



Editor:
**Wahyuddin S.
Dwiny Meidelfi**

DATA LAKE INSIGHTS

Tim Penulis:

**Widyastuti Andriyani, Aisyah Mutia Dawis, Rakhmat Purnomo,
Fajar Ciputra Daeng Bani, Mohammad Santosa Mulyo Diningrat,
Yusuf Wahyu Setiya Putra, Novi**

DATA LAKE INSIGHTS

Tim Penulis:

**Widyastuti Andriyani, Aisyah Mutia Dawis, Rakhmat Purnomo,
Fajar Ciputra Daeng Bani, Mohammad Santosa Mulyo Diningrat,
Yusuf Wahyu Setiya Putra, Novi**

DATA LAKE INSIGHTS

Tim Penulis:

**Widyastuti Andriyani, Aisyah Mutia Dawis, Rakhmat Purnomo,
Fajar Ciputra Daeng Bani, Mohammad Santosa Mulyo Diningrat,
Yusuf Wahyu Setiya Putra, Novi**

Desain Cover:

Septian Maulana

Sumber Ilustrasi:

www.freepik.com

Tata Letak:

Handarini Rohana

Editor:

**Wahyuddin S.
Dwiny Meidelfi**

ISBN:

978-623-459-624-3

Cetakan Pertama:

Agustus, 2023

Tanggung Jawab Isi, pada Penulis

Hak Cipta Dilindungi Oleh Undang-Undang

by Penerbit Widina Media Utama

Dilarang keras menerjemahkan, memfotokopi, atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari Penerbit.

PENERBIT:

WIDINA MEDIA UTAMA

Komplek Puri Melia Asri Blok C3 No. 17 Desa Bojong Emas
Kec. Solokan Jeruk Kabupaten Bandung, Provinsi Jawa Barat

Anggota IKAPI No. 360/JBA/2020

Website: www.penerbitwidina.com

Instagram: [@penerbitwidina](https://www.instagram.com/penerbitwidina)

Telepon (022) 87355370

KATA PENGANTAR

Buku ini merupakan hasil kerja keras dan kolaborasi antara sejumlah ahli data yang berpengalaman, yang berbagi pengetahuan dan wawasan mereka tentang *Data Lake*. Dalam era di mana data menjadi komoditas yang sangat berharga, penting bagi kita untuk memiliki pemahaman yang mendalam tentang cara mengelola dan menganalisis data dengan efektif. *Data Lake* telah muncul sebagai paradigma yang kuat dalam pengelolaan data, memungkinkan kita untuk menyimpan, mengolah, dan menggali potensi data dalam skala yang belum pernah terbayangkan sebelumnya.

Melalui buku ini, kami berusaha untuk memberikan gambaran komprehensif tentang konsep, prinsip, dan implementasi *Data Lake*. Kami menjelaskan tentang arsitektur *Data Lake* yang fleksibel, infrastruktur yang diperlukan, serta alat dan teknologi yang dapat digunakan untuk membangun dan mengelola *Data Lake* dengan efisiensi dan efektivitas maksimal. Selain itu, kami juga membahas tentang bagaimana *Data Lake* dapat memberikan wawasan yang berharga melalui analisis data yang mendalam. Kami membahas teknik dan praktik terbaik untuk memproses, menggali, dan menganalisis data di dalam *Data Lake*, dengan contoh kasus nyata yang memberikan ilustrasi konkret tentang manfaat yang dapat diperoleh.

Selain aspek teknis, kami juga tidak melupakan pentingnya aspek keamanan, privasi, dan kepatuhan dalam pengelolaan *Data Lake*. Kami memberikan wawasan tentang kerangka kerja keamanan data yang kokoh, serta praktik terbaik dalam menjaga kerahasiaan dan integritas data yang disimpan di dalam *Data Lake*. Kami ingin menyampaikan rasa terima kasih kepada tim penulis yang berdedikasi dan berkomitmen untuk menyampaikan pengetahuan mereka dalam buku ini. Tanpa upaya mereka, buku ini tidak akan menjadi kenyataan.

Akhir kata, kami berharap buku "*Data Lake Insights*" ini memberikan pemahaman yang mendalam dan praktis kepada Anda tentang bagaimana memanfaatkan *Data Lake* secara efektif. Semoga buku ini menginspirasi dan memberikan wawasan baru bagi Anda dalam menjelajahi dunia yang kaya potensi dari data.

Salam,

Tim Penulis

DAFTAR ISI

KATA PENGANTAR	iii
DAFTAR ISI	v
BAB 1 DATA LAKE	1
A. Mengenal <i>Data Lake</i>	1
B. <i>Data Lake Maturity</i>	5
C. Membuat <i>Data Lake</i> yang Berhasil	6
D. <i>Roadmap Data Lake</i>	9
E. <i>Arsitektur Data Lake</i>	11
F. Kesimpulan	13
BAB 2 STRATEGI PENAMBANGAN DATA LAKE	15
A. Penambangan <i>Data Lake</i>	15
B. Pentingnya Penambangan <i>Data Lake</i>	18
C. Strategi Penambangan <i>Data Lake</i>	20
D. Penutup	24
BAB 3 PEMROSESAN DATA LAKE	27
A. Penambangan <i>Data Lake</i>	27
B. Metode Pemrosesan <i>Data Lake</i>	28
C. <i>Tools</i> Pemrosesan <i>Data Lake</i>	30
D. Studi Kasus Penerapan <i>Data Lake</i> di <i>ruparupa.com</i>	31
E. Menggunakan Layanan AWS	34
F. Penutup	37
BAB 4 MANAJEMEN METADATA PADA DATA LAKE	39
A. Pendahuluan	39
B. <i>Standard</i> Metadata	40
C. Proses Metadata	45
D. Penutup	50
BAB 5 KEAMANAN DATA DI DATA LAKE	53
A. Ancaman Terhadap Keamanan Data di <i>Data Lake</i> /Pendahuluan	53
B. Tantangan Keamanan <i>Data Lake</i>	55
C. Skala Data yang Besar	57
D. Keanekaragaman Sumber Data Keamanan di <i>Data Lake</i>	59

E. Kekurangan Kesadaran Keamanan.....	61
F. Tips Terhindar dari Masalah Keamanan di <i>Data Lake</i>	63
BAB 6 PRODUK PERANGKAT LUNAK DATA LAKE	65
A. Pengantar	65
B. Implementasi <i>Data Lake</i>	65
C. Produk <i>Data Lake</i>	72
D. Penutup	78
BAB 7 DATA LAKE DALAM SEKTOR BISNIS	79
A. Pengantar	79
B. Perusahaan Menghadapi Berbagai Tantangan Data	81
C. Bagaimana <i>Data Lake</i> dapat Memberikan Solusi.....	81
D. Bagaimana <i>Data Lake</i> dapat Memberikan Solusi.....	83
E. Apakah Semua Perusahaan Membutuhkan <i>Data Lake</i> ?	84
F. Manfaat <i>Data Lake</i> dalam Sektor Bisnis	86
G. Siapa yang Diuntungkan dari Layanan <i>Data Lake</i> ?	88
H. Perusahaan Teratas yang Menyediakan Layanan <i>Data Lake</i>	90
I. Penutup	93
DAFTAR PUSTAKA.....	94
PROFIL PENULIS	99



DATA LAKE

A. MENGENAL DATA LAKE

Data lake adalah repositori terpusat yang memungkinkan menyimpan semua data terstruktur dari *database* relasional (baris dan kolom) dan tidak terstruktur (csv, log, xml, json), data tidak terstruktur (email, dokumen, pdf), dan data biner (gambar, audio, video) dalam skala apa pun. Menyimpan data apa adanya, tanpa harus menyusun data terlebih dahulu, dan menjalankan berbagai jenis analitik mulai dari *dashboard* dan visualisasi hingga pemrosesan data dalam jumlah besar, proses modernisasi pada platform data menjadi upaya yang semakin kompleks, dalam beberapa tahun terakhir semakin banyak solusi bermunculan, para pengembang aplikasi dihadapkan pada banyaknya pilihan dan menghabiskan banyak waktu untuk memutuskan pendekatan yang mana yang terbaik untuk menyelesaikan suatu proyek (Inmon 2016). Pada tahap ini kita terlebih dahulu memahami peran dari *Data Storage* yaitu *Database*, *Data Warehouse*, *Data Lake*. *Data Storage* merupakan masalah besar pada suatu perusahaan, seiring berjalannya waktu perusahaan berusaha memaksimalkan *value* dari potensi data dalam skala besar. *Data Storage* biasanya ditangani dalam *database* tradisional. Namun, seiringan penambahan jumlah data pada suatu perusahaan, kemudian beralih menggunakan *Data Warehouse* dan *Data Lake* (Nambiar and Mundra, 2022). Seringkali *Data Lake* dibandingkan dengan *Data Warehouse*, keduanya sangat berbeda mulai dari struktur dan pemrosesan kemudian siapa yang akan menggunakannya dan kenapa demikian akan dijelaskan secara sederhana pada buku ini.

BAB

2

STRATEGI PENAMBANGAN *DATA LAKE*

A. PENAMBANGAN *DATA LAKE*

Penambangan *data lake* merupakan metode komputasi yang digunakan dalam analitik untuk memproses dan menggali informasi dari kumpulan data yang besar. Dengan menggunakan alat dan teknik penambangan data, perusahaan dapat mengidentifikasi pola dan relasi tersembunyi dalam data mereka. Proses ini mengubah data mentah menjadi pengetahuan yang dapat diterapkan secara praktis. Informasi yang dihasilkan digunakan oleh perusahaan untuk memecahkan masalah, menganalisis dampak keputusan bisnis di masa depan, dan meningkatkan profitabilitas mereka.



Gambar 2.1. Ilustrasi Penambangan Data

(Sumber: www.freepik.com)



PEMROSESAN *DATA LAKE*

A. PENAMBANGAN *DATA LAKE*

Ide awal *Data Lake* berkembang di dorong oleh tantangan kebutuhan yang dihadapi perusahaan atau organisasi ketika menangani data yang besar, mulai dari cara mengumpulkannya, menyimpan, dan memprosesnya (Inmon, 2016). Data yang besar ini bermakna memiliki *field* dan *record* yang banyak. Setiap bagian di perusahaan biasanya memiliki aplikasi masing-masing. Misalnya aplikasi penggajian, aplikasi gudang, aplikasi sumber daya manusia, dan aplikasi lain yang menghasilkan banyak data. Data-data yang dihasilkan tidak terintegrasi dan belum di olah menjadi informasi dan pengetahuan baru. Saat itu, ada solusi *Data Mart* yang dapat mengekspor data dari aplikasi ke dalam *Raw Data*, namun hanya menyelesaikan sebagian masalah. Masalah lainnya seperti kepemilikan data, kemudahan akses data, ketersediaan data ketika dibutuhkan, dan hubungan antar data pada bagian yang berbeda-beda sehingga menghasilkan informasi dan pengetahuan yang berguna bagi perusahaan. Teknologi *Data Warehouse* hadir untuk menyelesaikan masalah ini, tetapi hanya menangani data terstruktur. Jika datanya semi-terstruktur atau tidak terstruktur, *data warehouse* belum berfungsi optimal. Dari masalah ini, solusi *Data Lake* hadir. *Data Lake* dapat menangani data terstruktur, semi-terstruktur, dan data tidak terstruktur (Grossman, 2019). Data pada *Data Lake* dapat berasal dari banyaknya pengguna yang melakukan “klik” di sosial media, data yang dihasilkan dari perangkat sensor berbasis *Internet of Things*, atau berbagai bentuk data seperti teks, audio, video, dokumen, *file*, dan berbagai tipe data lainnya. Dalam *Data Lake*, bukan saja memastikan data tersebut tersedia tetapi



MANAJEMEN METADATA PADA DATA LAKE

A. PENDAHULUAN

Perkembangan teknologi pada era digitalisasi dapat menyebabkan pengguna data dengan jumlah besar setiap hari. Sebelum kita dapat melakukan manajemen data dalam jumlah yang sangat besar, kita memerlukan konsep penyimpanan data yang efisien dengan menerima semua jenis data dari yang terstruktur hingga yang tidak terstruktur. Data yang dianggap sebagai aset yang tidak memiliki bentuk dan dapat memberikan nilai yang besar dalam mencapai tujuannya. Sumber data disediakan oleh semua sumber data yang dapat disediakan dalam pengumpulan data (Pohl, Staegemann *and* Turowski, 2022). Konsep pemrosesan metadata akan dirancang untuk mengelola informasi metadata yang diperlukan yang dapat digunakan untuk menghasilkan nilai dengan mengimplementasikan logika dan melakukan operasi pencocokan (Daeng Bani *et al.*, 2018).

Dengan menggunakan standar metadata dalam *data lake* sangat penting untuk memastikan konsistensi, interoperabilitas, dan pemahaman yang baik tentang data yang dikandungnya. Jika data tidak terstruktur yang disimpan dalam *data lake* tidak dianalisis dengan benar, data tersebut dapat diisi dengan informasi yang tidak relevan, yang pada gilirannya sulit dikelola dan dapat mengakibatkan data tidak dapat digunakan. Semua informasi yang diberikan oleh organisasi disimpan dalam struktur data yang disebut *data lake*. Data disimpan dalam bentuk aslinya di dalam *data lake*.

BAB

5

KEAMANAN DATA DI *DATA LAKE*

A. ANCAMAN TERHADAP KEAMANAN DATA DI *DATA LAKE*/ PENDAHULUAN

Data Lake merupakan wadah penyimpanan yang dapat menampung berbagai jenis data dari berbagai sumber. Meskipun memiliki manfaat yang besar dalam hal skalabilitas dan kemampuan analisis, *Data Lake* juga rentan terhadap berbagai ancaman keamanan yang harus diwaspadai. Dalam pembahasan ini, kita akan menjelajahi beberapa ancaman yang paling umum dan sering terjadi terhadap keamanan data di *Data Lake*



Gambar 5.1 Ilustrasi *Data Lake*

(Sumber: freepik.com)



PRODUK PERANGKAT LUNAK DATA LAKE

A. PENGANTAR

Pada era revolusi industri digital saat ini yang sudah mencapai generasi 5.0, banyak perusahaan-perusahaan besar yang memanfaatkan sumber data besar yang digunakan untuk keperluan dan kepentingan sistem kerja serta keberlangsungan perusahaan. Salah satu sumber data yang diperlukan dan dipergunakan adalah *Data Lake* atau danau data, seperti yang sudah dijelaskan pada bab sebelumnya dimana *data lake* adalah sebuah tempat penyimpanan data yang besar dimana data-data didalamnya masih bersifat mentah dengan 3 (tiga) bentuk format data yaitu terstruktur, setengah terstruktur dan tidak terstruktur. Format data-data tersebut akan tetap atau tidak akan diubah selama data itu belum digunakan, dan akan diubah sesuai kebutuhan apabila data tersebut akan dipergunakan.

Dari definisi, fungsi, tujuan, arsitektur dan manfaat dari danau data atau *data lake* yang telah dijelaskan pada bab sebelumnya, maka akan dibahas implementasi dari *data lake* itu sendiri dalam bentuk produk perangkat lunak yang digunakan oleh perusahaan ataupun perorangan. Produk-produk perangkat lunak *data lake* akan dibahas pada sub bab berikut.

B. IMPLEMENTASI DATA LAKE

Data Lake adalah indikator yang sangat penting bagi sebuah perusahaan. Terdapat banyak perusahaan yang bekerja menggunakan danau data ini seperti di sektor industri, bisnis, dll. Setiap perusahaan yang bekerja menggunakan danau

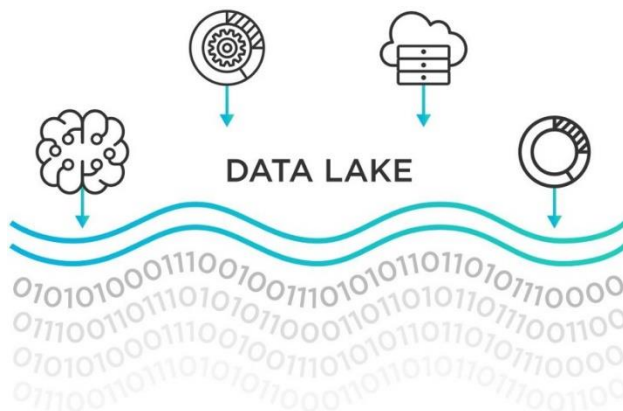
BAB 7

DATA LAKE DALAM SEKTOR BISNIS

A. PENGANTAR

Sebuah penelitian menunjukkan bahwa sekitar 2,5 triliun *byte* data dihasilkan setiap hari. Banyaknya volume data yang terkumpul adalah salah satu hal paling transformatif yang disaksikan dunia saat ini. Dan semua data tersebut sangat penting dalam keberhasilan bisnis, oleh sebab itu maka perusahaan harus dapat mengelolanya dengan baik.

Meskipun pengolahan data dimulai dengan *database*, tetapi banyak perusahaan mulai menggunakan data *warehouse* dan secara bertahap mengadopsi *data lake* untuk penyimpanan data dalam skala besar. *Data lake* disimpan dalam skenario IoT, karena dengan metode penyimpanan ini data relasional dan *nonrelasional* dapat disimpan dalam jumlah besar tanpa melakukan transformasi atau definisi skema terlebih dahulu.



Gambar 7.1

(sumber : <https://www.tibco.com>)

DAFTAR PUSTAKA

- Abdel-Basset, M., Manogaran, G., Chilamkurti, N., Mohamed, M., & Gupta, B. (2020). Blockchain-based secure data storage services for big data applications in the fog-cloud computing environment
- Aberdeen+Research++Angling+for+Insights+in+Today's+Data+Lake*. (n.d.). Scribd. <https://www.scribd.com/document/615300737/Aberdeen-Research-Angling-for-Insights-in-Today-s-Data-Lake>
- Action. (2022, September 25). *Data Lake Analytics - what is it and why does it matter?* | Action. <https://www.action.com/data-lake-analytics/>
- Admin. (2023). How a Data Lake Positively Impacts Your Business? *Intellectyx*. <https://www.intellectyx.com/blog/how-a-data-lake-positively-impacts-your-business/>
- Ashish, M. and Patil, N. (2022) 'A Novel Approach For Understanding Ideology Behind Managing Data Within Data Lake In Era Of Big Data', (July).
- AWS (2023) *rupa-rupa.com, Amazone Web Service*. Available at: <https://aws.amazon.com/id/blogs/indonesia/bagaimana-ruparupa-mendapatkan-insight-bisnis-terbaru-menggunakan-data-lake-amazon-s3-aws-glue-apache-hudi-dan-amazon-quicksight/>.
- AWS Prescriptive Guidance Designing a data lake for growth and scale on the AWS Cloud* (2023).
- AWS, Building Big Data Storage Solutions (Data Lakes) for Maximum Flexibility, 2019.
- Azeroual, O. et al. (2022) *Combining Data Lake and Data Wrangling for Ensuring Data Quality in CRIS*. Available at: <https://hal.science/hal-03694519v2>.
- Beheshti, A., Benatallah, B., Nouri, R., Chhieng, V.M., Xiong, H. and Zhao, X., 2017, November. Coredb: a data lake service. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 2451-2454).
- Biddle, K. (2023, February 7). The 5 leading companies in the Data Lake Solutions Industry. *BCC Research*. <https://blog.bccresearch.com/the-5-leading-companies-in-the-data-lake-solutions-industry>

- Brackenbury, W. *et al.* (2018) 'Draining the data swamp: A similarity-based approach', in *Proceedings of the Workshop on Human-In-the-Loop Data Analytics, HILDA 2018*. Association for Computing Machinery, Inc. Available at: <https://doi.org/10.1145/3209900.3209911>.
- Cloudera. (n.d.). Logo Gallery. Diakses pada 26 Mei 2023, dari <https://www.cloudera.com/about/company/brand-assets.html>
- Daeng Bani, F. C. *et al.* (2018) 'Implementation of Database Massively Parallel Processing System to Build Scalability on Process Data Warehouse', *Procedia Computer Science*, 135, pp. 68–79. doi: 10.1016/j.procs.2018.08.151.
- Data Lakes: What They Are and Why Companies Use Them*. (n.d.). Segment. <https://segment.com/blog/data-lakes/Databricks>. (n.d.). Brand Guidelines. Diakses pada 26 Mei 2023, dari <https://databricks.com/brand>
- Fang, H., 2015, June. Managing data lakes in big data era: What's a data lake and why has it become popular in data management ecosystem. In *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)* (pp. 820-824). IEEE.
- Freepik. (n.d.). Brand Resources. Diakses pada 26 Mei 2023, dari <https://www.freepikcompany.com/brand>
- Giebler, C. *et al.* (2019) 'Modeling data lakes with data vault: Practical experiences, assessment, and lessons learned', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Science and Business Media Deutschland GmbH, pp. 63–77. Available at: https://doi.org/10.1007/978-3-030-33223-5_7.
- Giebler, C., Gröger, C., Hoos, E., Eichler, R., Schwarz, H. and Mitschang, B., 2021. The Data Lake Architecture Framework. *BTW 2021*.
- Gorelik, A., 2019. The enterprise big data lake: Delivering the promise of big data and data science. O'Reilly Media, Inc, Sebastopol, California, Amerika.
- Grossman, R.L. (2019) 'Data Lakes , Clouds , and Commons : A Review of Platforms for Analyzing and Sharing Genomic Data', *Trends in Genetics*, xx, pp. 1–12. Available at: <https://doi.org/10.1016/j.tig.2018.12.006>.
- Hai, R., Quix, C. and Jarke, M., 2021. Data lake concept and systems: a survey. *arXiv preprint arXiv:2106.09592*.
- Inmon, B. (2016) *Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump*. Basking Ridge, USA: Technics Publication.

- Inmon, B. (no date) *Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump*.
- John, T. and Misra, P., 2017. *Data lake for enterprises*. Packt Publishing Ltd.
- John, T., & Misra, P., 2017. *Data lake for enterprises*. Packt Publishing Ltd, Birmingham, Britania Raya.
- Khine, P. P. and Wang, Z. S. (2018) 'Data lake: a new ideology in big data era', *ITM Web of Conferences*, 17(December), p. 03025. doi: 10.1051/itmconf/20181703025.
- Khine, P.P. and Wang, Z.S., 2018. Data lake: a new ideology in big data era. In *ITM web of conferences* (Vol. 17, p. 03025). EDP Sciences.
- Kuzma, W. (2022, November 4). *What is a data lake, and why does my company need one?* - *dyvenia*. Dyvenia. <https://dyvenia.com/blog/what-is-a-data-lake-and-why-does-my-company-need-one/#:~:text=Data%20lakes%20are%20repositories%20designed,%2C%20enrichment%2C%20or%20storing%20required>
- Laurent, A., Madera, C., Laurent, D., 2020. *Data Lakes : Volume 2*. Wiley, Hoboken, New Jersey, Amerika.
- Michael Lock (2017) *ANGLING FOR INSIGHT IN TODAY'S DATA LAKE*.
- Mohanty, S., Jagadeesh, G., & Srivatsa, B. (2017). *Large-Scale Data Analytics: Discovering, Analyzing, and Visualizing Data*. Apress.
- Nambiar, A. and Mundra, D. (2022) 'An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management', *Big Data and Cognitive Computing*. MDPI. Available at: <https://doi.org/10.3390/bdcc6040132>.
- Nargesian, F., Zhu, E., Miller, R.J., Pu, K.Q. and Arocena, P.C., 2019. Data lake management: challenges and opportunities. *Proceedings of the VLDB Endowment*, 12(12), pp.1986-1989.
- Nypl, J., & Nypl, N. (2016). *Data Lake for Enterprises: The What, Why, and How*. Apress.
- Orogat, A. (2021) 'Data Lakes for Big Data [Report on Existing Data Lakes]', (March), pp. 2–8. doi: 10.13140/RG.2.2.17599.84643.
- Perera, L. P. and Mo, B. (2017) 'Visual Analytics in Ship Performance and Navigation Information for Sensor Specific Fault Detection', (July). doi: 10.1115/omae2017-61118.
- Plunkett, T., & Prentice, B. (2015). *Data Lakes: Practical Solutions for Building and Managing Data Lakes for Big Data*. Addison-Wesley Professional.

- Pohl, M., Staegemann, D. G. and Turowski, K. (2022) 'The Performance Benefit of Data Analytics Applications', *Procedia Computer Science*, 201(C), pp. 679–683. doi: 10.1016/j.procs.2022.03.090.
- Prasada. (2023, January 16). *Danau data - Azure Architecture Center*. Microsoft Learn. <https://learn.microsoft.com/id-id/azure/architecture/data-guide/scenarios/data-lake>
- Ramakrishnan, R., Sridharan, B., Douceur, J.R., Kasturi, P., Krishnamachari-Sampath, B., Krishnamoorthy, K., Li, P., Manu, M., Michaylov, S., Ramos, R. and Sharman, N., 2017, May. Azure data lake store: a hyperscale distributed file service for big data analytics. In *Proceedings of the 2017 ACM International Conference on Management of Data* (pp. 51-63).
- Rao, B. (2020). *Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump*. Apress.
- Ravat, F. and Zhao, Y., 2019. Data lakes: Trends and perspectives. In *Database and Expert Systems Applications: 30th International Conference, DEXA 2019, Linz, Austria, August 26–29, 2019, Proceedings, Part I* 30 (pp. 304-313). Springer International Publishing.
- Rooney, S. et al. (2021) 'Pathfinder: Building the Enterprise Data Map', *Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021*, (December), pp. 1909–1919. doi: 10.1109/BigData52589.2021.9671608.
- Russom, P. (2017) *Data Lakes Purposes, Practices, Patterns, and Platforms BEST PRACTICES REPORT*.
- Samuk, S. (n.d.). *What is a Data Lake? Does My Business Need It?* <https://www.to-increase.com/analytics-and-business-intelligence/blog/does-my-business-need-data-lake#:~:text=A%20data%20lake%20is%20essentially,even%20streamlining%20a%20supply%20chain>.
- Sawadogo, N. et al. (2021) 'On data lake architectures and metadata management', *Journal of Intelligent Information Systems*, 56(1). Available at: <https://doi.org/10.1007/s10844-020-00608>.
- Segner, M. (2023). Top Data Lake Vendors In 2023 (Quick Reference Guide). *Monte Carlo Data*. <https://www.montecarlodata.com/blog-top-data-lake-vendors/>
- Shiraly, K. (2022) *Best Practices in Data Lake Implementation*, *width.ai*. Available at: <https://www.width.ai/post/data-lake-implementation>.

- Snowflake. (n.d.). Brand Assets. Diakses pada 26 Mei 2023, dari <https://www.snowflake.com/brand-assets/>
- Sukhobokov, A.A. *et al.* (2022) 'The concept of an intelligent data lake management system: Machine consciousness and a universal data model', in *Procedia Computer Science*. Elsevier B.V., pp. 407–414. Available at: <https://doi.org/10.1016/j.procs.2022.11.085>.
- Team, E. (2022, December 3). *5 Key Benefits Of Building A Centralized Data-Lake | eCloudChain*. eCloudChain. <https://www.ecloudchain.com/5-key-benefits-of-data-lake/>
- What is a Data Lake?* (n.d.). TIBCO Software. <https://www.tibco.com/reference-center/what-is-a-data-lake>
- Woodall, P. *et al.* (2016) 'Data quality problems in ETL: The state of the practice in large organisations', *Proceedings of the 21st International Conference on Information Quality, ICIQ 2016*, (June).
- Yaddow, The Process of Data Mapping for Data Integration Projects Data Mapping - A Key The Process of Data Mapping for Data Integration Projects Data Mapping - A Key Work Product for Data Warehouse , Data Integration, and D. M. P. (2019) 'The Process of Data Mapping for Data Integration Projects Data Mapping -A Key The Process of Data Mapping for Data Integration Projects Data Mapping - A Key Work Product for Data Warehouse , Data Integration , and Data Migration Projects', (October). doi: 10.13140/RG.2.2.10352.81925.

PROFIL PENULIS

Dr. Widyastuti Andriyani, S.Kom., M.Kom.



Penulis lahir di Karanganyar 17 Maret, saat ini menjadi dosen pada Program Studi Magister Teknologi Informasi di Universitas Teknologi Digital Indonesia Yogyakarta, sejak tahun 2012. Riset yang saat ini dikerjakan *focus* dan *scope* dibidang Sistem Cerdas. Penulis dapat di hubungi melalui email: widya@utdi.ac.id

Aisyah Mutia Dawis, S.Kom., M.Kom.



Penulis lahir di Kota Surakarta pada tanggal 3 Juni 1991. Ia Lulus S1 pada tahun 2013 hingga mendapat gelar Sarjana Komputer di Universitas Muhammadiyah Surakarta serta mendapatkan penghargaan sebagai lulusan terbaik Se Fakultas Komunikasi dan Informatika, selanjutnya ia lulus S2 pada tahun 2019 hingga mendapat gelar Magister Komputer di Universitas Amikom Yogyakarta. Saat ini ia tercatat sebagai dosen tetap di Program Studi Sarjana Sistem dan Teknologi Informasi di Universitas 'Aisyiyah Surakarta. Selain mengajar ia aktif dalam kegiatan tridharma lainnya diantaranya ialah penelitian dan pengabdian kepada masyarakat. Kegiatan penelitian internal dan eksternal pernah dilakukannya. Beberapa penelitian berjudul: "*Evaluation of The Website 'Aisyiyah Surakarta of University Performance Based on Search Engine Optimization Using Automated Software Testing Gtmetrix*" dan "*Utilization of Virtual Reality Technology in Knowing the Symptoms of Acrophobia and Nyctophobia*". Ia sering menjadi *invited speaker* diantaranya menjadi tenaga pengajar BPPTIK Kementerian Kominfo. Serta sebagai bentuk pengabdian kepada masyarakat, ia pun pernah terlibat aktif sebagai *trainer* bimbingan teknis bantuan TIK SMP, SMA, Mahasiswa, Tenaga Pendidik, Guru dan Dosen dari tahun 2010 hingga saat ini.

Rakhmat Purnomo, S.Pd., S.Kom., M.Kom.



Penulis dosen Tetap Prodi Informatika di Universitas Bhayangkara Jakarta Raya yang pernah mengambil Ilmu Pendidikan di Universitas Negeri Jakarta. Saat ini sudah Tersertifikasi Dosen sejak 2019 dan Jabatan Fungsional Lektor 300. Mata Kuliah yang diajarkan Robotika, Jaringan Komputer, Sistem Operasi, dan Riset Teknologi Informasi. Pernah mendapatkan Hibah Penelitian Dosen Pemula sebagai ketua pada tahun 2018. Pernah menjadi Kepala Program Studi Informatika di Universitas Bhayangkara Jakarta Raya. Pernah mengelola jurnal Ilmiah. Saat ini fokus melakukan tri dharma. Penulis dapat dihubungi melalui email : rakhmat.purnomo@dsn.ubharajaya.ac.id.

Fajar Ciputra Daeng Bani, S.Kom., M.TI.



Penulis lahir di Kota Bandung pada tanggal 18 September 1989. Ia Lulus S1 pada tahun 2012 hingga mendapat gelar Sarjana Komputer di Institute Perbanas Jakarta, selanjutnya ia lulus S2 pada tahun 2017 hingga mendapat gelar Magister Teknik Informatika di Universitas Bina Nusantara. Saat ini ia tercatat sebagai dosen tetap di Program Studi Perdagangan Internasional Wilayah Asean dan RRT, Politeknik APP Jakarta yang merupakan salah satu perguruan tinggi di lingkungan Kementerian Perindustrian Republik Indonesia yang berada di bawah Badan Pengembangan Sumber Daya Industri (BPSDMI). Selain mengajar ia aktif dalam kegiatan tridharma lainnya diantaranya ialah penelitian dan pengabdian kepada masyarakat. Kegiatan penelitian pernah dilakukannya. Ia sebagai pemenang Penelitian Terapan tahun 2022 tingkat Nasional pada bidang *industry 4.0* yang berjudul: *"Perancangan Enterprise Resource Planning (ERP) berbasis Customized Business Process dengan Implementasi pada Fashion & Apparel Sebagai Solusi Mencapai Operational Excellence"*. Serta sebagai bentuk pengabdian kepada masyarakat, ia pun pernah terlibat aktif sebagai tim pengelolaan penyelenggara pada unit inkubator bisnis di Politeknik APP Jakarta hingga saat ini.

Mohammad Santosa Mulyo Diningrat, S.Kom., M.Kom.



Penulis lahir di Kota Metro pada tanggal 12 Juli 1994. Ia Lulus S1 pada tahun 2015 hingga mendapat gelar Sarjana Komputer di STMIK Amikom Yogyakarta, selanjutnya ia lulus S2 pada tahun 2019 hingga mendapat gelar Magister Komputer di Universitas Amikom Yogyakarta. Saat ini ia tercatat sebagai dosen tetap di Program Studi Manajemen Logistik Industri Elektronika, Politeknik APP Jakarta yang merupakan salah satu perguruan tinggi di lingkungan Kementerian Perindustrian Republik Indonesia yang berada di bawah Badan Pengembangan Sumber Daya Industri (BPSDMI). Selain mengajar ia aktif dalam kegiatan tridharma lainnya diantaranya ialah penelitian dan pengabdian kepada masyarakat.

Yusuf Wahyu Setiya Putra, S.Kom., M.Kom.



Penulis lahir di Kota Magelang pada tanggal 2 Mei. Ia Lulus jenjang pendidikan Sarjana pada tahun 2015 dengan masa studi 3 tahun di Program Studi Sistem Informasi. Dan Lulus jenjang Magister pada tahun 2019 di Universitas Amikom Yogyakarta. Berkariier sebagai dosen sejak tahun 2019, ia memulai kariernya di salah satu Perguruan Tinggi Swasta di Kabupaten Kendal hingga tahun 2022. Dan tahun 2022 hingga saat ini ia sudah kembali ke Kota asalnya dan tercatat sebagai dosen tetap pada Program Studi Sistem Informasi di Sekolah Tinggi Manajemen Informatika dan Komputer Bina Patria Magelang. Selain mengajar ia aktif dalam kegiatan tridharma lainnya diantaranya ialah penelitian dan pengabdian. Sesuai dengan konsentrasi bidang ilmu Sistem Informasi yang dipelajarinya, telah terbit beberapa publikasi karya ilmiah dan salah satunya adalah penelitian dengan judul, Analisis Penerimaan Sistem Informasi Rumah Sakit pada RS Padma Lalita Magelang Menggunakan Metode *Technology Acceptance Model* (TAM). Selain itu ia juga membuat penelitian yang dimanfaatkan oleh salah satu perusahaan yang cukup besar di Kota Semarang yaitu Penerapan Metode *Analytical Hierarchy Process* pada Sistem Pendukung Keputusan Penerimaan Karyawan PT. SDN. Dan masih ada beberapa publikasi penelitian dan pengabdian yang terindeks di Google Scholar, Sinta, dll.

Novi, S.Kom., M.T.



Penulis lahir di Kota Bukittinggi pada tanggal 01 November. Ia Lulus pada tahun 2017 dan mendapat gelar Magister Teknik Elektro di Universitas Andalas. Saat ini ia tercatat sebagai dosen tetap di Politeknik Negeri Padang.

DATA LAKE INSIGHTS

Buku "*Data Lake Insights*" adalah sebuah karya yang bertujuan untuk memberikan wawasan mendalam tentang konsep, strategi, dan manfaat penggunaan *Data Lake* dalam pengelolaan dan analisis data. Buku ini ditulis oleh sekelompok ahli data yang berpengalaman, yang telah berhasil mengimplementasikan solusi *Data Lake* dalam berbagai lingkungan bisnis.

Dalam buku ini, para penulis menjelaskan dengan jelas dan sistematis tentang apa itu *Data Lake* dan bagaimana membangun dan mengelolanya. Mereka memperkenalkan konsep arsitektur *Data Lake* yang fleksibel dan *scalable*, yang mampu menampung berbagai jenis data dari berbagai sumber, baik terstruktur maupun tidak terstruktur. Selain itu, para penulis juga membahas teknologi dan alat yang digunakan dalam implementasi *Data Lake*, seperti *Hadoop*, *Apache Spark*, dan sistem penyimpanan berbasis *cloud*.

Selain aspek teknis, buku ini juga menyoroti manfaat dan keuntungan yang dapat diperoleh dengan mengadopsi pendekatan *Data Lake*. Para penulis menjelaskan bagaimana *Data Lake* dapat meningkatkan kemampuan analisis data, memfasilitasi eksplorasi dan penemuan informasi baru, serta mendukung pengambilan keputusan yang lebih baik. Mereka juga menyoroti pentingnya keamanan dan kepatuhan data dalam konteks *Data Lake*, serta memberikan panduan praktis untuk mengatasi tantangan yang mungkin muncul. Dengan bahasa yang jelas dan penjelasan yang rinci, "*Buku Data Lake Insights*" cocok untuk para profesional data, pengembang perangkat lunak, arsitek sistem, dan manajer yang tertarik dalam memanfaatkan potensi *Data Lake*. Buku ini memberikan pemahaman yang mendalam tentang konsep, strategi, dan praktek terbaik dalam membangun dan mengelola *Data Lake*, dan akan menjadi sumber referensi yang berharga bagi pembaca yang ingin menggali lebih dalam di bidang ini.