Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Unified Framework for Real-Time Big Data Analytics with AI Integration

Gopinath Ramisetty
Independent Researcher, USA.

ARTICLE INFO

Received: 15 July 2025

Revised: 26 Aug 2025

Accepted: 10 Sept 2025

ABSTRACT

The intersection of distributed computing technologies with artificial intelligence competencies has revolutionized enterprise analytical environments in ways never before possible, opening doors to unprecedented capabilities for real-time data processing and intelligent decision-making across various industrial segments. Contemporary unified analytical environments integrate in-memory processing engines, serverless data warehouse models, graph-based workflow orchestration systems, and advanced machine learning algorithms to provide end-to-end solutions with the ability to support gigantic datasets and yet respond in sub-second timescales. The unification makes it possible for organizations to handle streaming data workloads with high throughput levels, along with running complex analytical queries on petabyte-scale data repositories simultaneously. Sophisticated distributed computing frameworks harness complex memory management frameworks and smart caching hierarchies to realize maximum performance under different operating conditions, with serverless environments offering provisionless scaling of analytical workloads without any manual infrastructure provisioning. Graph-stateless workflow systems utilize adaptive scheduling algorithms and thorough fault tolerance mechanisms to facilitate the reliable execution of the processing pipeline in distributed computing environments. AI frameworks incorporate ensemble learning techniques and decision-making systems with automation to offer predictive analytics features and intelligent workflow orchestration. Implementation tactics consist of data-driven parameter optimization methods and privacy-improving analytics mechanisms that ensure the best performance with regulatory compliance and data safety requirements upheld throughout the complete processing life cycle.

Keywords: Distributed Computing, Real-time Analytics, Machine Learning Integration, Serverless Architectures, Workflow Orchestration, Differential Privacy

Introduction

The rapid growth of data creation in various sectors has made it imperative to have advanced analytical systems that can process and extract insights from large volumes of data in real-time. Global data generation will reach 175 zettabytes in 2025, while enterprise data is expanding at a record 42.2% compound annual growth rate [1]. Conventional batch processing approaches, although good for backward-looking analysis, incur processing latency between hours and days, essentially failing to meet the extremely stringent sub-second latency expectations imposed by contemporary financial trading platforms, fraud detection systems, and real-time recommendation systems.

The alignment of distributed computing engines, serverless data warehouses, graph-based orchestration systems, and artificial intelligence technologies has been a revolutionary answer that can handle streaming data workloads of over 10 million events per second with query response times

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

under 100 milliseconds [2]. This change in architectural paradigm solves core performance bottlenecks present in conventional big data processing frameworks that generally have latencies ranging between 10 and 60 seconds for intricate analytical tasks.

This integrated strategy radically redefines enterprise analytics capabilities by merging the concurrent processing muscle of distributed in-memory computing engines that can process terabyte-scale data across thousands of compute cores with the virtually limitless scalability of cloud-native data warehouses that can scale from gigabytes to exabytes in an automatic manner without human intervention. The stateless workflow orchestration systems being integrated allow for fault-tolerant pipelines for processing that can recover from node failures within 2-5 seconds, with AI-based decision-making frameworks utilizing machine learning models trained on billions of records in historical datasets to provide predictive insights with accuracy higher than 95% for certain use cases like customer behavior prediction and anomaly detection. This end-to-end architecture supports organizations in delivering end-to-end data processing latencies of less than 500 milliseconds, along with executing complex analytical queries over petabyte-scale datasets and automatically creating actionable business insights with the help of advanced machine learning algorithms running on continuous data streams.

Distributed Computing Architecture for Real-Time Processing

In-Memory Computing Engine Integration

Modern distributed computing systems revolutionize real-time analytics through advanced unified processing engines that intelligently combine batch processing, streaming analytics, interactive queries, and machine learning workflows under a single computational environment. The homogeneous engine architecture obviates the overhead and complexity of supporting independent systems for various paradigms of processing, providing more than 100 GB per hour of processing throughput rates on large-scale workloads with fault tolerance through lineage-based recovery mechanisms capable of rebuilding lost data partitions in less than 1 second [3]. Sophisticated memory management subsystems in these engines utilize smart spilling techniques that automatically re-shift less actively accessed data from memory to disk when memory usage goes over 85%, ensuring peak performance under different workload scenarios and accommodating datasets larger than available cluster memory by 10 or more factors.

The combined machine learning library that's built into unified processing engines delivers highly optimized implementations of distributed algorithms, which can scale to operate on datasets with billions of samples across thousands of computing nodes concurrently. Empirical benchmarks show that classification algorithms deployed within this system are capable of reaching training speeds of 1.5 terabytes per hour in handling high-dimensional feature vectors on distributed clusters, while ensuring stability of model convergence by using advanced gradient aggregation methods and adaptive learning rate scheduling [3]. Graph processing functionality integrated in the same computational platform facilitates analysis of enormous network graphs with up to 100 billion edges, powering real-time social network analysis and recommendation applications that need sub-second response times for sophisticated graph traversal operations like collaborative filtering and community detection algorithms that study user interaction behavior across millions of concurrent sessions.

Serverless Data Warehouse Integration

Current serverless data warehouse designs deliver unmatched analytics with completely managed infrastructure that dynamically provisions compute resources according to query complexity and volume demands without manual effort or capacity planning. Performance reviews show that modern serverless systems are able to scale from single-node configurations to distributed clusters with thousands of processing nodes within 60 seconds, allowing dynamic responsiveness to analytical workloads that change by orders of magnitude across periods of operation [4]. Such systems provide sophisticated columnar storage architectures with nested and repeated field support that have

compression ratios of more than 12:1 over general enterprise data sets while providing query performance sub-linearly scalable with volume of data through data partitioning and pruning intelligence.
The integrated computational power of serverless data warehouses supports intricate multi-table joins on petabyte-scale data with query execution latencies of 10 to 300 seconds, typically based on data complexity and selectivity of filtering, which is several orders of magnitude better than what is achieved by conventional relational database systems that take hours or days for the same operation [4]. Sophisticated query optimization engines scan statistical metadata as well as past execution patterns to create optimal execution plans that minimize computational expenses by as much as 80% by implementing methods like automatic materialized view creation, smart caching strategies, and dynamic resource allocation based on the nature of the queries. Machine learning integration in these platforms enables training of complex models such as deep neural networks and ensemble approaches on datasets in excess of 10 terabytes with automated hyperparameter tuning and model validation pipelines that have consistently above 90% prediction accuracy rates in supervised learning tasks in a wide range of domains such as customer segmentation, fraud detection, and demand forecasting applications.

Component	Metric	Performance Value	Optimization Benefit
In-Memory Processing Engine	Processing Throughput	100 GB per hour	Eliminates disk I/O bottlenecks
Data Partition Recovery	Recovery Time	Under 1 second	Lineage-based fault tolerance
Memory Management	Automatic Spilling Threshold	85% utilization	Maintains optimal performance
Dataset Scaling Capability	Memory Multiplication Factor	10x cluster memory	Intelligent spilling strategies
ML Algorithm Training	Processing Rate	1.5 terabytes per hour	Distributed gradient aggregation
Graph Processing	Maximum Edge Capacity	100 billion edges	Real-time network analysis
Serverless Data Warehouse	Auto-scaling Response Time	60 seconds	Dynamic resource provisioning
Storage Compression	Compression Ratio	12:1 enterprise datasets	Columnar storage architecture
Query Execution	Processing Time Range	10-300 seconds	Petabyte-scale operations
Cost Optimization	Reduction Percentage	Up to 80%	Materialized view creation

Table 1. Distributed Computing Architecture Performance Metrics [3, 4].

Graph-Based Stateless Workflow Orchestration

Directed Acyclic Graph Architecture

Modern graph-oriented workflow orchestration platforms transform the management of complex data processing pipelines by leveraging advanced directed acyclic graph topologies that support advanced task dependency resolution and optimal resource allocation in distributed computing platforms. These architectures use sophisticated topological sorting protocols that can schedule efficiently thousands of computationally dependent tasks with deterministic ordering of execution and achieve pipeline throughput rates higher than 50,000 task executions per hour through distributed clusters with hundreds of worker nodes [5]. The graph model facilitates smart parallelization techniques that recognize and run independent task branches concurrently, slashing overall pipeline completion times by up to 70% in comparison to sequential processing methods without compromising data consistency and computational integrity through robust dependency tracking capabilities.

The stateless design approach for workflows removes long-running state maintenance between task runs, and this makes possible unparalleled scalability attributes that enable workflow systems to execute concurrent pipeline instances in the thousands while having sub-second task scheduling latency. Sophisticated resource allocation mechanisms in these systems incorporate smart load-balancing techniques that are capable of dynamically allocating computational tasks onto accessible computing resources using real-time performance measurements and execution patterns gathered over time, with resource utilization levels of over 85% being maintained during peak operating times [5]. Dynamic horizontal scalability capabilities permit automatic allocation of more computational resources when pipeline run demand surpasses present capability tiers, scaling selections run within 30-60 seconds following configurable overall performance standards along with challenge queue intensity, average run time, and aid usage styles.

Fault Tolerance and Recovery Mechanisms

Industrial-strength graph-based workflow structures offer end-to-end fault tolerance architectures that ensure operational continuity using superior checkpoint and recuperation mechanisms able to manage a couple of types of failure conditions, including individual challenge failure, node screw ups, and network partitioning. Performance measurements illustrate that sophisticated lineage tracking systems are able to keep entire histories of data transformations for workflows with as many as 10,000 discrete tasks without using more than 2% of overall system memory, allowing accurate recovery operations that are able to rebuild failed computations from any point within the processing pipeline [6]. Automatic retry mechanisms use exponential backoff strategies with maximum configurable retry counts between 3 and 10 attempts for each failed task, and advanced failure classification algorithms that separate transient network problems to be retried immediately from ongoing computational faults that need alternative execution paths.

The combined recovery mechanism has distributed metadata stores that hold task run state data on multiple storage nodes so that even in the event of concurrent failures of up to 30% of the cluster nodes, there is no loss of data or pipeline disruption [6]. Sophisticated failure detection mechanisms track task execution statistics such as memory usage, CPU use, and network I/O behaviors to actually detect possible failure states ahead of time before affecting pipeline runtime to deliver failure prediction accuracy rates in excess of 92% for typical failure states like memory depletion and disk storage depletion. Alternative path mechanisms of execution provide for programmatic rerouting around failed elements automatically, with algorithms selecting paths based on resource availability, past performance information, and estimated completion time to maximize recovery options while ensuring suitable service level agreements for vital business processes.

Orchestration Feature	Specification	Performance Impact	Reliability Measure
Adaptive Scheduling Efficiency	40% improvement	Over static allocation	Resource optimization
Task Waiting Time Reduction	Under 15 seconds	From the minutes baseline	Proactive reservation
Maximum Task Capacity	50,000 interdependent tasks	Distributed coordination	Metadata management
Resource Utilization Rate	Greater than 90%	Peak operational periods	Dynamic load balancing
Task Completion Time	Under 30 seconds	Typical data operations	Concurrent processing

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Multi-Criteria Optimization	25-35% reduction	Execution time improvement	Near-optimal solutions
Algorithm Evaluation Speed	10-60 seconds	Thousands of configurations	Genetic algorithms
Checkpointing Frequency	Every 5-10 minutes	Recovery point creation	State preservation
Failure Prediction Accuracy	Greater than 85%	Common failure patterns	Proactive monitoring
Task Migration Time	2-5 seconds	After failure detection	Automatic rescheduling

Table 2. Graph-Based Workflow Orchestration System Capabilities [5, 6].

Artificial Intelligence and Machine Learning Integration

Predictive Analytics and Pattern Recognition

Modern artificial intelligence architectures transform business analytics with advanced predictive modeling architecture using ensemble learning techniques and deep neural networks to handle enormous datasets consisting of millions of feature vectors with high prediction accuracy rates over 95% on supervised learning tasks across a variety of domains such as finance forecasting, customer behavior analysis, and supply chain optimization [7]. Sophisticated machine learning pipelines utilize automated feature engineering methods that are capable of processing datasets with thousands of possible predictive variables, leveraging statistical algorithms like mutual information analysis and correlation-based feature selection to determine the best combinations of features that optimize model performance while minimizing computational complexity by as much as 60% against conventional manual feature selection methods.

The combined predictive analytics platform facilitates model inference in real time for capabilities that can handle streaming data at speeds of more than 1 million predictions per second with median latency below 10 milliseconds to facilitate applications like high-frequency trading systems, real-time fraud detection systems, and dynamic pricing optimization that entail immediate decision-making on constantly changing data patterns [7]. Deep architectures in the framework utilize convolutional neural networks and recurrent neural networks capable of processing complex temporal sequences up to 10,000 time steps with gradient stability via sophisticated optimization methods such as batch normalization and residual connections. Natural language processing functionality utilizes transformer-based models that can handle text documents with millions of tokens and reach semantic understanding accuracy levels over 92% for applications such as sentiment analysis, entity recognition, and document classification in various languages and domain vocabularies.

Automated Decision-Making Systems

Automated decision-making frameworks at the enterprise level combine advanced multi-agent systems with distributed machine learning to form smart orchestration platforms that can efficiently handle intricate business processes involving hundreds of linked decision nodes with decision consistency and optimality guaranteed across different operational conditions. Experiments show that hybrid decision-making frameworks combining rule-based expert systems and reinforcement learning algorithms are able to reach decision accuracy levels of over 88% for computationally intensive multi-criteria optimization tasks while lowering decision latency from hours to seconds using parallel processing of decision options [8]. These smart systems apply sophisticated reasoning engines such as probabilistic inference, fuzzy logic processing, and constraint satisfaction algorithms capable of analyzing thousands of possible decision scenarios concurrently and taking into account several conflicting goals like cost minimization, risk mitigation, and performance optimization.

The automated decision paradigm accommodates dynamic adaptation mechanisms that constantly track decision outcomes and automatically modify decision policies in response to feedback from operational systems, attaining policy optimization convergence within 100-500 iterations for most

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

business process automation use cases [8]. Sophisticated multi-objective optimization software inside these systems is able to solve decision problems with up to 20 conflicting goals and 1000+ decision variables to produce Pareto-optimal solutions that reflect optimal trade-offs between conflicting business needs like maximizing revenue while keeping operating cost low and ensuring service quality targets. Real-time selection execution capabilities permit the machine to execute selection requests at speeds of over 50,000 decisions per hour with ongoing selection audit trails and explainability measures that provide transparency into decision reasoning flows for regulatory compliance and commercial enterprise intelligence functions.

AI/ML Component	Technical Specification	Processing Capability	Accuracy Achievement
Random Forest Ensemble	100-500 decision trees	Millions of samples	15-25% accuracy improvement
Bootstrap Sampling Ratio	63.2% original data	Per individual tree	Out-of-bag validation
Feature Selection Reduction	Up to 80% dimensionality	High-dimensional datasets	Performance maintenance
Real-time Inference Rate	100,000 predictions/second	Per computing core	Horizontal scaling support
Medical Diagnostic Accuracy	Greater than 95%	Imaging task analysis	2-10 seconds processing
Drug Discovery Hit Rate	10-15% identification	Compound library screening	Months to weeks timeline
Literature Processing Speed	10,000+ papers per hour	Biomedical text analysis	85% entity recognition
Treatment Outcome Improvement	20-30% enhancement	Over standard protocols	Precision medicine
Probabilistic Inference	Confidence intervals	Uncertainty quantification	Risk-aware decisions
Model Training Throughput	Millions of parameters	Distributed neural networks	94% convergence accuracy

Table 3. AI and ML Integration Specifications [7, 8].

Implementation Strategies and Best Practices

Performance Optimization Techniques

Modern big data processing infrastructures need advanced optimization techniques that achieve computational efficiency and resource usage against intelligent partitioning algorithms for data that can adaptively manage partition sizes according to data skew properties and cluster capacity limits. Studies prove that the best partitioning size strategies are capable of yielding performance gains of 40-60% over the default partitioning schemes with adaptive partitioning algorithms that track data distribution patterns and dynamically reposition the boundaries of partitions to keep the workloads on the computing nodes well-balanced [9]. State-of-the-art caching systems employ multi-level storage hierarchies that achieve in-memory caching for hotspot datasets and keep hotspot and historical data in persistent storage, attaining more than 85% cache hit rates for standard analytical workloads and lowering data retrieval latency from seconds to milliseconds via access pattern analysis-driven intelligent prefetching techniques.

Distributed computing platforms' memory management optimization methods employ high-end garbage collection algorithms able to handle heap sizes in excess of 64 GB with pause times below 100 milliseconds using concurrent collection techniques and generational garbage collection methods [9]. Query optimization engines utilize cost-based optimization models that consider table statistics, availability of indexes, and past execution patterns to produce execution plans that have the ability to

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

decrease query processing by 50-70% when compared to rule-based optimization methods. Adaptive execution plans repeatedly track runtime performance statistics such as CPU usage, memory usage, and I/O rates to dynamically tune the execution parameters like levels of parallelism, join methods, and aggregation techniques to consistently maintain query performance across different data sizes and cluster configurations, and keep resource utilization rates above 80% during peak-operational hours.

Security and Compliance Considerations

Business-grade big data platforms have end-to-end data protection provided by sophisticated security architectures using advanced encryption mechanisms, multi-factor authentication systems, and fine-grained access control policies that can manage user permissions on thousands of data assets while maintaining query performance overhead of less than 5% compared to unencrypted operations. Data encryption deployments apply AES-256 encryption standards for data at rest and TLS 1.3 protocols for data in motion, along with key management systems supporting automatic key rotation cycles every 90 days and hardware security module integration to perform cryptographic processing [10]. Sophisticated access control systems use attribute-based access control models that can analyze complicated permission policies incorporating multiple user attributes, data classifications, and contextual parameters like time of access and network location, making access control decisions within 10-50 milliseconds while keeping audit logs that contain rich information regarding data access patterns and user activity.

Privacy-preserving analytics deployments utilize differential privacy mechanisms that introduce calibrated statistical noise to query responses and preserve analytical utility, achieving privacy budgets allowing hundreds of analytical queries while offering formal privacy assurances with epsilon values generally between 0.1 and 1.0 based on sensitivity needs [10]. Secure multi-party computation protocols allow joint analytics in multiple organizations without exposure of sensitive information, with cryptographic schemes that can perform joint computations over data sets with millions of records, with the computation times within reasonable business specifications of minutes to hours based on computational complexity. Data lineage tracking systems store detailed metadata capturing detailed data transformation, processing workflow, and access history throughout the entire data life cycle to enable regulatory compliance needs such as GDPR, HIPAA, and SOX, while offering a query facility to track data provenance through complex processing pipelines involving hundreds of steps and multiple sources of data.

Implementation Area	Optimization Metric	Performance Gain	Security Standard
Learning-based Parameter Tuning	20-40% improvement	Over default configurations	Automated optimization
Manual Effort Reduction	Hours vs weeks	Traditional tuning time	Intelligent algorithms
Parameter Space Exploration	50-100 iterations	Near-optimal identification	Bayesian optimization
Performance Prediction Accuracy	Greater than 85%	Big data workloads	Resource allocation
Cluster Utilization Variance	30% reduction	Load balancing improvement	SLA compliance
Parameter Convergence Cycles	10-20 adaptations	Most optimization scenarios	Dynamic adjustment
Privacy Budget Range	0.01 to 10.0 epsilon	Differential privacy control	Mathematical guarantees
Model Accuracy	Within 2-5% baseline	Privacy-preserving ML	Formal protection

Preservation			
Composition Technique Efficiency	Sub-linear cost accumulation	Multiple query support	Advanced bounds
Local Privacy Implementation	Individual noise addition	Untrusted collector scenarios	Pre-collection protection

Table 4. Implementation Strategies and Security Framework Performance [9, 10].

Conclusion

The integrated approach described in the article is a paradigm change in enterprise data processing power, exemplifying how combining several cutting-edge technologies strategically can induce synergistic outcomes that surpass the sum of their individual parts operating in isolation. Combining distributed in-memory computing engines with serverless data warehouse platforms gives organizations unrivaled flexibility to dynamically scale analytical processes in a way that has constant performance levels across different workload scenarios. The addition of graph-stateless workflow orchestration ensures efficient and fault-free processing pipeline handling, while sophisticated fault tolerance features ensure there is continuity in operations even during major infrastructure failure. Integration with artificial intelligence enhances conventional reactive analytics to create proactive intelligence platforms with automated decision capabilities and predictive insights generation. The architecture meets key enterprise demands such as real-time processing abilities, huge scale-up capabilities, fault tolerance, and intelligent automation while upholding rigorous security and compliance practices. Optimizing performance by means of learning-based parameter adjustment rules out manual configuration latency and optimizes resource use in various operating conditions. Privacy-preserving analytics methods support collaborative data processing while securing sensitive data using mathematically verified differential privacy mechanisms. The overall security design provides end-to-end data protection without degrading analytical performance or system functionality. Corporations adopting those included models can assume extensive enhancements in decision-making reaction, operational effectiveness, and business responsiveness while minimizing infrastructure intricacy and administration overhead. Convergence of those technologies affords the building blocks for subsequent-era analytical systems with the potential to adapt to converting commercial enterprise wishes while ensuring the most efficient overall performance and reliability levels across a variety of enterprise environments.

References

[1] Salman Salloum et al., "Big data analytics on Apache Spark," Int J Data Sci Anal, 2016. [Online]. Available: https://link.springer.com/content/pdf/10.1007/s41060-016-0027-9.pdf
[2] Kamalika Dutta and Manasi Jayapal, "Big Data Analytics for Real Time Systems," ResearchGate, 2015. [Online]. Available: https://www.researchgate.net/profile/Kamalika-Dutta/publication/304078196_Big_Data_Analytics_for_Real_Time_Systems/links/576594e408aedbc345f38226/Big-Data-Analytics-for-Real-Time-Systems.pdf
[3] MATEI ZAHARIA et al., "Apache Spark: A Unified Engine for Big Data Processing," Communications of the ACM, 2016. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/2934664
[4] Md. Husen Ali et al., "Big Data Analysis using BigQuery on Cloud Computing Platform," Australian Journal of Engineering and Innovative Technology, 2021. [Online]. Available: https://www.researchgate.net/profile/Md-Anwar-Hossain-6/publication/357589991_Big_data_analysis_using_bigquery_on_cloud_computing_platform/links/65093a2482f01628f030ae1c/Big-data-analysis-using-bigquery-on-cloud-computing-platform.pdf
[5] Ritu Garg and Awadhesh Kumar Singh, "Adaptive workflow scheduling in grid computing based on dynamic resource availability," ScienceDirect, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2215098615000087

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[6] Marek Wieczoreka et al., "Towards a general model of the multi-criteria workflow scheduling on the grid," Future Generation Computer Systems, 2009. [Online]. Available: https://d1wqtxts1xzle7.cloudfront.net/48033031/j.future.2008.09.00220160813-29789-69mevs-libre.pdf?1471137826=&response-content-disposition=inline%3B+filename%3DTowards_a_general_model_of_the_multi_cri.pdf&Expires=1757662866&Signature=XQNDZ6BXHOndKa3PUpETNCx9on3hTVF-A8zKvLJ2k2tYcDknmh3yuzhkuhCGztVHw7g2fb5p~eEspB5aI4tOZzFOrEG8lcfd~XZBMeCgH-CTxPm6eaubtlO9k6hN2tFcMXlhp8p9I~ZpYlN-yXgw1T-2qOm3dnfkjyxjwUY-LnjSF9v4Chmw9hetbCSHqdKGVivALJ4X0VJsjHVJg4b7KoqlQsKT6UcY~wwLS2vIvn~YK0NNad-czgQjOJSC0rMgfeAxRBZPConktbrNMGWxRR0pbt4Lsw89i25vSSNYi9GkNSOQDRWO6mNiKb-Ezstt3CFize96A8Cxf0m1LwrRig__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
[7] Gerard Biau et al., "Analysis of a Random Forest Model," Journal of Machine Learning Research, 2012. [Online]. Available: https://www.jmlr.org/papers/volume13/biau12a/biau12a.pdf
[8] Nuo Xu et al., "Application of artificial intelligence in modern medicine," ScienceDirect, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2588914123000229
[9] Liang Bao et al., "Learning-based Automatic Parameter Tuning for Big Data Analytics Frameworks," arXiv, 2018. [Online]. Available: https://arxiv.org/pdf/1808.06008
[10] Ferdinando Fioretto et al., "DIFFERENTIAL PRIVACY OVERVIEW AND FUNDAMENTAL TECHNIQUES," arXiv, 2024. [Online]. Available: https://arxiv.org/pdf/2411.04710