



Perbandingan Teknik Normalisasi dan Denormalisasi dalam Pengelolaan Data Skala Besar

Zainudin Sidik ¹, Abdul Rahman Ismail ^{2*}

Program Studi Teknologi Rekayasa Perangkat Lunak, Sekolah Vokasi,

Universitas Negeri Gorontalo, Gorontalo, Indonesia

Email: zainudsidik@ung.ac.id¹, abdulrahmanismail@ung.ac.id²

Abstrak

Penelitian ini bertujuan menganalisis perbandingan teknik normalisasi dan denormalisasi dalam pengelolaan data skala besar untuk mengidentifikasi karakteristik performa dan *trade-offs* yang terlibat. Metode penelitian menggunakan pendekatan eksperimental komparatif dengan mengimplementasikan dua sistem basis data yang identik namun berbeda dalam penerapan normalisasi dan denormalisasi, menggunakan dataset sintetis berukuran 500GB hingga 2TB untuk simulasi big data. Pengujian dilakukan dengan mengukur berbagai metrik performa meliputi waktu respons *query*, *throughput system*, utilisasi *resource*, dan *scalability* dalam berbagai skenario *workload*. Hasil penelitian menunjukkan denormalisasi memberikan keunggulan signifikan dalam operasi baca dengan peningkatan *throughput* 84,3% dan pengurangan waktu respons 73,8%, namun memerlukan konsumsi memori lebih besar 46% dibandingkan normalisasi. Evaluasi *scalability* menunjukkan denormalisasi mempertahankan performa konsisten pada beban tinggi, sementara normalisasi mengalami degradasi hingga 84% pada kondisi peak load. Normalisasi tetap unggul dalam integritas data dan efisiensi operasi tulis. Penelitian ini merekomendasikan penerapan strategi *hybrid* yang menggabungkan normalisasi untuk data transaksional dan denormalisasi selektif untuk data analitik, disesuaikan dengan karakteristik *workload* spesifik sistem untuk mencapai optimasi performa yang optimal.

Kata Kunci: Normalisasi *Database*; Denormalisasi *Database*; Big Data Management; Database Performance; Scalability

ABSTRACT

This research aims to analyze the comparison between normalization and denormalization techniques in large-scale data management to identify performance characteristics and trade-offs involved. The research method employs a comparative experimental approach by implementing two identical database systems that differ in normalization and denormalization application, using synthetic datasets sized 500GB to 2TB for big data simulation. Testing was conducted by measuring various performance metrics including query response time, system throughput, resource utilization, and scalability across different workload scenarios. Results demonstrate that denormalization provides significant advantages in read operations with 84.3% throughput improvement and 73.8% response time reduction, but requires 46% higher memory consumption compared to normalization. Scalability evaluation shows denormalization maintains consistent performance under high loads, while normalization experiences degradation up to 84% under peak load conditions. Normalization remains superior in data integrity and write operation efficiency. This research recommends implementing hybrid strategies combining normalization for transactional data and selective denormalization for analytical data, tailored to specific system workload characteristics to achieve optimal performance optimization.

Keywords: Database Normalization; Database Denormalization; Big Data Management; Database Performance; Scalability

1. PENDAHULUAN

Era digital saat ini telah menghasilkan ledakan data yang luar biasa, dengan volume data global mencapai 120 *zettabytes* pada tahun 2023 dan diproyeksikan akan mencapai 181 *zettabytes* pada akhir tahun 2025 (*Softteco*, 2025). Pertumbuhan eksponensial data ini menimbulkan tantangan kompleks dalam pengelolaan dan optimasi basis data, khususnya dalam konteks *big data* dan sistem skala besar. Organisasi moderen menghadapi dilema fundamental dalam merancang arsitektur basis data: memilih antara pendekatan normalisasi (*normalization*) yang memprioritaskan integritas data atau denormalisasi (*denormalization*) yang mengutamakan performa *query* (*Couchbase*, 2025).

Normalisasi merupakan teknik fundamental dalam desain basis data yang bertujuan untuk meminimalkan redundansi data dan meningkatkan integritas melalui strukturisasi data ke dalam tabel-tabel yang terorganisir dengan baik (*CelerData*, 2025). Proses ini mengikuti seperangkat aturan formal yang dikenal sebagai *normal forms*, yang dirancang untuk menghilangkan anomali data seperti *insertion*, *update*, dan *deletion anomalies*. Namun, dalam konteks sistem skala besar dengan volume transaksi tinggi, pendekatan normalisasi yang ketat dapat mengakibatkan performa *query* yang lambat karena kompleksitas *join operations* yang diperlukan untuk mengakses data yang tersebar di multiple tabel (*Pure Storage*, 2024). Sebaliknya, denormalisasi merupakan strategi optimasi basis data yang secara sengaja menambahkan redundansi data untuk meningkatkan performa baca dan mengurangi kompleksitas *query* (*Secoda*, 2025). Teknik ini melibatkan penggabungan tabel-tabel yang telah dinormalisasi atau duplikasi data untuk menghindari *costly joins* dalam operasi basis data relasional (*GeeksforGeeks*, 2025). Pendekatan denormalisasi sangat relevan dalam konteks *data warehousing*, *business intelligence*, dan sistem *real-time analytics* di mana kecepatan akses data menjadi prioritas utama dibandingkan dengan integritas data yang ketat. Dalam pengelolaan data skala besar, pemilihan strategi antara normalisasi dan denormalisasi tidak dapat dilakukan secara sembarang. Setiap pendekatan memiliki *trade-offs* yang signifikan dalam hal performa, integritas data, kompleksitas sistem, dan biaya *maintenance* (*Visual Paradigm*, 2023). Normalisasi cenderung menghasilkan operasi tulis yang cepat namun operasi baca yang lambat, sementara denormalisasi menghasilkan operasi baca yang cepat namun operasi tulis yang lambat (*TechTarget*, 2024). Pemahaman mendalam tentang karakteristik masing-masing teknik menjadi krusial untuk membuat keputusan arsitektural yang tepat.

Penelitian terdahulu menunjukkan bahwa efektivitas normalisasi dalam konteks *big data* dan *machine learning* sangat bergantung pada karakteristik dataset dan tujuan analisis

yang ingin dicapai. Studi mengenai efektivitas normalisasi dalam preprocessing *big data* untuk performa *machine learning* menunjukkan hasil yang bervariasi tergantung pada algoritma yang digunakan dan struktur data yang dianalisis (Taipalus, 2024). Hal ini mengindikasikan bahwa tidak ada solusi universal dalam pemilihan teknik normalisasi atau denormalisasi. Kompleksitas pengelolaan data skala besar semakin meningkat dengan munculnya berbagai paradigma basis data seperti *NoSQL*, *NewSQL*, dan sistem terdistribusi. Dalam konteks ini, prinsip-prinsip tradisional normalisasi perlu dievaluasi ulang untuk disesuaikan dengan karakteristik sistem moderen yang memprioritaskan *scalability*, *availability*, dan *partition tolerance* sesuai dengan teorema CAP. Denormalisasi menjadi semakin relevan dalam sistem terdistribusi karena kemampuannya mengurangi komunikasi antar node dan meningkatkan *data locality*.

Aspek performa menjadi pertimbangan utama dalam pemilihan teknik pengelolaan data. Dalam sistem *OLTP* (*Online Transaction Processing*), normalisasi umumnya lebih efektif karena karakteristik operasi yang didominasi oleh *insert*, *update*, dan *delete* yang memerlukan konsistensi data tinggi. Sebaliknya, dalam sistem *OLAP* (*Online Analytical Processing*) dan *data warehousing*, denormalisasi cenderung memberikan performa yang superior karena fokus pada operasi baca yang kompleks dan analisis data historis (Empirical Edge, 2025). Faktor skala sistem juga mempengaruhi efektivitas masing-masing teknik. Pada sistem dengan volume data kecil hingga menengah, *overhead* dari normalisasi mungkin tidak signifikan dan manfaat integritas data dapat lebih terasa. Namun, pada sistem dengan volume data yang sangat besar dan tingkat *concurrency* yang tinggi, *overhead* dari *join operations* dalam normalisasi dapat menjadi *bottleneck* yang signifikan, sehingga denormalisasi menjadi pilihan yang lebih pragmatis. Perkembangan teknologi *in-memory computing*, *columnar storage*, dan *distributed computing frameworks* seperti *Apache Spark* dan *Hadoop* juga mempengaruhi relevansi masing-masing teknik. Teknologi ini memungkinkan optimasi performa yang tidak mungkin dilakukan dengan arsitektur basis data tradisional, sehingga *trade-offs* antara normalisasi dan denormalisasi perlu dievaluasi ulang dalam konteks teknologi moderen.

Berdasarkan latar belakang yang telah diuraikan, penelitian ini merumuskan beberapa permasalahan utama yang perlu dijawab. Pertama, bagaimana perbandingan performa *query* antara teknik normalisasi dan denormalisasi dalam konteks pengelolaan data skala besar, khususnya dalam hal *throughput*, *latency*, dan *resource utilization*. Kedua, faktor-faktor apa saja yang mempengaruhi efektivitas masing-masing teknik dalam berbagai skenario penggunaan, seperti karakteristik *workload*, ukuran dataset, dan pola akses data. Ketiga, bagaimana *trade-offs* antara integritas data, performa sistem, dan kompleksitas *maintenance*

dapat dievaluasi secara objektif untuk memandu pengambilan keputusan arsitektural dalam sistem pengelolaan data skala besar. Penelitian ini bertujuan untuk melakukan analisis komprehensif terhadap teknik normalisasi dan denormalisasi dalam konteks pengelolaan data skala besar. Tujuan utama adalah mengidentifikasi karakteristik performa masing-masing teknik melalui evaluasi empiris yang mencakup berbagai metrik performa seperti waktu eksekusi *query*, *resource consumption*, dan *scalability*. Selain itu, penelitian ini bertujuan untuk mengembangkan *framework* evaluasi yang dapat digunakan untuk menentukan teknik yang paling sesuai berdasarkan karakteristik spesifik dari sistem dan *use case* yang dihadapi. Manfaat dari penelitian ini dapat dilihat dari perspektif teoritis dan praktis. Secara teoritis, penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan *body of knowledge* dalam bidang pengelolaan basis data skala besar, khususnya dalam hal pemahaman mendalam tentang *trade-offs* antara normalisasi dan denormalisasi. Secara praktis, hasil penelitian ini dapat menjadi panduan bagi praktisi *database engineering*, *data architects*, dan *system designers* dalam membuat keputusan arsitektural yang optimal untuk sistem pengelolaan data skala besar, sehingga dapat meningkatkan efisiensi operasional dan mengurangi biaya infrastruktur teknologi informasi.

2. METODE

Penelitian ini mengadopsi pendekatan metode eksperimental komparatif untuk menganalisis performa teknik normalisasi dan denormalisasi dalam pengelolaan data skala besar (Munte et al., 2024). Jenis penelitian ini dipilih karena kemampuannya dalam mengukur dan membandingkan efektivitas dua atau lebih perlakuan berbeda dalam kondisi yang terkontrol, sesuai dengan karakteristik penelitian yang bertujuan mengidentifikasi perbedaan performa antara kedua teknik basis data tersebut. Metodologi eksperimental komparatif memungkinkan peneliti untuk mengisolasi variabel-variabel yang mempengaruhi performa sistem dan mengukur dampak spesifik dari setiap teknik terhadap metrik performa yang telah ditetapkan (*Enago Academy*, 2023). Subjek penelitian dalam studi ini terdiri dari dua sistem basis data yang diimplementasikan dengan struktur yang identik namun menggunakan pendekatan yang berbeda dalam hal normalisasi dan denormalisasi. Implementasi eksperimental dilakukan menggunakan platform PostgreSQL versi 15.2 sebagai sistem basis data relasional utama, dipilih berdasarkan kemampuan enterprise-grade dan dukungan komprehensif terhadap optimasi query kompleks. Konfigurasi sistem meliputi penggunaan server dengan spesifikasi Intel Xeon Gold 6248R (24-core, 3.0GHz), 128GB DDR4 RAM, dan

storage NVMe SSD 2TB untuk memastikan konsistensi lingkungan eksperimental. Kedua implementasi database dikonfigurasi dengan parameter identik meliputi buffer pool size (32GB), checkpoint interval (300 detik), dan work memory allocation (4GB) untuk mengisolasi pengaruh variabel konfigurasi terhadap hasil pengukuran performa, sehingga perbedaan yang terobservasi dapat diatribusikan secara spesifik kepada perbedaan pendekatan normalisasi dan denormalisasi.

Dataset yang digunakan dalam eksperimen ini merupakan data sintetis yang dibangkitkan dengan karakteristik yang mencerminkan skenario *real-world* pengelolaan data skala besar, dengan volume data mencapai beberapa *terabyte* untuk memastikan validitas hasil dalam konteks *big data*. Pemilihan dataset sintetis dilakukan untuk memastikan kontrol penuh terhadap variabel-variabel yang dapat mempengaruhi hasil eksperimen, serta untuk menghindari bias yang mungkin timbul dari penggunaan dataset yang telah ada. Meskipun penggunaan dataset sintetis memberikan kontrol eksperimental yang optimal terhadap variabel-variabel penelitian, penelitian ini mengakui keterbatasan inherent dalam hal representasi kompleksitas data aktual. Untuk memitigasi limitasi ini, dataset sintetis dirancang dengan mengadopsi pola distribusi data, struktur relasional, dan karakteristik workload yang diobservasi dari studi empiris sebelumnya terhadap sistem enterprise-scale database management. Validasi eksternal dilakukan melalui benchmarking dengan dataset standar industri seperti TPC-H dan TPC-C untuk memastikan relevansi hasil terhadap skenario implementasi praktis dalam lingkungan produksi.

Teknik pengumpulan data dalam penelitian ini melibatkan pengukuran sistematis terhadap berbagai metrik performa sistem basis data dalam kondisi *workload* yang beragam. Metrik yang diukur meliputi waktu respons *query*, *throughput* sistem, utilisasi *resource* (CPU, memori, dan I/O), serta *scalability system* dalam menghadapi peningkatan beban kerja. Pengumpulan data dilakukan menggunakan *performance monitoring tools* yang terintegrasi dengan sistem basis data untuk memastikan akurasi dan konsistensi pengukuran. Setiap skenario pengujian dijalankan dalam *multiple iterations* dengan kondisi sistem yang identik untuk mengurangi variabilitas hasil yang disebabkan oleh faktor eksternal (*Full Scale*, 2025). Analisis data dalam penelitian ini menggunakan pendekatan statistik komparatif yang melibatkan *statistical significance testing* untuk menentukan apakah perbedaan performa yang terobservasi merupakan hasil yang signifikan secara statistik atau hanya variasi random. Teknik analisis yang digunakan meliputi *paired t-test* untuk membandingkan means performa antara kedua teknik, serta *analysis of variance* (ANOVA) untuk menganalisis perbedaan performa

dalam berbagai kondisi *workload*. Selain itu, dilakukan juga analisis regresi untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap performa masing-masing teknik (*TestGrid*, 2025). Validitas internal penelitian dijaga melalui implementasi *controlled experimental environment* yang memastikan bahwa semua variabel eksternal yang dapat mempengaruhi hasil dikontrol dengan ketat. Hal ini meliputi penggunaan *hardware* yang identik, konfigurasi sistem operasi yang sama, serta isolasi *network environment* untuk menghindari interferensi eksternal. Validitas eksternal penelitian diperkuat melalui penggunaan berbagai skenario *workload* yang mencerminkan kondisi penggunaan yang beragam dalam *real-world applications*, mulai dari operasi *OLTP* intensif hingga *analytical queries* kompleks yang karakteristik dalam sistem *OLAP*. Alur penelitian yang sistematis dalam studi ini digambarkan dalam Gambar 1, yang menunjukkan tahapan metodologis mulai dari persiapan dataset sintetis, implementasi sistem basis data dengan pendekatan normalisasi dan denormalisasi, hingga evaluasi komprehensif terhadap metrik kinerja sistem.



Gambar 1. Alur Penelitian

Keseluruhan tahapan penelitian mengikuti alur metodologis yang terstruktur sebagaimana digambarkan dalam Gambar 1, yang menunjukkan lima fase sequential dimulai dari persiapan infrastruktur eksperimental, generasi dataset sintetis dengan karakteristik *big data*, implementasi dual-system database dengan pendekatan normalisasi dan denormalisasi, hingga evaluasi kinerja komprehensif. Flowchart tersebut mengilustrasikan interdependensi antar tahapan penelitian yang dirancang untuk memastikan validitas internal melalui kontrol variabel eksperimental yang ketat, mulai dari standardisasi konfigurasi *hardware* hingga isolasi lingkungan pengujian untuk mengeliminasi bias eksternal. Setiap komponen dalam alur penelitian ini berkontribusi terhadap pencapaian objektif studi melalui pendekatan eksperimental komparatif yang memungkinkan pengukuran empiris terhadap metrik kinerja

spesifik dan identifikasi *trade-offs* fundamental antara kedua teknik pengelolaan basis data dalam konteks sistem skala besar.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Performa Query Normalisasi dan Denormalisasi.

Hasil pengujian eksperimental menunjukkan perbedaan signifikan dalam performa *query* antara sistem yang menerapkan normalisasi penuh dan denormalisasi. Berdasarkan analisis yang dilakukan terhadap berbagai skenario *workload*, sistem denormalisasi menunjukkan keunggulan dalam operasi baca dengan waktu respons yang lebih cepat dibandingkan sistem normalisasi. Hal ini sejalan dengan temuan (Ajie, 2025) yang menunjukkan bahwa optimasi struktur basis data dapat meningkatkan *throughput* sistem hingga 40% dan mengurangi *latency* akses hingga 30%. Dalam konteks normalisasi, kompleksitas *join operations* yang diperlukan untuk mengakses data yang tersebar di *multiple* tabel menjadi faktor utama yang mempengaruhi performa, terutama pada *query* yang melibatkan agregasi data dari berbagai entitas.

Pengujian yang dilakukan dengan menggunakan dataset berukuran 500GB hingga 2TB menunjukkan bahwa sistem denormalisasi mempertahankan performa yang konsisten bahkan dengan peningkatan volume data, sementara sistem normalisasi mengalami degradasi performa yang signifikan. Temuan ini diperkuat oleh penelitian (Aulia et al., 2023) yang menunjukkan bahwa pemilihan strategi *query* yang tepat sangat diperlukan untuk efisiensi, terutama ketika menangani data dalam jumlah besar. Dalam konteks *big data management*, (Shethiya, 2025) menekankan pentingnya optimasi performa untuk memastikan aplikasi dapat menangani peningkatan demand pengguna sambil mempertahankan kecepatan dan efisiensi optimal. Tabel 1 menunjukkan perbandingan detail performa *query* antara kedua teknik dalam berbagai skenario pengujian:

Tabel 1. Perbandingan Performa *Query* Normalisasi vs Denormalisasi

Metrik Performa	Normalisasi	Denormalisasi	Selisih (%)
Avg Response Time (ms)	485.6	127.3	-73.8%
Max Response Time (ms)	1,247.8	398.5	-68.1%
Throughput (queries/sec)	1,856	3,421	+84.3%
CPU Utilization (%)	78.4	52.7	-32.8%
Memory Usage (GB)	12.6	18.4	+46.0%
I/O Operations/sec	8,945	3,267	-63.5%

3.2. Evaluasi Scalability dan Distribusi Data

Analisis *scalability* menunjukkan bahwa kedua teknik memiliki karakteristik yang berbeda dalam menghadapi peningkatan beban kerja dan volume data. Sistem denormalisasi menunjukkan *scalability* yang lebih baik dalam skenario *read-heavy workloads*, namun mengalami *bottleneck* dalam operasi tulis karena kompleksitas *data synchronization*. Sebaliknya, sistem normalisasi menunjukkan konsistensi yang lebih baik dalam operasi tulis namun menghadapi limitasi dalam *concurrent read operations*. (Murti et al., 2024) menjelaskan bahwa implementasi sistem basis data terdistribusi menjadi solusi untuk mengatasi tantangan *scalability*, dengan berbagai metode pengembangan yang memiliki keunggulan dalam aspek-aspek tertentu. Dalam konteks sistem terdistribusi, (Andronie et al., 2023) menekankan pentingnya *remote big data management tools* dan teknologi *sensing and computing* untuk mengelola data dalam skala besar. Hasil eksperimen menunjukkan bahwa denormalisasi lebih cocok untuk implementasi dalam arsitektur terdistribusi karena mengurangi kompleksitas komunikasi antar node dan meningkatkan *data locality*. Namun, tantangan utama yang dihadapi adalah *consistency management* dan efisiensi resource, yang memerlukan strategi khusus untuk memastikan integritas data *across distributed nodes*. Tabel 2 menunjukkan hasil evaluasi *scalability* kedua teknik dalam berbagai kondisi beban kerja:

Tabel 2. Evaluasi Scalability Sistem Database

Beban Kerja	Concurrent Users	Normalisasi (QPS)	Denormalisasi (QPS)	Degradasi Normalisasi (%)	Degradasi Denormalisasi (%)
Light	100	1,856	3,421	-	-
Medium	500	1,542	3,198	-16.9%	-6.5%
Heavy	1,000	1,187	2,879	-36.1%	-15.8%
Extreme	2,000	743	2,245	-60.0%	-34.4%
Peak	5,000	298	1,567	-84.0%	-54.2%

3.3. Trade-offs dan Faktor Determinan

Analisis mendalam terhadap *trade-offs* antara kedua teknik menunjukkan bahwa pemilihan strategi yang optimal sangat bergantung pada karakteristik *workload* dan tujuan sistem. (Wijaya, 2025) dalam analisisnya terhadap struktur basis data *e-commerce* menunjukkan bahwa normalisasi hingga bentuk normal ketiga (3NF) efektif untuk menjaga integritas dan konsistensi data, namun tantangan muncul pada performa sistem akibat kompleksitas *query*. Hal ini mendukung temuan bahwa strategi *hybrid* yang menggabungkan normalisasi untuk integritas data dan denormalisasi terbatas untuk optimasi performa dapat menjadi solusi yang optimal.

Faktor-faktor yang mempengaruhi efektivitas masing-masing teknik meliputi karakteristik *workload* (*OLTP* vs *OLAP*), pola akses data, dan *requirements* sistem terhadap konsistensi data. (Khomsi Pane et al., 2022) menekankan bahwa desain basis data yang baik berperan sangat penting dalam performa dan kelancaran operasi suatu organisasi, yang memerlukan pendekatan normalisasi menggunakan *entity relationship diagram* (ERD) untuk memastikan struktur yang optimal. (Santika et al., 2024) menjelaskan bahwa evaluasi kinerja *database* menjadi krusial untuk memastikan responsivitas, keandalan, dan efisiensi sistem. Dalam konteks ini, *systematic literature review* menjadi pendekatan yang relevan untuk memahami tren dan metodologi dalam evaluasi performa *database*. Hasil penelitian menunjukkan bahwa tidak ada solusi universal dalam pemilihan teknik, melainkan diperlukan evaluasi komprehensif terhadap kebutuhan spesifik sistem.

3.4. Analisis Kuantitatif Biaya Pemeliharaan dan Kompleksitas Operasional

Evaluasi komprehensif terhadap aspek ekonomis dan operasional menunjukkan perbedaan signifikan dalam total biaya kepemilikan antara pendekatan normalisasi dan denormalisasi. Pengukuran kuantitatif dilakukan terhadap berbagai komponen biaya operasional yang mencakup beban pemeliharaan konsistensi data, kompleksitas proses ekstraksi-transformasi-pemuatan data, dan kebutuhan sumber daya untuk sinkronisasi data. Sistem denormalisasi menunjukkan peningkatan biaya pemeliharaan sebesar 67% dibandingkan normalisasi, terutama disebabkan oleh kompleksitas dalam menjaga konsistensi data di seluruh tabel yang mengalami redundansi dan beban komputasional untuk proses pembaruan berkelompok.

Analisis mendalam terhadap proses ekstraksi-transformasi-pemuatan data menunjukkan bahwa sistem denormalisasi memerlukan waktu eksekusi yang 156% lebih lama dibandingkan normalisasi, dengan kompleksitas algoritma transformasi yang meningkat secara eksponensial seiring dengan tingkat redundansi data. Pengukuran dilakukan terhadap beban operasional yang mencakup proses validasi data, mekanisme pemeriksaan integritas, dan prosedur pengembalian untuk mempertahankan konsistensi data. Sistem denormalisasi memerlukan tambahan beban penyimpanan sebesar 43% untuk mempertahankan jejak audit dan redundansi cadangan, sementara sumber daya komputasional untuk operasi pemeriksaan konsistensi mengonsumsi 78% lebih banyak siklus pemrosesan dibandingkan sistem yang dinormalisasi. Tabel 3 menunjukkan analisis kuantitatif biaya operasional dan pemeliharaan.

Tabel 3. Analisis Kuantitatif Biaya Operasional dan Pemeliharaan

Komponen Biaya	Normalisasi	Denormalisasi	Beban Tambahan (%)
Waktu Pemrosesan Data (jam)	4,2	10,8	+156,7%
Validasi Konsistensi Data (jam-pemrosesan)	12,4	22,1	+78,2%
Beban Penyimpanan Tambahan (TB)	2,8	4,0	+42,9%
Durasi Jendela Pemeliharaan (jam)	6,5	14,2	+118,5%
Skor Kompleksitas Pengembalian (1-10)	3,2	8,1	+153,1%
Waktu Staf Operasional (orang-jam/minggu)	16,8	28,4	+69,0%

Evaluasi terhadap kompleksitas pemeliharaan menunjukkan bahwa sistem denormalisasi memerlukan keahlian khusus dan jendela pemeliharaan yang diperpanjang yang berdampak pada ketersediaan sistem. Waktu pemulihan dalam skenario kerusakan data menunjukkan perbedaan substansial, dengan sistem denormalisasi memerlukan waktu 118% lebih lama untuk pemulihan sistem secara lengkap dibandingkan sistem yang dinormalisasi. Temuan ini mengindikasikan bahwa meskipun denormalisasi memberikan keunggulan kinerja dalam operasi baca, biaya tersembunyi dalam pemeliharaan operasional dapat mengimbangi keuntungan tersebut dalam siklus hidup sistem jangka panjang, khususnya untuk organisasi dengan sumber daya teknis terbatas dan persyaratan ketersediaan yang ketat.

3.5. Implikasi Praktis dan Rekomendasi

Berdasarkan temuan eksperimental, beberapa rekomendasi praktis dapat dirumuskan untuk *database architects* dan *system designers*. Untuk sistem dengan karakteristik *read-heavy workloads* dan kebutuhan performa tinggi, implementasi denormalisasi selektif dapat memberikan manfaat signifikan. Sebaliknya, untuk sistem dengan kebutuhan integritas data yang ketat dan operasi tulis yang intensif, pendekatan normalisasi tetap menjadi pilihan yang tepat. Strategi *hybrid* yang menggabungkan kedua teknik dapat menjadi solusi optimal, dengan menerapkan normalisasi pada *core transactional data* dan denormalisasi a data yang digunakan untuk *reporting* dan *analytics*. Implementasi *caching mechanisms*, *indexing strategies*, dan *content delivery networks* dapat lebih lanjut meningkatkan performa sistem tanpa mengorbankan integritas data.

Penelitian ini juga mengidentifikasi kebutuhan akan *framework* pengambilan keputusan yang dapat membantu praktisi dalam memilih teknik yang paling sesuai berdasarkan karakteristik spesifik sistem mereka. *Framework* tersebut harus mempertimbangkan faktor-faktor seperti volume data, pola akses, *requirements* konsistensi, dan *constraints resource* untuk memberikan rekomendasi yang optimal. Implementasi kerangka kerja ini memerlukan pengembangan algoritma penilaian multikriteria yang mampu melakukan evaluasi bobot relatif

terhadap setiap parameter sistem, termasuk proyeksi pertumbuhan data, distribusi beban kerja temporal, dan toleransi latensi aplikasi untuk menghasilkan skor kelayakan yang terukur. Selain itu, kerangka kerja tersebut perlu dilengkapi dengan mekanisme adaptif yang dapat melakukan reevaluasi berkala terhadap strategi basis data yang dipilih berdasarkan perubahan pola penggunaan sistem dan evolusi kebutuhan bisnis, sehingga dapat memberikan rekomendasi penyesuaian arsitektural yang proaktif untuk mempertahankan kinerja optimal dalam jangka panjang.

Untuk mengatasi keterbatasan rekomendasi strategi hibrida yang bersifat umum, penelitian ini mengusulkan metodologi terstruktur untuk implementasi denormalisasi selektif berdasarkan analisis kuantitatif beban kerja sistem. Metodologi ini melibatkan tiga tahapan sistematis: pertama, klasifikasi data berdasarkan frekuensi akses dan pola penggunaan menggunakan analisis log transaksi dengan threshold 80% operasi baca sebagai indikator kelayakan denormalisasi; kedua, penentuan titik optimal denormalisasi melalui algoritma cost-benefit yang membandingkan gain kinerja terhadap overhead pemeliharaan menggunakan formula: Skor Kelayakan = (Peningkatan Throughput × Bobot Kinerja) - (Biaya Pemeliharaan × Bobot Operasional); ketiga, implementasi bertahap dengan monitoring berkelanjutan terhadap metrik kinerja untuk validasi efektivitas strategi yang dipilih. Pembagian data transaksional dan analitik dilakukan berdasarkan karakteristik temporal akses, di mana data dengan pola akses harian dan memerlukan konsistensi real-time dipertahankan dalam bentuk normalisasi, sementara data historis dengan pola akses sporadis dan toleransi latensi tinggi diimplementasikan menggunakan denormalisasi selektif.

4. PENUTUP

Simpulan dan Saran

Penelitian ini berhasil menjawab permasalahan utama yang dirumuskan melalui analisis eksperimental komprehensif yang menunjukkan bahwa perbandingan kinerja kueri antara teknik normalisasi dan denormalisasi dalam pengelolaan data skala besar menghasilkan karakteristik yang kontras, di mana denormalisasi menunjukkan keunggulan signifikan dalam throughput dengan peningkatan 84,3%, pengurangan latensi hingga 73,8%, namun memerlukan utilisasi sumber daya memori 46% lebih besar dibandingkan normalisasi. Faktor-faktor yang mempengaruhi efektivitas masing-masing teknik telah diidentifikasi secara empiris, meliputi karakteristik beban kerja dengan sistem denormalisasi yang optimal untuk skenario *read-heavy workloads*, ukuran dataset di mana degradasi kinerja normalisasi mencapai 84% pada kondisi

puncak beban kerja dengan 5.000 pengguna konkuren, dan pola akses data temporal yang menentukan kelayakan implementasi strategi hibrida. Evaluasi objektif terhadap *trade-offs* antara integritas data, kinerja sistem, dan kompleksitas pemeliharaan menunjukkan bahwa meskipun denormalisasi memberikan keunggulan kinerja operasional, biaya pemeliharaan tersembunyi meningkat 67% dengan kompleksitas proses ekstraksi-transformasi-pemuatan data yang memerlukan waktu 156% lebih lama, sehingga pengambilan keputusan arsitektural optimal memerlukan implementasi metodologi terstruktur yang mengintegrasikan analisis kuantitatif beban kerja, algoritma *cost-benefit* dengan *threshold* 80% operasi baca sebagai indikator kelayakan denormalisasi, dan pembagian sistematis antara data transaksional yang dipertahankan dalam bentuk normalisasi untuk menjaga konsistensi *real-time* dan data analitik historis yang diimplementasikan menggunakan denormalisasi selektif untuk mengoptimalkan kinerja sistem pengelolaan data skala besar.

DAFTAR PUSTAKA

- Ajie, R. (2025). *Optimasi Kinerja Basis Data Terdistribusi Menggunakan Algoritma Replication dan Partitioning*. 3(2), 371–378.
- Andronie, M., Lăzăroiu, G., Karabolevski, O. L., Ștefănescu, R., Hurloiu, I., Dijmărescu, A., & Dijmărescu, I. (2023). *Remote Big Data Management Tools, Sensing and Computing Technologies, and Visual Perception and Environment Mapping Algorithms in the Internet of Robotic Things*. *Electronics* (Switzerland), 12(1). <https://doi.org/10.3390/electronics12010022>
- Aulia, C. P., Pratama, M. Y., & Dewi, H. L. (2023). Perbandingan Performa *Query Select Dasar , View , Dan Stored Procedure Pada Database Mysql Performance Comparison of Basic Select Queries , Views , and*. Seminar Nasional Teknologi Dan Sistem Informasi, September, 6–7.
- Khomsi Pane, S. Y., Ramadhan, N. G., & Adhinata, F. D. (2022). Perancangan Basis Data Menggunakan Normalisasi Tabel Pada Perusahaan Dagang Barokah Abadi. *Journal of Dinda : Data Science, Information Technology, and Data Analytics*, 2(2), 90–96. <https://doi.org/10.20895/dinda.v2i2.563>
- Munte, R. S., Jailani, M. S., & Siregar, I. (2024). Analisis Dokumenter Praktisi Pendidikan Islam : Pendekatan Eksperimen dan Noneksperimen (Design Kausal Komparatif dan Design Korelasional). 8, 4370–4375.
- Murti, H., Supriyanto, E., Redjeki, R., Lestariningsih, E., & Ardhianto, E. (2024). Studi Perkembangan dan Implementasi Sistem Basis Data Terdistribusi dalam Studi Literatur Review. *Jurnal Informatika Polinema*, 10(2), 249–256. <https://doi.org/10.33795/jip.v10i2.4549>

- Santika, K. Y., Hostiadi, D. P., Desiana, P., & Ayu, W. (2024). *Systematic Literature Review Analisis Dalam Metode Performance Database*. 1(2), 834–839.
- Shethiya, A. S. (2025). *Scalability and Performance Optimization in Web Application Development*. 2(1), 1–7.
- Taipalus, T. (2024). Database management system performance comparisons: A systematic literature review. *Journal of Systems and Software*, 208(October), 111872. <https://doi.org/10.1016/j.jss.2023.111872>
- Wijaya, A. K. (2025). Analisis Struktur Basis Data pada Aplikasi E-Commerce Tokopedia : Studi Kualitatif terhadap Desain dan Optimalisasi Skema Relasional. 50–57.
- CelerData. (2025). *Normalization vs Denormalization: The Trade-offs You Need to Know*. Retrieved from <https://celerdata.com/glossary/normalization-vs-denormalization-the-trade-offs-you-need-to-know>
- Couchbase. (2025). *Data Normalization vs. Denormalization Comparison*. Retrieved from <https://www.couchbase.com/blog/normalization-vs-denormalization/>
- Empirical Edge. (2025). *Understanding Database Normalization and Denormalization*. Retrieved from <https://empiricaledge.com/blog/understanding-database-normalization-and-denormalization/>
- GeeksforGeeks. (2025). *Denormalization in Databases*. Retrieved from <https://www.geeksforgeeks.org/dbms/denormalization-in-databases/>
- Pure Storage. (2024). *Denormalized vs. Normalized Data*. Retrieved from <https://blog.purestorage.com/purely-educational/denormalized-vs-normalized-data/>
- Secoda. (2025). *Data Denormalization - Explanation & Examples*. Retrieved from <https://www.secoda.co/glossary/data-denormalization>
- Sofiteco. (2025). *What Is Data Normalization? Optimize Your Data Efficiently*. Retrieved from <https://softteco.com/blog/what-is-data-normalization>
- Visual Paradigm. (2023). *Balancing Data Integrity and Performance: Normalization vs. Denormalization in Database Design*. Retrieved from <https://guides.visual-paradigm.com/balancing-data-integrity-and-performance-normalization-vs-denormalization-in-database-design/>
- Enago Academy. (2023). *Experimental Research Designs: Types, Examples & Advantages*. Retrieved from <https://www.enago.com/academy/experimental-research-design/>
- Full Scale. (2025). *Your Ultimate Guide to Better Performance Testing in 2025*. Retrieved from <https://fullscale.io/blog/performance-testing/>
- TestGrid. (2025). *The Best Performance Testing Tools in 2025*. Retrieved from <https://testgrid.io/blog/performance-testing-tools/>