



Big Data dan Manajemen Basis Data Terdistribusi



**Adryan Rachman, Indah Dwi Mumpuni, Weda Adistianaya Dewa,
Dinny Wahyu Widarti, Farida Islamiah, Esa Kurniawan,
A.Kachsyfur Djasim Ilyas Paenrongi, Mustika Kurnia Mayangsari,
Timotius Victory, Dwi Diana Wazaumi, M. Rhifky Wayahdi, Siti Aisyah,
Deyana Kusuma Wardani, Thesya Atarezcha Pangruruk,
Kurnia Yahya, Nariza Wanti Wulan Sari,
Erlyne Nadhilah Widyaningrum, Phie Chyan, Triana Dewi Salma,
Fari Katul Fikriah, Rosa Eliviani, Amelia Devi Putri Ariyanto,
Luthfi Atikah**

Big Data dan Manajemen Basis Data Terdistribusi

**Adryan Rachman, Indah Dwi Mumpuni, Weda Adistianaya
Dewa, Dinny Wahyu Widarti, Farida Islamiah, Esa
Kurniawan, A.Kachsyfur Djasim Ilyas Paenrongi, Mustika
Kurnia Mayangsari, Timotius Victory, Dwi Diana Wazaumi,
M. Rhifky Wayahdi, Siti Aisyah, Deyana Kusuma Wardani,
Thesya Atarezcha Pangruruk, Kurnia Yahya, Nariza Wanti
Wulan Sari, Erlyne Nadhilah Widyaningrum, Phie Chyan,
Triana Dewi Salma, Fari Katul Fikriah, Rosa Eliviani, Amelia
Devi Putri Ariyanto, Luthfi Atikah**



PT. MIFANDI MANDIRI DIGITAL

Undang-Undang No. 28 Tahun 2014 Tentang Hak Cipta:

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam pasal 9 ayat (1) huruf i untuk penggunaan secara komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp. 100.000.000,- (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk penggunaan secara komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp. 500.000.000,- (lima ratus juta rupiah).
3. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk penggunaan secara komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp. 1.000.000.000,- (satu miliar rupiah).
4. Setiap Orang yang memenuhi unsur sebagaimana dimaksud pada ayat (3) yang dilakukan dalam bentuk pembajakan, dipidana dengan pidana penjara paling lama 10 (sepuluh) tahun dan/atau pidana denda paling banyak Rp. 4.000.000.000,- (empat miliar rupiah).

Big Data dan Manajemen Basis Data Terdistribusi

Adryan Rachman, Indah Dwi Mumpuni, Weda Adistianaya Dewa, Dinny Wahyu Widarti, Farida Islamiah, Esa Kurniawan, A.Kachsyfur Djasim Ilyas Paenrongi, Mustika Kurnia Mayangsari, Timotius Victory, Dwi Diana Wazaumi, M. Rhifky Wayahdi, Siti Aisyah, Deyana Kusuma Wardani, Thesya Atarezcha Pangruruk, Kurnia Yahya, Nariza Wanti Wulan Sari, Erlyne Nadhilah Widyaningrum, Phie Chyan, Triana Dewi Salma, Fari Katul Fikriah, Rosa Eliviani, Amelia Devi Putri Ariyanto, Luthfi Atikah

ISBN: 978-634-7226-33-4

Editor : Sarwandi, M.Pd.T
Layout : Miftahul Jannah, M.Kom
Desain sampul : Rifki Ramadan

Penerbit
PT. Mifandi Mandiri Digital

Redaksi & Distributor Tunggal
PT. Mifandi Mandiri Digital
Komplek Senda Residence Jl. Payanibung Ujung D Dalu
Sepuluh-B Tanjung Morawa Kab. Deli Serdang Sumatera Utara

Cetakan Pertama, Agustus 2025

Hak Cipta © 2025 by PT. Mifandi Mandiri Digital

Hak cipta Dilindungi Undang-Undang
Dilarang memperbanyak karya tulis ini dalam bentuk dan
dengan cara apapun tanpa izin tertulis dari penerbit.

Kata Pengantar

Perkembangan teknologi informasi yang begitu pesat telah membawa umat manusia memasuki era baru yang dikenal sebagai era Big Data. Di era ini, data tumbuh dengan kecepatan dan volume yang luar biasa besar, melintasi berbagai domain kehidupan — mulai dari bisnis, pendidikan, kesehatan, hingga pemerintahan. Data tidak lagi hanya menjadi produk sampingan dari proses digital, melainkan telah menjelma menjadi aset strategis yang memiliki peran sentral dalam pengambilan keputusan dan inovasi.

Buku ini hadir sebagai upaya menjembatani kebutuhan pemahaman yang komprehensif mengenai pengelolaan data dalam skala besar dan terdistribusi. Dengan cakupan yang luas dan sistematis, buku Big Data dan Manajemen Basis Data Terdistribusi mengurai berbagai aspek penting yang dibutuhkan oleh akademisi, praktisi, maupun pengambil kebijakan. Mulai dari konsep dasar dan karakteristik Big Data, tantangan teknis dalam penyimpanan dan replikasi, hingga penerapan teknologi seperti Hadoop, Apache Spark, NoSQL, dan Cloud Computing.

Lebih lanjut, buku ini juga menyajikan pembahasan aplikatif dalam berbagai sektor — seperti kesehatan, keuangan, bisnis, smart city, dan Internet of Things (IoT) — yang menunjukkan betapa krusialnya peran manajemen data dalam ekosistem digital saat ini. Bab-bab penutup pun mengajak

pembaca untuk melihat tren masa depan dalam pengelolaan data, seperti integrasi kecerdasan buatan, edge computing, green computing, hingga quantum computing.

Kami berharap buku ini dapat menjadi referensi utama dan sumber inspirasi bagi para pembaca yang ingin memahami dan menguasai strategi pengelolaan data dalam lanskap teknologi masa kini dan mendatang. Terima kasih kepada semua pihak yang telah berkontribusi dalam penyusunan buku ini. Semoga kehadiran buku ini dapat memberi manfaat luas dan mendorong kolaborasi inovatif di era digital berbasis data.

Medan, Juni 2025

Penulis

Daftar Isi

Kata Pengantar	i
Daftar Isi	iii
BAB 1 ERA BIG DATA DAN TANTANGAN PENGELOLAAN DATA	1
Pendahuluan	1
Evolusi Pengelolaan Data	5
Karakteristik Unik Big Data	8
Tantangan Pengelolaan Big Data	12
Peluang Besar dalam Mengelola Big Data	15
BAB 2 KONSEP DASAR BIG DATA	18
Pendahuluan	18
Definisi Big Data	19
Karakteristik Big Data	19
Sumber Big Data	21
Teknologi Pendukung Big Data	22
Manfaat Big Data	24
BAB 3 ARSITEKTUR DAN MODEL BASIS DATA TERDISTRIBUSI	28
Pendahuluan	28
Kategori Sistem Basis Data Terdistribusi	28
Arsitektur Basis Data Terdistribusi	31
Teknik Desain Basis Data Terdistribusi	33
Model Basis Data Terdistribusi	35
BAB 4 PERBANDINGAN BASIS DATA TERPUSAT DAN BASIS DATA TERDISTRIBUSI	41
Pendahuluan	41
Basis Data Terpusat	41
Basis Data Terdistribusi	43
Perbandingan Basis Data Terpusat dan Basis Data Terdistribusi	47
BAB 5 JENIS DAN KLASIFIKASI BASIS DATA DALAM BIG DATA	48
Pendahuluan	48
Jenis-jenis Basis Data dalam Big Data	49
Klasifikasi Basis Data Berdasarkan Karakteristik Big Data	56

BAB 6 MODEL PENYIMPANAN DATA DALAM SISTEM BIG DATA	59
Pendahuluan	59
Karakteristik Umum Penyimpanan dalam Sistem Big Data	60
Arsitektur Umum Sistem Penyimpanan Big Data	62
Model Penyimpanan Data	64
 BAB 7 TEKNOLOGI CLOUD COMPUTING UNTUK BIG DATA	 68
Pendahuluan	68
Konsep Dasar Cloud Computing dan Big Data	71
Penerapan Cloud Computing untuk Big Data di Sektor Kesehatan	72
Penerapan Cloud Computing untuk Big Data di Sektor Pendidikan	74
Penerapan Cloud Computing untuk Big Data di Sektor Bisnis	75
Penerapan Cloud Computing untuk Big Data di Sektor Transportasi	77
Prospek Masa Depan Cloud Computing dan Big Data	79
 BAB 8 HADOOP DAN EKOSISTEMNYA DALAM PENGOLAHAN BIG DATA	 81
Pendahuluan	81
Sejarah dan Pengertian Hadoop	81
Arsitektur dan Ekosistem Hadoop	82
Hadoop Distributed File System	84
MapReduce	87
Yet Another Resource Negotiator (YARN)	88
 BAB 9 FRAMEWORK APACHE SPARK UNTUK BIG DATA PROCESSING	 90
Pendahuluan	90
Sejarah dan Evolusi Apache Spark	93
Arsitektur dan Alur Kerja Spark	95
Fitur Utama Pada Spark	98
Integrasi Spark dengan Sistem dan Platform Lainnya	100
Studi Kasus Penggunaan Spark	103
 BAB 10 MANAJEMEN DATA DENGAN NOSQL DATABASES	 108
Pendahuluan	108
Jenis Database NoSQL	109
MongoDb	110
Cassandra	114
HBase	119
 BAB 11 ARSITEKTUR CLIENT-SERVER DALAM BASIS DATA TERDISTRIBUSI	 124
Pendahuluan	124

Prinsip Dasar Arsitektur Client-Server	126
Evolusi Arsitektur	128
Penerapan dalam Basis Data Terdistribusi	131
Tantangan dan Pertimbangan Khusus	133
 BAB 12 REPLIKASI DATA	136
Pendahuluan	136
Konsep dan Tujuan Replikasi Data	137
Jenis-jenis Replikasi	140
Teknik Replikasi Data	142
Implementasi dan Tantangan Replikasi Data	144
Teknologi Replikasi Data yang Umum Digunakan	145
Contoh Implementasi dan Tantangan Replikasi Data	147
 BAB 13 KONSISTENSI DATA DALAM SISTEM TERDISTRIBUSI	148
Pengenalan Sistem Terdistribusi	148
Penerapan Replikasi Data Pada Sistem Terdistribusi	151
Teorema CAP Pada Sistem Terdistribusi	153
Ancaman Umum Konsistensi Data Pada Sistem Terdistribusi	155
 BAB 14 KEAMANAN DAN PRIVASI DALAM BIG DATA	158
Pendahuluan	158
Keamanan Big Data	159
Privasi Big Data	162
 BAB 15 MANAJEMEN AKSES DAN PROTEKSI DATA DALAM BASIS DATA TERDISTRIBUSI	165
Pendahuluan	165
Konsep Dasar Manajemen Akses	165
Komponen Utama Manajemen Akses	167
Konsep Dasar Proteksi Data	169
Tantangan Proteksi Data dalam Sistem Terdistribusi	171
 BAB 16 TEKNIK DATA MINING UNTUK BIG DATA	173
Pendahuluan	173
Teknik-Teknik Data Mining	176
 BAB 17 PENGENALAN MACHINE LEARNING DALAM BIG DATA	188
Pendahuluan	188
Machine Learning	189
Jenis-jenis Machine Learning	190
Aplikasi Machine Learning dalam Analisis Data	192

Mengukur Keberhasilan	194
Generalisasi, Overfitting, dan Underfitting	195
Tantangan dan Batasan Machine Learning	197
 BAB 18 ANALISIS DATA REAL TIME MENGGUNAKAN STREAMING	
PROCESSING	198
Pendahuluan	198
Pengertian dan Konsep Dasar Streaming Processing	199
Arsitektur Streaming Processing	201
Perbandingan Batch Processing dan Streaming Processing	202
Teknologi dan Platform Streaming Processing	203
Studi Kasus Implementasi Streaming Data	204
Tantangan dan Solusi dalam Streaming Data	205
Masa Depan Analisis Data Real Time	206
 BAB 19 ANALISIS VISUALISASI DATA UNTUK BIG DATA	
Pendahuluan	208
Konsep Dasar Visualisasi Data	209
Teknik Visualisasi Big Data	210
Tools dan Teknologi Visualisasi Data	216
Tantangan dan Strategi Mengatasi Visualisasi Data	218
 BAB 20 MANAJEMEN BIG DATA DALAM BISNIS DAN INDUSTRI	
Pendahuluan	220
Manajemen Big Data dalam Konteks Bisnis	221
Peran Strategis Big Data dalam Bisnis Modern	222
Sumber dan Jenis Data dalam Dunia Bisnis	224
Tantangan Big Data dalam Bisnis	225
Manajemen Big Data dalam Konteks Industri	227
Ciri Khas Data dalam Dunia Industri	228
Penerapan Big Data pada Industri Spesifik	230
Teknologi Pendukung Big Data di Dunia Industri	232
Tantangan Industri dalam Menerapkan Big Data	234
 BAB 21 IMPLEMENTASI BIG DATA DALAM SMART CITY DAN IOT	
Pendahuluan	236
Konsep Big Data dalam Konteks Smart City dan IoT	237
Arsitektur Sistem Big Data untuk Smart City dan IoT	239
Big Data berbasis IoT untuk Edge–Cloud	241
Studi Kasus Implementasi	242
Tantangan dan Masa Depan	245

BAB 22 BIG DATA DALAM HEALTHCARE DAN KEUANGAN	249
Pendahuluan	249
Big Data dalam Healthcare	251
Sumber dan Jenis Big Data dalam Healthcare	252
Manfaat Strategis Big Data dalam Healthcare	254
Tantangan dan Kompleksitas Data Kesehatan	255
Teknologi dan Analisis untuk Big Data Healthcare	255
Big Data dalam Keuangan	258
Sumber dan Jenis Big Data dalam Keuangan	259
Manfaat Strategis Big Data dalam Keuangan	259
Tantangan Big Data dalam Keuangan	260
Teknologi dan Analisis untuk Big Data dalam Keuangan	262
 BAB 23 TREN MASA DEPAN DALAM BIG DATA DAN DATA TERDISTRIBUSI	 264
Masa Depan Big Data	264
Integrasi Kecerdasan Buatan dan Pembelajaran Mesin	264
Real Time Data Processing dan Streaming Analytic	266
Edge Computing dan Federated Learning	266
DataOps dan Otomatisasi Pipeline	268
Cloud Native Big Data dan Hybrid Cloud	268
Graph Processing dan Analitik	269
Green Computing dan Efisiensi Energi	270
Quantum Computing dan Big Data	271
Tantangan Big Data Pengelolaan Big Data	272
 Daftar Pustaka	 275
Tentang Penulis	300

BAB 1 ERA *BIG DATA* DAN TANTANGAN PENGELOLAAN DATA

Pendahuluan

Dalam dua dekade terakhir, dunia telah menyaksikan transformasi digital yang begitu cepat dan revolusioner, mengubah cara manusia berinteraksi, bekerja, dan mengambil keputusan. Kemajuan teknologi informasi, ditandai dengan perkembangan internet, kecerdasan buatan, dan komputasi awan, telah mendorong pertumbuhan eksponensial dalam jumlah data yang dihasilkan setiap detiknya. Setiap klik di *e-commerce*, transaksi perbankan, unggahan media sosial, sensor IoT, dan rekam medis digital telah menciptakan jejak data yang terus bertambah dengan kecepatan yang luar biasa. Fenomena ini, yang dikenal sebagai ledakan data global, telah mengubah paradigma tradisional dalam pengelolaan dan analisis informasi, menuntut pendekatan yang lebih inovatif dan skalabel.

Transformasi digital tidak hanya memengaruhi sektor teknologi, tetapi juga membentuk ulang berbagai industri, mulai dari keuangan, kesehatan, manufaktur, hingga pemerintahan. Organisasi kini beralih ke analitik berbasis data untuk mengoptimalkan operasional, meningkatkan efisiensi, serta menciptakan strategi yang lebih akurat dalam menjawab kebutuhan pasar. Namun, dengan *volume* data yang terus meningkat, tantangan baru muncul: bagaimana mengelola, menyimpan, dan menganalisis informasi dalam skala besar tanpa kehilangan nilai dan keakuratannya? Di sinilah konsep *Big Data* memainkan peran sentral dalam mendefinisikan masa

BAB 3 ARSITEKTUR DAN MODEL BASIS DATA TERDISTRIBUSI

Pendahuluan

Seiring dengan pertumbuhan pesat teknologi informasi dan kebutuhan akan pengelolaan data yang efisien, sistem basis data terdistribusi menjadi solusi yang semakin relevan. Dalam era digital saat ini, organisasi sering kali memiliki data yang tersebar di berbagai lokasi geografis. Penggunaan basis data terdistribusi memungkinkan pengelolaan data secara efisien, konsisten, dan aman di seluruh jaringan yang luas. Hal ini mendukung operasional yang lebih fleksibel dan responsif terhadap kebutuhan pengguna di berbagai lokasi. Basis data terdistribusi adalah kumpulan data yang secara logis saling berhubungan dan tersebar secara fisik pada beberapa komputer yang terhubung melalui jaringan komunikasi. Salah satu motivasi utama di balik penggunaan sistem basis data adalah keinginan untuk mengintegrasikan data operasional perusahaan dan menyediakan akses terpusat, sehingga terkontrol ke data tersebut.

Kategori Sistem Basis Data Terdistribusi

Secara umum, ada dua kategori Sistem Basis Data Terdistribusi:

1. **Sistem Basis Data Terdistribusi Homogen:** Dalam Sistem ini, data tersebar di server yang menjalankan skema yang sama dari Database fisik (skema yang sama dari bagian-bagiannya) dan perangkat lunak Sistem Manajemen Basis

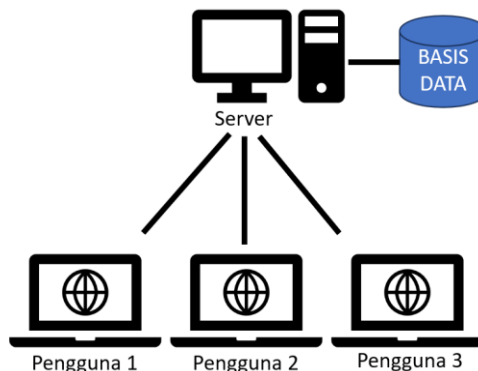
BAB 4 PERBANDINGAN BASIS DATA TERPUSAT DAN BASIS DATA TERDISTRIBUSI

Pendahuluan

Penggunaan basis data dapat memilih diantara basis data terpusat dan basis data terdistribusi. Perbedaan pemilihan menggunakan basis data terpusat (*centralized*) atau basis data terdistribusi (*distributed*), baiknya pilih yang mana? Sebelum kita memilihnya, kita perlu mempelajari keduanya, dengan mengetahui kelebihan dan kekurangan masing-masing.

Basis Data Terpusat

Pengertian dari basis data terpusat adalah sistem di mana semua data disimpan dalam satu lokasi atau server utama, dan semua pengguna mengakses data dari lokasi tersebut (Usman, A., dkk, 2025). Dengan basis data terpusat, tidak perlu lagi menyimpan duplikat data di berbagai lokasi (Prananingrum, L., 2024).



Gambar 4.1. Basis data terpusat

BAB 5 JENIS DAN KLASIFIKASI BASIS DATA DALAM *BIG DATA*

Pendahuluan

Basis data adalah kumpulan data yang terorganisir secara sistematis sehingga dapat diakses, dikelola, dan diperbarui dengan mudah. Dalam ekosistem *Big Data*, basis data memainkan peran krusial dalam menyimpan dan mengolah data dalam jumlah besar secara cepat dan efisien. Berbeda dengan basis data konvensional (relasional) yang menggunakan skema tetap dan tabel, basis data *Big Data* harus mampu menangani data tidak terstruktur dan semi-terstruktur serta beroperasi dalam lingkungan terdistribusi.

Basis data pada *Big Data* dirancang khusus untuk menyimpan kuantitas data terstruktur, semi-terstruktur, dan tidak terstruktur yang sangat besar tanpa skema yang kaku. Kemampuan ini memungkinkan organisasi untuk memperoleh pemahaman yang lebih komprehensif tentang tren dan pola dengan mengintegrasikan beragam set data untuk membentuk gambaran yang lengkap.

Arsitektur *Big Data* dirancang secara khusus untuk menangani penyerapan, pemrosesan, dan analisis data yang volumenya terlalu besar atau kompleks untuk sistem basis data tradisional. Konsep ini melibatkan tiga pilar utama yang saling terkait yaitu integrasi data, pengelolaan data (penyimpanan) dan analisis data. Kebutuhan untuk mengintegrasikan, mengelola, dan menganalisis *Big Data* secara inheren mendorong adopsi arsitektur terdistribusi. Ini bukan hanya tentang membagi data

BAB 6 MODEL PENYIMPANAN DATA DALAM SISTEM *BIG DATA*

Pendahuluan

Kemajuan pesat teknologi digital telah menyebabkan ledakan besar data dari berbagai sumber, seperti media sosial, perangkat *Internet of Things*, transaksi bisnis, dan aktivitas *online* sehari-hari. Kondisi ini mendorong ide tentang *Big Data*, yang merupakan kumpulan data yang sangat besar, beragam, dan berkembang dengan kecepatan tinggi. Sistem penyimpanan dan manajemen data konvensional tidak lagi memadai untuk menangani jumlah data sebesar ini. Ini tidak lagi memenuhi syarat dalam hal kapasitas, kemudahan akses, dan fleksibilitas struktur data. Oleh karena itu, pendekatan baru diperlukan untuk merancang model penyimpanan yang mampu menangani sifat kompleks dari sistem *Big Data*. Sistem *Big Data* berbeda dengan sistem basis data tradisional, yang biasanya terpusat dan berbasis skema relasional; mereka menggunakan penyimpanan terdistribusi, yang memungkinkan pengelolaan data dalam berbagai format, termasuk yang terstruktur, semi-terstruktur, dan tidak terstruktur. Sistem tradisional biasanya menekankan pada konsistensi data dan inkonsistensi data. Pergeseran paradigma ini membutuhkan model penyimpanan yang tangguh secara teknis dan fleksibel untuk memenuhi kebutuhan analisis data skala besar secara *batch* maupun *real-time*.

Tujuan dari bab ini adalah untuk memberikan pemahaman menyeluruh tentang berbagai model penyimpanan data yang digunakan dalam ekosistem *Big Data*. Pembahasan

BAB 7 TEKNOLOGI *CLOUD COMPUTING* UNTUK *BIG DATA*

Pendahuluan

Kemajuan teknologi informasi telah membawa perubahan mendasar pada berbagai aspek kehidupan manusia. Dalam era digital saat ini, hampir setiap aktivitas manusia sehari-hari tersentuh oleh teknologi komputasi dan data. *Cloud Computing* (komputasi awan) dan *Big Data* merupakan dua inovasi penting yang berperan besar dalam transformasi digital tersebut. *Cloud Computing* menyediakan infrastruktur dan layanan komputasi melalui internet secara fleksibel, sementara *Big Data* merujuk pada kumpulan data berukuran sangat besar dan beragam yang memerlukan metode pengolahan canggih. Kolaborasi antara *Cloud Computing* dan *Big Data* menciptakan peluang baru untuk meningkatkan efisiensi dan efektivitas di berbagai bidang kehidupan, mulai dari kesehatan hingga transportasi.

Aktivitas manusia sehari-hari secara garis besar dapat dikelompokkan menjadi dua kategori utama, yaitu aktivitas pemenuhan kebutuhan (konsumsi) dan aktivitas produktif (produksi) (Satmoko, 2022). Kategori pertama mencakup kegiatan-kegiatan seperti menjaga kesehatan pribadi dan menempuh pendidikan, yang bertujuan memenuhi kebutuhan dasar dan pengembangan diri individu. Kategori kedua mencakup kegiatan dalam lingkup sosial-ekonomi, misalnya menjalankan usaha bisnis dan mengelola transportasi, yang berorientasi pada produksi barang/jasa dan mobilitas untuk mendukung kehidupan masyarakat. Kedua kategori aktivitas ini

BAB 8 HADOOP DAN EKOSISTEMNYA DALAM PENGOLAHAN *BIG DATA*

Pendahuluan

Lonjakan aplikasi baru menghasilkan *volume* data kompleks (media sosial, belanja *online*, dll.) yang sulit dikelola dengan alat konvensional (Chen dkk., 2014; Lam, 2010). Teknologi *Big Data*, khususnya *Hadoop*, muncul sebagai solusi efisien dan hemat biaya untuk menyimpan, memproses, dan menganalisis data masif, terutama yang tidak terstruktur, dengan memanfaatkan kluster komputer terjangkau (White, 2015). *Hadoop* menggunakan komputasi terdistribusi dan toleran kesalahan (misalnya melalui MapReduce dan HDFS) untuk mengolah data besar secara paralel, dengan komputasi dan penyimpanan yang terintegrasi di seluruh *node* (Dean & Ghemawat, 2008; White, 2015).

Sejarah dan Pengertian *Hadoop*

Hadoop dikembangkan di awal tahun 2000-an oleh Doug Cutting dan Mike Cafarella, terinspirasi oleh makalah Google MapReduce (Dean & Ghemawat, 2008) dan Google *File System* (Ghemawat dkk., 2003) yang ditulis oleh Jeffrey Dean dari Google. *Hadoop* menggunakan komputasi terdistribusi dan toleran kesalahan (misalnya melalui MapReduce dan HDFS) untuk mengolah data besar secara paralel, dengan komputasi dan penyimpanan yang terintegrasi di seluruh *node* (White,

BAB 9 *FRAMEWORK APACHE SPARK UNTUK BIG DATA PROCESSING*

Pendahuluan

Munculnya era *Big Data* telah merevolusi cara organisasi dalam mengelola, menganalisis, dan memanfaatkan data. *Big Data* tidak hanya ditandai oleh jumlah data yang sangat besar (*volume*), tetapi juga oleh laju pertumbuhan data yang sangat cepat (*velocity*), serta keberagaman format dan tipe data yang semakin luas (*variety*)—tiga karakteristik utama yang dikenal dengan istilah 3V (Gandomi & Haider, 2015). Ketiga aspek utama ini dikenal sebagai 3V—*Volume*, *Velocity*, dan *Variety*—yang kini menjadi tantangan tersendiri dalam pengelolaan data modern. Di satu sisi, organisasi dituntut untuk dapat memanfaatkan data secara maksimal, namun di sisi lain, karakteristik 3V membuat proses penyimpanan, pengolahan, dan analisis data menjadi jauh lebih kompleks dibandingkan era sebelumnya (Chen & Zhang, 2014).

Sistem komputasi konvensional ternyata tidak mampu mengikuti laju perkembangan kebutuhan analisis data dalam era *Big Data*. Ketika harus menangani data dalam jumlah yang sangat besar, keterbatasan pada skalabilitas, kapasitas memori, dan kecepatan pemrosesan menjadi hambatan utama. Performa sistem tradisional semakin tertinggal, terutama ketika beban kerja meningkat secara signifikan. Selain itu, waktu yang dibutuhkan untuk mengeksekusi *query* atau menjalankan proses analitik sering kali melampaui ekspektasi pengguna dan kebutuhan organisasi yang serba cepat (Chen & Zhang, 2014). Di

BAB 10 MANAJEMEN DATA DENGAN NOSQL DATABASES

Pendahuluan

Dalam era *Big Data*, kebutuhan untuk mengelola *volume* data yang sangat besar dan beragam dengan kecepatan tinggi telah melahirkan paradigma baru dalam dunia basis data, yaitu NoSQL (Fadli et al., 2020). NoSQL, singkatan dari Not Only SQL, merupakan pendekatan alternatif terhadap sistem basis data relasional (RDBMS) yang konvensional (Sadalage & Fowler, n.d.). Teknologi ini menawarkan fleksibilitas skema, skalabilitas horizontal, serta performa tinggi, sehingga sangat cocok digunakan dalam berbagai aplikasi modern seperti media sosial, *Internet of Things* (IoT), dan sistem rekomendasi. Perbandingan antara basis data relasional dan NoSQL melibatkan beberapa faktor utama, seperti model data yang digunakan, kemampuan skalabilitas, fleksibilitas dalam pengelolaan data, serta kemampuan dalam melakukan kueri dan analitik secara efisien (Hasibuan et al., 2023).

Berbeda dengan RDBMS yang kaku dalam struktur tabel dan relasi, NoSQL dirancang untuk mendukung data semi-terstruktur dan tidak terstruktur, serta mampu menangani partisi dan replikasi secara efisien di banyak *node*. Dengan berbagai keunggulan tersebut, NoSQL menjadi pilihan utama bagi pengembang dan organisasi yang membutuhkan sistem basis data yang lebih fleksibel dan *scalable* untuk menangani kompleksitas data modern.

Oleh karena itu, bab ini akan membahas manajemen data

BAB 11 ARSITEKTUR *CLIENT-SERVER* DALAM BASIS DATA TERDISTRIBUSI

Pendahuluan

Arsitektur *client-server* merupakan fondasi komputasi modern yang menopang hampir semua aplikasi, termasuk sistem manajemen basis data terdistribusi/*database management system* (DDBMS). Pada dasarnya, model ini memisahkan peran antara penyedia layanan (*server*) dan pemintanya (*client*). Namun, kesederhanaan konsep ini diuji ketika data tidak lagi tersimpan di satu lokasi terpusat, melainkan tersebar di berbagai *node* yang terhubung oleh jaringan. Dalam lingkungan terdistribusi, evaluasi keberhasilan arsitektur ini tidak lagi sesederhana mengukur kecepatan respons, melainkan menuntut analisis metrik yang lebih kompleks. Salah satu metrik yang paling krusial adalah *distribution transparency*, yaitu sejauh mana kompleksitas distribusi data berhasil disembunyikan dari pengguna, sehingga mereka dapat berinteraksi dengan sistem seolah-olah semua sumber daya bersifat lokal.

Pentingnya *distribution transparency* tidak hanya bersifat teoretis, tetapi juga berdampak langsung pada *usability* dan efisiensi operasional. Ketika pengguna tidak perlu lagi dipusingkan oleh kerumitan teknis di balik layar, mereka bisa fokus pada tugas utamanya. Sebagai contoh nyata, sebuah arsitektur *client-server* yang dirancang untuk skenario simulasi resusitasi neonatal menunjukkan bahwa transparansi menjadi komponen vital dalam evaluasi sistem. Transparansi yang tinggi secara inheren mengatasi kelemahan klasik model *client-server*,

BAB 12 REPLIKASI DATA

Pendahuluan

Di tengah perkembangan teknologi digital yang semakin pesat, data telah menjadi bagian tak terpisahkan dari kehidupan manusia modern. Hampir semua aktivitas, mulai dari berkomunikasi, bertransaksi, hingga menyimpan informasi pribadi atau profesional, melibatkan data. Kebutuhan terhadap sistem yang mampu menjaga data agar tetap aman dan selalu tersedia menjadi semakin penting (Silberschatz, Galvin, & Gagne, 2018).

Salah satu solusi utama dalam menjaga keandalan data adalah dengan menerapkan teknik replikasi data. Replikasi data adalah proses menyalin data dari satu tempat ke tempat lain untuk memastikan data tetap tersedia, meskipun terjadi kerusakan sistem, gangguan teknis, atau bencana. Proses ini menjadi dasar dari sistem yang tahan terhadap kegagalan (*fault-tolerant systems*) dan sangat umum digunakan di berbagai sektor, baik sektor publik maupun swasta (Tanenbaum & Steen, 2007).

Teknologi ini tidak hanya digunakan oleh perusahaan besar atau instansi pemerintah saja. Dalam kehidupan sehari-hari, kita pun sering memanfaatkan replikasi data secara tidak sadar. Misalnya, saat kita mencadangkan dokumen di Google Drive, mengaktifkan sinkronisasi WhatsApp, atau menyimpan data penting di *flashdisk* dan *hard disk* eksternal. Replikasi data juga menjadi pondasi utama dalam sistem layanan publik digital, layanan keuangan, pendidikan, hingga usaha kecil yang mulai

BAB 13 KONSISTENSI DATA DALAM SISTEM TERDISTRIBUSI

Pengenalan Sistem Terdistribusi

Dari tahun ke tahun, perkembangan dunia teknologi informasi sangatlah pesat (Khoirunnisya & Nasution, 2024). Dampak dari perkembangan ini menyebabkan kebutuhan sistem komputasi yang kompleks, skalable, dan andal mengalami peningkatan di berbagai sektor. Sistem yang dikembangkan diharapkan mampu menangani *volume* data yang besar serta melayani pengguna di berbagai lokasi. Sebagai contoh, kita dapat mengambil salah satu sistem dari sektor Pendidikan, yaitu Dapodik (Data Pokok Pendidikan). Dapodik adalah sistem pendataan secara *real-time* yang dikelola langsung oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi yang digunakan untuk mengumpulkan dan menyimpan data peserta didik, pendidik, tenaga kependidikan, sumber daya Pendidikan, substansi Pendidikan, dan capaian Pendidikan (Menteri Pendidikan, Kebudayaan, Riset & Indonesia, 2022).

Pendekatan yang dilakukan agar sistem dapat menangani data berskala besar dan dapat diakses di berbagai lokasi (seperti Dapodik) menggunakan arsitektur sistem terdistribusi. Arsitektur dari sistem terdistribusi melibatkan banyaknya komputeryang saling berhubungan melalui jaringan dan bekerja sama menyelesaikan sebuah tugas (seperti melayani permintaan pengguna memproses dan menyimpan data yang memiliki *volume* besar) (Ajismanto et al., 2023). Setiap komputer yang terhubung memungkinkan berbagi penyimpanan, daya

BAB 14 KEAMANAN DAN PRIVASI DALAM *BIG DATA*

Pendahuluan

Dunia kini memiliki istilah “*Everything is digital*”, karena pada abad ini memasuki sebuah era digitalisasi yang disebut era revolusi industri 4.0, yakni sebuah era yang menekankan pada pola *digital economy* (Pujianto et al., 2018). *Big Data* merupakan satu terminologi yang merupakan ketersediaan data sosial yang terekam secara digital. *Big Data* adalah kumpulan data dalam skala yang sangat besar, baik yang tersusun rapi maupun yang acak, yang umumnya tersimpan di *cloud* atau pusat data (Munawar et al., 2023). Banyak pihak yang memanfaatkan data tersebut, seperti perusahaan hingga lembaga pemerintahan, guna mengambil keputusan yang lebih cerdas dan strategis. Di era digital yang kian canggih, *Big Data* memegang peranan sentral dalam merumuskan berbagai keputusan krusial dari ranah pemerintahan hingga aktivitas sehari-hari. Data yang telah dikumpulkan dari media sosial, transaksi digital dan berbagai sumber lainnya, membentuk pola dan wawasan yang luar biasa kaya. Namun, seiring dengan potensinya yang begitu besar, hadir pula tantangan yang tak kalah penting: bagaimana memastikan bahwa data yang dikumpulkan dan dianalisis itu tetap aman, dan lebih dari itu, menghormati hak-hak privasi individu.

Pengelolaan *Big Data* dalam ranah keamanan dan privasi melampaui urusan teknis semata, ini merupakan persoalan etis yang tak terelakkan. Bayangkan jutaan informasi pribadi

BAB 16 TEKNIK DATA MINING UNTUK *BIG DATA*

Pendahuluan

Big Data telah menjadi fondasi utama pengambilan keputusan modern. *Volume* data yang terus meningkat memaksa kita untuk menggunakan pendekatan analitik canggih, salah satunya adalah data *mining*. Data *mining* mengacu pada proses ekstraksi pengetahuan tersembunyi dari data berukuran besar menggunakan teknik statistik, *Machine Learning*, dan algoritma komputasi skala besar (Aljumily & Barakat, 2021). Dalam konteks *Big Data*, pendekatan ini harus dilakukan secara efisien di lingkungan terdistribusi seperti Apache *Spark*, *Hadoop*, dan *Cloud Computing platforms*. Pekerjaan dalam bidang data *mining* secara umum dapat dikategorikan ke dalam empat jenis utama, yaitu: pemodelan prediktif (*prediction modeling*), analisis klaster (*cluster analysis*), analisis asosiasi (*association analysis*), dan pendeteksian anomali (*anomaly detection*). Proses *data mining* itu sendiri umumnya terdiri atas tiga tahapan utama yang berlangsung secara sistematis:

1. Eksplorasi dan Pra-pemrosesan data. Tahap awal mencakup berbagai kegiatan seperti pembersihan data, normalisasi, transformasi, penanganan data yang tidak valid, reduksi dimensi, dan seleksi fitur. Semua aktivitas ini bertujuan untuk menyiapkan data agar siap dianalisis secara lebih lanjut.
2. Pembangunan dan validasi model. Pada fase ini, dilakukan penerapan berbagai metode analitik untuk

BAB 17 PENGENALAN *MACHINE LEARNING* DALAM *BIG DATA*

Pendahuluan

Di era digital yang semakin berkembang pesat, Data yang dihasilkan dari berbagai sumber seperti media sosial, sensor IoT, transaksi digital, dan perangkat *mobile* ini sering kali disebut sebagai *Big Data*. *Big Data* tidak hanya memiliki *volume* besar, tetapi juga memiliki kecepatan tinggi dalam perolehan data (*velocity*), variasi dalam jenis data (*variety*), serta tantangan dalam validitas dan nilai data tersebut. Seiring dengan pertumbuhan *Big Data*, muncul kebutuhan akan metode yang mampu mengekstrak informasi dan wawasan dari data dalam skala besar secara efisien dan efektif. Di sinilah peran *Machine Learning* menjadi sangat penting. *Machine Learning* merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma dan model yang dapat belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit.

Penerapan *Machine Learning* dalam konteks *Big Data* membuka banyak peluang, mulai dari analisis prediktif, klasifikasi pola perilaku pengguna, deteksi anomali, hingga sistem rekomendasi yang personal. Namun, penggabungan dua konsep besar ini juga menimbulkan tantangan tersendiri, seperti skala komputasi, pemrosesan data tidak terstruktur, dan perlunya model yang adaptif terhadap data yang terus berubah. Bab ini akan memberikan gambaran umum tentang konsep dasar *Machine Learning*, karakteristik *Big Data*, serta

BAB 18 ANALISIS DATA *REAL TIME* MENGUNAKAN *STREAMING PROCESSING*

Pendahuluan

Di era digital saat ini, informasi memegang peranan yang sangat penting dalam pengambilan keputusan, inovasi teknologi, dan pengembangan bisnis. Peningkatan *volume* data yang muncul dari berbagai sumber—baik internal maupun eksternal—menimbulkan kebutuhan mendesak akan pemrosesan informasi yang cepat dan akurat agar dapat merespons perubahan dengan tepat waktu. Hal inilah yang menjadi dasar lahirnya paradigma baru pengolahan data, yakni pemrosesan data secara *real time* melalui *streaming processing*.

Perubahan drastis dalam teknologi sensor, komunikasi data, dan penyimpanan telah membuat data mengalir terus menerus dalam bentuk *stream* atau aliran tanpa henti. Data tersebut layaknya gelombang informasi yang sangat dinamis dan memerlukan sistem yang mampu menangkap dan mengolahnya secara simultan tanpa menunggu seluruh data terkumpul terlebih dulu seperti pada sistem pemrosesan *batch* tradisional. Oleh karenanya, pemrosesan *streaming* muncul sebagai solusi yang memungkinkan pemanfaatan data secara simultan dan berkelanjutan.

Buku ini menawarkan panduan komprehensif mengenai berbagai aspek pemrosesan data *real time* dengan memfokuskan pada *streaming processing*. Tujuan utama adalah menyediakan landasan teoretis dan praktis yang dapat digunakan oleh mahasiswa dan akademisi yang memiliki dasar pengetahuan

BAB 19 ANALISIS VISUALISASI DATA UNTUK *BIG DATA*

Pendahuluan

Kita hidup di zaman ketika data membanjiri hampir setiap aspek kehidupan. Di era *Big Data* ini, informasi dalam kehidupan sehari-hari hadir dalam jumlah besar, bergerak dengan cepat, dan memiliki struktur yang sangat bervariasi (Salma, dkk., 2021). Namun, tidak semua data yang dikumpulkan serta-merta menjadi pengetahuan yang berguna. Di sinilah peran visualisasi data menjadi sangat penting, sebagai jembatan antara kompleksitas data dan kemampuan manusia untuk memahaminya secara cepat dan intuitif. Dalam konteks *Big Data*, visualisasi bukan sekadar pemanis tampilan, melainkan alat vital untuk mengungkap makna tersembunyi dari data yang sangat besar, cepat berubah, dan sangat bervariasi. Dalam konteks *Big Data*, visualisasi menghadirkan tantangan dan peluang unik.

Dengan jumlah data yang mencapai jutaan *entri* per detik, metode visualisasi tradisional tak lagi memadai. Diperlukan pendekatan baru yang adaptif dan interaktif, mulai dari pemilihan *grafik* yang tepat hingga pemanfaatan teknologi berbasis *cloud* dan *real-time rendering*. Visualisasi modern bukan hanya soal tampilan, tetapi juga tentang bagaimana menjembatani interaksi manusia dengan mesin melalui representasi data yang cerdas. Bab ini membahas secara mendalam berbagai aspek visualisasi *Big Data*, termasuk teknik visualisasi terbaru, pemilihan alat bantu yang sesuai, serta

BAB 20 MANAJEMEN *BIG DATA* DALAM BISNIS DAN INDUSTRI

Pendahuluan

Dalam era digital yang terus berkembang, data telah berubah menjadi aset strategis yang sangat penting bagi dunia bisnis dan industri. Setiap aktivitas digital—mulai dari interaksi pelanggan, transaksi keuangan, hingga operasional mesin—menghasilkan data dalam jumlah besar dengan kecepatan tinggi dan beragam bentuk. Pertumbuhan data yang sangat masif ini telah melampaui kemampuan sistem basis data tradisional, sehingga memunculkan kebutuhan akan pendekatan baru yang dikenal sebagai *Big Data*.

Konsep *Big Data* bukan sekadar soal *volume* data yang besar, melainkan juga bagaimana data tersebut dikelola dan dimanfaatkan untuk mendukung proses bisnis secara cerdas dan efektif. Dalam praktiknya, perusahaan-perusahaan memanfaatkan *Big Data* untuk membaca pola perilaku konsumen, meningkatkan efisiensi kerja, serta menciptakan inovasi berbasis informasi. Di sektor industri, peran *Big Data* semakin penting untuk melakukan prediksi, pemeliharaan sistem secara *real-time*, dan meningkatkan ketahanan proses produksi terhadap perubahan yang tidak terduga (S.C. Huang, S. McIntosh, S. Sobolevsky, & P. C. K. Hung, 2017).

Pengelolaan *Big Data* melibatkan rangkaian tahapan penting seperti pengumpulan data, penyimpanan dalam sistem terdistribusi, pemrosesan data berskala besar, hingga visualisasi informasi guna mendukung pengambilan keputusan. Oleh

BAB 21 IMPLEMENTASI *BIG DATA* DALAM *SMART CITY* DAN IOT

Pendahuluan

Didorong oleh kemajuan pesat dalam teknologi dan *data science*, transformasi besar sedang melanda seluruh kota di dunia. Kemajuan teknologi ini membawa perubahan dalam pengelolaan kota dan kehidupan masyarakat. Teknologi yang cukup berperan dalam perubahan ini adalah pemanfaatan *Big Data*. Sebagaimana kita tahu bahwa saat ini setiap detik akan terus ada data baru yang bertambah sehingga data yang disimpan tidak lagi dalam jumlah yang relatif kecil. *Big Data* dapat diartikan sebagai bidang yang didedikasikan untuk analisis, pemrosesan, dan penyimpanan kumpulan data besar yang sering kali berasal dari berbagai sumber (Erl et al., 2016). Dalam konteks ini, *Big Data* tidak hanya berkaitan dengan besarnya *volume* data, tetapi juga mencakup kemampuan untuk menggali informasi dan wawasan berharga dari data dalam jumlah yang sangat besar (Iskandar et al., 2024).

Pemanfaatan *Big Data* ini punya dampak yang besar untuk pengelolaan kota dan kehidupan masyarakat sehingga dapat membentuk konsep *Smart City*. *Big Data* dapat sangat bermanfaat bagi *Smart City* jika informasi penting dapat diekstraksi dan dianalisis dari *volume* data (Dey et al., 2018). *Smart City* merupakan istilah yang berasal dari adopsi dan penerapan sistem komputasi bergerak melalui jaringan manajemen data di antara semua komponen dan lapisan kota itu sendiri (Kirimtat et al., 2020). Dalam konteks ini,

BAB 22 *BIG DATA* DALAM *HEALTHCARE* DAN KEUANGAN

Pendahuluan

Perkembangan teknologi digital telah melahirkan gelombang transformasi besar dalam berbagai sektor, termasuk layanan kesehatan dan industri keuangan. Di tengah arus digitalisasi ini, istilah *Big Data* menjadi kunci utama dalam menjawab tantangan kompleks serta mendorong inovasi yang lebih cerdas dan terukur. *Big Data* tidak hanya sekadar kumpulan data dalam jumlah besar, tetapi merupakan kerangka kerja yang mampu menangkap, menyimpan, mengelola, dan menganalisis informasi dalam *volume* yang sangat masif, beragam bentuk, dan terus bertambah dengan cepat. Keunggulan ini telah mengubah cara organisasi membuat keputusan, merancang layanan, dan menjangkau pengguna secara lebih personal serta efisien.

Dalam dunia layanan kesehatan, *Big Data* hadir sebagai landasan baru dalam membangun sistem yang lebih tanggap, prediktif, dan personal. Data yang berasal dari rekam medis elektronik, hasil pemeriksaan laboratorium, perangkat *wearable*, hingga interaksi di media sosial kini dikumpulkan dan dianalisis untuk memahami kondisi pasien secara menyeluruh. Transformasi ini tidak hanya mempercepat proses diagnosis dan pengobatan, tetapi juga memperkuat pengembangan pengobatan presisi (*precision medicine*), meningkatkan efisiensi operasional rumah sakit, serta memungkinkan pemantauan kesehatan masyarakat secara *real-time*. *Big Data* juga

BAB 23 TREN MASA DEPAN DALAM *BIG DATA* DAN DATA TERDISTRIBUSI

Masa Depan *Big Data*

Big Data akan terus berkembang dan menjadi fondasi utama dalam transformasi digital serta inovasi di berbagai bidang. *Big Data* sendiri menggambarkan *volume* informasi yang sangat besar, yang dihasilkan, disimpan, dan dianalisis oleh berbagai industri untuk meningkatkan layanan dan pengambilan keputusan. Latar belakang munculnya *Big Data* didorong oleh keterbatasan sistem pemrosesan data tradisional yang tidak mampu menangani skala data yang masif, sehingga dibutuhkan berbagai teknik khusus untuk mengelola dan menganalisis data yang terus berkembang dengan cepat. Ciri khas *Big Data* mencakup berbagai data terstruktur, semi terstruktur, dan tidak terstruktur (Bansal et al., 2021). Saat ini dibagi menjadi 10 (Sepuluh) tingkat yaitu: *Volume*, *Velocity*, *Variety*, *Veracity*, *Variability*, *Visualization*, *Vulnerability*, *Volatility*, *Validity*, dan *Value* (Niu et al., 2021). Berbagai terobosan hadir untuk menangani data besar diantaranya:

Integrasi Kecerdasan Buatan dan Pembelajaran

Mesin

Big Data dan Kecerdasan Buatan (*Artificial Intelligence*) merupakan dua bidang yang terus berkembang dan saling melengkapi. *Big Data* merujuk pada kumpulan data berukuran sangat besar dengan variasi jenis yang kompleks, sedangkan

Daftar Pustaka

- Abernathey, R. P., Augspurger, T., Banihirwe, A., Blackmon-Luca, C. C., Crone, T. J., Gentemann, C. L., Hamman, J. J., Henderson, N., Lepore, C., McCaie, T. A., Robinson, N. H., & Signell, R. P. (2021). Cloud-native Repositories for Big Scientific Data. *Computing in Science and Engineering*, 23(2), 26–35. <https://doi.org/10.1109/MCSE.2021.3059437>
- Adje, K. D. C., Letaifa, A. Ben, Haddad, M., & Habachi, O. (2023). Smart City Based on Open Data: A Survey. In *IEEE Access* (Vol. 11, pp. 56726–56748). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2023.3283436>
- Ajie, R. (2025). Optimasi Kinerja Basis Data Terdistribusi Menggunakan Algoritma Replication dan Partitioning. 3(2), 371–378.
- Ajismanto, F., Mahmud, M., Barovih, G., Rupilele, F. G. J., Guntoro, G., Octafian, D. T., Aprizal, Y., Siregar, M. T., Afnarius, S., Krisnanik, E., & Widiastiwi, Y. (2023). Sistem Terdistribusi. WIDINA BHAKTI PERSADA BANDUNG.
- Akidau, T., et al. (2015). The Dataflow Model: A Practical Approach to Balancing Correctness, Latency, and Cost in Massive-Scale, Unbounded, Out-of-Order Data Processing. *Proceedings of the VLDB Endowment*.
- Al-Sai, Z. A., Husin, M. H., Syed-Mohamad, S. M., Abdin, R. M., Damer, N., Abualigah, L., & Gandomi, A. H. (2022). Explore Big Data Analytics Applications and Opportunities: A Review. In *Big Data and Cognitive Computing* (Vol. 6, Issue 4). <https://doi.org/10.3390/bdcc6040157>

- Albahri, O., Alamleh, A., Al-Quraishi, T., & Thakkar, R. (2023). Smart real-time iot mhealth-based conceptual framework for healthcare services provision during network failures. *Applied Data Science and Analysis*, 2023, 110-117.
- Ali, I., Rehman, H., & Shahid, A. (2022). A performance comparison of Machine Learning models for large-scale classification. *Journal of Big Data*, 9(1), 44.
- Aljumily, H., & Barakat, S. (2021). Data mining in Big Data environment: Challenges and opportunities. *International Journal of Advanced Computer Science and Applications*, 12(3), 99–105.
- Almassabi, A., Bawazeer, O., & Adam, S. (2018). Top NewSQL Databases and Features Classification. *International Journal of Database Management Systems*, 10(2).
- American Public Transportation Association (APTA). (2020). Leveraging Big Data in the Public Transportation Industry (Policy Brief). Washington, DC: APTA..
- Apache Spark Documentation 4.0.0. (2025). Apache Software Foundation. <https://Spark.apache.org/docs/latest/>
- Arhami, M., & Nasir, M. (2020). Data Mining Algoritma dan Implementasi. Penerbit Andi.
- Ariyani, R. (n.d.). Basis Data Apache Cassandra Sebagai Media Penyimpanan Data Sensor Dan Backend Internet Of Things (IOT). *JUIT*, 1(2).
- Ashish MK. 2014. “Security and Concurrency Control in Distributed Database System”. *International Journal of Scientific Research and Management*. 2014 Dec 5;2(12)
- Aswal, S. (2023). Distributed database systems for large-scale data management. *Turkish Online Journal of Qualitative Inquiry (TOJQI)*, 11(4), 2260-2269.
- Aumont, O., Pierre, U., Marie, E. T., Paris, C., Gómez-Letona, M.,

- Ramos, A. G., Coca, J., Arístegui, J., Machu, E., France, L. P. O. B., Auger, P. A., Gorgues, T., Machu, E., Aumont, O., Brehmer, P., 1.1, Lathuilière, C., Echevin, V., Lévy, M., ... Hilmi, K. (2018). PANDUAN MENGHADAPI SERANGAN DENIAL OF SERVICE (DOS) Untuk Badan Usaha Kecil dan Menengah. *Biogeosciences*, 43(5), 679–694. <http://www.unil.ch/ssp/page34569.html>
- Ayvaz, S., & Alpaz, K. (2021). Predictive maintenance system for production lines in manufacturing: A Machine Learning approach using IoT data in real-time. *Expert Systems with Applications*, 173. <https://doi.org/10.1016/j.eswa.2021.114598>
- Babar, M., Jan, M. A., He, X., Tariq, M. U., Mastorakis, S., & Alturki, R. (2023). An Optimized IoT-Enabled Big Data Analytics Architecture for Edge-Cloud Computing. *IEEE Internet of Things Journal*, 10(5), 3995–4005. <https://doi.org/10.1109/JIOT.2022.3157552>
- Bach, M. P., Krstič, Ž., Seljan, S., & Turulja, L. (2019). Text mining for Big Data analysis in financial sector: A literature review. *Sustainability (Switzerland)*, 11(5). <https://doi.org/10.3390/su11051277>
- Bahri, S., Zoghlami, N., Abed, M., & Tavares, J. M. R. S. (2019). BIG DATA for Healthcare: A Survey. *IEEE Access*, 7, 7397–7408. <https://doi.org/10.1109/ACCESS.2018.2889180>
- BAMS. (2024). Big Data analysis for medical systems. *Biomedical Analytics Management Systems*, 12(1), 44–55.
- Bansal, R., Obaid, A. J., Gupta, A., Singh, R., & Pramanik, S. (2021). Impact of Big Data on Digital Transformation in 5G Era. *Journal of Physics: Conference Series*, 1963(1). <https://doi.org/10.1088/1742-6596/1963/1/012170>

- Bintoro, P., Ratnasari, Wihardjo, E., Putri, I. P., & Asari, A. (2024). Pengantar Machine Learning. PT MAFY MEDIA LITERASI INDONESIA.
- BPS Indonesia. (2020). Kajian Big Data sebagai Pelengkap Data dan Informasi Statistik Ekonomi.
- Buyya, R., Calheiros, R. N., & Dastjerdi, A. V. (2016). Big Data: Principles and paradigms. Morgan Kaufmann.
- C. P. Peralta, J. Z. Mondragon, L. S. Terrones, & J. R. Villacorta. (2024). Big Data analysis and its impact on the marketing industry: a systematic review. *Indonesian Journal of Electrical Engineering and Computer Science*, 1032-1040.
- Carvalho, G., Cabral, B., Pereira, V., & Bernardino, J. (2021). Edge computing: current trends, research challenges and future directions. *Computing*, 103(5), 993–1023. <https://doi.org/10.1007/s00607-020-00896-5>
- Carvalho, I., Sá, F., & Bernardino, J. (2023). Performance Evaluation of NoSQL Document Databases: Couchbase, CouchDB, and MongoDB. *Algorithms*, 16(2). <https://doi.org/10.3390/a16020078>
- Ceri, S., Pernici, B., and Wiederhold, G. (1987). Distributed database design methodologies. *Proc. IEEE*, 75(5):533–546. 38, 73, 125
- Chambers, B., & Zaharia, M. (2018). Spark: The definitive guide – Big Data processing made simple. O'Reilly Media. <https://www.oreilly.com/library/view/Spark-the-definitive/9781491912201/>
- Chen, C. L. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314–347. <https://doi.org/10.1016/j.ins.2014.01.015>
- Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile*

- Networks and Applications, 19(2), 171–209.
<https://doi.org/10.1007/s11036-013-0489-0>
- Chen, Z., Yi, W., Nallanathan, A., & Chambers, J. (2024). Distributed digital twin migration in multi-tier computing systems. *Ieee Journal of Selected Topics in Signal Processing*, 18(1), 109-123.
- Clark, L. (2018, May 16). Apache Spark speeds up Big Data decision-making. *Computer Weekly*.
<https://www.computerweekly.com/feature/Apache-Spark-speeds-up-big-data-decision-making>
- Cloud, G. (2025). Google Distributed Cloud overview.
<https://cloud.google.com/distributed-cloud/docs/overview>
- Cremonini, M. (2024). Data visualization in R and Python. John Wiley & Sons.
- Cybertrend Intrabuana. (2020). Business Analytics & Big Data In Cloud Computing. Diakses dari <https://cybertrend-intra.com/article/business-analytics-big-data-in-cloud-computing/>
- D’Alberto, R., & Giudici, H. (2023). A sustainable smart mobility? Opportunities and challenges from a Big Data use perspective. *Sustainable Futures*, 6.
<https://doi.org/10.1016/j.sftr.2023.100118>
- Damji, J. (2016, July 14). A tale of three Apache Spark APIs: RDDs, DataFrames, and *Datasets*. Databricks Blog.
<https://databricks.com/blog>
- Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904-920.
- Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big Data in healthcare: management, analysis and future

- prospects. *Journal of Big Data*, 6(1).
<https://doi.org/10.1186/s40537-019-0217-0>
- Date, C. J., 2004. *An Introduction to Database Systems* (Edisi ke-8), Pearson Education: Boston.
- Dean, J., & Ghemawat, S. (2008). MapReduce. *Communications of the ACM*, 51(1), 107–113.
<https://doi.org/10.1145/1327452.1327492>
- Dey, N., Hassanien, A. E., Bhatt, C., Ashour, A. S., & Satapathy, S. C. (2018). Internet of Things and Big Data Analytics Toward Next-Generation Intelligence.
<http://www.springer.com/series/11970>
- Dhaniawaty, R. P. (2016). *Pengantar Sistem Informasi*.
- Dimitropoulos, N., Michalos, G., Arkouli, Z., Kokotinis, G., & Makris, S. (2024). Industrial collaborative environments integrating AI, Big Data and Robotics for smart manufacturing. *Procedia CIRP*, 128, 858–863.
<https://doi.org/10.1016/j.procir.2024.04.027>
- Dorogush, A. V., Ershov, V., & Gulin, A. (2021). CatBoost: gradient boosting with categorical features support. *arXiv:1810.11363*.
- DQLab. (2021). *Memahami Hadoop dan Implementasinya dalam Big Data*.
- Du, Y., Cai, Z., & Ding, Z. (2024). Query optimization in distributed database based on improved artificial bee colony algorithm. *Applied Sciences*, 14(2), 846.
- Eliviani, R., & Bandung, Y. (2023). WSN-IoT Forecast: Wireless Sensor Network Throughput Prediction Framework in Multimedia Internet of Things. *Journal of ICT Research and Applications*, 17(3), 336–355.
<https://doi.org/10.5614/itbj.ict.res.appl.2023.17.3.4>
- Elmasri, R., & Navathe, S. B., 2016. *Fundamentals of Database*

- Systems (Edisi ke-7), Pearson Education: Boston.
- Emara, T. Z., & Huang, J. Z. (2020). Distributed data strategies to support large-scale data analysis across geo-distributed data centers. *IEEE Access*, 8(October), 178526–178538. <https://doi.org/10.1109/ACCESS.2020.3027675>
- Erl, T., Khattak, W., & Buhler, P. (2016). *Big Data fundamentals: Concepts, drivers & techniques*. Prentice Hall.
- Ezechiel. K. K., dkk. 2019. “A Systematic Review On Distributed Databases Systems And Their Techniques”. Volume 96 Nomor 1. *Journal of Theoretical and Applied Information Technology (JATIT)*. ISSN: 1992-8645
- F. Madani, S. V. R. Nooraei, & M. M. Parast. (2024). *Big Data Analytics (BDA): Principles, Premises, and Applications in Organizational Research*. *International Journal of Data Science and Big Data Analytics*, 79-91.
- Fadli, A., Zulfa, M. I., Widhi Nugraha, A. W., Taryana, A., & Aliim, M. S. (2020). Analisis Perbandingan Unjuk Kerja Database SQL dan Database NoSQL Untuk Mendukung Era Big Data. *JURNAL NASIONAL TEKNIK ELEKTRO*, 9(3). <https://doi.org/10.25077/jnte.v9n3.774.2020>
- Fathansyah, 2018. “Basis Data” Revisi ketiga, Informatika: Bandung.
- Fawzy, D., Moussa, S. M., & Badr, N. L. (2022). The Internet of Things and Architectures of Big Data Analytics: Challenges of Intersection at Different Domains. In *IEEE Access* (Vol. 10, pp. 4969–4992). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2022.3140409>
- Fitri, M. O. (2013). Trend penggunaan NoSQL untuk basis data non-relasional. *Jurnal Teknosains*, 7(1), 120–127.
- Freitas, N., Rocha, A. D., & Barata, J. (2025). *Data management*

- in industry: concepts, systematic review and future directions. *Journal of Intelligent Manufacturing*, 1. <https://doi.org/10.1007/s10845-025-02570-z>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
- Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. *Proceedings of the nineteenth ACM symposium on Operating systems principles*, 29–43. <https://doi.org/10.1145/945445.945450>
- Gonzalez, J. E., Xin, R. S., Dave, A., Crankshaw, D., Franklin, M. J., & Stoica, I. (2014). GraphX: Graph processing in a distributed dataflow framework. In *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14)* (pp. 599–613). USENIX Association. <https://www.usenix.org/system/files/conference/osdi14/osdi14-paper-gonzalez.pdf>
- Google Cloud. (2024). Big Data dan Referensi Analisis. Cloud Architecture Center.
- Gorbenko, A., Romanovsky, A., & Tarasyuk, O. (2020). Interplaying cassandra nosql consistency and performance: A benchmarking approach. *Communications in Computer and Information Science*, 1279 CCIS, 168–184. https://doi.org/10.1007/978-3-030-58462-7_14
- Guo, L. (2023). Research on federated learning data sharing scheme based on differential privacy. *Computers Materials & Continua*, 74(3), 5069–5085.
- Hajjaji, Y., Boulila, W., Farah, I. R., Romdhani, I., & Hussain, A. (2021). Big Data and IoT-based applications in smart

- environments: A systematic review. In *Computer Science Review* (Vol. 39). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.cosrev.2020.100318>
- Hamzah. M., 2018. *Transformasi Digital di Pedesaan: Strategi dan Tantangan*, UII Press: Yogyakarta.
- Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of Big Data on finance. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00291-z>
- Hasan, M., Hoque, A., & Le, T. (2023). Big Data-driven Banking Operations: Opportunities, Challenges, and Data Security Perspectives. *FinTech*, 2(3), 484–509. <https://doi.org/10.3390/fintech2030028>
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., et al. (2015). The rise of “Big Data” on Cloud Computing: Review and open research issues. *Information Systems*, 47, 98–115.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of “Big Data” on Cloud Computing: Review and open research issues. *Information Systems*, 47, 98–115. <https://doi.org/10.1016/j.is.2014.07.006>
- Hasibuan, S. H., Irwan, M., & Nasution, P. (2023). A Comparative Study of Relasional and NoSQL database for Big Data Analytics (Vol. 02, Issue 03). <http://jurnal.minartis.com/index.php/jpst/>
- He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212, 106622. <https://doi.org/10.1016/j.knosys.2020.106622>
- Hesse, G., & Lorenz, D. (2019). *A Stream Processing Perspective on Big Data*. Springer. <https://www.tutorialspoint.com>
- Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). *Toward Scalable*

- Systems for Big Data Analytics: A Technology Tutorial. IEEE Access, 2, 652–687. <https://doi.org/10.1109/ACCESS.2014.2332453>
- Huang, L. (2025). Creating a dashboard for interactive data visualization with Dash in Python. *Programming Historian*, 14. <https://doi.org/10.46430/phen0124>
- Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big Data for dummies*. John Wiley & Sons.
- IBM. (2025a). Apa itu data replication? IBM. <https://www.ibm.com/id-id/topics/data-replication>
- IBM. (2025b). Apa itu latensi? IBM. <https://www.ibm.com/id-id/topics/latency>
- IBM. (2025c). Apa yang dimaksud dengan teorema CAP? IBM. <https://www.ibm.com/id-id/topics/cap-theorem>
- idBigData. (2024). *Mengenal Apache Spark: Komponen dan Fungsinya*.
- Iskandar, A. P. S., Setiawan, H., Judijanto, L., Mahendra, G. S., Ardi, M., Putri, N. A. R., Handika, I. P. S., Ratih, Mandowen, S. A., Wazaumi, D. D., Winahyu, R. R. K. K., Atikah, L., Wardani, D. K., Eliviani, R., Kusumawati, A., Ratwastuti, N., & Bowo, I. T. (2024). *Teknologi Big Data Pengantar dan Penerapan Teknologi Big Data di Berbagai Bidang* (E. Rianty, Ed.). PT. Green Pustaka Indonesia. www.greenpustaka.com
- J, J. (2022). Cassandra is a Better Option for Handling Big Data in a No-SQL Database. *International Journal of Research Publication and Reviews*, 880–883. <https://doi.org/10.55248/gengpi.2022.3.9.27>
- Jain, P., Gyanchandani, M., & Khare, N. (2016). Big Data privacy: a technological perspective and review. *Journal of Big Data*, 3(1), 1–25.

- Jaya, H., Djawad, Y. A., Suhaeb, S., Saharuddin, & Idhar. (2017). Embedded systems and robotics. In UNM. UNM.
- John, B. (2025). Data Processing in Data Lakes vs. Data Warehouses.
- Kamalina, A. R. (2025). Sistem Penukaran Uang Pintar BI Sempat Error, Bank Indonesia Sebut Ada Hacker DDoS. Finansial Bisnis. <https://finansial.bisnis.com/read/20250319/11/1862969/sistem-penukaran-uang-pintar-bi-sempat-eror-bank-indonesia-sebut-ada-hacker-ddos#:~:text=Bisnis.com%2C JAKARTA — Bank,-of-Service alias DDoS>.
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in Big Data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561–2573. <https://doi.org/10.1016/j.jpdc.2014.01.003>
- Karr, A. F., Kohnen, C. N., Oganian, A., Reiter, J. P., & Sanil, A. P. (2006). A framework for evaluating the utility of data altered to protect confidentiality. *American Statistician*, 60(3), 1–9.
- Khan, M. A., Siddiqui, M. S., Rahmani, M. K. I., & Husain, S. (2022). Investigation of Big Data Analytics for Sustainable Smart City Development: An Emerging Country. *IEEE Access*, 10, 16028–16036. <https://doi.org/10.1109/ACCESS.2021.3115987>
- Khine, P. P., & Wang, Z. 2019. A Review of Polyglot Persistence in the Big Data World. *Applied Sciences*, 9(8), 1599.
- Khoirunnisya, G. S., & Nasution, M. I. P. (2024). Perkembangan Teknologi Informasi di Indonesia: Tantangan dan Peluang. *Jurnal Sains Student Research*, 3(1).
- KHONDE Noel, B., MANGATA Bopatriciat, B., MUKENDI

- Eugène, M., & CHRISTIAN Parfum, B. (n.d.). STUDY AND IMPROVEMENT OF PERFORMANCE OF NoSQL DATABASES: MongoDB, HBase and OrientDB. <http://www.oracle.com/technetwork/java/ja>
- Kirimtat, A., Krejcar, O., Kertesz, A., & Tasgetiren, M. F. (2020). Future Trends and Current State of Smart City Concepts: A Survey. In IEEE Access (Vol. 8, pp. 86448–86467). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2020.2992441>
- Kleppmann, M. (2017). Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. O'Reilly Media.
- Kreps, J., Narkhede, N., & Rao, J. (2017). Kafka: The Definitive Guide. O'Reilly Media.
- Kumar, R., Grover, N., Singh, R., Kathuria, S., Kumar, A., & Bansal, A. (2023). Imperative Role of Artificial Intelligence and Big Data in Finance and Banking Sector. 2nd International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2023 - Proceedings, 523–527. <https://doi.org/10.1109/ICSCDS56580.2023.10105062>
- Kusnadi, D., 2021. Digitalisasi Pendidikan dan Tata Kelola Data Akademik, Alfabeta: Bandung.
- L. Judijanto, S. Uhai, & I. Suri. (2024). The Influence of Business Analytics and Big Data on Predictive Maintenance and Asset Management . The Eastasouth Journal of Information System and Computer Science, 123-135.
- Lam, C. (2010). *Hadoop in Action* (1st ed.). Manning.
- Laney, D. (2001). 3-D data management: Controlling data volume, velocity and variety. META Group (Research Note).

- Lawton, G. (2022, February 17). *Hadoop* vs. Spark: An in-depth Big Data framework comparison. TechTarget SearchDataManagement.
<https://www.techtarget.com/searchdatamanagement/feature/Hadoop-vs-Spark-Comparing-the-two-big-data-frameworks>
- Lee, Y. J., Park, S., & Han, K. (2021). Scalable deep embedded clustering with parallel optimization. *Information Sciences*, 577, 212–228.
- Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). Mining of massive *datasets* (3rd ed.). Cambridge University Press.
- Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692.
- Lnenicka, M., Hervert, P., & Horak, O. (2024). Understanding Big Data and data protection measures in Smart City strategies: An analysis of 28 cities. *Urban Governance*.
<https://doi.org/10.1016/j.ugj.2024.12.008>
- Lundberg, S. M., & Lee, S.-I. (2020). A unified approach to interpreting model predictions. *NeurIPS*, 33, 4765–4774.
- M. K. Hidayat, E. Priyanti, D. Putri dan R. Ardianto, 2022. “Sistem Manajemen Basis Data”, Yogyakarta: Teknosain.
- M. N. O. Sadiku, P. O. Adebo, & S. M. Musa. (2018). Big Data in Business. *International Journal of Advance Research in Computer Science and Software Engineering*, 160-162.
- Ma, X., Li, J., Guo, Z., & Wan, Z. (2024). Role of Big Data and technological advancements in monitoring and development of smart cities. *Heliyon*, 10(15).
<https://doi.org/10.1016/j.heliyon.2024.e34821>
- Malik, N. K., Tasya, A., Ayu, P., Karlina, N. A., & Lestiyarini, D. R. (2020). Replikasi Database Dengan Model Master-Master dan *Master-slave* MariaDB l. December.

- Mandala, N. R. (2023). Advances in Distributed Storage Systems for Big Data. *Journal of Mathematical & Computer Applications*, 1–8. [https://doi.org/10.47363/JMCA/2023\(2\)E136](https://doi.org/10.47363/JMCA/2023(2)E136)
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Maryanto, B. 2017. Big Data Dan Pemanfaatannya Dalam Berbagai Sektor. *Media Informatika*, 16(2).
- Marz, N. (2015). *Big Data: Principles and Best Practices of Scalable Real-time Data Systems*. Manning Publications.
- McAfee, A., & Brynjolfsson, E. (2012). Big Data: the management revolution. *Harvard Business Review*, 90(10), 60-68.
- McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform manifold approximation and projection. *arXiv:1802.03426*.
- Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing (SP 800-145)*. Gaithersburg, MD: National Institute of Standards and Technology.
- Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., ... & Zadeh, R. (2016). MLlib: Machine Learning in Apache Spark. *Journal of Machine Learning Research*, 17(34), 1–7. <http://jmlr.org/papers/v17/15-237.html>
- MENTERI PENDIDIKAN, KEBUDAYAAN, RISET, D. T., & INDONESIA, R. (2022). *Keputusan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia. Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia*, 33(1), 1–12.
- Mhlanga, D. (2024). The role of Big Data in financial technology toward financial inclusion. *Frontiers in Big Data*, 7.

- <https://doi.org/10.3389/fdata.2024.1184444>
- Microsoft. (2024). Data SQL vs. NoSQL. Microsoft. <https://learn.microsoft.com/id-id/dotnet/architecture/cloud-native/relational-vs-nosql-data>
- Microsoft. (2025, April 9). Tutorial - Gunakan Apache HBase di Microsoft Azure HDInsight. <https://learn.microsoft.com/id-id/azure/hdinsight/hbase/apache-hbase-tutorial-get-started-linux>.
- Moussa, S. (2021). Are FT50 journals really leading? A comment on Fassin. *Scientometrics*, 126. <https://doi.org/10.1007/s11192-021-04158-9>
- Munawar, Z., Muliantara, Agus Kmurswak, R. M., Reba, A. F., Sroyer, A. M., Sukmawan Dede, Asianingrum, A. R. A. H., Insany, G. P., Mandowen, S. A., Toyib Kamdan Wildan, Yoga S, T. A., Kharisma, I. L., & Beno, I. S. (2023). *Big Data Analytics Konsep, Implementasi, dan Aplikasi Terkini*. Kaizen Media Publishing.
- Naheman, W., & Wei, J. (2013). Review of NoSQL databases and performance testing on HBase. *Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, 2304–2309. <https://doi.org/10.1109/MEC.2013.6885425>
- Nguyen, T. and Kawashima, H. (2025). Fairly decentralizing a hybrid concurrency control protocol for real-time database systems. *Concurrency and Computation Practice and Experience*, 37(4-5).
- NIST. (2015). *NIST Big Data Interoperability Framework: Volume 4, Security and Privacy*. NIST Special Publication, 4, 75.

- Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C. B. (2021). Organizational business intelligence and decision making using Big Data analytics. *Information Processing and Management*, 58(6).
<https://doi.org/10.1016/j.ipm.2021.102725>
- NoLimit. (2024). Menjelaskan Konsep 5V dalam Big Data.
- O. Placida., dan Asagba., P.O. 2021. “Distributed Database Management System (Dbms) Architectures And Distributed Data Independence”. *IJCSMC*, Vol. 10, Issue. 1, January 2021, pg.23 – 48.
- Ognyanova, K. (2024). Network visualization with R. Kateto. Retrieved May 18, 2025, from <https://kateto.net/network-visualization>
- Oluwaferanmi, A. (2025). Integrating MLOps and DataOps for Scalable and Resilient Machine Learning Deployment Pipelines: Challenges, Frameworks, and Best Practices. <https://www.researchgate.net/publication/390694834>
- Özsu, M. Tamer, & Valduriez, Patrick. 2020
- Ozsu., M.T Dan Valduriez. P. 2011. “Principles of Distributed Database System”. Springer New York Dordrecht Heidelberg London. ISBN 978-1-4419-8833-1. DOI 10.1007/978-1-4419-8834-8.
- Pan, X., Luo, Z., & Zhou, L. (2023). Navigating the Landscape of Distributed File Systems: Architectures, Implementations, and Considerations. *Innovations in Applied Engineering and Technology*, 1–12.
<https://doi.org/10.62836/iaet.v2i1.157>
- Pavlo, A., et al. 2021. “Distributed Database Systems: Current Trends and Challenges. *Communications of the ACM*”, 64(12), 72–81.
- Permana, A. A., S, W., Santoso, L. W., Wibowo, G. W. N.,

- Wardhani, A. K., Rahmadden, Wahidin, A. J., Yulastuti, G. E., Elisawati, Wijayanti, R. R., & Abdurrasyid. (2023). Machine Learning. In P. G. E. Teknologi (Ed.), Machine Learning (Vol. 45, Issue 13). PT Global Eksekutif Teknologi.
- Prabowo, A., dkk: Optimalisasi Manajemen Perpustakaan: Sistem Basis Data Terdistribusi Menggunakan Metode Rapid Application Development (RAD), Volume 4, No. 2, STORAGE – Jurnal Ilmiah Teknik dan Ilmu Komputer, Mei, 2023, hal 137-147.
- Prananingrum., L., dkk, 2024. “Sistem Basis Data”, CV. Gita Lentera: Padang-Sumatera Barat.
- Prasetyo, Eko. (2014). Mengolah Data Menjadi Informasi menggunakan Matlab. Penerbit ANDI Yogyakarta.
- Principles of Distributed Database Systems (4th Edition). Springer.
- Pujianto, A., Mulyati, A., & Novaria, R. (2018). Pemanfaatan BigData dan Perlindungan Privasi Konsumen Di Era Ekonomi Digital. Majalah Ilmiah BIJAK, 15(2), 127–137.
- Qu, Z., Li, X., Xu, J., Tang, B., Lu, Z., & Liu, Y. (2023). On the convergence of multi-server federated learning with overlapping area. Ieee Transactions on Mobile Computing, 22(11), 6647-6662.
- Raghupathi, W., & Raghupathi, V. (2014). Big Data analytics in healthcare: promise and potential. Health Information Science and Systems, 2(3), 1-10.
- Raharjo, B. (2021). Pembelajaran Mesin (Machine Learning). In Angewandte Chemie International Edition, 6(11), 951–952. Yayasan Prima Agus Teknik.
- Raja, S. P. (2021). Green Computing and Carbon Footprint Management in the IT Sectors. IEEE Transactions on

- Computational Social Systems, 8(5), 1172–1177.
<https://doi.org/10.1109/TCSS.2021.3076461>
- Rajpurkar, P., et al. (2022). Benchmarking Large Language Models for Medical Text Classification. *Nature Machine Intelligence*.
- Ramakrishnan, M. R., Sujithra, R., & Naranjalin, A. J. (2024). Privacy Challenges and Solutions in Big Data Analytics: A Comprehensive Review. *International Journal for Research in Applied Science and Engineering Technology*, 12(5), 2435–2440.
- Redmond, R. (n.d.). NoSQL database. Free-Ebooks.net.
- Rizky, F., & Arsyad, M., 2019. Keamanan Data dan Replikasi di Sistem Terdistribusi, ITS Press: Surabaya.
- Rodrigues, J. F., Florea, L., de Oliveira, M. C. F., Diamond, D., & Oliveira, O. N. (2021). Big Data and Machine Learning for materials science. In *Discover Materials* (Vol. 1, Issue 1). Discover. <https://doi.org/10.1007/s43939-021-00012-0>
- Rozzi Kesuma Dinata, & Novia, H. (2020). Machine Learning. In *Machine Learning*. UNIMAL PRESS.
- S.C. Huang, S. McIntosh, S. Sobolevsky, & P. C. K. Hung. (2017). Big Data Analytics and Business Intelligence in Industry. *Inf Syst Front*, 1229–1232.
- Sadalage, P. J., & Fowler, M. (n.d.). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence.
- Sainstekno. (2024). Mengenal NoSQL dan Penerapannya.
- Sakr, S., Bonifati, A., Voigt, H., Iosup, A., Ammar, K., Angles, R., Aref, W., Arenas, M., Besta, M., Boncz, P. A., Daudjee, K., Valle, E. Della, Dumbrava, S., Hartig, O., Haslhofer, B., Hegeman, T., Hidders, J., Hose, K., Iamnitchi, A., ... Yoneki, E. (2021). The future is big graphs. *Communications of the ACM*, 64(9), 62–71.

<https://doi.org/10.1145/3434642>

- Salma, T. D., Saptawati, G. A. P., & Rusmawati, Y. (2021, September). Text Classification Using XLNet with Infomap Automatic Labeling Process. In 2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA) (pp. 1-6). IEEE.
- Salsabila, F., Bukhari, F., & Nurdianti, S. (2024). Proving the correctness of the extended serial graph-validation queue scheme in the client-server system. *Barekeng Jurnal Ilmu Matematika Dan Terapan*, 18(2), 1359-1368.
- Salwa, N. D. K. (2024). Simaka Penyebab Umum Kerusakan Data Digital dan Cara Mencegahnya. *Cloud Computing Indonesia*.
<https://www.cloudcomputing.id/pengetahuan-dasar/penyebab-kerusakan-data#:~:text=Penyebab utama kerusakan data meliputi,masing dapat mempengaruhi integritas data.>
- Santosa, Budi. (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. *Graha Ilmu*.
- Santoso, J. T. (2020). Analisis Big Data. In Penerbit Yayasan Prima Agus Teknik.
<https://penerbit.stekom.ac.id/index.php/yayasanpat/article/view/155>
- Sarker, I. H. (2021). Machine Learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Sasaki, Y. (2022). A Survey on IoT Big Data Analytic Systems: Current and Future. *IEEE Internet of Things Journal*, 9(2), 1024–1036.

<https://doi.org/10.1109/JIOT.2021.3131724>

- Satmoko, N. D. (2022). *Produksi dan Distribusi*. Dalam G. C. A. Indarto (Ed.), *Pengantar Bisnis: Manajemen, Pembiayaan, Pemasaran, dan Operasional*. Yogyakarta: Diandra Kreatif.
- Satmoko. B., 2022. *Ekonomi dan Kegiatan Sosial Masyarakat*, Pustaka Rakyat: Yogyakarta.
- Services, A. W. (2025a). Apa Itu Komputasi Terdistribusi? <https://aws.amazon.com/id/what-is/distributed-computing/>
- Services, A. W. (2025b). Apa itu latensi jaringan? <https://aws.amazon.com/id/what-is/latency/>
- Sesetuperekiye E. 2020. “Framework for Client-Server Distributed Database System for Federal Road Safety Commission Nigeria”. *Advances in Social Sciences Research Journal* – Vol.7, No.5. DOI:10.14738/assrj.75.8237
- Shaikh, T. A., & Ali, R. (2016). Quantum Computing in Big Data Analytics: A Survey. 2016 IEEE International Conference on Computer and Information Technology (CIT), 112–115. <https://doi.org/10.1109/CIT.2016.79>
- Shostack, A. (2014). *Threat Modeling: Designing for Security*.
- Silberschatz. A., Galvin. P. B., & Gagne. G., 2018. *Operating System Concepts* (Edisi ke-10), Wiley: Hoboken, New Jersey.
- Singh, T., Solanki, A., Sharma, S. K., Nayyar, A., & Paul, A. (2022). A Decade Review on Smart Cities: Paradigms, Challenges and Opportunities. In *IEEE Access* (Vol. 10, pp. 68319–68364). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2022.3184710>

- Smolić, E., Boras, B., Horvat, M., & Jaguš, T. (2024). Smartphone-enabled interaction on large displays—a web-technology-based approach. *Electronics*, 13(5), 929.
- Stonebraker, M., & Hellerstein, J. M., 2005. “What Goes Around Comes Around”, dalam *Readings in Database Systems* (Edisi ke-4), MIT Press: Cambridge, Massachusetts.
- Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(04), 687–719. <https://doi.org/10.1142/S0218001409007326>
- Supermicro. (2025). What is Distributed Computing? <https://www.supermicro.com/en/glossary/distributed-computing#:~:text=Key Characteristics of Distributed Computing,one or more nodes fail>.
- Talebkhah, M., Sali, A., Gordan, M., Hashim, S. J., & Rokhani, F. Z. (2023). Comprehensive Review on Development of Smart Cities Using Industry 4.0 Technologies. In *IEEE Access* (Vol. 11, pp. 91981–92030). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2023.3302262>
- Talebkhah, M., Sali, A., Marjani, M., Gordan, M., Hashim, S. J., & Rokhani, F. Z. (2021). IoT and Big Data Applications in Smart Cities: Recent Advances, Challenges, and Critical Issues. *IEEE Access*, 9, 55465–55484. <https://doi.org/10.1109/ACCESS.2021.3070905>
- Tanenbaum, A. S., & Van Steen, M. (2006). *Distributed Systems: Principles and Paradigms* (2nd ed.). Pearson Education.
- Tanenbaum, A. S., & Van Steen, M. (2007). *Distributed System Principles and Paradigms*, second edition. In Pearson Education, Inc. (Vol. 18, Issue 1). [https://doi.org/10.1016/0140-3664\(95\)90078-0](https://doi.org/10.1016/0140-3664(95)90078-0)

- Tanenbaum, A. S., & Steen, M. V., 2007. *Distributed Systems: Principles and Paradigms* (Edisi ke-2), Prentice Hall: Upper Saddle River, New Jersey.
- Tham, J. C. K., & Verhulsdonck, G. (2023). Smart Education in Smart Cities: Layered Implications for Networked and Ubiquitous Learning. *IEEE Transactions on Technology and Society*, 4(1), 87–95. <https://doi.org/10.1109/tts.2023.3239586>
- Tsai, C. P., Chang, C. W., Hsiao, H. C., & Shen, H. (2022). The Time Machine in Columnar NoSQL Databases: The Case of Apache HBase. *Future Internet*, 14(3). <https://doi.org/10.3390/fi14030092>
- Ullman, J. D., & Garcia-Molina, H. (2002). *Database Systems: The Complete Book*. Prentice Hall.
- Usino, W., Kusumawardhani, D. A., Ramadhan, T., Pratiangga, A., & O. Qurotulain. (2024). Big Data Analytics: Transforming Business Intelligence and Decision Making. *Journal of Computer Science and Technology Application (CORISINTA)*, 154-163.
- Usman, A., dkk, 2025. “Konsep Dasar Basis Data”, PT. Serasi Media Teknologi: Payakumbuh-Sumatera Barat.
- Vanny Andini, Lipur Sugiyanta, & Bachren Zaini. (2020). Analisis Kinerja Parameter Throughput Dan Delay Akses Inetrnet Di Smk Karyaguna Jakarta Selatan. *PINTER : Jurnal Pendidikan Teknik Informatika Dan Komputer*, 4(2), 41–44. <https://doi.org/10.21009/pinter.4.2.8>
- Vavilapalli, V. K., Murthy, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S., Saha, B., Curino, C., O'Malley, O., Radia, S., Reed, B., & Baldeschwieler, E. (2013). *Apache Hadoop YARN*. Proceedings of the 4th annual Symposium on Cloud

- Computing, 1–16.
<https://doi.org/10.1145/2523616.2523633>
- W. Usino, D. A. R. Kusumawardhani, T. Ramadhan, A. Pratiangga, & O. Qurotulain. (2024). Big Data Analytics: Transforming Business Intelligence and Decision Making. *Journal of Computer Science and Technology Application (CORISINTA)*, 154-163.
- Wali, M., Efitra, Iwan Sudipa, I. G., Heryani, D. A., Hendriyani, D. C., Rahman, R., Santika, P. P., Indarto, S. L., Dr Tanwir, Ibrahim, M. B., Iskandar, D. A., Kertati, D. I., Nainggolan, H., & Sepriano. (2022). Penerapan & Implementasi Big Data di Berbagai Sektor (Pembangunan Berkelanjutan Era Industri 4.0 dan Society 5.0) (Vol. 1, Issue 1). SONPEDIA.
- Wang, X., Li, Y., & Zhang, T. (2022). Learning Association Rules with Neural Networks. *Neural Computing and Applications*, 34(12), 9575–9591.
- Warner, S. L. (1965). Randomized Response: a Survey Technique for Eliminating EVasive Answer Bias. *Journal of the American Statistical Association*, 60(309), 63–69.
- White, T. (2015). *Hadoop: The Definitive Guide* (4th ed.). O'Reilly Media.
- Widodo. A., 2020. *Manajemen Data untuk Pemerintahan Digital*, Kominfo Press: Jakarta.
- Wiryanto, M. (2023). *Big Data dan Perilaku Konsumen*. Academia.edu.
- Wu, J., Xiong, J., Dai, H., Wang, Y., & Xu, C. (2022). MIX-RS: A Multi-Indexing System Based on HDFS for Remote Sensing Data Storage. *Tsinghua Science and Technology*, 27(6), 881–893.
<https://doi.org/10.26599/TST.2021.9010082>

- Wu, Y., et al. (2021). Semi-supervised anomaly detection using hybrid autoencoder and contrastive learning. *IEEE Access*, 9, 123456–123467.
- Xie, J., et al. (2016). Unsupervised deep embedding for clustering analysis. *ICML*.
- Xin, R., Rosen, J., Zaharia, M., Franklin, M., Shenker, S., & Stoica, I. (2013). Shark: SQL and rich analytics at scale. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data* (pp. 13–24). ACM. <https://dl.acm.org/doi/10.1145/2463676.2465288>
- Zaharia, M. (2016, July 28). Structured streaming in Apache Spark: A high-level API for continuous data processing. *Databricks Blog*. <https://databricks.com/blog>
- Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauley, M., ... & Stoica, I. (2012). Resilient distributed *datasets*: A fault-tolerant abstraction for in-memory cluster computing. In *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 2012)* (pp. 15–28). USENIX Association. <https://www.usenix.org/system/files/conference/nsdi12/nsdi12-final138.pdf>
- Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauley, M., ... & Stoica, I. (2016). Apache Spark: A unified engine for Big Data processing. *Communications of the ACM*, 59(11), 56–65. <https://doi.org/10.1145/2934664>
- Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster computing with working sets. In *2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud'10)*. USENIX Association. <https://www.usenix.org/legacy/event/hotcloud10/tech/f>

ull_papers/Zaharia.pdf

- Zenati, H., et al. (2020). Efficient GAN-Based Anomaly Detection. arXiv:1802.06222.
- Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems, 216. <https://doi.org/10.1016/j.knosys.2021.106775>
- Zhang, D. (2018). Big Data Security and Privacy Protection. Atlantis Press, 77, 275–278.
- Zhou, J., et al. (2021). Parallel FP-Growth algorithm for large-scale association rule mining. Information Sciences, 563, 254–268.
- Zhou, Z.-H. (2021). Machine Learning (2nd ed.). Springer.
- Zikopoulos, P. C., Eaton, C., DeRoos, D., Deutsch, T., & Lapis, G. (2012). Understanding Big Data (Analytics for Enterprise Class *Hadoop* and Streaming Data) (Vol. 4, Issue 1).

Tentang Penulis



Dr. Adryan Rachman, S.Ip., M.M., seorang dosen berdedikasi di bidang *Electronics Word Of Mouth (eWOM)*, telah mengabdikan selama sembilan tahun di Fakultas Manajemen dan Humaniora, Universitas Pradita Tangerang. Dengan pengalaman luas dan keahlian mendalam dalam Ilmu Manajemen, beliau dipilih untuk berkonsentrasi pada penulisan ilmiah yang memperkaya literatur akademik. Karya-karya yang telah terbit meliputi buku-buku manajemen ber-ISBN dalam bidang: *Digital marketing*, *Personal Branding*, *Artificial Intelligence*, *Metaverse*, dan *SDGs*. Beliau juga aktif mempublikasikan di jurnal ilmiah nasional bereputasi serta jurnal internasional terindeks *Scopus* Q2 hingga Q4, dan Jurnal SINTA 2, 3, dan 4. Dr. Adryan merupakan penerima dana hibah penelitian eksternal *PDP LLDIKTI IV* pada tahun 2019 hingga 2021, serta aktif dalam Pengabdian kepada Masyarakat melalui berbagai kegiatan. Beliau juga aktif dalam berbagai forum dan asosiasi, seperti Forum Manajemen Indonesia (FMI) dan Ikatan Sarjana Ilmu Ekonomi (ISEI). Dengan pengalaman manajerial struktural melalui penugasan di berbagai posisi penting akademik dan administratif, saat ini beliau juga menjabat sebagai Direktur Utama PT. Multi Edu Creation, dan terus memberikan kontribusi penting bagi pengembangan strategi pemasaran serta kemajuan ilmu pengetahuan di Indonesia. Kepakaran, dedikasi, dan semangat yang dihadirkan dalam setiap peran menjadikan beliau sumber inspirasi dan panutan dalam dunia akademik dan profesional.



Dr. Indah Dwi Mumpuni, S.Kom, MM, adalah dosen tetap Program Studi Sistem Informasi di STMIK PPKIA Pradnya Paramita, bidang keilmuan sistem informasi dan salah satu matakuliah yang penulisampu yaitu Sistem Basis Data. Buku ini merupakan karya ketiga penulis, dan insyaa Allah akan diikuti oleh buku-buku berikutnya yang masih berkaitan erat dengan mata kuliah dan bidang yang penulisampu. Setiap subjek yang dituangkan dalam buku kolaborasi ini disusun bukan semata-mata sebagai bahan ajar, tetapi sebagai wujud nyata dari komitmen penulis dalam berbagi ilmu dan pengalaman kepada para pembaca.



Weda Adistianaya Dewa, lahir di Madiun pada tanggal 25 November 1984. Seorang Dosen di STMIK PPKIA Pradnya Paramita Malang. Buku ini adalah salah satu karya dan nantinya secara konsisten akan disusul dengan buku-buku berikutnya. Pokok bahasan buku yang ditulis semata-mata untuk berbagi ilmu pengetahuan.



Dinny Wahyu Widarti, lahir di kota Bandung tanggal 06 Juli 1979. Setelah lulus SMA, penulis menempuh pendidikan pada jenjang D3 program studi Manajemen Informatika di STMIK PPKIA Pradnya Paramita Malang lulus pada tahun 2000, lalu melanjutkan S1 program studi Sistem Informasi di STMIK PPKIA Pradnya Paramita Malang, pada tahun 2013 penulis melanjutkan studi S2 jurusan Perangkat Lunak Sistem Informasi di Universitas Gunadarma Jakarta lulus pada tahun 2015. Saat ini penulis menjadi Dosen tetap di kampus STMIK PPKIA Pradnya

Paramita pada program studi Teknologi Informasi. Buku ini adalah karya buku kedua saya, yang digunakan sebagai bahan ajar serta semata-mata untuk berbagi pengetahuan.



Farida Islamiah, lahir di Tanah Grogot pada tanggal 21 Mei 1990. Jenjang Pendidikan S1 Statistika FMIPA Universitas Mulawarman di Samarinda lalu melanjutkan pendidikan S2 Statistika Institut Teknologi Sepuluh Nopember di Surabaya. Bekerja sebagai Dosen tetap program studi Manajemen di STIE Widya Praja Tanah Grogot tahun 2015-2023. Saat ini sebagai Dosen pada Program Studi Bisnis Digital Universitas Negeri Makassar. Selain itu, Penulis juga sebagai *reviewer* pada beberapa jurnal.



Esa Kurniawan, S.Kom., M.M., M.Kom, lahir di Sungai Penuh pada tahun 1994 dan saat ini mengabdikan sebagai Dosen pada Program Studi Sistem Informasi di Universitas LIA. Ia menyelesaikan pendidikan Magister Manajemen dengan konsentrasi Sumber Daya Manusia pada tahun 2021, kemudian meraih gelar Magister Ilmu Komputer dengan spesialisasi Teknik Informatika pada tahun 2023 dari UPI YPTK Padang. Penulis juga memiliki sertifikasi *Trainer* bidang pendidikan dengan gelar *Certified Education Trainer* (C.ET). Sejak tahun 2018, ia memulai karier profesional di Bank Rakyat Indonesia dan kemudian menjabat sebagai staf *Human Resource Development* (HRD) di Yayasan Abdurrahingga tahun 2023. Pengalamannya ini memperkaya kompetensinya dalam mengajar dan melakukan riset, terutama di bidang sistem informasi dan manajemen. Penulis aktif berpartisipasi dalam berbagai seminar tingkat nasional maupun

internasional serta telah mempublikasikan sejumlah artikel di jurnal-jurnal terakreditasi. Beberapa karya ilmiahnya antara lain: 1) Penerapan Metode *Preference Selection Index* (PSI) dalam Rekrutmen Karyawan di Yayasan XYZ, 2) Peningkatan Pengalaman Pengguna pada Aplikasi *Mobile Whoosh*, 3) Analisis Sistem Peramalan Produksi dan Permintaan Telur, 4) Penerapan Metodologi *Scrum* untuk Pengembangan Sistem Informasi di Apotek Qilah Sehat. Di samping artikel ilmiah, penulis juga turut berkontribusi dalam penulisan beberapa *book chapter*, seperti Pengantar Bisnis Digital (2024), Pendidikan di Era Digital (2025), dan Pengantar Teknologi Informasi (2025). Profil ORCID penulis dapat diakses melalui tautan berikut: <https://orcid.org/0009-0003-2992-6587>



A.Kachsyfur Djasim Ilyas Paenrongi, buku ini adalah salah satu karya dan insyaa allah secara konsisten akan disusul dengan buku-buku berikutnya. Pokok bahasan buku yang ditulis semata-mata untuk berbagi ilmu pengetahuan.



Mustika Kurnia Mayangsari, S.Kom., M.Tr.Kom adalah dosen dan peneliti di bidang teknik informatika. Ia menyelesaikan pendidikan S1 Teknik Informatika ITS dan S2 Teknik Informatika dan Komputer PENS. Fokus kajiannya meliputi rekayasa perangkat lunak, data *mining*, dan Natural Language *Processing*.

Buku ini ditulis sebagai kontribusi dalam pengembangan literatur ilmiah yang aplikatif dan kontekstual di bidang ilmu

pengetahuan dan teknologi.



Timotius Victory, S.Kom., M.Kom., Dosen Prodi Teknologi Rekayasa Perangkat Lunak di AstraTech yang mempunyai minat penelitian di bidang *Machine Learning*, *Deep Learning* dan *Big Data*. Penulis juga aktif dalam kegiatan pengabdian kepada Masyarakat, penulisan buku, maupun seminar-seminar di bidang ilmu

komputer.



Dwi Diana Wazaumi, S.Tr.Kom., M.T., lahir di Kota Bengkulu pada 4 Juni 1997. Di usia 28 tahun, penulis masih terus belajar dan tumbuh dalam perannya sebagai dosen. Baginya, menjadi pengajar bukan berarti sudah selesai belajar, justru setiap interaksi di kelas, diskusi riset, hingga tantangan baru di dunia teknologi

adalah kesempatan untuk terus berkembang bersama mahasiswa dan rekan sejawat. Penulis menyelesaikan pendidikan magisternya di Institut Teknologi Bandung (ITB), dengan minat khusus di bidang *Smart City*, *Data Modeling*, *Big Data*, dan *Internet of Things* (IoT). Saat ini, penulis juga tengah mendalami riset tentang *sustainability*, mencoba mencari irisan antara teknologi dan keberlanjutan dalam menjawab tantangan masa depan. Penulis berharap buku ini bisa menjadi pintu masuk bagi pembaca yang ingin mengenal dunia *NoSQL*.



M. Rhifky Wayahdi, lahir di Medan, 05 Februari 1993, merupakan anak pertama dari tiga bersaudara. Penulis merupakan alumni Program Sarjana (S-1) di Universitas Potensi Utama pada Jurusan Sistem Informasi dan lulus tahun 2015. Penulis melanjutkan studi Program Magister (S-2) Teknik Informatika di Universitas Sumatera Utara dan lulus tahun 2019. Kemudian saat ini Penulis sedang melanjutkan Pendidikan Doktoral (S-3) Ilmu Komputer di Universitas Sumatera Utara mulai 2024 sampai sekarang (on-going). Berkarir sebagai dosen dimulai dari tahun 2020 di Universitas Battuta. Mulai tahun 2023-sekarang diberi kepercayaan menjadi Dekan Fakultas Teknologi di Universitas Battuta.



Siti Aisyah, S.Tr., M.Sc., menempuh pendidikan D4 Teknik Informatika di Politeknik Negeri Jakarta (PNJ) pada tahun 2015. Pada tahun 2020 melanjutkan pendidikan *Master of Science Big Data* di *University of Stirling, Scotland, UK*. Saat ini penulis menjadi Dosen tetap di Universitas Insan Cita Indonesia (UICI). Buku ini adalah karya buku kedua saya, yang digunakan sebagai bahan ajar serta semata-mata untuk berbagi pengetahuan.



Deyana Kusuma Wardani, S.Tr.Kom., M.Tr.Kom. adalah dosen dan peneliti di bidang Teknik Informatika. Fokus penelitiannya meliputi *Data Mining, Natural Language Processing*, Rekayasa Perangkat Lunak, serta *Big Data*. *Book chapter* berjudul “*Big Data* dan

Manajemen Basis Data Terdistribusi” ini merupakan karya kedua yang ia tulis dalam bidang tersebut. Sebelumnya, ia juga telah menulis *book chapter* dengan fokus serupa, yaitu mengenai *Big Data*. Melalui karya keduanya ini, ia berharap dapat memberikan kontribusi dan memperluas wawasan di bidang Teknik Informatika khususnya di *Big Data*.



Thesya Atarezcha Pangruruk, S.Stat., M.Si adalah dosen di Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman. Ia menyelesaikan pendidikan magister di bidang Statistika dengan fokus pada pengembangan regresi *time series*. Ketertarikannya terhadap

data dan pemodelan kuantitatif membawanya memahami berbagai pendekatan analitik dalam skala besar. Seiring perkembangan teknologi informasi dan ledakan data digital, ia mulai mendalami isu-isu seputar pengelolaan data besar (*Big Data*), khususnya dari sisi keamanan dan privasi. Topik tersebut menjadi perhatian baru yang dianggap penting untuk menjembatani antara pemanfaatan data dan perlindungan individu maupun institusi. Dalam *book chapter* ini, ia mengulas konsep dasar dan tantangan praktis dalam menjaga keamanan dan privasi pada sistem *Big Data*, dengan menggabungkan pendekatan statistika, prinsip teknologi informasi, dan perspektif kebijakan.



Kurnia Yahya, SE., M.Cs., Ph. D, buku ini adalah salah satu karya dan Inshaa Allah secara konsisten akan disusul dengan buku-buku berikutnya. Pokok bahasan buku yang ditulis semata-mata untuk berbagi ilmu pengetahuan.



Nariza Wanti Wulan Sari, S.Si, M.Si, buku ini ditulis sebagai media berbagi dan menuangkan pengetahuan terutama berkaitan dengan statistika komputasi. Ia juga Dosen Program Studi Statistika FMIPA Universitas Mulawarman, Laboratorium Statistika Komputasi.



Erlyne Nadhilah Widyaningrum, S.Mat., M.Stat, Lahir di Balikpapan 21 Agustus 1999. Erlyne menyelesaikan Pendidikan S1 pada Program Studi Matematika, Institut Teknologi Sepuluh Nopember pada tahun 2021, lalu melanjutkan program magister pada Program Studi Statistika Institut Teknologi Sepuluh Nopember dan menyelesaikannya pada Tahun 2023, kemudian pada Tahun 2024 ia menjadi Dosen Tetap Program Studi Statistika Universitas Mulawarman, Samarinda.



Dr. Phie Chyan, ST, M.Cs memperoleh gelar Sarjana Teknik Elektro dari Universitas Atma Jaya Makassar, Indonesia, pada tahun 2003, gelar Magister Ilmu Komputer dari Universitas Gadjah Mada, Yogyakarta, Indonesia, pada tahun 2011, dan Gelar Doktor Teknik Elektro, Universitas Hasanuddin pada tahun 2024. Saat

ini aktif melakukan penelitian di bidang visi komputer, *Speech Processing*, dan *Natural Language Processing*. Saat ini juga aktif sebagai staf pengajar di Departemen Informatika, Fakultas Teknologi Informasi, Universitas Atma Jaya Makassar. Hingga saat ini ia telah menuliskan beberapa buku teks dan referensi antara lain: Pengantar *Machine Learning*, 2023; Pengantar Data *Mining*, 2023; dan Sistem Operasi Konsep Dasar dan Penerapan Modern, 2024.



Triana Dewi Salma, M.T., Penulis lahir di Tuban pada September 1994. Saat ini, penulis menjabat sebagai Dekan Fakultas Sains dan Bisnis serta dosen tetap pada Program Studi Informatika di Universitas LIA. Penulis memperoleh gelar Sarjana pada bidang Informatika dari Universitas Muhammadiyah

Surakarta dan melanjutkan pendidikan Magister Informatika di Institut Teknologi Bandung. Dorongan utama dalam berkarya adalah *keyakinan* bahwa ilmu pengetahuan harus terus berkembang dan memberi manfaat jangka panjang. Buku ini adalah salah satu karya dan Inshaa Allah secara konsisten akan disusul dengan buku-buku berikutnya. Pokok bahasan buku yang ditulis semata-mata untuk berbagi ilmu pengetahuan.



Fari Katul Fikriah, merupakan seorang akademisi dan peneliti yang memiliki latar belakang pendidikan dari Universitas Dian Nuswantoro, Jawa Tengah. Sejak menyelesaikan pendidikannya, ia menekuni dunia akademik dengan menjadi dosen sejak tahun 2020. Selama berkarier di bidang pendidikan tinggi, ia aktif dalam pengajaran, penelitian, serta pengabdian kepada masyarakat, khususnya dalam bidang teknologi dan ilmu data. Sebagai seorang dosen dan peneliti, Fari Katul Fikriah memiliki fokus utama pada bidang *Data Mining*, *Artificial Intelligence* (AI), dan *Image Processing*. Dalam kajian akademiknya, ia banyak mengeksplorasi berbagai metode dan teknik dalam analisis data serta penerapan kecerdasan buatan untuk membantu pengambilan keputusan yang lebih akurat dan efisien. Keahliannya dalam pemrosesan citra digital juga membawanya pada berbagai penelitian yang bertujuan untuk mengembangkan teknologi berbasis gambar dan visualisasi data, baik dalam dunia medis, industri, maupun sektor lainnya.



Rosa Eliviani, S.Kom., M.T., adalah seorang dosen dan peneliti di bidang Informatika dengan fokus pada *information system*, *Machine Learning*, dan *Internet of Things*. Memiliki pengalaman 2 tahun dalam mengajar dan melakukan penelitian terkait transformasi digital dan pemanfaatan data untuk solusi perkotaan. Penulis aktif dalam berbagai kegiatan pengabdian masyarakat, seminar nasional, dan telah menulis beberapa artikel ilmiah terkait pengembangan sistem cerdas berbasis data.



Amelia Devi Putri Ariyanto, merupakan seorang dosen dan peneliti di bidang komputer dan teknologi informasi. Ia menyelesaikan pendidikan magister pada program studi Informatika di Institut Teknologi Sepuluh Nopember (ITS) Surabaya, Indonesia, pada tahun 2022. Saat ini, Amelia sedang melanjutkan studi doktoralnya di institusi yang sama. Di samping kegiatan akademiknya sebagai mahasiswa doktoral, ia juga menjalankan peran sebagai pengajar di Universitas Widya Husada Semarang. Aktif dalam dunia riset, Amelia telah mempublikasikan berbagai artikel ilmiah di jurnal dan konferensi internasional yang berfokus pada ilmu komputer. Minat penelitiannya mencakup bidang kecerdasan buatan (*artificial intelligence*), pemrosesan bahasa alami (*natural language processing*), serta penambangan data (*data mining*). Dengan kombinasi antara latar belakang akademis yang kuat dan pengalaman riset yang mendalam, ia terus berkontribusi dalam pengembangan teknologi dan keilmuan di ranah komputasi.



Luthfi Atikah, ia seorang dosen dan penulis, lahir di Madiun 19 Januari 1995. Saat ini aktif mengajar di Politeknik Astra mengampu mata kuliah Basis Data, Struktur Data, dan Sistem Informasi serta terlibat dalam berbagai penelitian terkait *Medical Imaging*, *Information Retrival*, dan *Machine Learning*.



Buku Big Data dan Manajemen Basis Data Terdistribusi membahas secara menyeluruh bagaimana data dalam jumlah besar (big data) dikelola secara efisien melalui sistem basis data yang tersebar di berbagai lokasi (terdistribusi). Buku ini diawali dengan penjelasan tentang karakteristik unik big data, tantangan pengelolaannya, dan peluang yang dapat dimanfaatkan di berbagai sektor. Pembaca akan diperkenalkan dengan konsep dasar, teknologi pendukung, serta manfaat strategis dari big data dalam kehidupan modern. Selanjutnya, buku ini menggali aspek teknis dari manajemen data, termasuk arsitektur basis data terdistribusi, teknik replikasi, konsistensi data, serta penggunaan NoSQL, Hadoop, dan Apache Spark dalam pengolahan big data. Disertakan pula pembahasan mendalam mengenai integrasi big data dengan cloud computing, analisis real-time, visualisasi data, dan penerapan machine learning. Buku ini juga menampilkan studi kasus dan implementasi nyata dalam bidang bisnis, kesehatan, transportasi, pendidikan, dan smart city. Sebagai penutup, buku ini menawarkan wawasan mengenai tren masa depan dalam dunia big data dan sistem terdistribusi, seperti edge computing, federated learning, green computing, dan quantum computing. Dengan pendekatan teoritis dan praktis yang seimbang, buku ini menjadi referensi penting bagi mahasiswa, peneliti, profesional IT, dan pengambil kebijakan yang ingin memahami serta menerapkan manajemen data secara optimal di era digital.

DITERBITKAN OLEH
PT. MIFANDI MANDIRI DIGITAL



Jln Payanibung Ujung D
Dalu Sepuluh-B, Tanjung Morawa
Kab. Deli Serdang Sumatera Utara

