



**Derivatif: Jurnal Manajemen Ekonomi dan Akuntansi**  
Yayasan Salmiah Education Global International  
(YSEGI)

Jl. Pendidikan, Kec. Percut Sei Rotan, Kab. Deli Serdang, Sumatera Utara, 21333

Website: <https://glonus.org/index.php/derivatif> Email: [glonus.info@gmail.com](mailto:glonus.info@gmail.com)

## **Manajemen Data di Google: Praktik, Infrastruktur, dan Tantangan**

**Fahrezi Arman<sup>1</sup>, Muhammad Irwan Padli Nasution<sup>2</sup>**

<sup>1,2</sup>Universitas Islam Negeri Sumatra Utara, Indonesia

<sup>1</sup>fahreziarman627@gmail.com

### **Abstrak**

Penelitian ini bertujuan untuk mengkaji secara mendalam praktik manajemen data di Google dengan meninjau infrastruktur teknologi yang digunakan serta tantangan yang dihadapi dalam pengelolaan data berskala besar. Sebagai perusahaan teknologi global, Google memproses dan menyimpan data dalam jumlah masif setiap detik, sehingga memerlukan sistem manajemen data yang efisien, aman, dan skalabel. Studi pustaka ini menganalisis berbagai sumber ilmiah, termasuk jurnal internasional, artikel teknis, dan dokumentasi resmi Google, untuk mengidentifikasi praktik terbaik dalam pengumpulan, penyimpanan, pemrosesan, dan analitik data. Hasil penelitian menunjukkan bahwa Google menggunakan sistem infrastruktur canggih seperti Google File System (GFS), Bigtable, Spanner, dan Dremel, yang mendukung efisiensi dan keandalan pemrosesan data secara real-time. Namun, tantangan signifikan masih dihadapi, terutama dalam aspek privasi pengguna, kepatuhan terhadap regulasi global, serta efisiensi energi pusat data. Penelitian ini memberikan kontribusi terhadap pemahaman manajemen data modern dan dapat menjadi acuan bagi organisasi lain dalam merancang sistem data yang tangguh dan berkelanjutan.

**Kata Kunci:** Manajemen Data, Google, Infrastruktur Data, Big Data, Tantangan Teknologi

### **Pendahuluan**

Perkembangan teknologi informasi telah membawa transformasi besar dalam cara organisasi mengelola dan memanfaatkan data. Di era digital, data bukan hanya aset pendukung, tetapi menjadi fondasi utama dalam pengambilan keputusan strategis dan inovasi layanan. Google, sebagai salah satu perusahaan teknologi terbesar di dunia, memiliki sistem manajemen data yang kompleks dan canggih untuk mendukung operasionalnya yang berskala global. Manajemen data di Google mencakup berbagai proses mulai dari pengumpulan, penyimpanan, pengolahan, hingga analitik data dalam jumlah yang sangat besar dan terus bertambah secara real-time (Dean & Ghemawat, 2004; Chang et al., 2008).

Google dikenal sebagai pelopor dalam pengembangan infrastruktur data yang inovatif, seperti Google File System (GFS), MapReduce, Bigtable, dan Spanner, yang kini menjadi standar dalam teknologi big data (Ghemawat et al., 2003; Corbett et al., 2012). Sistem ini memungkinkan Google mengelola miliaran permintaan pencarian, video YouTube, dan layanan lainnya dengan efisien dan reliabel. Namun, di balik keberhasilan tersebut, terdapat berbagai tantangan serius yang dihadapi, seperti pengelolaan privasi pengguna, kepatuhan terhadap regulasi data lintas negara, serta efisiensi energi di pusat data berskala besar (Zaharia et al., 2010; Wu et al., 2016).

Dalam era big data, kemampuan untuk mengelola data secara efektif menjadi keunggulan kompetitif yang krusial bagi perusahaan teknologi. Google merupakan contoh nyata entitas yang sangat tergantung pada efektivitas sistem pengelolaan datanya. Infrastruktur teknis seperti GFS (Google File System), Bigtable, dan Spanner bukan hanya memungkinkan penyimpanan data dalam skala petabyte, tetapi juga menjamin konsistensi, kecepatan, dan ketersediaan tinggi (Dean & Ghemawat, 2004; Chang et al., 2008; Corbett et al., 2012).

Namun demikian, kompleksitas sistem ini juga membawa tantangan yang tidak sedikit. Salah satu isu utama adalah bagaimana menjaga privasi dan keamanan data pengguna di tengah tekanan regulasi yang semakin ketat, seperti General Data Protection Regulation (GDPR) di Uni Eropa (Tikkinen-Piri et al., 2018). Selain itu, pengoperasian pusat data yang tersebar di berbagai belahan dunia membutuhkan konsumsi energi yang tinggi, sehingga menimbulkan isu keberlanjutan dan efisiensi energi (Shehabi et al., 2016).

Permasalahan lainnya adalah bagaimana memastikan bahwa sistem data yang besar dan tersebar ini tetap adaptif terhadap perubahan teknologi, kebutuhan bisnis, serta volume data yang terus meningkat. Dengan melakukan studi pustaka yang komprehensif, penelitian ini mencoba menggambarkan praktik manajemen data Google secara menyeluruh, mengidentifikasi tantangan kritis, serta menawarkan perspektif teoritis dan praktis terhadap pengelolaan data berskala besar.

Sejumlah penelitian terdahulu telah membahas infrastruktur dan teknologi manajemen data yang digunakan oleh Google, seperti Google File System (GFS), MapReduce, Bigtable, dan Spanner (Ghemawat et al., 2003; Dean & Ghemawat, 2004; Chang et al., 2008; Corbett et al., 2012). Penelitian-penelitian tersebut secara teknis menjelaskan bagaimana sistem ini bekerja untuk menangani data berskala besar secara efisien dan terdistribusi. Selain itu, beberapa studi juga telah menyoroti penggunaan sistem analitik seperti Dremel dan Pregel untuk kebutuhan analisis data real-time dan pemrosesan graf berskala besar (Melnik et al., 2010; Malewicz et al., 2010).

Namun demikian, sebagian besar kajian tersebut hanya berfokus pada aspek teknis dan arsitektural, tanpa mengintegrasikan pembahasan yang holistik mengenai tantangan manajemen data modern seperti isu privasi data pengguna, kepatuhan terhadap regulasi internasional (misalnya GDPR), serta efisiensi energi dan keberlanjutan operasional pusat data (Tikkinen-Piri et al., 2018; Shehabi et al., 2016). Selain itu, belum banyak penelitian studi pustaka yang mengkaji praktik dan strategi manajemen data Google secara menyeluruh, dari hulu ke hilir, yang mengaitkan antara arsitektur teknologi, operasional bisnis, dan tanggung jawab sosial korporasi dalam satu kerangka analisis terpadu.

Dengan demikian, terdapat gap riset dalam hal pendekatan multidisipliner yang menggabungkan perspektif teknologi, kebijakan data, dan tantangan lingkungan dalam konteks pengelolaan data masif oleh perusahaan teknologi raksasa seperti Google. Penelitian ini mencoba untuk mengisi kekosongan tersebut dengan menelaah secara komprehensif praktik manajemen data Google dari berbagai sudut pandang, berdasarkan studi pustaka yang terkini dan relevan.

Kebaruan dalam penelitian ini terletak pada pendekatannya yang interdisipliner dan komprehensif dalam menganalisis manajemen data di Google. Tidak seperti penelitian sebelumnya yang cenderung bersifat fragmentaris dan teknosentris, studi pustaka ini menghadirkan integrasi antara aspek teknis (seperti GFS, Spanner, BigQuery), aspek kebijakan dan regulasi data (seperti GDPR, CCPA), serta aspek keberlanjutan operasional (seperti efisiensi energi dan pengurangan emisi karbon di pusat data) sebagai satu kesatuan sistem manajemen data yang saling terkait.

Selain itu, penelitian ini juga menawarkan pendekatan evaluatif terhadap bagaimana praktik manajemen data Google dapat dijadikan model adaptif untuk organisasi lain yang ingin menerapkan sistem pengelolaan data berskala besar dengan prinsip efisiensi, keamanan, dan keberlanjutan. Studi ini juga menambahkan perspektif kritis mengenai batas-batas teknologi Google dalam menghadapi dinamika regulasi global dan ekspektasi publik terhadap transparansi serta tanggung jawab etis dalam pengelolaan data (Zuboff, 2019; Pasquale, 2015).

Studi pustaka ini bertujuan untuk mengeksplorasi secara sistematis bagaimana Google membangun dan mengelola sistem data raksasanya, serta mengidentifikasi tantangan dan pendekatan solusi yang telah atau sedang diupayakan. Dengan menelaah berbagai literatur ilmiah dan dokumentasi teknis, penelitian ini diharapkan dapat memberikan wawasan kritis mengenai praktik manajemen data modern yang dapat dijadikan rujukan oleh institusi maupun pengembang teknologi di bidang serupa.

## Metode

Penelitian ini menggunakan pendekatan studi pustaka (literature review) yang bersifat kualitatif-deskriptif dengan tujuan untuk mengkaji dan menganalisis berbagai literatur ilmiah serta dokumen teknis yang relevan terkait manajemen data di Google. Studi pustaka dipilih karena sesuai untuk menghimpun dan mensintesis informasi dari berbagai sumber terdokumentasi guna memperoleh pemahaman yang komprehensif terhadap topik yang diteliti (Snyder, 2019).

Sumber data yang digunakan dalam penelitian ini berasal dari jurnal ilmiah internasional yang terindeks Scopus dan Web of Science, prosiding konferensi teknologi (seperti ACM, IEEE, VLDB), buku-buku akademik, white paper teknis Google, serta kebijakan privasi dan dokumen legal seperti regulasi GDPR dan CCPA. Kriteria inklusi data mencakup: (1) publikasi antara tahun 2003 hingga 2024, (2) berfokus pada sistem penyimpanan, pemrosesan, keamanan, dan etika data, serta (3) relevan dengan konteks pengelolaan data berskala besar di Google. Sementara itu, literatur yang tidak bersumber dari publikasi ilmiah terpercaya atau bersifat opini tanpa dasar metodologis dieksklusi dari analisis.

Langkah-langkah dalam studi pustaka ini mengikuti panduan metodologis dari Kitchenham (2004), yaitu: (1) perumusan pertanyaan penelitian, (2) identifikasi sumber literatur, (3) seleksi dan evaluasi literatur berdasarkan kualitas dan relevansi, (4) ekstraksi dan sintesis informasi inti, serta (5) interpretasi temuan secara tematik. Analisis dilakukan dengan teknik analisis isi tematik (*thematic content analysis*), di mana data dikategorisasi berdasarkan tema utama, yaitu: praktik manajemen data, infrastruktur teknologi, serta tantangan etis, hukum, dan keberlanjutan dalam pengelolaan data.

Peneliti juga menggunakan pendekatan triangulasi sumber untuk meningkatkan validitas temuan, dengan membandingkan informasi dari jurnal ilmiah, publikasi teknis resmi Google, dan laporan kebijakan privasi dari lembaga regulator. Metode ini memberikan gambaran yang lebih utuh dan mendalam mengenai kompleksitas manajemen data Google dalam perspektif teknologi dan sosial. Dengan pendekatan ini, penelitian diharapkan dapat memberikan kontribusi teoretis dan praktis terhadap pemahaman pengelolaan data berskala besar serta menjadi rujukan dalam pengembangan sistem data korporasi yang beretika dan berkelanjutan.

## Hasil dan Pembahasan

Hasil studi pustaka menunjukkan bahwa Google menerapkan pendekatan manajemen data yang sangat terstruktur, inovatif, dan berskala global. Google telah mengembangkan

sistem infrastruktur data internal yang tidak hanya mampu menangani data dalam skala petabyte, tetapi juga memastikan kecepatan, konsistensi, dan reliabilitas tinggi dalam setiap prosesnya (Ghemawat et al., 2003; Dean & Ghemawat, 2004). Sistem utama yang menjadi tulang punggung manajemen data Google meliputi:

1. Google File System (GFS) – dirancang untuk menangani file berukuran besar dan distribusi data yang luas, dengan toleransi terhadap kegagalan sebagai bagian dari desain utamanya (Ghemawat et al., 2003).
2. MapReduce – framework pemrosesan paralel yang memungkinkan pemrosesan data dalam jumlah besar secara efisien dengan prinsip *divide and conquer* (Dean & Ghemawat, 2004).
3. Bigtable – sistem penyimpanan data terstruktur berskala besar yang mendukung berbagai aplikasi Google seperti Gmail, Search, dan Maps (Chang et al., 2008).
4. Spanner – database relasional terdistribusi global yang menjamin konsistensi kuat dan kemampuan sinkronisasi lintas benua (Corbett et al., 2012).
5. Dremel dan BigQuery – digunakan untuk analisis data interaktif dan berskala besar dengan latensi rendah, mendukung pengambilan keputusan berbasis data real-time (Melnik et al., 2010).

Infrastruktur ini dibangun di atas prinsip *horizontal scalability*, *fault tolerance*, dan *high availability*, memungkinkan Google untuk memproses lebih dari 100 miliar pencarian per bulan dan menyimpan eksabita data pengguna di seluruh dunia. Dengan menerapkan sistem internal yang dirancang sendiri, Google tidak hanya mengandalkan teknologi komersial, tetapi juga mendorong batas kemampuan komputasi awan dan data science (Vahdat, 2010). Namun demikian, temuan juga menunjukkan adanya sejumlah tantangan besar dalam manajemen data Google.

### **Isu Privasi dan Etika Data**

Dengan dominasi dalam pengumpulan data pengguna melalui produk seperti Gmail, YouTube, dan Chrome, Google menghadapi tekanan besar terkait perlindungan data pribadi. Penerapan regulasi seperti General Data Protection Regulation (GDPR) di Uni Eropa dan California Consumer Privacy Act (CCPA) di AS memaksa Google untuk mengembangkan kebijakan internal yang ketat dalam hal transparansi, persetujuan pengguna, dan hak untuk dilupakan (*right to be forgotten*) (Tikkinen-Piri et al., 2018).

### **Tantangan Energi dan Keberlanjutan**

Pusat data Google yang tersebar di seluruh dunia menggunakan energi dalam jumlah

besar. Meskipun Google telah berkomitmen pada energi terbarukan dan net-zero emissions, pengelolaan data dalam skala besar tetap menjadi ancaman terhadap keberlanjutan ekologis global (Shehabi et al., 2016). Bahkan, laporan Greenpeace menyebut pusat data sebagai salah satu kontributor emisi karbon digital terbesar secara global.

### **Skalabilitas dan Evolusi Teknologi**

Pertumbuhan eksponensial volume data dan kompleksitas aplikasi mendorong Google untuk terus berinovasi. Sistem yang dahulu revolusioner seperti MapReduce kini dianggap kurang fleksibel dibandingkan dengan teknologi baru seperti Apache Spark dan Flink (Zaharia et al., 2010). Oleh karena itu, Google harus terus menyesuaikan diri dengan dinamika teknologi dan kebutuhan pengguna yang berubah dengan cepat.

### **Risiko Monopoli dan Black Box Algorithm**

Selain tantangan teknis dan legal, Google juga dikritik karena kurangnya transparansi dalam sistem algoritmiknya yang bersifat tertutup (Pasquale, 2015). Ini menimbulkan kekhawatiran akan bias algoritma dan konsentrasi kekuasaan informasi dalam satu korporasi global, yang berdampak pada keadilan sosial digital.

Dari hasil temuan tersebut, dapat disimpulkan bahwa keberhasilan manajemen data Google tidak hanya terletak pada kekuatan teknologinya, tetapi juga pada kemampuannya mengintegrasikan sistem data dengan kebutuhan bisnis dan kepatuhan terhadap kebijakan global. Google menjadi contoh nyata bagaimana perusahaan dapat merancang sistem informasi berkelas dunia yang efisien, adaptif, dan berbasis inovasi internal. Namun, tantangan yang dihadapi juga memperlihatkan pentingnya pendekatan multidisipliner dalam manajemen data modern di mana aspek teknis harus diseimbangkan dengan dimensi etis, hukum, dan lingkungan.

Studi pustaka ini mempertegas pentingnya membangun sistem data yang tidak hanya kuat secara teknologis, tetapi juga akuntabel secara sosial dan berkelanjutan secara lingkungan, terutama bagi perusahaan yang bergerak dalam skala global. Oleh karena itu, temuan ini dapat menjadi rujukan penting bagi organisasi, pengembang sistem, maupun pembuat kebijakan dalam merancang dan mengelola ekosistem data yang bertanggung jawab.

### **Kesimpulan**

Berdasarkan hasil telaah literatur ilmiah dan dokumentasi teknis, dapat disimpulkan bahwa Google telah membangun sistem manajemen data yang sangat kompleks, efisien, dan

inovatif guna menangani volume data dalam skala eksabita. Praktik manajemen data Google didukung oleh infrastruktur canggih seperti Google File System (GFS), Bigtable, MapReduce, Spanner, hingga Dremel dan BigQuery yang menjamin performa tinggi, skalabilitas global, dan konsistensi data. Infrastruktur ini memungkinkan pemrosesan data secara real-time, terdistribusi, dan andal dalam berbagai produk dan layanan Google. Namun demikian, di balik keunggulan teknis tersebut, Google juga menghadapi sejumlah tantangan serius. Tantangan tersebut mencakup isu privasi dan keamanan data pengguna yang semakin kompleks seiring dengan penerapan regulasi global seperti GDPR dan CCPA, konsumsi energi yang tinggi di pusat data yang menimbulkan dampak ekologis, serta tekanan etis dan sosial terkait dominasi algoritma dan monopoli data. Oleh karena itu, manajemen data tidak hanya dituntut unggul secara teknologi, tetapi juga harus adaptif terhadap aspek hukum, sosial, dan keberlanjutan. Penelitian ini menunjukkan bahwa Google adalah contoh perusahaan yang berhasil mengintegrasikan manajemen data dengan strategi bisnis dan inovasi teknologi, namun tetap membutuhkan pendekatan yang lebih inklusif, transparan, dan bertanggung jawab dalam menjawab tantangan masa depan. Organisasi lain yang ingin meniru model manajemen data Google perlu memahami bahwa teknologi saja tidak cukup. Perlu penguatan kebijakan privasi, tata kelola data yang transparan, serta kesiapan infrastruktur berkelanjutan. Adaptasi teknologi seperti Bigtable atau Spanner harus disertai dengan analisis konteks kebutuhan dan kemampuan sumber daya organisasi. Diperlukan pengembangan lanjutan terhadap sistem manajemen data yang lebih efisien energi dan bersifat open-source untuk mendemokratisasi akses terhadap teknologi canggih. Peneliti juga disarankan melakukan eksplorasi lebih lanjut terhadap interaksi antara arsitektur data dan regulasi lintas negara dalam konteks globalisasi digital. Perlu kebijakan yang lebih progresif dan adaptif dalam menghadapi perkembangan sistem manajemen data berskala global. Regulasi harus mampu menjembatani kepentingan privasi pengguna, hak atas data pribadi, dan inovasi teknologi, dengan menekankan prinsip akuntabilitas, keadilan data, dan keberlanjutan ekosistem digital. Edukasi literasi data perlu ditingkatkan agar masyarakat memahami bagaimana data mereka dikumpulkan, diproses, dan dimanfaatkan. Kesadaran ini penting untuk membangun kontrol individu terhadap data pribadi dan mendorong keterlibatan aktif dalam tata kelola data digital yang etis dan transparan.

## Daftar Pustaka

Abadi, D. J. (2009). *Data management in the cloud: Limitations and opportunities*. *IEEE Data Engineering Bulletin*, 32(1), 3–12.

- Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). *Big data security and privacy in healthcare: A review*. *Procedia Computer Science*, 113, 73–80.
- Al-Roomi, M., Al-Ebrahim, S., Buqrais, S., & Ahmad, I. (2013). *Cloud computing pricing models: A survey*. *International Journal of Grid and Distributed Computing*, 6(5), 93–106.
- Barroso, L. A., Clidaras, J., & Hölzle, U. (2013). *The datacenter as a computer: An introduction to the design of warehouse-scale machines*. *Synthesis Lectures on Computer Architecture*, 8(3), 1–154.
- Chang, F., Dean, J., Ghemawat, S., et al. (2008). *Bigtable: A distributed storage system for structured data*. *ACM Transactions on Computer Systems (TOCS)*, 26(2), 1–26.
- Corbett, J. C., Dean, J., Epstein, M., et al. (2012). *Spanner: Google's globally distributed database*. *ACM Transactions on Computer Systems (TOCS)*, 31(3), 1–22.
- Dean, J., & Ghemawat, S. (2004). *MapReduce: Simplified data processing on large clusters*. *OSDI*, 137–150.
- Ghemawat, S., Gobioff, H., & Leung, S. T. (2003). *The Google file system*. *ACM SIGOPS Operating Systems Review*, 37(5), 29–43.
- Google Cloud. (2022). *Building data pipelines with BigQuery and Dataflow*. *Google Cloud Technical Whitepaper*.
- Malewicz, G., Austern, M. H., Bik, A. J., et al. (2010). *Pregel: A system for large-scale graph processing*. *ACM SIGMOD*, 135–146.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. *Houghton Mifflin Harcourt*.
- Melnik, S., Gubarev, A., Long, J. J., et al. (2010). *Dremel: Interactive analysis of web-scale datasets*. *Proceedings of the VLDB Endowment*, 3(1–2), 330–339.
- Pasquale, F. (2015). *The Black Box Society: The secret algorithms that control money and information*. *Harvard University Press*.
- Reinsel, D., Gantz, J., & Rydning, J. (2017). *Data Age 2025: The Evolution of Data to Life-Critical*. *IDC White Paper*.
- Ristenpart, T., Tromer, E., Shacham, H., & Savage, S. (2009). *Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds*. *ACM CCS*, 199–212.
- Shehabi, A., Smith, S. J., Sartor, D. A., et al. (2016). *United States data center energy usage report*. *Lawrence Berkeley National Laboratory*.
- Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). *EU GDPR: Changes and implications for personal data collecting companies*. *Computer Law & Security Review*, 34(1), 134–153.
- Vahdat, A. (2010). *Google's infrastructure for the cloud: The datacenter as a computer*.

*Communications of the ACM*, 53(1), 72–79.

Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). *Spark: Cluster computing with working sets*. *HotCloud*, 10(10–10), 95.

Zuboff, S. (2019). *Surveillance capitalism and the challenge of digital governance*. *Journal of Information Technology*, 34(1), 1–20.