teknologi pada sistem kearsipan adalah dimungkinnnya pengiriman dan penyampaian informasi menjadi semakin cepat, apalagi jika mempunyai sistem yang terstruktur. Untuk itu diperlukannya sebuah sistem yang sangat besar yang dapat menampung segala informasi dari berbagai belahan dunia manapun dan dapat diakses secara cepat dan mudah oleh khalayak. Sistem teknologi *Big Data* memungkinkan sebuah kearsipan digital menjadi nyata.

Seiring berkembangnya zaman, sistem kearsipan mulai merambah ke dunia teknologi. Sistem kearsipan mulai berintegrasi dengan teknologi informasi, maka dari itu konsekuensinya adalah

harus mengikuti peraturan yang ada dalam dunia sistem informasi. Pada sistem informasi dikenal sebuah sertifikasi bernama ISO 27001. ISO 27001 merupakan sebuah dokumen standar Sistem Manajemen Keamanan Informasi (SMKI) yang memberikan gambaran secara umum mengenai kerangka kerja dari sebuah organisasi dalam rangka mengimplementasikan konsep keamanan informasi (Itg.id, 2020). Sertifikasi ISO 27001 dikembangkan paralel dengan meningkatnya kebutuhan dan penggunaan teknologi informasi dan komunikasi (TIK) dalam menunjang aktivitas bisnis. Semakin bertambahnya aktivitas bisnis, semakin meningkat pula resiko gangguan keamanan dari penggunaan teknologi informasi.

ISO 27001 terdiri dari 14 klausa yang mencakup 113 kontrol (Itg.id, 2020). 14 klausa tersebut diantaranya, (1) *Information security policies*, (2) *How information security is organised*, (3) *human resources security*, (4) *Asset management*, (5) *Access control and managing user access*, (6) *Cryptographic technology*, (7) *Physical security of the organization's sites and equipment*, (8) *Operational security*, (9) *Secure communications and data transfer*, (10) *Secure acquisition, development, and support of information systems*, (11) *Security for suppliers and third parties*, (12) *Incident management*, (13) *Business continuity/disaster recovery*, (14) *Compliance*. Keseluruhan klausa dan

kontrol ini dalam implementasinya menyesuaikan dengan relevansi dari kondisi yang sedang dialami.

b. Big Data

Big Data merupakan istilah yang diberikan pada kumpulan data yang berukuran sangat besar dan kompleks, sehingga tidak bisa lagi menggunakan perangkat pengolah data yang konvensional (Maryanto, 2017). Menurut *TechAmerica Foundation's Federal Big Data Commission* (Gandomi., Amir., dan Murtaza Haider, 2014) *Big Data* adalah istilah yang menggambarkan volume besar data yang berkecepatan tinggi, kompleks dan variabel yang membutuhkan teknik dan teknologi canggih untuk memungkinkan penangkapan, penyimpanan, distribusi, Manajemen, dan analisis informasi. Sementara menurut *Gartner IT Glossary* (dalam Gandomi., Amir., dan Murtaza Haider., 2014) *Big Data* adalah aset informasi volume tinggi, kecepatan tinggi, dan beragam yang menuntut bentuk-bentuk inovatif dari proses informasi yang hemat biaya untuk peningkatan wawasan dan pengambilan keputusan.

Laney (dalam Gandomi., Amir., dan Murtaza Haider, 2014) menyebutkan bahwa tantangan dalam manajemen data terdapat tiga hal yaitu *volume*, *variety*, dan *velocity* (volume, variety, dan kecepatan). Kemudian tiga tantangan manajemen data tersebut oleh Chen, Chiang & Storey (2012), Kwon, Lee & Shin (2014) (dalam Gandomi., Amir., dan

Murtaza Haider, 2014) menggunakan *volume*, *variety*, dan *velocity* sebagai kerangka kerja umum untuk menggambarkan mengenai *Big Data*. *Volume* mengacu pada ukuran atau besarnya data dalam satuan *byte*. *Variety* mengacu pada heterogenitas struktural dalam dataset. *Velocity* (kecepatan) mengacu pada tingkat dimana data dihasilkan dan kecepatan dalam data tersebut harus dianalisis dan ditindaklanjuti (Gandomi. Amir., dan Murtaza Haider, 2014).

Melalui beberapa penjelasan sebelumnya, dapat disintesis jika *Big Data* adalah istilah baru dalam penggambaran data yang didalamnya memiliki banyak data dalam satuan *byte*. *Big Data* tercipta dari bentuk reaksi pada melimpahnya data di masa sekarang yang sudah tidak mampu ditangani oleh aplikasi pengolah data konvensional. *Big Data* tersebut digunakan dalam banyak sektor tidak hanya di sektor informasi tetapi juga merambah ke sektor kesehatan, sektor publik, sektor ritel, sektor manufaktur, dan sektor data pribadi (Colissodin, 2016). Tanpa adanya *Big Data* di masa sekarang, teknologi informasi, data, arsip dan hal-hal lain yang terkandung banyak data di dalamnya akan mengalami hambatan karena sulitnya mengakses data secara konvensional.

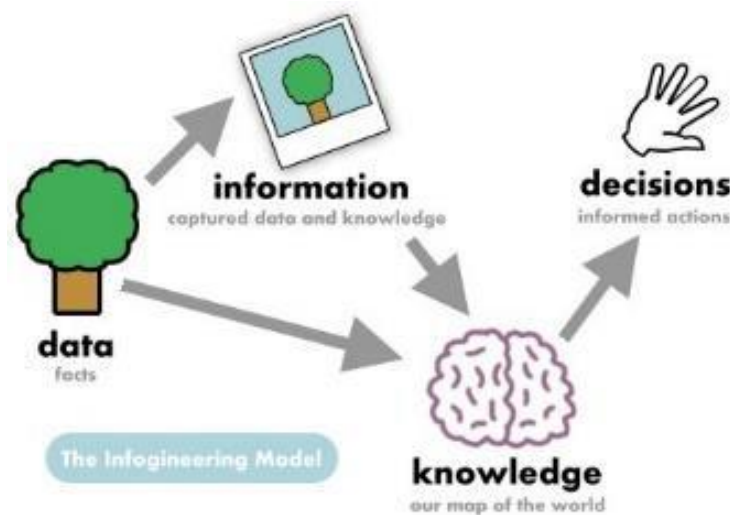
Gambar 1 merupakan perumpamaan yang tepat untuk menggambarkan apa itu *Big Data*, dan keempatnya adalah elemen penting dalam *Big Data*. Data adalah

catatan dan informasi atas kumpulan fakta atau keterangan dari suatu hal yang didapat dari pengamatan dan pencarian terhadap sumber-sumber tertentu. Informasi adalah data yang telah diolah menjadi bentuk yang mempunyai arti bagi si Penerima dan bermanfaat dalam pengambilan keputusan di masa kini maupun masa yang akan datang. Pengetahuan adalah apa yang kita ketahui yang ada dalam otak kita ataupun yang ada di alam semesta, untuk membangun ilmu pengetahuan diperlukannya informasi dan data. Kemudian, setelah pengetahuan itu diolah maka diperlukan pengambilan keputusan atas data yang diolah.

Agar mempermudah penjelasan tentang *Big Data*, dapat dilihat definisi dari Hurwitz. Menurut Hurwitz, et al., (2013), *Big Data* merupakan istilah untuk sekumpulan data yang begitu besar atau

kompleks dimana tidak bisa ditangani lagi dengan sistem teknologi komputer konvensional. Suatu data bisa dikatakan sebagai *Big Data* adalah ketika data-data tersebut memiliki tiga karakteristik yang harus ada dalam *Big Data* yaitu *volume*, *velocity* dan *variety*.

Volume adalah seberapa besar data ditampung dalam suatu sistem, seperti yang diketahui, *Big Data* haruslah mempunyai *volume* data yang sangat besar yang tidak bisa ditampung oleh sebuah komputer konvensional, Google sendiri memiliki sekitar 15 *Exabytes* data dan volumenya terus meningkat pada tingkatan yang belum pernah terjadi sebelumnya. *Velocity* adalah kecepatan data yang masuk, dimana membuat data, memproses data, menganalisis data kecepatannya terus meningkat dan ada aliran data yang terjadi. *Variety* adalah tentang mengelola kumpulan dari



Gambar 1. Gambaran Umum *Big Data*
Sumber: Cholissodin, I. dan Riyandani, E., 2016)

berbagai macam data yang kompleks baik yang terstruktur, semi terstruktur maupun yang tidak terstruktur, kumpulan data yang tidak terstruktur lebih mendominasi. Suatu perusahaan perlu menganalisis berbagai tipe data dari berbagai sumber informasi dan data yang dihasilkan pun tak terhitung jumlahnya. Seiring dengan berkembangnya teknologi, karakteristik *Big Data* pun berubah, ada dua karakteristik lagi yang menjadi acuan apakah data yang dikelola itu adalah sistem *Big Data*. *Variability* adalah bagaimana *user* menafsirkan data dari perubahan yang terjadi dalam struktur data. *Value* adalah sebuah nilai dari suatu data, nilai bisnis ini memberikan keuntungan bagi suatu perusahaan.

HASIL DAN PEMBAHASAN

Sejarah Google

Google Inc. adalah sebuah perusahaan *search engine* yang didirikan oleh Sergey Brin dan Larry Page pada tahun 1998. Perusahaan ini telah memonopoli sebesar 70% dari pencarian online dunia. Google mempunyai pusat di Mountain View, California. Pada awalnya, Google adalah sebuah perusahaan yang bergerak di bidang pencarian saja, tetapi seiring berjalannya waktu, Google mengembangkan lebih dari 50 produk dan jasa untuk handphone ataupun komputer, seperti *electronic mail* (Gmail), Google Meets, Google Maps dan banyak lagi. Walaupun begitu, mesin pencarian Google tetap menjadi kunci kesuksesan dari Google.

Perjalanan Google dimulai pada tahun 1995, ketika Larry dan Sergey masih berada di Stanford University. Mereka awalnya menciptakan sebuah program komputer yang bernama BackRub, sebuah mesin pencarian yang memanfaatkan analisis *backlink* untuk melacak dan merekam data di internet. Mesin pencari ini sangat unik karena menggunakan teknologi yang mereka kembangkan bernama Pagerank, yang menetapkan peringkat pada situs web dengan menghitung jumlah halaman, mengolahnya kemudian menautkan ke laman yang asli. Kemudian hasil yang muncul dalam mesin pencari itu berdasarkan seberapa sering kata yang dicari muncul di halaman *web*. Sistem PageRank inilah yang mengantarkan Google menjadi mesin pencarian *web* yang dibisniskan. PageRank adalah teknologi yang belum pernah dimiliki *search engine* lain yang ada sebelum Google dan membuat Google menjadi *search engine* unik di awal masa kemunculannya.

Pada 1998 Google menerima bantuan dari pendiri Microsystem, Andy Bechtolsheim sebagai modal awal mereka mendirikan perusahaan. Sejak saat itulah Google meluncur menjadi mesin pencari internet yang populer. Pada awalnya, Perusahaan Google bernama Googol adalah sebuah istilah matematika yang berarti sebutan untuk nomor satu dan diikuti oleh 100 nol. Pada tahun 2000, Google menjadi klien dari *search engine* (mesin pencari) situs *web* yang paling populer. Pada tahun 2004, ketika Yahoo dan Google sudah tak tak bermitra, mesin pencarian Google

menerima pencarian lebih dari 200 juta kali sehari. Bahkan hingga saat ini Google menangani sekitar tiga milyar pencarian per hari (Gambar 2).

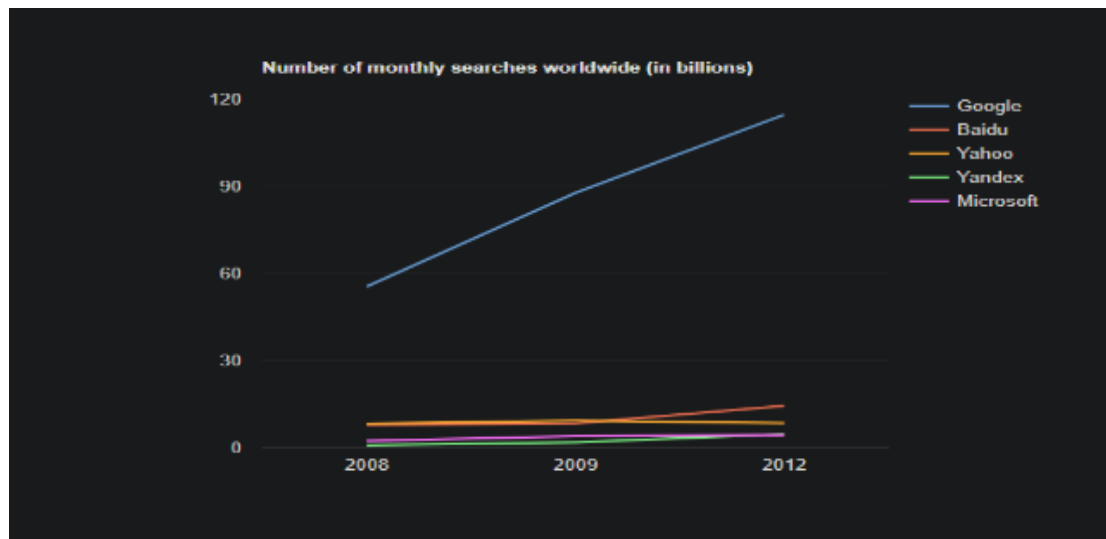
Terlampau banyaknya data yang masuk perhari, Google membangun 11 pusat data di seluruh dunia. Setiap pusatnya mengandung lebih dari ratusan ribu server untuk menunjang perusahaan tersebut. Google membangun tiga sistem kode yang menjadi jantung dari Google itu sendiri yaitu Google File System, Bigtable dan Mapreduce. Mereka saling berkait membangun pondasi besar.

Google telah berkembang selama bertahun-tahun dan menjadi perusahaan besar multinasional. Pada tahun 2001, para investor menunjuk Eric Schmidt untuk menjadi CEO dari Google karena kekhawatiran mereka terhadap Brin dan Page yang tidak berpengalaman dalam memimpin perusahaan. Selama Schmidt

menjadi CEO, Larry Page ditunjuk menjadi *President of Product* dan Sergey Brin menjadi *President of Technology*. Pada 2015, Google mengalami restrukturisasi divisi dan menjadi anak perusahaan dari Perusahaan Induk Alphabet Inc. Produk-produk Google seperti Mesin Pencarian, Youtube, Google Maps, Google Android, Google Advertising, tetap berada dibawah perusahaan Google. Tetapi, beberapa Google Ventures yang lain seperti Longevity Research Company Calico, Laboratorium penelitian Google X dan beberapa lainnya berada dibawah Alphabet Inc. Kemudian Larry Page menjadi CEO dari Alphabet Inc., Sergey Brin menjadi Presiden Eksekutif dan Eric Schmidt menjadi Ketua Eksekutifnya.

Google dan Teknologi *Big Data*

Big Data bisa dijadikan sebagai bisnis yang menjanjikan. Pada era teknologi



Gambar 2. Data Pencarian oleh beberapa Mesin Pencari (*Search Engine*)
Sumber: www.internetlivestats.com

seperti sekarang ini. data merupakan aset yang berharga, maka diperlukan sebuah sistem untuk menyimpan data-data tersebut untuk kemudian dijual sebagai bisnis. Google memonetisasi mesin pencari mereka dengan mencari cara untuk mengambil data yang dikumpulkannya dari seseorang saat menjelajahi Web, membangun pendapatan yang besar dengan menjadi penjual iklan online terbesar di dunia. Di sektor ini, Google adalah pelopor. Pada tahun 2006 mereka pernah memperkenalkan Google Bigtable, yang merupakan sistem *database* berskala besar dan cepat yang digunakan Google untuk mengolah berbagai jenis data dari berbagai layanan Google, termasuk layanan dari mesin pencarian Google yang berbasis internet.

Kapasitas dari teknologi dunia untuk menyimpan informasi dan data telah meningkat sebanyak dua kali lipat setiap 40 bulan sejak tahun 1980. Contohnya saja, pada tahun 2012 sudah ada lebih dari 2.5 quintillion bytes data yang dibuat setiap harinya. Sejak 2008, Google dengan menggunakan Big Data Analysis telah mengantongi 2 petabytes perhari dari Mesin Pencari saja. Hal itu terus meningkat dan dalam 10 tahun terakhir jumlah data yang ada dalam mesin pencari Google sebanyak 160 petabytes perhari. Oleh karena itu, Volume data ini tidak dapat ditangani lagi oleh Proses manajemen data tradisional, jadi Google menggunakan teknologi lain dalam menangani lonjakan data yang sangat besar setiap harinya, teknologi tersebut adalah teknologi Big Data. Google memakai Big Data untuk memelihara sistem file nya.

Google juga menggunakan Cloud Bawaan untuk menyimpan data-data. Google memproses 3,5 Milyar permintaan per hari dan setiap permintaan menanyakan basis data 20 milyar halaman web.

Google telah mampu mengelola dan memberdayakan data yang melimpah pada era Ledakan Informasi. Teknologi crawler yang dimiliki oleh Google mampu mendownload web page yang ada di dunia secara terus menerus tanpa berhenti bahkan mengikuti update yang ada pada web-web tersebut. Google membangun tiga sistem kode yang menjadi jantung dari Google. Google File System (GFS). GFS adalah media penyimpanan data yang menyimpan data dalam jumlah besar secara terdistribusi. Mereka membagi beberapa indeks situs web mereka di ribuan node penyimpanan, dan tersebar di antara banyak pusat komputasi di seluruh dunia. Kapasitas penyimpanan dari GFS bisa terus diperbesar dan diperbanyak hanya dengan menambahkan jumlah komputer di dalam sistem nya. Jantung yang kedua adalah MapReduce, yang termasuk kedalam teknologi pengolahan data. MapReduce mampu mengolah data yang disimpan oleh GFS secara paralel yang mampu mengolah data ukuran besar dalam waktu yang ratusan kali lipat lebih cepat dibanding komputer konvensional dengan bantuan ribuan computer. Kemudian yang terakhir ada Bigtable, merupakan sebuah program basis data yang dimiliki Google. Bigtable mampu menyajikan data berukuran besar yang tidak mampu ditangani oleh sistem database yang sudah ada.

Google sebagai Perpustakaan Universal

Google adalah sebuah perpustakaan universal yang memuat berbagai macam data dan sejarah dari berbagai zaman. Perpustakaan digital yang menampung lebih dari milyaran pengetahuan tentang masa lalu, masa kini dan untuk masa yang akan datang. Miliaran arsip tersimpan rapi dalam jutaan server yang berada di seluruh dunia, menjaganya untuk kemudian menjadi sebuah saksi bisu untuk masa yang akan datang. Sebenarnya konsep perpustakaan universal sudah dipopulerkan pada satu abad yang lalu. Dimulai pada tahun 1939, melalui essay dari Jorge Luis Borges yang berjudul “La Biblioteca Total” atau “Perpustakaan Total” yang isinya adalah tentang sejarah singkat mengenai gagasan untuk membuka sebuah perpustakaan yang universal dan bisa diakses oleh siapa pun, kapan pun dan dimanapun. Kemudian, Borges menyebutkan Sejarah Korespondensi di antara gagasan ini dan teori Kuno Atomisme serta Teorinya Renaisans analisis kombinatori. Dan Borges pun juga menyetujui saran dari Theodor Wolff bahwa perpustakaan total itu adalah turunan dari mesin berpikir yang dirancang oleh Polymath pada abad pertengahan yang fungsinya untuk membantu umat islam menjadi umat Kristen yang ‘Ta’at’.

Teori Atomisme yang dirumuskan oleh Aristoteles menjadi salah satu poin dari terbentuknya Essay Perpustakaan Total. Penjelasan universal Atomisme tentang realitas bergantung pada kepercayaan bahwa semua benda fisik, terlepas dari bentuk individu, didasari pada stoicheia, perbedaan

di antara benda-benda ini sepenuhnya berasal “dari posisi, keteraturan, atau bentuk” (Borges, dalam Hillis. 2012). Filsafat atomis, sebagian, merupakan perpanjangan dari kepercayaan Pythagoras bahwa "benda yang kasat mata, teragregasi dari sejumlah unit yang sama-sama dianggap sebagai titik-titik geometri, atom benda, dan satuan aritmatika". Dari sini dapat diketahui bahwa sebenarnya teori Atomisme dapat fantasi dari setiap orang yang tidak berkaitan dengan bentuk (tubuh/jasad) karena pada dasarnya kita hanya terdiri dari sekumpulan informasi.

James Gleick membahas tentang warisan dari teori Atomisme, dan dia mengaitkannya dengan membuat hubungan antara Atomisme dan bit (*binary digit*) yang mana itu adalah unit informasi terkecil. Hal inilah yang menjadi sebab terjadinya sebuah mitos asal yang mengaitkan munculnya teori informasi dan ilmu komputer. Teori ini seperti blok bangunan yang bisa digunakan untuk membayangkan bagaimana individu dapat mengurangi kerumitan dan ukuran dari semua informasi yang tersimpan dalam memori atau merepresentasikan sebuah pengetahuan ke dalam sebuah pola yang dapat diakses oleh para pencari informasi.

Teori Atomisme dan Prinsip-prinsip pertama Pythagoras adalah indikasi awal terhadap kode yang terus mendorong budaya rekayasa abad dua puluh satu seperti Google. Seperti yang dilihat, kesamaan biner dari tongkat dan batu yang para atomist pakai sebagai *placeholder* untuk mencatat nilai dan posisi nomor apapun dalam sebuah tempat yang lebih luas sama

seperti kode biner yang dipakai dalam komputer. Ini adalah perangkat penyimpanan data awal yang terletak dalam inti khusus Google dari konsep *granularity* universal yang menjadi sarana untuk menghasilkan bentuk pencarian yang dipersonalisasi.

Misi dari Google adalah “Memfasilitasi akses informasi ke seluruh dunia dan dalam setiap bahasa, dengan lebih dari 130 bahasa dengan menggunakan alat terjemahan dari Google, setiap orang dapat mengakses informasi dari sisi lain dunia dalam bahasa yang mereka gunakan.” (Google dalam Hillis, dkk. 2012). Sejatinya hal ini merupakan bentuk manifestasi dari sebuah kisah menara Babel yang dijelaskan oleh Kitab Kejadian dan merupakan inspirasi untuk Borges menyelesaikan essay Perpustakaan Total. Menara Babel ditafsirkan sebagai keangkuhan manusia yang ingin menyaingi Tuhan. Diceritakan, setelah Banjir Besar, sebuah suku bermigrasi dari timur ke Shinar (Irak). Seorang Penguasa besar memerintahkan orang-orang untuk membangun sebuah kota dan sebuah menara yang puncaknya di surga untuk meyakini diri mereka bisa lebih daripada tuhan. Namun, Tuhan yang melihat itu dan kemudian marah lalu meruntuhkannya. Melalui kisah ini, keragaman bahasa adalah hasil dari kemarahan Tuhan. Dan hal itu lah yang menginspirasi Google untuk mempersatukan manusia melalui informasi agar mereka bisa mengerti informasi dari dunia yang berbeda.

Melalui beberapa pemaparan sebelumnya Google kemudian terinspirasi

sehingga bisa membuat perusahaannya seperti sekarang. Data-data yang Google miliki seperti buku, jurnal, artikel dan sebagainya tersipkan seperti perpustakaan pada umumnya namun lebih canggih. Melalui sistemnya Google mampu ‘memanggil’ sesuatu dengan mudah tanpa perlu menunggu lama. Uniknyanya seluruh data tersebut tidak sepenuhnya dimiliki oleh Google melainkan terintegrasi dengan laman penyedia buku, jurnal artikel atau arsip lainnya. Laman penyedia ini hanya dapat diakses melalui mesin pencari Google, atau mengetik secara spesifik url yang menjadi tujuan. Sayangnya tidak ada satu orangpun yang ingin repot mengetik url lengkap sehingga lebih memilih mencari melalui mesin pencari Google. Tidak hanya itu, data-data yang terdapat di Google dapat diakses tanpa memperhatikan dimensi ruang dan waktu sehingga sangat fleksibel. Seseorang dapat dengan mudah mencari buku-buku atau arsip yang sudah dibuat puluhan tahun sebelumnya dan arsip tersebut akan tahan tinggal puluhan tahun kemudian. *Big Data* yang terintegrasi dengan perusahaan Google dirasa mampu menghasilkan sebuah terobosan gemilang dalam hal pengarsipan sehingga lebih memudahkan dan lebih efisien.

Pengelolaan Data Google Melalui *Bigtable*

Google menggunakan sistem penyimpanan terdistribusi untuk mengelola data terstruktur yang disebut *Bigtable*. *Bigtable* adalah sistem penyimpanan terdistribusi untuk mengelola data terstruktur yang dirancang untuk menskala

ke ukuran yang sangat besar : petabyte data di ribuan server komoditas (Chang, 2008). Lebih dari 60 proyek Google yang menyimpan data pada *Bigtable*, Contohnya, Google Earth, Google Finance, Google Analytics, Okrut dan Pengindeksan Web. *Big Table* menyerupai basis data dalam banyak hal karena *Bigtable* memiliki banyak strategi implementasi dengan basis data (Chang, dkk. 2006). *Model data* dari *Bigtable* sendiri adalah sebuah peta yang diurutkan secara multidimensi yang terdistribusi dan persisten. *Bigtable* akan menyimpan data dalam skala masif. *Row Keys*, *Column Keys* dan *Timestamps* terindeks pada peta yang setiap nilainya adalah *array bytes* yang tidak terinterpretasikan.

Row Keys berfungsi untuk memungkinkan *Bigtable* menyimpan data dalam urutan leksikografis. Setiap rentang dalam *row* tersebut disebut dengan tablet yang merupakan unit distribusi dan penyeimbangan beban. Hal ini menyebabkan bacaan dari *range row* menjadi lebih pendek dan lebih efisien, sehingga hanya memerlukan bentuk komunikasi yang sederhana. Jadi, pengguna jasa dapat dengan mudah mendapatkan lokasi akses untuk data yang mereka cari. Dalam *wehtable* biasanya website dengan domain yang sama dikelompokkan dalam satu baris yang sama dan komponen hostname URL nya dibalik. Hal ini membuatnya menjadi lebih efisien. Contohnya, Google menyimpan data untuk *maps.google.com/index.html* di bawah URL *com.google.maps/index.html*.

Column Keys merupakan data yang dikelompokkan pada set yang bernama *Column Families*. *Column Families* ini membentuk unit dasar kontrol akses. Biasanya data yang tersimpan adalah data yang sama jenisnya. Yang terakhir dari Model Data adalah *Timestamps*. *Timestamps Bigtable* Google adalah bilangan bulat 64bit dan mengindeks beberapa versi data yang sama yang ada dalam setiap sel *Bigtable*. Mereka dapat ditugaskan oleh *Bigtable* untuk mewakili waktu dalam ukuran mikrodetik atau secara lebih spesifik ditugaskan oleh aplikasi dari pengguna jasa. Versi lain dari data yang sama akan disimpan terlebih dahulu agar pengguna jasa dapat membaca versi yang lebih baru dahulu.

Tabel terdiri dari *Row* yang bermanifestasikan entitas tunggal, dan *Column* yang berisi nilai individual untuk *Row*. Setiap baris di indeks dengan satu *Row Keys*, dan kolom yang berkaitan satu sama lain biasanya akan dikelompokkan menjadi satu *Column family*. Kombinasi *Column Family* dengan *Column Qualifier* inilah yang nantinya akan diidentifikasi oleh setiap *Column*. Setiap persimpangan *Row* atau *Column* dapat diisi oleh beberapa sel, atau versi pada *timestamps* yang berbeda. Hal ini memberikan catatan tentang perubahan data tersimpan dari waktu ke waktu.

Contoh pada Gambar 3 adalah mengenai jejaring sosial media presiden Amerika Serikat yang disebut Prezzy (hanya contoh). Dari penelusuran, setiap presiden yang berbeda dapat mengikuti dan melihat

pos dari presiden lainnya. Tabel diatas lah yang menjadi gambaran tentang siapa yang mengikuti siapa. Pada tabel tersebut, terdapat satu *Column Family* yang berisi beberapa *Column Qualifier*. Ini disebut dengan *follows family*, siapa yang *mem-follows*. Kemudian ada *Row key* sebagai *username* dengan asumsi bahwa *username* ini tersebar secara merata pada setiap alpabet. Jadi akses datanya akaj cukup seragam di seluruh tablet. Kemudian, *Column Qualifier* digunakan sebagai data untuk mengetahui siapa yang memfollow siapa.

Bigtable menggunakan *Google File System* (GFS) yang didistribusikan untuk menyimpan log dan data. Biasanya prosesnya sering berbagi mesin dengan aplikasi lain. *Bigtable* bergantung pada sistem manajemen cluster untuk menjadwalkan sebuah pekerjaan untuk

sistemnya, mengelola sumber daya pada *sharing-pool*, menangani kegagalan dari mesin itu serta memantau status dari mesin, apakah berjalan dengan baik atau tidak. Chubby adalah layanan *distributed lock* yang *highly-available* dan *persistent* yang digunakan oleh *Bigtable*. Layanan ini terdiri dari lima replica aktif yang salah satunya menjadi master dan aktif melayani permintaan. Chubby menggunakan algoritma Paxos untuk menjaga replica itu tetap konsisten ketika mengalami kegagalan sistem. *Bigtable* menggunakan Chubby untuk memastikan ada satu master yang aktif setiap saat, untuk menyimpan lokasi bootstrap data *Bigtable*, untuk menyimpan informasi skema *Bigtable* dan untuk menyimpan daftar kolom akses (Chang, dkk. 2006).

Ada tiga komponen utama untuk pengimplementasian *Bigtable*, yaitu sebuah

"follows" column family

	Follows			
Row Key	gwasington	jadams	tjefferson	wmckinley
gwasington		1		
jadams	1		1	
tjefferson	1	1		1
wmckinley			1	

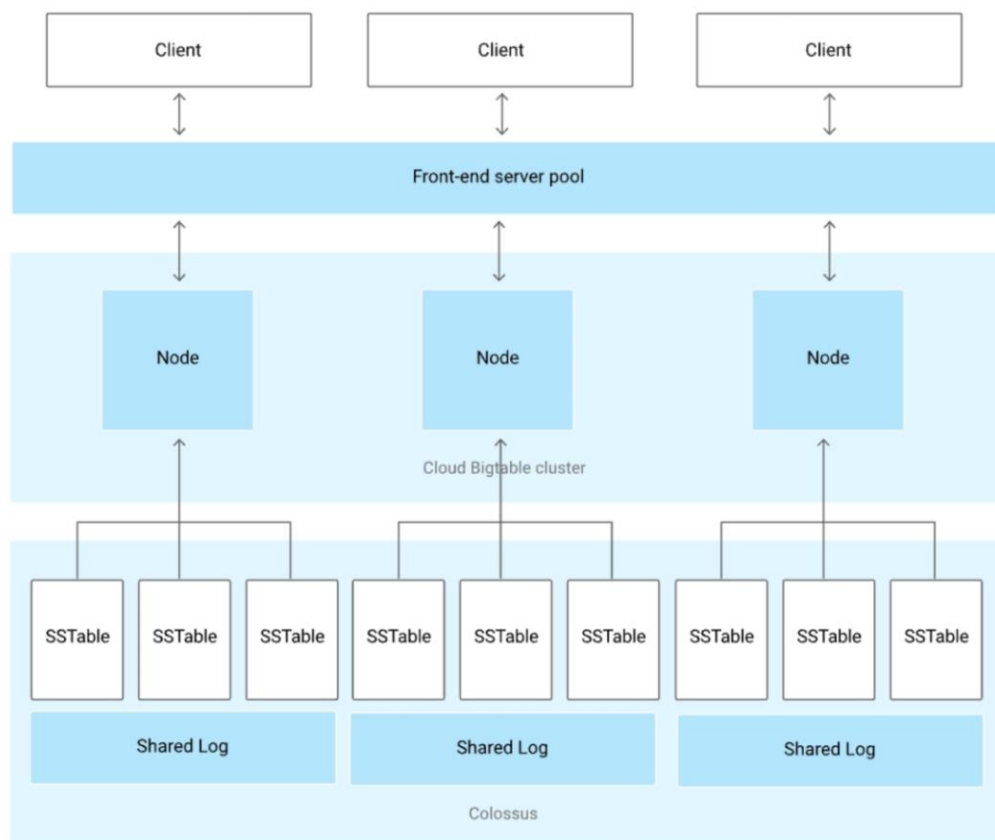
Multiple versions

Gambar 3. Contoh kasus penggunaan *Column Key*, *Row Key*, dan *Timestamps*
Sumber: cloud.google.com/bigtable

perpustakaan data yang terhubung ke setiap pengguna jasa, satu server master dan banyak server tablet. Master disini adalah komponen yang bertanggung jawab untuk menempatkan tablet ke tablet server, mendeteksi penambahan dan waktu berakhirnya server tablet, menyeimbangkan beban yang ada pada server tablet dan pengumpulan koleksi file di GFS. Setiap server tablet mengelola satu set tablet, google memiliki antara sepuluh hingga seribu tablet per server.

Gambar 4 merupakan arsitektur yang sederhana dari keseluruhan arsitektur

Bigtable Google yang rumit. Dari gambar diatas terlihat *Client* adalah yang pertama. Semua permintaan *Client* akan melalui server *front-end* terlebih dahulu sebelum dikirimkan ke node. Node yang dimaksud adalah server tablet. Kemudian, kotak biru yang melingkupi node disebut sebagai *Cloud Bigtable Cluster*, fungsinya adalah untuk mengatur server tablet yang berantakan. Hal ini juga digunakan sebagai kontainer untuk memuat berbagai macam data dari *Client*. Kemudian setelah itu setiap server tablet di cluster menangani subset dari permintaan *Client* ke cluster. Hal ini



Gambar 4. Diagram Penggambaran Sederhana *Bigtable*
Sumber: cloud.google.com/bigtable

dimaksudkan untuk meningkatkan jumlah simultan yang dapat ditangani oleh Cluster ketika menambahkan server tablet ke cluster.

Dalam diagram juga terlihat bahwa tabel dari Cloud Big Data dibagi menjadi beberapa tablet, tablet ini adalah blok yang berbaris berdekatan untuk membantu menyeimbangkan beban kerja kueri. Warna biru muda yang membentengi berbagai tablet disebut Colossus dan itu adalah Google File System yang berformat SSTable. Dalam SSTable disediakan peta yang persisten dan teratur merujuk kepada *Key* dan *Value*, Yang mana merencanakan adalah *string byte arbiter*. Setiap tablet akan dikaitkan dengan server tablet tertentu. Untuk peningkatan ketahanan pada *Bigtable*, semua penulisan selain SSTable di simpan di log bersama Colossus setelah ada pengakuan dari Cloud *Bigtable*. Data tidak pernah disimpan di server tablet, melainkan setiap server tablet memiliki petunjuk ke satu set tablet yang disimpan di Colossus.

Contoh Penggunaan *Bigtable* dalam Platform Google

Bigtable mulai diperkenalkan oleh Google pada tahun 2006. Pada saat itu sudah ada 60 produk Google yang memakai sistem ini. Ada sekitar 388 klaster *Bigtable* yang berjalan dengan total gabungan sekitar 24.500 server tablet (Chang, dkk. 2006). Berikut adalah Contoh Produk Google yang menggunakan Sistem Kearsipan *Bigtable*.

a. Google Analytics

Google Analytics adalah layanan analisa gratis dari google yang diperuntukkan untuk membantu seorang webmaster untuk menganalisis pola lalu lintas web yang ada di situs web mereka. Google Analytics merupakan bagian dari Google Marketing Platform. Layanan yang disediakan meliputi laporan realtime, laporan pengunjung, laporan akuisisi pengunjung, laporan konversi dan laporan perilaku pengunjung.

Untuk mengaktifkan layanan Google Analytics biasanya webmaster menanamkan program JavaScript di halaman web mereka. Kemudian program ini mencatat berbagai informasi tentang permintaan di Google Analytics, seperti informasi tentang halaman yang dibuka ataupun tentang pengindetifikasian pengguna. Data itu yang nantinya akan diserahkan kepada webmaster dalam bentuk diagram.

Terdapat dua tabel yang digunakan oleh Google Analytics. Yaitu *raw click table* (200 TB) yang berfungsi untuk *maintains* sebuah row untuk setiap sesi pengguna akhir. Nama row ini adalah tuple yang berisi nama situs web dan waktu pembuatan sesi. Skema ini memastikan sesi itu yang mengunjungi website yang sama berdekatan dan diurutkan secara kronologis. *Summary Table* (20TB) berisi berbagai ringkasan untuk setiap situs web yang telah ditentukan sebelumnya. Tabel ini dibuat dari *raw click table* yang di jadwalkan

secara berkala oleh MapReduce. Pekerjaan dari MapReduce adalah mengekstrak data sesi terbaru dari *raw click table*.

b. Google Earth

Google Earth merupakan sebuah program virtual globe. Program ini membuat Salinan bumi dalam bentuk digital. Produk ini memungkinkan penggunaanya untuk bernavigasi di seluruh permukaan dunia yang diambil lewat pemetaan satelit, fotografi udara dan globe GIS 3D. Penggunaannya melalui program antarmuka Googlemaps berbasis web (*maps.google.com*) dan melalui Google Earth (*earth.google.com*).

Produk ini menggunakan satu tabel untuk memproses datanya dan satu tabel lainnya untuk melayani data yang masuk dari pengguna jasa. Tabel yang pertama adalah *preprocessing pipeline* yang digunakan untuk menyimpan citra yang mentah. *preprocessing pipeline* ini sangat bergantung pada MapReduce untuk mentransformasikan datanya. Kemudian ada tabel yang digunakan untuk melayani pengguna jasa. Tabel itu digunakan untuk mengindeks data yang disimpan dalam GFS. Besaran data ini relative kecil, hanya sekitar 500 GB, tetapi harus melayani puluhan ribu kueri per detik. Dalam hal ini Google menyimpan data jalan-jalan yang ada di seluruh dunia dalam satu tabel dan mengarsipkannya

untuk kemudian akan ditampilkan ketika pengguna jasa mencari data.

c. Personalized Search

Layanan ini merupakan layanan fitur pencarian yang dipersonalisasi yang diperkenalkan oleh Google pada tahun 2004. Bisa dibilang, semua pencarian di Google Search pasti dikaitkan dengan *cookie browser*. Hal ini karena fungsi dari *Personalized Search* adalah untuk merekam pencarian web, gambar dan berita kemudian pengguna jasa dapat menelusurinya kembali riwayat pencarian mereka dalam *Google Search*. Hal ini didasari dari pergerakan dan tindakan pengguna jasa disimpan dalam sebuah tabel yang bernama *userid*. Segala pencarian yang dipersonalisasi menyimpan data masing-masing pengguna jasanya dalam *userid*. Melalui *Personalize Search* seseorang dapat dengan mudah membuka data sebelumnya dengan cepat sebab data tersebut telah tersimpan sebelumnya.

KESIMPULAN

Big Data merupakan sebuah pengembangan baru dari teknologi internet. Melalui *Big Data*, setiap informasi atau data dapat tersimpan dalam jumlah yang besar tanpa batasan waktu. Masa sekarang, *Big Data* dimanfaatkan secara cerdas oleh sebuah perusahaan bernama Google.Inc yang didirikan oleh Sergey Brin dan Larry Page pada tahun 1998. Perusahaan tersebut bergerak di bidang mesin pencari (*search*

engine) yang kini telah merambah ke banyak sektor. Dibandingkan dengan mesin pencari lainnya Google jauh lebih unggul dan lebih diminati oleh pengguna. Sehingga menimbulkan banyaknya data yang terdapat pada server Google.

Dalam rangka mempermudah pencarian informasi melalui Google, mereka kemudian mengeluarkan sebuah platform bernama Google Bigtable. Google Bigtable merupakan sistem database berskala besar dan cepat yang digunakan Google untuk mengolah berbagai jenis data dari berbagai layanan google, termasuk layanan dari mesin pencarian google yang berbasis internet. Pada penerapannya Bigtable dapat dilihat dalam tiga aspek yaitu Google Analyst yang merupakan layanan analisis data gratis bagi seorang Webmaster. Kedua adalah Google Earth yaitu merupakan sebuah tiruan dunia virtual yang dapat melihat penampakan dunia secara jelas. Terakhir adalah Personalize Search yang mempermudah seseorang dalam mencari data di platform Google. Melalui Bigtable ini Google menjadi perusahaan yang bergerak dibidang mesin pencari dan *Big Data* tersukses di dunia hingga memiliki data yang hampir tidak terhitung dalam servernya. Jika Google mengalami kebangkrutan, pengelolaan *Big Data* yang selama ini diakuisisi oleh google tidak akan menghilang. Google berada dibawah naungan Alphabet Inc, dan mempunyai banyak anak perusahaan. Pertama-tama yang harus Google lakukan adalah menjual aset-aset dan akuisisi yang dimiliki. Setiap

anak perusahaan yang berada di bawah google akan bertanggung jawab untuk data-data yang berada dalam database mereka, apakah itu cloud, keamanan internet, peta dan lainnya. Kemudian mereka akan menjadi sebuah perusahaan yang akan berdiri sendiri sebagai perusahaan yang terpisah baik di bawah nama perusahaan yang membeli mereka atau menjadi perusahaan yang tidak dinaungi perusahaan lain.

Jika Google memang berada pada kebangkrutan yang penuh, dalam arti kehilangan nama atau semua aset yang dimiliki, semua data yang berada dalam database Google tidak akan pernah hilang. Tetapi, akan ada banyak perusahaan lain yang ingin berinvestasi dan membeli semua data yang Google punya. Karena, akan ada banyak perusahaan yang ingin membayar untuk akses penuh, kontrol dan kepemilikan dari informasi digital yang sangat berharga. Data-data digital yang berada dalam database akan terus ada, jika bangkrut akan berpindah tangan lagi dan kepemilikan hingga jam ke-n atau waktu yang tak terhingga.

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada kehadiran yang maha kuasa karena telah memberikan kekuatan untuk menyelesaikan artikel ini. Hambatan serta halangan mampu penulis netralisir berkat bantuan yang maha kuasa. Sehingga tercipta karya yang tidak terkira tenaga serta usaha yang telah diberikan.

Penulis juga mengucapkan terimakasih kepada pihak-pihak yang telah membantu melancarkan publikasi artikel ini. Baik dari pihak KSM Eka Prasetya UI yang telah menjalin hubungan baik dengan ANRI, serta untuk Bapak Hary yang telah memberikan kesempatan kepada KSM Eka Prasetya UI. Tidak lupa ucapan terimakasih diberikan kepada seluruh redaksi jurnal ANRI yang telah memberi masukan kepada penulis sehingga akhirnya artikel ini dapat dipublikasikan.

DAFTAR PUSTAKA

- Abdullah, N., Ismail, S.A., Sophiayati, S. dan Sam, S.M., 2015. Data Quality in Big Data: A Review. *International Journal Advance Soft Computer. Applied*, Vol. 7, No. 3, hlm 17-27.
- Admindoxa. 2019. *Apa itu Google Analytics? Tools Wajib Yang Pemilik Situs.* (Online). <https://www.doxadigital.com/opti-masi-web/apa-itu-google-analytics/>. (Diakses pada 20 Mei 2020).
- Alfi. 2019. *Inilah Total Jumlah Data di Internet.* (Online). <https://teknorus.com/jumlah-data-di-internet/>. (Diakses pada 14 Mei 2020).
- Ali, Irhamni. 2015. Big Data: Apa dan Pengaruhnya pada Perpustakaan. *Jurnal Media Pustakawan*, Vol. 12, No. 1, hlm 15-19.
- APJII. 2019. *Laporan Survei Penetrasi & Profil Perilaku Pengguna Internet Indonesia.* (Online). <https://apjii.or.id/survei2018s/download/RYAmnx58wdrPs21VaMiB0qlytF7OKQ>. (Diakses pada 15 Mei 2020).
- Archieve. 2020. *Internet Archieve.* (Online). <https://archive.org/>. (Diakses pada 14 Mei 2020).
- Battelle, J., 2011. *The search: How Google and Its rivals Rewrote the Rules of Business and Transformed Our Culture.* United Kingdom: Hachette.
- Bellis, Mary. 2020. *The History of Google and How It Was Invented.* (Online). <https://www.thoughtco.com/who-invented-google-1991852>. (Diakses 20 Mei 2020).
- Bollier, D. dan Firestone, C.M., 2010. *The Promise and Peril of Big Data* (hlm. 1-66). Washington: Aspen Institute, Communications and Society Program.
- Chang, F., Dean, J., Ghemawat, S., Hseih, W.C., Wallach, D. A., Burrows, M., ... & Gruber, R.E. 2008. Bigtable : A Distributed Storage System for Structured Data. *ACM*

- Transactions on Computer System*, Vol 26, No 2, hlm 1-26.
- Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C., Wallach, D.A., Burrows, M., Chandra, T., Fikes, A. and Gruber, R.E., 2006. Bigtable: A Distributed Storage System for Structured Data. *ACM Transaction on Computer System*, Vol 26 No. 4, hlm 1-4.
- Cholissodin, I. dan Riyandani, E., 2016. *Analisis Big Data*. Malang: Fakultas Ilmu Komputer (Fikom), Universitas Brawijaya (UB).
- CNN Indonesia. 2019. *Mengenal Sejarah Internet*. (Online). <https://www.cnnindonesia.com/teknologi/20190312125646-185-376484/mengenal-sejarah-internet>. (Diakses pada 14 Mei 2020).
- Davenport, T.H. dan Dyché, J., 2013. *Big Data in Big Companies*. International Institute for Analytics, hlm 1-31.
- Dulal, E.R., *Big Data and Google File System*. Lalitpur: Nepal College of Information Technology (NCIT)
- Marr, B., Willey. *Big Data Case Study Collection*.
- Gandomi, Amir., dan Murtaza Haider. 2014. Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of*
- Information Management* Vol. 35, No 2, hlm 137-144.
- Google Cloud. 2007. *Overview of Cloud Bigtable*. <https://cloud.google.com/bigtable/docs/overview>. (10 Januari 2021).
- Hall, William L. 2020. *Google*. (Online). <https://www.britannica.com/topic/Google-Inc>. (Diakses pada 29 Mei 2020).
- Hillis, K., Petit, M. dan Jarrett, K., 2012. *Google and the Culture of Search*. Oxfordshire: Routledge.
- Hurwitz, J.S., Nugent, A., Halper, F. dan Kaufman, M., 2013. *Big Data for Dummies*. New Jersey: John Wiley & Sons.
- Itg.id. (2020). *ISO 27001 adalah Ikon Standarisasi Manajemen Keamanan Informasi*. <https://itgid.org/iso-27001-adalah/>. (Diakses pada 10 Januari 2021).
- Krishnan, K., 2013. *Data Warehousing in the Age of Big Data*. Oxford: Newnes.
- Kumar, M., 2016. Google Cloud Platform: a Powerful Big Data Analytics Cloud Platform. *International Journal for Research Applied Science & Engineering Technology*, Vol. 4 No. 11, hlm 387-392.

- Kune, R., Konugurthi, P.K., Agarwal, A., Chillarige, R.R. dan Buyya, R., 2016. The Anatomy of Big Data Computing. *Software: Practice and Experience*, Vol. 46 No. 1, hlm 79-105.
- Maestro. 2010. *Apa itu Google Earth*. (Online) <https://maestro.unud.ac.id/apa-itu-google-earth/>. (Diakses pada 20 Mei 2020).
- Martinez-Urbe, L., 2019. Digital Archives as Big Data. *Mathematical Population Studies*, Vol. 26, No. 2, hlm 69-79.
- Maryanto, B., 2017. Big Data dan Pemanfaatannya dalam Berbagai Sektor. *Media Informatika*, Vol. 16, No. 2, hlm 14-19.
- Milenkovic, Jovan. 2019. *30 Eye-Opening Big Data Statistics for 2020: Patterns are Everywhere*. Komando Tech. (Online) <https://kommandotech.com/statistics/big-data-statistics/>. (Diakses pada 7 Januari 2021)
- Muljono, P. 2010. *Manajemen Arsip dengan Sistem Modern*. (Online). <https://202.124.205.241/bitstream/handle/123456789/33842/KPMpj-m-Artik8-Manajemen%20arsip.pdf?sequence=1&isAllowed=y>.
- O'Connell, Brian. 2018. *History of Google: How It Began and What's Happening Beyond 2019*. (Online) <https://www.thestreet.com/technology/history-of-google-14820930>. (Diakses 20 Mei 2020).
- Ramadhan, H.A. dan Putri, D.A., 2018. *Big Data, Kecerdasan Buatan, Blockchain, dan Teknologi Finansial di Indonesia*. Jakarta: Direktorat Jenderal Aplikasi Informatika Kementerian Komunikasi dan Informasi.
- Rustam, Muhammad. 2014. *Pengelolaan Arsip Elektronik*. (Online). <http://repository.ut.ac.id/4158/1/A-SIP4429-M1.pdf>. (Diakses pada 14 Mei 2020).
- Schmidt, E. dan Rosenberg, J., 2014. *How Google Works*. United Kingdom: Hachette.
- Sirait, E.R.E., 2016. Implementasi Teknologi Big Data di Lembaga Pemerintahan Indonesia. *Jurnal Penelitian Pos dan Informatika*, Vol. 6, No, hlm 113-136.
- Teknologi Bigdata. 2013. *Big Data dan Rahasia Kejayaan Google*. (Online). <http://www.teknologi-bigdata.com/2013/02/big-data-dan-rahasia-kejayaan-google.html>. (Diakses pada 21 Juni 2020).

Zinchenko, Y., 2017. *Big Data and Google Cloud Platform: Usage to Create Interactive Reports*. Helsinki: Helsinki Metropolia University of Applied Sciences.